Ю. В. ЛИННИК

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ И ОСНОВЫ ТЕОРИИ ОБРАБОТКИ НАБЛЮДЕНИЙ

Ю. В. ЛИННИК

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ И ОСНОВЫ МАТЕМАТИКО-СТАТИСТИЧЕСКОЙ ТЕОРИИ ОБРАБОТКИ НАБЛЮДЕНИЙ

ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ

МОСКВА 1958

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

Метод наименьших квадратов в настоящее время широко применяется при обработке количественных результатов естественно-научных опытов, технических данных, астрономических и геодезических наблюдений и измерений. После классического труда А. А. Маркова [37] на русском языке появилось некоторое число руководств, излагающих теорию этого метода (см., например, А. С. Чеботарёв [51], Н. И. Идельсон [19], А. П. Ющенко [56]). В них подробно изложена вычислительная сторона метода и указано много применений и примеров. Однако математико-статистическая сторона дела в них трактуется, в основном, в классическом духе прошлого века, без учета современных идей и достижений математической статистики, в которую метод наименьших квадратов может быть естественно включен как часть теории оценивания параметров. Исключение составляет небольшая книга В. И. Романовского [44], но она затрагивает лишь часть метода наименьших квадратов (и к тому же в изложении ее имеется ряд досадных теоретических неточностей).

Между тем применение современной математической статистики позволяет более полно и точно использовать информацию, извлекаемую из наблюдений методом наименьших квадратов, и глубже понять смысл и значение данных, полученных этим методом. Таким образом, оно полезно как практически, так и теоретически.

Настоящая книга представляет изложение теории метода наименьших квадратов с упором на математико-статистический смысл получаемых по этому методу данных (что, разумеется, имеет смысл лишь при естественном предположении о том, что погрешности измерений можно рассматривать как случайные величины).

Излагаются также основные вычислительные приемы метода (в § 12 гл. VI и в многочисленных примерах).

После введения, содержащего ряд практических примеров и догматическое применение к ним метода наименьших квадратов, в главах I, II, III вкратце излагаются необходимые сведения из алгебры, теории вероятностей и математической статистики, включая элементы современной теории оценивания параметров.

Теория обработки прямых измерений излагается в главах IV и V; здесь, в частности, изложен полезный прием обработки группы прямых измерений с извлечением информации о точности измерений из всех групп в совокупности. Глава VI трактует косвенные безусловные измерения, главы VIII и IX — уравнивание с помощью элементов и с помощью коррелят. Всюду проводится построение доверительных интервалов для оценивания измеряемых величин и точности измерений. В главе VII излагаются работы Ю. Неймана и Ф. Дэвид по оцениванию линейных форм от параметров. Глава X посвящена некоторым задачам геодезии, а глава XI — специально теории засечек. Некоторым новшеством здесь является применение доверительных эллипсов в теории засечек. В главе XII излагаются основы теории параболического интерполирования по П. Л. Чебышеву, в главе XIII — результаты А. Вальда о выравнивании ряда точек по прямой линии, когда не только ординаты, но и абсциссы измеряются с погрешностями. Последняя (четырнадцатая) глава дает некоторые дополнительные сведения о методе наименьших квадратов — общую конструкцию доверительных эллипсоидов, теоремы о тесной связи метода наименьших квадратов с нормальным законом погрешности, теорему А. Н. Колмогорова, А. А. Петрова, Ю. М. Смирнова и понятие о методе Коши обработки наблюдений.

Изложение теории ведется в матричной форме; нужные сведения о матрицах собраны в главе I. Надо заметить, что такое изложение в настоящее время стало общепринятым в соответствующих математических работах благодаря краткости и удобству записей; оно проникает также в астрономию и геодезию (см., например, финскую работу по геодезии [50]). Однако с учетом того, что матричный счет еще не для всех привычен, в конце некоторых глав настоящей книги помещены параграфы, содержащие сводку результатов и правила без применения матриц: кроме того, приведено сравнительно большое число примеров, где матрицы применяются мало.

Наиболее важные теоретические результаты выделены в перенумерованные теоремы.

При составлении настоящей книги большая помощь была оказана А. П. Хусу; ею написан § 12 главы VI, сделаны вычисления в многочисленных примерах и подготовлены таблицы, за что выражаю ей глубокую благодарность.

Ю. В. Линник

ВВЕДЕНИЕ

§ 1. Постановка задач и характерные примеры

Для того чтобы наиболее естественным образом подойти к проблематике и математическому аппарату метода наименьших квадратов, разберем некоторые конкретные примеры.

Пример 1 (А. А. Марков [37]). В „Основах химии“ Д. И. Менделеева (изд. 1906 г.) приводятся данные о растворимости азотнокислого натрия $NaNO_3$ в зависимости от температуры воды. В 100 частях воды растворяется следующее число условных частей $NaNO_3$ при соответствующих температурах:

0°	4°	10°	15°	21°	29°	36°	51°	68°
66,7	71,0	76,3	80,6	85,7	92,9	99,4	113,6	125,1

Теоретические соображения позволяют думать, что количественная сторона этого явления довольно точно описывается линейной зависимостью

$$y = a + bx, \tag{0.1.1}$$

где x — температура в градусах, а y — растворимость в условных частях на 100 частей воды. Однако, если мы попытаемся определить коэффициенты a и b в (0.1.1), полагая $x = x_i = 0°, 4°, 10°, \ldots$ и $y = y_i = 66{,}7;\ 71{,}0;\ 76{,}3, \ldots$, то придем к 9 уравнениям

$$y_i = a + bx_i \qquad (i = 1, 2, \ldots, 9) \tag{0.1.2}$$

относительно двух величин a и b, причем эти уравнения окажутся несовместными. Эта несовместность уравнений может объясняться либо несовершенством теории, предписывающей линейность зависимости, либо погрешностями наблюдений, либо тем и другим вместе. Мы можем, однако, думать, что количественные поправки к теории и наблюдениям невелики, и пытаться если не точно, то приближенно выразить наблюдения линейной зависимостью (0.1.1). Естественно тогда постараться найти такие значения коэффициентов a и b, для которых абсолютные величины „ошибок“

$$\eta_i = y_i - (a + bx_i) \qquad (i = 1, 2, \ldots, 9), \tag{0.1.3}$$

были бы в каком-то смысле „малыми в совокупности“. Способ расчета коэффициентов a и b и их значения будут зависеть от того смысла, в каком мы будем понимать „малость $|\eta_i|$ в совокупности". Мы можем, например, искать a и b из условия

$$\sum_{i=1}^{9} |\eta_i| = \min \quad (\text{минимальна}), \tag{0.1.4}$$

но тотчас же обнаруживаем, что такая задача неудобна и громоздка в вычислительном отношении, так как вводит в расчеты аналитически неудобную (не всюду дифференцируемую) функцию $|y-(a+bx)|$. Метод наименьших квадратов в данном частном случае состоит в определении a и b из условия

$$Q = \sum_{i=1}^{9} \eta_i^2 = \min. \tag{0.1.5}$$

Без дополнительной информации относительно теории и погрешностей наблюдений предписание (0.1.5) будет столь же догматичным, как, например, (0.1.4).

В дальнейшем будет выяснено, что для весьма широкого класса задач предписание (0.1.5) будет в определенном смысле наилучшим. Мы увидим тогда, что разбираемый пример (и все последующие примеры) вполне разумно причислить к этому классу задач. Пока же мы можем убедиться, что расчет по методу наименьших квадратов удобен в вычислительном отношении. Для нахождения точек (a, b), удовлетворяющих (0.1.5), составим уравнения:

$$\left.\begin{aligned} \frac{\partial Q}{\partial a} &= -2\sum_{i=1}^{9}(y_i - a - bx_i) = 0, \\ \frac{\partial Q}{\partial b} &= -2\sum_{i=1}^{9} x_i\,(y_i - a - bx_i) = 0. \end{aligned}\right\} \tag{0.1.6}$$

Для дальнейшего удобно ввести обозначения

$$\bar{x} = \frac{1}{9}\sum_{i=1}^{9} x_i; \quad \bar{y} = \frac{1}{9}\sum_{i=1}^{9} y_i; \quad S_1 = \frac{1}{9}\sum_{i=1}^{9} x_i y_i; \quad S_2 = \frac{1}{9}\sum_{i=1}^{9} x_i^2.$$

После этого (0.1.6) приводит к уравнениям

$$a + b\bar{x} = \bar{y}; \quad a\bar{x} + bS_2 = S_1. \tag{0.1.7}$$

Отсюда

$$b = \frac{S_1 - \bar{x}\bar{y}}{S_2 - \bar{x}^2}; \quad a = \bar{y} - b\bar{x}. \tag{0.1.8}$$

Мы видим, что a и b определяются из (0.1.6) однозначно, так что минимум Q, если он существует, должен быть единственным и задаваться значениями (a, b) из (0.1.8). Существование же минимума непосредственно очевидно из записи: $Q = \sum_{i=1}^{9} (y_i - a - bx_i)^2$.

Легко проверить, что формулы (0.1.8) можно переписать в виде

$$b = \frac{\sum_{i=1}^{9} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{9} (x_i - \bar{x})^2}; \quad a = \bar{y} - b\bar{x}. \tag{0.1.9}$$

Вычитая из (0.1.1) второе из равенств (0.1.9), находим искомую прямую по наименьшим квадратам" в виде

$$y - \bar{y} = b(x - \bar{x}), \tag{0.1.10}$$

где b берем из (0.1.9). Расчет можно производить по схеме, указанной в табл. 1.

Таблица 1

	x_i	y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
	0	66,7	—26	676	—23,4	608,4
	4	71,0	—22	484	—19,1	420,2
	10	76,3	—16	256	—13,8	220,8
	15	80,6	—11	121	—9,5	104,5
	21	85,7	—5	25	—4,4	22,0
	29	92,9	3	9	2,8	8,4
	36	99,4	10	100	9,3	93,0
	51	113,6	25	625	23,5	587,5
	68	125,1	42	1764	35,0	1470,0
Сумма	234	811,3		4060		3534,8

Следовательно,

$$\bar{x} = \frac{234}{9} = 26, \quad \bar{y} = \frac{811,3}{9} = 90,1,$$

откуда имеем

$$y - 90,1 = \frac{3534,8}{4060}(x - 26),$$

или

$$y = 67,5 + 0,87x \tag{0.1.11}$$

(что и было указано Д. И. Менделеевым еще в 1881 г.).

Получив прямую (0.1.11), мы можем обсудить, пока лишь путем наглядного сравнения с данными опыта, насколько удовлетворительно она изображает явление. Обозначим через y'_i данные, вычисленные по x_i ($i=1, 2, \ldots, 9$) на основании (0.1.11). Получаем таблицу:

x_i	y_i	y'_i
0	66,7	67,5
4	71,0	71,0
10	76,3	76,2
15	80,6	80,6
21	85,7	85,8
29	92,9	92,5
36	99,4	98,8
51	113,6	111,87
68	125,1	126,6

Можно заметить, что согласование, по-видимому удовлетворительное при $i=2, 3, \ldots, 7$, начинает „портиться" на краях интервала температур. Не имея дальнейших сведений и определенной, скажем математико-статистической, схемы явления, мы не можем более точно оценить качество согласования.

В данном примере мы имели дело, по-видимому, с достаточно жесткой функциональной связью температуры и растворимости; лишь погрешности наблюдений могли вносить сюда фактор случайного (в смысле теории вероятностей). В следующем примере этот фактор присутствует в самом явлении.

Пример 2 (М. Эзекиель [54]; данные переведены из английской системы мер в метрическую). При изучении движения уличного транспорта производились наблюдения над расстоянием, пройденным автомобилем по инерции после сигнала остановиться в зависимости от скорости (табл. 2).

Обозначая величину скорости через x_i, а пройденное расстояние через y_i, мы видим, что явление не выражается однозначной функциональной (и тем более линейной) зависимостью (при $x_i=6{,}44$ имеем $y_i=0{,}61$; $3{,}05$; при $x_i=11{,}26$ имеем $y_i=1{,}22$; $6{,}71$ и т. д.). Мы можем, однако, пытаться для изучения зависимости пройденного после сигнала расстояния от скорости приближенно представить наблюдения линейной зависимостью $y=ax+b$, найти a и b по методу наименьших квадратов и затем сравнить полученные данные с наблюдениями, оставляя более глубокое обсуждение на будущее. В дальнейшем будут указаны и наиболее удобные схемы расчета. Пока же заметим, что часто в подобных расчетах удобно использовать готовые таблицы квадратов. Формулы (0.1.8) и (0.1.9)

Таблица 2

Скорость в *км/час*	Пройденное расстояние в метрах	Выравненное значение	Разность
6,44	0,61; 3,05	—0,56	1,17; 3,61
11,26	1,22; 6,71	3,03	—1,81; 3,68
12,87	4,88	4,23	0,65
14,48	3,05	5,43	—2,38
16,09	7,93; 5,49; 10,37	6,63	1,30; —1,14; 3,74
17,70	8,54; 5,18	7,83	0,71; —2,65
19,31	6,10; 4,27; 7,32; 8,54	9,03	—2,93; —4,76; —1,71; —0.49
20,92	10,37; 7,93; 10,37; 14,03	10,23	0,14; —2,30; 0,14; 3,80
22,53	10,98; 7,93; 18,30; 24,40	11,43	—0,45; —3,50; 6,87; 12,97
24,14	16,47; 7,93; 6,10	12,63	3,84; —4,70; —6,53
25,74	9,76; 12,20	13,82	—4,06; —1,62
27,35	15,25; 12,20; 9,76	15,02	0,23; —2,82; —5,26
28,96	17,08; 25,62; 23,18; 12,81	16,22	0,86; 9,40; 6,96; —3,41
30,57	20,74; 14,03; 10,98	17,42	3,32; —3,39; —6,44
32,18	14,64; 17,08; 19,52; 15,86; 9,76	18,62	—3,98; —1,54; 0,90; —2,76
35,40	20,13	21,03	—0,90
37,01	16,47	22,23	—5,76
38,62	28,36; 21,35; 36,60; 28,06	23,43	4,93; —2,08; 13,17; 4,63
40,23	25,92	24,63	1,29

с заменой числа наблюдений $N=9$ на $N=50$, разумеется, годны и здесь. Из (0.1.8) находим

$$b=\frac{50\sum x_iy_i-\sum x_i\sum y_i}{50\sum x_i^2-\left(\sum x_i\right)^2}, \tag{0.1.12}$$

где суммирование идет от $i=1$ до $i=50$, считая повторяющиеся данные (например, $x_1=x_2=6,44$; $x_3=x_4=11,26$ и т. д.).

Мы имеем

$$\sum x_i=1238,98;\ \sum y_i=655,43;\ \sum x_i^2=34\,248,4;\ \sum x_iy_i=18\,885,2;$$
$$\bar{x}=24,780;\quad \bar{y}=13,109;\quad b=0,7454;\quad a=-5,362,$$

поэтому

$$y=-5,362+0,7454x. \tag{0.1.13}$$

Сравнение с наблюдениями приведено в табл. 2.

Мы видим, что характер связи в целом приближенно описывается (0.1.13) (в дальнейшем будет указан более точный смысл такого описания). Некоторые же данные не только плохо согласовываются с наблюдениями, но и вообще абсурдны, как, например, отрицательное значение y_i для $i=1$.

Тем не менее приведенные расчеты, по уточнении их реального смысла, оказываются полезными.

В предыдущих примерах разбиралось описание наблюдений с помощью линейной зависимости или, как это часто называют, — выравнивание наблюдений по прямой линии. Разберем теперь более сложный случай выравнивания по кривой линии, в данном примере — по параболе.

Пример 3 (М. Эзекиель [54]). Производится наблюдение над двумя переменными — содержанием протеина x в зернах пшеницы и процентным содержанием y „темных стекляннистых ядрышек" в тех же зернах. Обе переменные характеризуют качество пшеницы, но определение x требует сложного химического анализа, а определение y может быть сделано гораздо проще и без приборов. В табл. 3 приведены результаты 20 наблюдений.

Таблица 3

№	Содержание (%)	
	протеина	стекляннистых ядрышек
1	10,3	6
2	12,2	75
3	14,5	87
4	11,1	55
5	10,9	34
6	18,1	98
7	14,0	91
8	10,8	45
9	11,4	51
10	11,0	17
11	10,2	36
12	17,0	97
13	13,8	74
14	10,1	24
15	14,4	85
16	15,8	96
17	15,6	92
18	15,0	94
19	13,3	84
20	19,0	99

Попробуем произвести выравнивание этих наблюдений по параболе

$$y = a + bx + cx^2. \qquad (0.1.14)$$

В данном случае метод наименьших квадратов приводит к условию: a, b, c нужно определить так, чтобы

$$Q = \sum_i \left[y_i - \left(a + bx_i + cx_i^2\right)\right]^2 = \min. \qquad (0.1.15)$$

Снова замечаем, что минимум Q достигается, и составляем уравнения

$$\frac{\partial Q}{\partial a} = 0; \quad \frac{\partial Q}{\partial b} = 0; \quad \frac{\partial Q}{\partial c} = 0. \qquad (0.1.16)$$

Введем обозначения:

$$x_i^2 = z_i; \quad S_1 = \sum x_i y_i; \quad S_2 = \sum x_i^2; \quad T_1 = \sum x_i z_i; \quad T_2 = \sum z_i^2;$$

$$V_1 = \sum y_i z_i; \quad \bar{x} = \frac{1}{N}\sum x_i; \quad \bar{y} = \frac{1}{N}\sum y_i; \quad \bar{z} = \frac{1}{N}\sum z_i$$

$$(i = 1, 2, \ldots, 20); \quad N = 20.$$

Тогда

$$y_i = a + bx_i + cz_i, \qquad (0.1.17)$$

и из (0.1.16) легко получить следующие уравнения для b, c, a:

$$\left.\begin{aligned} S_2 b + T_1 c &= S_1; \\ T_1 b + T_2 c &= V_1; \\ a &= \bar{y} - b\bar{x} - c\bar{z}. \end{aligned}\right\} \tag{0.1.18}$$

Это частный случай так называемых нормальных уравнений. Удобный способ решения нормальных уравнений (по Э. Дулиттлю) будет изложен далее. Приведем лишь окончательные результаты расчета:

$$a = 11{,}56; \quad b = -0{,}088; \quad c = 0{,}00144,$$

так что

$$y = 11{,}56 - 0{,}088x + 0{,}00144x^2. \tag{0.1.19}$$

Согласование расчетных данных y' по (0.1.19) с наблюдениями видно из табл. 4.

Таблица 4

Содержание в (%)		Выравненное значение содержания протеина, % (минус 10)	Разность
стекловидных ядрышек	протеина (минус 10)		
6	0,3	1,08	—0,78
75	2,2	3,06	—0,86
87	4,5	4,80	—0,30
55	1,1	1,08	0,02
34	0,9	0,23	0,67
98	8,1	6,79	1,31
91	4,0	5,50	—1,50
45	0,8	0,52	0,28
51	1,4	0,83	0,57
17	1,0	0,48	0,52
36	0,2	0,26	—0,06
97	7,0	6,60	0,40
74	3,8	2,95	0,85
24	0,1	0,28	—0,18
85	4,4	4,51	—0,11
96	5,8	6,41	—0,61
92	5,6	5,68	—0,08
94	5,0	6,04	—1,04
84	3,3	4,35	—1,05
99	9,0	6,99	2,01

Согласование производит впечатление удовлетворительного и оправдывающего выбор параболы (0.1.19). Просматривая расчеты, можно заметить, что для них не было существенным то обстоятельство, что $z_i = x_i^2$; числа z_i в (0.1.17) могли быть набором заданных чисел, вообще не зависящим от набора чисел x_i. Взяв вместо квадратичной параболы (0.1.14) кубическую параболу

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3, \tag{0.1.20}$$

можно пытаться выравнивать данные табл. 3 по такой параболе; полагая $x_i = x_{1i}$; $x_i^2 = x_{2i}$; $x_i^3 = x_{3i}$, пришли бы к системе несовместных уравнений

$$y_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + a_3 x_{3i} \qquad (i = 1,\ 2,\ \ldots,\ 20), \tag{0.1.21}$$

и могли бы найти (0.1.20) по методу наименьших квадратов. Таким же образом можно действовать в примере 2. Однако в тех случаях, когда, как в примере 2, нет однозначной функциональной зависимости данных, это сильно усложнит расчеты.

Пример 4. Измерения трех углов A_1, A_2, A_3 плоского треугольника привели к значениям:

$$A_1 = 54°5'; \qquad A_2 = 50°1'; \qquad A_3 = 76°6'.$$

Сумма углов $A_1 + A_2 + A_3 = 180°12'$ дает невязку в $12'$, происходящую от погрешностей наблюдений. Постараемся ликвидировать невязку, заменив A_i на $A_i + \lambda_i = A_i'$ $(i = 1,\ 2,\ 3)$ так, чтобы $A_1' + A_2' + A_3' = 180°$. Тогда должно быть

$$\lambda_1 + \lambda_2 + \lambda_3 = -0°12'. \tag{0.1.22}$$

Этим на величины λ_1, λ_2, λ_3 накладывается связь (0.1.22), а в остальном они остаются неопределенными. В данном случае метод наименьших квадратов сводится к предписанию

$$Q = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 = \min \tag{0.1.23}$$

в дополнение к условию (0.1.22). Это — задача на условный минимум, в общем случае разрешаемая с помощью вспомогательных множителей Лагранжа. В данном случае ее решение можно провести без дифференциального исчисления. Положим $\bar{\lambda} = \frac{1}{3}(\lambda_1 + \lambda_2 + \lambda_3)$. Тогда, как легко проверить,

$$Q = \lambda_1^2 + \lambda_2^2 + \lambda_3^2 = \sum_{i=1}^{3} (\lambda_i - \bar{\lambda})^2 + 3\bar{\lambda}^2.$$

Из (0.1.22) видим, что $\bar{\lambda} = -0°4'$, так что

$$Q = \sum_{i=1}^{3} (\lambda_i + 0°4')^2 + 3 \cdot (0°4')^2.$$

Отсюда видно, что Q достигает единственного минимума при $\lambda_i = -0°4'$, так что

$$A_1' = 54°1'; \quad A_2' = 49°57'; \quad A_3' = 76°2'. \qquad (0.1.24)$$

Здесь мы имеем пример применения метода наименьших квадратов при условии связанных наблюдений. Углы A_1', A_2', A_3' не дают невязки; реальный смысл введенных поправок будет выяснен в дальнейшем.

Пример 5. Из нескольких точек на плоскости $M_1, M_2, \ldots, M_s$ производятся засечки на некоторую точку O (рис. 1), скажем, визуальным, звукометрическим или иным способом. Засечки приводят к ряду направлений M_iN_i. В силу погрешностей наблюдений, прямые M_iN_i не пересекаются в одной точке. Требуется найти приближение к искомой точке O на основании наблюденных направлений M_iN_i.

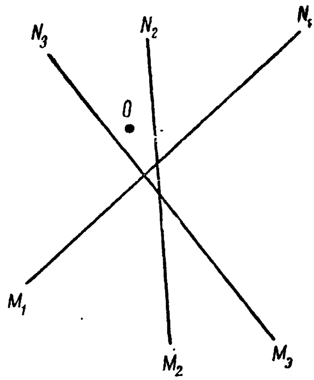

Рис. 1.

Можно было бы применять метод наименьших квадратов в форме следующего предписания. Пусть $d_i(X)$ обозначает расстояние некоторой точки X от прямой M_iN_i. Будем искать X из условия

$$\sum d_i^2(X) = \min. \qquad (0.1.25)$$

Как мы увидим далее (гл. XI), целесообразнее другое предписание также типа наименьших квадратов, которое будет изложено ниже.

Аналогично можно поставить задачу о засечках в пространстве (например, засечка самолета).

Пример 6. При математико-статистическом описании неровностей профиля поверхности при производственной обработке важную роль играет понятие средней линии случайного профиля (см. Ю. В. Линник и А. П. Хусу [33]). В весьма упрощенной форме задачу нахождения средней линии профиля можно поставить следующим образом. На плоскости XOY задан ряд точек (x_i, y_i) $(i = 1, 2, \ldots, s)$. Требуется провести прямую $y = ax + b$ так, чтобы сумма квадратов расстояний точек (x_i, y_i) от этой прямой была минимальной.

Эту задачу можно интерпретировать механически. Припишем точкам (x_i, y_i) единичные массы и будем искать ось вращения, дающую данной системе точек наименьший момент инерции (наименьшую кинетическую энергию вращения при заданной угловой скорости). Из элементарного курса механики известно, что такая ось проходит через центр тяжести системы и что перпендикулярная ей ось, проходящая через центр тяжести, будет давать для той же системы максимальный момент инерции. Легко заметить, что искомых осей может быть несколько (скажем, если точки расположены по вершинам углов правильного многоугольника или вообще система точек самосовмещается после некоторого поворота на острый угол).

Далее, из основных свойств непрерывных функций явствует, что по крайней мере одна такая ось существует. Докажем, что она проходит через центр тяжести системы точек (x_i, y_i), который, очевидно, имеет координаты $(\overline{x}, \overline{y})$ *).

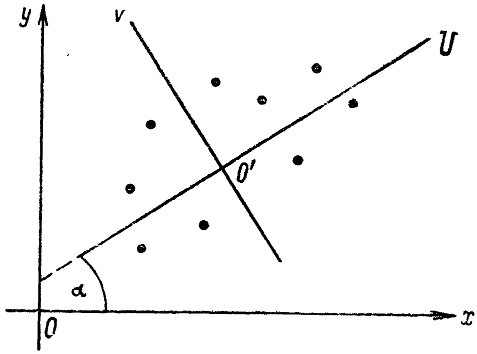

Рис. 2.

Обозначим нашу ось $O'U$ и проведем перпендикулярную ей ось $O'V$ (рис. 2). Пусть координаты точек (x_i, y_i) в этой новой системе координат будут (u_i, v_i). Из равенства

$$\sum (v_i - \overline{v})^2 = \sum v_i^2 - n\overline{v}^2$$

заключаем, что если $\overline{v} \neq 0$, то, заменяя ось $O'U$ на ось $v = \overline{v}$, мы уменьшим момент инерции, что невозможно. Итак, $\overline{v} = 0$, так что ось $O'U$ проходит через центр тяжести $(\overline{u}, \overline{v})$, который имеет старые координаты $(\overline{x}, \overline{y})$. Будем считать, что начало отсчета O' системы $(O'U, O'V)$ совпадает с $(\overline{x}, \overline{y})$. Остается найти условие для угла наклона α новой системы к старой.

Имеем

$$Q = Q(\alpha) = \sum v_i^2 = \min, \tag{0.1.26}$$

откуда

$$\sum v_i \frac{dv_i}{d\alpha} = 0. \tag{0.1.27}$$

Далее, полагая

$$x'_i = x_i - \overline{x}; \qquad y'_i = y_i - \overline{y},$$

получаем по формулам преобразования координат

$$\left.\begin{aligned} u_i &= x'_i \cos\alpha + y'_i \sin\alpha, \\ v_i &= -x'_i \sin\alpha + y'_i \cos\alpha. \end{aligned}\right\} \tag{0.1.28}$$

*) Это утверждение равносильно известной теореме Штейнера из механики.

Отсюда $\frac{dv_i}{d\alpha} = -u_i$, так что (0.1.27) переписывается в виде

$$\sum u_i v_i = 0.$$

Подставляя сюда (0.1.28), имеем

$$(\cos^2\alpha - \sin^2\alpha)\sum x_i' y_i' - \sin\alpha\cos\alpha\sum(x_i'^2 - y_i'^2) = 0.$$

Отсюда

$$\operatorname{tg} 2\alpha = \frac{2\sum x_i' y_i'}{\sum x_i'^2 - \sum y_i'^2} = \frac{2\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2 - \sum(y_i - \bar{y})^2}. \qquad (0.1.29)$$

Как видим, $\operatorname{tg} 2\alpha$, вообще говоря, определяется из (0.1.29) однозначно, так что угол α получает всего четыре значения. На первый взгляд кажется, что это трудно примирить с указанными выше случаями наличия многих осей симметрии. Но это можно объяснить тем, что в подобных случаях выражение (0.1.29) должно образовывать неопределенность типа $\frac{0}{0}$.

В случаях, интересных для теории качества обработки поверхностей, точки (x_i, y_i) мало удалены от некоторой прямой, и угол α из (0.1.29) определяется однозначно.

Сравним формулу (0.1.29) с формулой (0.1.9). В (0.1.9) имеем $b = \operatorname{tg}\alpha'$, где α' — угол наклона оси, проведенной по методу наименьших квадратов, к оси OX. В силу того, что

$$\operatorname{tg} 2\alpha' = \frac{2\operatorname{tg}\alpha'}{1 - \operatorname{tg}^2\alpha'},$$

находим из (0.1.9)

$$\operatorname{tg} 2\alpha' = \frac{2\sum(x_i - \bar{x})(y_i - \bar{y})\sum(x_i - \bar{x})^2}{\left(\sum(x_i - \bar{x})^2\right)^2 - \left(\sum(x_i - \bar{x})(y_i - \bar{y})\right)^2}. \qquad (0.1.30)$$

Найденная прямая называется прямой ортогональной регрессии.

Рассмотрев приведенные выше примеры, можем убедиться, что математическая постановка большинства из них приводится к следующей форме (метод наименьших квадратов): задается система чисел x_{ki} $(k = 1, 2, \ldots, n;\ i = 1, 2, \ldots, N)$; $w_i > 0$ $(i = 1, 2, \ldots, N)$ и система чисел η_i $(i = 1, 2, \ldots, N)$. Ищутся числа $a_1, a_2, \ldots, a_n$, для которых

$$Q = \sum_{i=1}^{N} w_i\left(\eta_i - \sum_{k=1}^{n} a_k x_{ki}\right)^2 = \min \qquad (0.1.31)$$

при возможном наличии дополнительных условий:

$$F_j(a_1, a_2, \ldots, a_n) = 0 \qquad (j = 1, 2, \ldots, m). \qquad (0.1.32)$$

где F_j заданные функции n аргументов.

В большинстве случаев при этом предполагается наличие определенной вероятностной схемы, в которой величины η_i (а иногда и x_{ki}) рассматриваются как случайные, подчиняющиеся определенным требованиям. При этом и найденные по методу наименьших квадратов величины $a_1, a_2, \ldots, a_n$ оказываются случайными. Описание их поведения и выяснение их экстремальных свойств среди других допустимых в определенном смысле величин составляет вероятностную часть теории метода наименьших квадратов и поясняет реальный вероятностный смысл расчетов по этому методу. Вопросы же, связанные с нахождением $a_1, a_2, \ldots, a_n$ из (0.1.31) при заданных x_{ki}, η_i, w_i, F_j, составляют алгебраическую и вычислительную часть теории. При этом полезны и некоторые геометрические представления.

Ввиду этого перед изложением общей теории, в гл. I, II и III будут вкратце изложены необходимые сведения по алгебре, теории вероятностей и математической статистике.

§ 2. Краткий исторический обзор

Первое изложение элементов метода наименьших квадратов дано в 1806 г. А. М. Лежандром *) [30] в связи с вопросом о вычислениях кометных орбит. Ему же принадлежит название: „метод наименьших квадратов".

В 1809 г. К. Ф. Гаусс **) [7] дал первое вероятностное обоснование метода наименьших квадратов, а в 1810 г. он же [8] глубоко разработал вычислительную сторону вопроса и ввел символы и обозначения, сохранившиеся и поныне. Ряд новых важных результатов найден К. Ф. Гауссом в 1821 г. [9].

В 1812 г. П. С. Лаплас ***) в фундаментальном трактате по теории вероятностей [29] получил ряд важных результатов и применил их к методу наименьших квадратов. Дальнейшие важные результаты были получены в теории метода наименьших квадратов в 1859 г. П. Л. Чебышевым ****), разработавшим теорию интерполирования по методу наименьших квадратов с помощью ортогональных полиномов, носящих его имя (см. гл. XII).

А. А. Марков *****) в 1898 г. в работе [36] и в уже упоминавшемся в § 1 курсе теории вероятностей [37] внес в математическую статистику ряд весьма важных идей, пояснивших суть метода наименьших квадратов.

Много сделано для развития применений метода наименьших квадратов в астрономии и геодезии Ф. Гельмертом ******) в конце

*) А. М. Лежандр (1752—1833) — французский математик.
**) К. Ф. Гаусс (1777—1855) — немецкий математик.
***) П. С. Лаплас (1749—1827) — французский математик.
****) П. Л. Чебышев (1821—1894) — русский математик.
*****) А. А. Марков (1856—1922) — русский математик.
******) Ф. Гельмерт (1843—1917) — немецкий геодезист.

прошлого века. После работ А. А. Маркова с двадцатых годов нынешнего века метод наименьших квадратов включился в математическую статистику как важная и естественная часть теории оценивания параметров (см. гл. III). В этой связи ряд интересных и важных результатов получен Ю. Нейманом и Ф. Дэвид (см. об этом гл. VII), Эйткеном [55] и С. Рао [41].

В 1946 г. А. Н. Колмогоровым [22] дано изящное геометрическое изложение метода наименьших квадратов.

За последнее время в метод наименьших квадратов и его применения все более и более проникает матричное изложение. Оно позволяет удобную и короткую запись выкладок и результатов и в дальнейшем будет систематически использоваться.

ГЛАВА I

НЕОБХОДИМЫЕ СВЕДЕНИЯ ИЗ АЛГЕБРЫ

§ 1. Векторы

В настоящей главе будут коротко изложены без доказательств сведения из алгебры, нужные для выводов и удобной записи результатов метода наименьших квадратов. Доказательства приводимых теорем и другие полезные теоремы можно найти, например, в книгах Ф. Р. Гантмахера [6], А. Г. Куроша [28], А. И. Мальцева [35].

Упорядоченную совокупность n действительных чисел $(x_1, x_2, \ldots, x_n)$ будем называть (n-мерным) *вектором* и обозначать X. Вектор можно умножать на скаляр c, причем получается новый вектор $cX = (cx_1, cx_2, \ldots, cx_n)$. Два вектора, $X = (x_1, x_2, \ldots, x_n)$ и $Y = (y_1, y_2, \ldots, y_n)$, можно, по определению, складывать, получая новый вектор $Z = X + Y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$; можно складывать и более двух векторов. *Скалярное произведение* двух векторов, которое мы будем обозначать в символике Гаусса $[X \cdot Y]$, определяется как число, равное

$$[X \cdot Y] = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n,$$

так что

$$[X \cdot Y] = [Y \cdot X].$$

При $n = 1, 2, 3$ это произведение имеет наглядный геометрический смысл. Векторы X и Y называются *ортогональными*, если $[X \cdot Y] = 0$; при $n = 2, 3$ получаются обычные взаимно перпендикулярные векторы.

Линейные комбинации векторов $X_1, X_2, \ldots, X_s$ вида $c_1 X_1 + \ldots + c_s X_s$, где $c_1, \ldots, c_s$ могут быть любыми вещественными числами, образуют *векторное пространство* $\mathcal{X}_s$.

Если из соотношения $c_1 X_1 + \ldots + c_s X_s = 0$ следует, что $c_1 = c_2 = \ldots = c_s = 0$, то векторы $X_1, \ldots, X_s$ называются *линейно независимыми*. Во всяком векторном пространстве $\mathcal{X}_s$ можно выбрать *базис* из линейно независимых векторов $Z_1, Z_2, \ldots, Z_t$ $(t \leqslant s)$, т. е. систему векторов Z_i $(i = 1, 2, \ldots, t)$ такую, что всякий элемент векторного пространства однозначно выражается в виде $b_1 Z_1 + \ldots + b_t Z_t$ (b_i — вещественные числа). Говорят, что пространство $\mathcal{X}_s$ имеет *размерность* t.

Если имеются две системы векторов $X_1, \ldots, X_s$ и $Y_1, \ldots, Y_v$, причем $[X_i \cdot Y_j] = 0$ ($i = 1, 2, \ldots, s$; $j = 1, 2, \ldots, v$), то пространства $\mathscr{X}_s$ и $\mathscr{Y}_v$, образованные этими векторами, называются *ортогональными*.

Если несколько векторов попарно ортогональны, то они линейно независимы.

Отметим еще важное *неравенство Буняковского — Коши* *):

$$[X \cdot Y]^2 \leqslant [X \cdot X] \cdot [Y \cdot Y], \tag{1.1.1}$$

т. е. скалярное произведение векторов по абсолютной величине не превосходит квадратного корня из произведения скалярных квадратов. Более подробная запись:

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leqslant \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} y_i^2\right). \tag{1.1.2}$$

§ 2. Линейные уравнения. Матрицы

Система m линейных уравнений с n неизвестными $x_1, \ldots, x_n$ имеет вид

$$\left.\begin{array}{r} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m. \end{array}\right\} \tag{1.2.1}$$

Если $b_1 = b_2 = \ldots = b_m = 0$, система называется *однородной*. Введя векторы $a_i = (a_{i1}, \ldots, a_{in})$ ($i = 1, 2, \ldots, m$) и $X = (x_1, \ldots, x_n)$, можем записать систему в виде

$$[a_i \cdot X] = b_i \qquad (i = 1, 2, \ldots, m). \tag{1.2.2}$$

Таким образом, для случая однородной системы ($b_i = 0$; $i = 1, 2, \ldots, m$) задача решения системы совпадает с задачей построения векторного пространства, ортогонального пространству, построенному на векторах $a_1, \ldots, a_m$.

С изобретением матриц А. Кэйли **) изучение систем линейных уравнений, как части теории линейных преобразований, чрезвычайно углубилось и упростилось. Так как в методе наименьших квадратов все время приходится иметь дело с линейными уравнениями, применение там матриц весьма целесообразно.

---

*) В. Я. Буняковский (1804—1889) — русский математик; О. Коши (1789—1857) — французский математик.

**) А. Кэйли (1821—1895) — английский алгебраист.

Матрица типа $m \times n$ есть прямоугольная таблица вещественных чисел a_{ik} вида

$$A = \begin{Vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{Vmatrix},$$

или, в сокращенной записи, $A = \|a_{ik}\|$.

Чтобы подчеркнуть, что матрица A имеет m строк и n столбцов, иногда будем писать $A = A_{mn}$. Матрицу $A_{11} = \|a_{11}\|$ будем отождествлять с обычным числом a_{11}. Элемент a_{rs} матрицы A будем обозначать символом $\{A\}_{rs}$ *).

Две матрицы A_{mn} и $B_{m_1 n_1}$ будут считаться *равными* тогда и только тогда, если $m = m_1$, $n = n_1$ и $\{A\}_{rs} = \{B\}_{rs}$ для всех возможных r и s.

Нулевая матрица $0 = 0_{mn}$ есть такая матрица, что $\{0\}_{rs} = 0$ для всех возможных r и s.

Матрицы $X_{1n} = \|x_{11}, \dots, x_{1n}\|$ и $X_{n1} = \begin{Vmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{n1} \end{Vmatrix}$ можно считать n-мерными векторами.

Элементы $\{A\}_{rr}$ $(r = s)$ называются *диагональными*, а диагональ, на которой они расположены, — *главной диагональю*.

Если матрицу

$$A_{mn} = \begin{Vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{Vmatrix}$$

повернуть вокруг главной диагонали, то получится матрица

$$B_{nm} = \begin{Vmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{Vmatrix},$$

в которой строки A_{mn} заменены столбцами и обратно. Матрица B_{nm} называется *транспонированной* по отношению к A_{mn} и обозначается следующим образом: $B_{nm} = A_{mn}^T$.

Очевидно

$$B_{nm}^T = \left(A_{mn}^T\right)^T = A_{mn}.$$

*) Такое обозначение впервые введено советским математиком И. А. Лаппо-Данилевским (1895—1931).

Для всех матриц A определено умножение на любое вещественное число α:

$$\alpha A = \alpha \| a_{rs} \| = \| \alpha a_{rs} \|.$$

Для матриц с одинаковым числом строк m и столбцов n определено сложение. Если

$$A_{mn} = \| a_{rs} \|, \quad B_{mn} = \| b_{rs} \|,$$

то

$$A_{mn} + B_{mn} = \| a_{rs} + b_{rs} \|.$$

В силу таких определений матрицы $A = A_{mn}$ образуют векторное пространство; в частности, с помощью матриц

$$X_{n1} = \begin{Vmatrix} x_1 \\ \vdots \\ x_n \end{Vmatrix} \quad \text{и} \quad X_{1n} = \| x_1, \ldots, x_n \|$$

можно записать n-мерные векторы.

Произведение двух матриц A_{mn} и B_{uv} определяется только для того случая, когда $u = n$ (число столбцов матрицы A_{mn} равно числу строк матрицы B_{uv}). Их произведение

$$C_{mv} = A_{mn} B_{nv}$$

определяется так:

$$C_{mv} = \| c_{rs} \|,$$

где

$$c_{rs} = a_{r1} b_{1s} + a_{r2} b_{2s} + \ldots + a_{rn} b_{ns} = \sum_{i=1}^{n} a_{ri} b_{is}.$$

Смысл такого определения произведения полностью выясняется при сопоставлении матриц с линейными преобразованиями, из которых они возникли. Если

$$X = \begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_v \end{Vmatrix}, \quad B = B_{nv} = \| b_{rs} \|, \quad Y = \begin{Vmatrix} y_1 \\ \vdots \\ y_n \end{Vmatrix},$$

то запись

$$Y = B \cdot X \tag{1.2.3}$$

в весьма удобной и сокращенной форме выражает линейное преобразование

$$\left. \begin{array}{l} y_1 = b_{11} x_1 + b_{12} x_2 + \ldots + b_{1v} x_v, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_{n1} x_1 + b_{n2} x_2 + \ldots + b_{nv} x_v. \end{array} \right\} \tag{1.2.4}$$

Если наряду с (1.2.4) имеется еще преобразование

$$Z = A \cdot Y, \tag{1.2.5}$$

где

$$Z = \left\| \begin{matrix} z_1 \\ \vdots \\ z_m \end{matrix} \right\|, \quad A = A_{mn} = \| a_{rs} \|,$$

соответствующее равенствам

$$\left. \begin{matrix} z_1 = a_{11} y_1 + \ldots + a_{1n} y_n, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ z_m = a_{m1} y_1 + \ldots + a_{mn} y_n, \end{matrix} \right\} \tag{1.2.6}$$

то результат подстановки (1.2.4) в (1.2.6) дает новое линейное преобразование

$$Z = C \cdot X. \tag{1.2.7}$$

Здесь

$$C = C_{mv} = A_{mn} B_{nv}, \tag{1.2.8}$$

что неудивительно, ибо произведение (1.2.8) как раз и было так определено, чтобы такое совпадение получилось. Операция перемножения матриц в точности соответствует последовательному выполнению двух линейных преобразований. Ввиду этого естественны следующие законы умножения матриц:

ассоциативный закон:

$$(A_{mn} B_{nv}) C_{vw} = A_{mn} (B_{nv} C_{vw});$$

дистрибутивный закон:

$$A_{mn} (B_{nv} + C_{nv}) = A_{mn} \cdot B_{nv} + A_{mn} C_{nv},$$

$$(A_{mn} + B_{mn}) C_{nv} = A_{mn} C_{nv} + B_{mn} C_{nv}.$$

Эти законы можно распространить на любое число сомножителей или слагаемых.

Также естественно, ввиду указанной выше связи с линейными преобразованиями, что коммутативный (перестановочный) закон, действующий для обычных вещественных чисел, вообще говоря, не имеет места для матриц, т. е., $AB \neq BA$. Если $A = A_{mn}$, $B = B_{nv}$, то произведение AB определено, а BA даже не определено, если $v \neq m$, если же $v = m$, то BA существует, но, вообще говоря, отличается от AB.

Правило транспонирования произведения имеет вид

$$(AB)^T = B^T A^T,$$

т. е. при транспонировании множители меняют порядок и транспонируются (отчего само перемножение становится возможным). Это правило распространяется на любое число сомножителей.

Квадратной матрицей $A = A_{nn}$ называется такая, у которой число строк равно числу столбцов.

Диагональной матрицей называется квадратная матрица вида

$$D = D_{nn} = \begin{Vmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & d_n \end{Vmatrix},$$

где элементы вне главной диагонали равны 0. Если $d_i \geqslant 0$ ($i = 1, 2, \ldots, n$), то мы можем ввести удобное для аппарата метода наименьших квадратов понятие *дробной степени диагональной матрицы*. Пусть $\alpha > 0$. Через D^α обозначим матрицу

$$D^\alpha = \begin{Vmatrix} d_1^\alpha & 0 & \ldots & 0 \\ 0 & d_2^\alpha & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & d_n^\alpha \end{Vmatrix}. \tag{1.2.9}$$

Если $\alpha > 0$ и $\beta > 0$, то имеем очевидное равенство

$$D^\alpha \cdot D^\beta = D^{\alpha+\beta}.$$

В частности, $D^{\frac{1}{2}} \cdot D^{\frac{1}{2}} = D$.

Единичной матрицей называется диагональная матрица, у которой все диагональные элементы равны единице: $d_i = 1$ ($i = 1, 2, \ldots, n$). Ее будем обозначать E или E_{nn}. Имеем

$$A_{mn}E_{nn} = A_{mn}; \quad E_{mm}A_{mn} = A_{mn}$$

для любой матрицы A_{mn}.

Иногда появляется нужда выделить внутри квадратной матрицы $Q = Q_{nn}$ отдельные матрицы, из которых она составлена, по схеме:

$$Q = Q_{nn} = \left\Vert \begin{array}{c:c} A_{rr} & B_{rk} \\ \hdashline C_{kr} & D_{kk} \end{array} \right\Vert; \quad n = r + k.$$

Такую матрицу будем называть *разделенной*; пунктир указывает, как произведено разделение. Если имеется еще одна матрица $Q' = Q'_{nn}$, разделенная так же, как предыдущая:

$$Q' = Q'_{nn} = \left\Vert \begin{array}{c:c} A'_{rr} & B'_{rk} \\ \hdashline C'_{kr} & D'_{kk} \end{array} \right\Vert,$$

то имеет место важная формула

$$QQ' = \left\| \begin{array}{c:c} A_{rr} & B_{rk} \\ \hdashline C_{kr} & D_{kk} \end{array} \right\| \cdot \left\| \begin{array}{c:c} A'_{rr} & B'_{rk} \\ \hdashline C'_{kr} & D'_{kk} \end{array} \right\| =$$

$$= \left\| \begin{array}{c:c} A_{rr}A'_{rr} + B_{rk}C'_{kr} & A_{rr}B'_{rk} + B_{rk}D'_{kk} \\ \hdashline C_{kr}A'_{rr} + D_{kk}C'_{kr} & C_{kr}B'_{rk} + D_{kk}D'_{kk} \end{array} \right\|, \qquad (1.2.10)$$

т. е. мы получим верный результат, обращаясь при перемножении с матрицами так, как если бы это были числа.

Важную роль в теории квадратных матриц играют их *определители*, или *детерминанты*. Если $A = A_{nn} = \|a_{rs}\|$, то определитель

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}$$

будем обозначать $\det(A)$. Основные правила действий с определителями будем считать известными.

Если в матрице $A = \|a_{rs}\|$ взять какой-либо элемент a_{rs} и вычеркнуть из этой матрицы r-ю строчку и s-й столбец (на пересечении которых стоит элемент a_{rs}), то получается новая квадратная матрица $A'_{n-1,\,n-1}$. Ее определитель называется *минором* элемента a_{rs} и обозначается $M^{(r,\,s)}$. Величина $D^{(r,\,s)} = (-1)^{r+s} \cdot M^{(r,\,s)}$ называется *алгебраическим дополнением* a_{rs}.

Имеет место формула

$$\det(A) = \sum_{s=1}^{n} a_{rs} D^{(r,\,s)} = \sum_{r=1}^{n} a_{rs} D^{(r,\,s)}.$$

Указанные понятия обобщаются следующим образом. Если в матрице $A = A_{nn}$ возьмем строки с номерами $i_1, i_2, \dots, i_k$ и столбцы с номерами $j_1, j_2, \dots, j_k$, то, выделяя элементы, стоящие на пересечении таких строк и столбцов, получаем новую матрицу A'_{kk}; ее определитель M' называется *минором* k-го порядка. Если при этом $i_1 = j_1, i_2 = j_2, \dots, i_k = j_k$, то минор называется *главным*. Если вычеркнуть строки с номерами $i_1, \dots, i_k$ и столбцы с номерами $j_1, \dots, j_k$, то получается новая квадратная матрица $A''_{n-k,\,n-k}$; ее определитель M'' называется *дополнительным минором* для минора M'.

Величина $(-1)^{i_1+i_2+\dots+i_k+j_1+j_2+\dots+j_k} \cdot M''$ называется *алгебраическим дополнением* для минора M'. Имеет место важная теорема:

Теорема Лапласа. *Пусть в матрице $A = A_{nn}$ произвольно выбраны k строк (или k столбцов), $k \leqslant n$. Тогда сумма произве-*

дений всех миноров k-го порядка, содержащихся в выбранных строках (или столбцах), на их алгебраические дополнения равна $\det(A)$.

Если при помощи матрицы $A_{nn} = A = \|a_{rs}\|$ составим новую матрицу $A^* = \|a^*_{rs}\|$, где $a^*_{rs} = D^{(s,r)}$ (алгебраическое дополнение элементa a_{sr}), называемую *взаимной* к A матрицей, то

$$AA^* = A^*A = (\det(A)) \cdot E \tag{1.2.11}$$

Если $\det(A) \neq 0$, то матрица A называется *неособенной*. Для неособенной матрицы A можно составить матрицу

$$B = \frac{1}{\det(A)} A^*. \tag{1.2.12}$$

В силу (1.2.11) имеем:

$$AB = BA = E.$$

Естественно B называть матрицей, *обратной* к A, и обозначать A^{-1}. Имеем $AA^{-1} = A^{-1}A = E$; далее,

$$(AB)^{-1} = B^{-1}A^{-1}, \tag{1.2.13}$$

причем последнее правило распространяется на любое число неособенных сомножителей.

Матричная символика и аппарат позволяют весьма просто записывать решение линейных уравнений и изучать их свойства, и в этом заключается их ценность, в частности для метода наименьших квадратов. На основании сказанного выше можно, например, систему уравнений (1.2.1) переписать в виде

$$AX = B, \tag{1.2.14}$$

где

$$A = A_{mn} = \|a_{rs}\|; \quad X = X_{n1} = \begin{Vmatrix} x_1 \\ \vdots \\ x_n \end{Vmatrix},$$

$$B = B_{m1} = \begin{Vmatrix} b_1 \\ \vdots \\ b_m \end{Vmatrix}.$$

Если $m = n$, так что $A = A_{nn}$, и $\det(A) \neq 0$, то существует матрица A^{-1}; умножая на нее (1.2.14) слева, находим единственное решение системы

$$A^{-1}(AX) = X = A^{-1} \cdot B. \tag{1.2.15}$$

Если $Z = Z_{1n} = \|z_1, \ldots, z_n\|$, то та же система уравнений

$$\begin{gathered} a_{11}z_1 + \ldots + a_{1n}z_n = b_1 \\ \cdots\cdots\cdots\cdots\cdots \\ a_{m1}z_1 + \ldots + a_{mn}z_n = b_m \end{gathered}$$

может быть записана в виде

$$ZA^T = B^T.$$

Умножая справа на $(A^T)^{-1}$, найдем

$$Z = B^T (A^T)^{-1}. \tag{1.2.16}$$

Но в силу единственности решения нашей системы матрица Z должна изображать ту же систему чисел, что и X, только записанную в строку вместо столбца, так что должно быть

$$Z = X^T = (A^{-1}B)^T = B^T (A^{-1})^T.$$

Это в самом деле выполняется в силу легко проверяемого соотношения:

$$(A^{-1})^T = (A^T)^{-1}. \tag{1.2.17}$$

Перейдем теперь к важному понятию ранга матрицы $A = A_{mn}$. *Рангом матрицы* A называется наивысший порядок отличных от нуля миноров, образованных из этой матрицы. Это число будем записывать в виде: ранг (A).

Пусть имеется матрица

$$A = \left\| \begin{matrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{matrix} \right\|.$$

Столбцы этой матрицы образуют n векторов $\left\| \begin{matrix} a_{1s} \\ \vdots \\ a_{ms} \end{matrix} \right\|$ $(s = 1, 2, \ldots, n)$, а строки — m векторов $\|a_{r1}, a_{r2}, \ldots, a_{rn}\|$ $(r = 1, 2, \ldots, m)$.

Имеет место важная теорема:

Максимальное число линейно независимых столбцов матрицы, рассматриваемых как векторы, равно максимальному числу линейно независимых строк ее и равно рангу матрицы.

Иначе это можно высказать так: n столбцов матрицы A образуют линейное векторное пространство размерности, равной рангу (A), и справедливо то же утверждение относительно m ее строк.

Далее, для нас будет важна следующая теорема.

Ранг произведения нескольких матриц не превосходит минимального из рангов отдельных сомножителей.

Если $A_1, A_2, \ldots, A_k$ — матрицы, которые можно перемножать, то это утверждение можно записать в виде

$$\text{ранг}\,(A_1A_2 \ldots A_k) \leqslant \min(\text{ранг}\,(A_1), \ldots, \text{ранг}\,(A_k)). \tag{1.2.18}$$

§ 3. Некоторые теоремы об определителях. Определитель Грама *)

Для нас будут важны некоторые теоремы об определителях. Часть из них хорошо известна и изложена в книгах по высшей алгебре, указанных в § 1; одна из этих теорем известна, но носит несколько специальный характер, и мы приведем ее доказательство.

Если $A_1, A_2, \ldots, A_k$ — квадратные матрицы типа $n \times n$, то

$$\det(A_1 A_2 \ldots A_k) = \det(A_1) \cdot \det(A_2) \ldots \det(A_k). \tag{1.3.1}$$

Пусть имеется система декартовых координат в n-мерном пространстве.

Рассмотрим n векторов $(\xi_{11}, \ldots, \xi_{1n}), \ldots, (\xi_{n1}, \ldots, \xi_{nn})$ и построим матрицу

$$\Xi = \left\| \begin{matrix} \xi_{11} & \ldots & \xi_{1n} \\ \cdot & \cdot\;\cdot & \cdot \\ \xi_{n1} & \ldots & \xi_{nn} \end{matrix} \right\|.$$

Тогда объем параллелепипеда, построенного на этих векторах, будет равен

$$|\det(\Xi)|.$$

В частности, для линейной зависимости наших n векторов необходимо и достаточно, чтобы $\det(\Xi) = 0$, что ясно и геометрически. Это обстоятельство важно при вычислении кратных интегралов.

Пусть имеется n бесконечно малых векторов в n-мерной декартовой системе координат:

$$(dx_1, 0, \ldots, 0); (0, dx_2, 0, \ldots, 0); \ldots; (0, \ldots, 0, dx_n), \tag{1.3.2}$$

причем $dx_i > 0$ $(i = 1, 2, \ldots, n)$. Объем параллелепипеда, построенного на них, будет равен, очевидно, $dx_1 \ldots dx_n$. Если вводится линейное неособенное преобразование

$$\left\| \begin{matrix} dy_1 \\ \vdots \\ dy_n \end{matrix} \right\| = A \cdot \left\| \begin{matrix} dx_1 \\ \vdots \\ dx_n \end{matrix} \right\|, \quad A = A_{nn}, \ \det(A) \neq 0,$$

то объем параллелепипеда, построенного на новых бесконечно малых векторах, в которые переходят старые векторы (1.3.2), будет равен

$$|\det(A)|\, dx_1 \ldots dx_n. \tag{1.3.3}$$

Пусть дано n векторов:

$$x_j = (x_{1j}, \ldots, x_{Nj}); \quad j = 1, 2, \ldots, n; \ N \geqslant n. \tag{1.3.4}$$

*) И. Грам (1850—1916) — датский математик.

Составим всевозможные скалярные произведения этих векторов

$$[x_i x_j] = \sum_{r=1}^{n} x_{ri} x_{rj}; \quad i, j = 1, \ldots, n, \tag{1.3.5}$$

и образуем определитель

$$G(x_1, \ldots, x_n) = \begin{vmatrix} [x_1 x_1] & [x_1 x_2] & \ldots & [x_1 x_n] \\ [x_2 x_1] & [x_2 x_2] & \ldots & [x_2 x_n] \\ \cdot & \cdot & \cdot & \cdot \\ [x_n x_1] & [x_n x_2] & \ldots & [x_n x_n] \end{vmatrix}. \tag{1.3.6}$$

Этот определитель называется *определителем Грама* системы векторов $x_1, \ldots, x_n$. Если ввести матрицу

$$X_{Nn} = X = \begin{Vmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ x_{N1} & x_{N2} & \ldots & x_{Nn} \end{Vmatrix}, \tag{1.3.7}$$

то легко проверить, что

$$X^T X = \begin{Vmatrix} [x_1 x_1] & [x_1 x_2] & \ldots & [x_1 x_n] \\ [x_2 x_1] & [x_2 x_2] & \ldots & [x_2 x_n] \\ \cdot & \cdot & \cdot & \cdot \\ [x_n x_1] & [x_n x_2] & \ldots & [x_n x_n] \end{Vmatrix}. \tag{1.3.8}$$

Таким образом, имеем следующее выражение для определителя Грама:

$$G(x_1, \ldots, x_n) = \det(X^T X). \tag{1.3.9}$$

Если $N = n$, то матрицы X и X^T квадратные, и в силу известного равенства $\det(X^T) = \det(X)$ мы получим из (1.3.9)

$$G(x_1, \ldots, x_n) = [\det(X)]^2. \tag{1.3.10}$$

Таким образом, для случая $n=N$ определитель Грама $G(x_1, \ldots, x_n)$ равен квадрату объема параллелепипеда, построенного на векторах $x_1, \ldots, x_n$, заданных в декартовых координатах n-мерного пространства.

Для метода наименьших квадратов важно обобщение этой теоремы на случай $N > n$.

Теорема 1.3.1. *При $N \geqslant n$ имеем равенство*

$$G(x_1, \ldots, x_n) = \sum_k \delta_k^2, \tag{1.3.11}$$

где δ_k пробегает все C_N^n миноров n-го порядка, содержащихся в матрице $X = X_{Nn}$.

Доказательство. Составим вспомогательную квадратную матрицу

$$Y_{N+n,\,N+n}=Y=\left\| \begin{array}{cccc:c} 1 & 0 & \dots & 0 & \\ 0 & 1 & \dots & 0 & X_{Nn} \\ \cdot & \cdot & \cdot & \cdot & \\ 0 & 0 & \dots & 1 & \\ \hdashline & (X_{Nn})^T & & & 0_{nn} \end{array} \right\|. \tag{1.3.12}$$

Будем вычислять величину $\det(Y)$ двумя способами. Первый способ будет состоять в том, что мы будем, пользуясь известными правилами вычисления определителей, умножать столбцы матрицы

$$\left\| \begin{array}{cccc} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 1 \\ \hdashline & (X_{Nn})^T & & \end{array} \right\| \tag{1.3.13}$$

соответственно: 1-й столбец на $-x_{11}, -x_{12}, \dots, -x_{1n}$; 2-й столбец на $-x_{21}, -x_{22}, \dots, -x_{2n}, \dots$, N-й столбец на $-x_{N1}, -x_{N2}, \dots$ $\dots, -x_{Nn}$ и прибавлять ко всем столбцам матрицы

$$\left\| \begin{array}{c} X_{Nn} \\ \hdashline 0_{nn} \end{array} \right\|.$$

Внимательно проследив за результатами этих операций, убедимся, что получится матрица

$$Y_1=\left\| \begin{array}{cccc:c} 1 & 0 & \dots & 0 & \\ 0 & 1 & \dots & 0 & 0_{Nn} \\ 0 & 0 & \dots & 1 & \\ \hdashline & (X_{Nn})^T & & & -(X_{Nn})^T X_{Nn} \end{array} \right\| \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\0\\0\end{array}}\right\} N \\ \} \; n \end{array}.$$

При этом

$$\det(Y_1)=\det(Y)=(-1)^n\det(X^T\cdot X). \tag{1.3.14}$$

С другой стороны, этот же детерминант можно вычислить при помощи теоремы Лапласа. Для этого будем разлагать $\det(Y)$ по минорам n последних строк. Пусть мы взяли минор в этих строках со столбцами $j_1, j_2, \dots, j_n$ $(j_1<j_2< \dots <j_n)$. Мы можем считать $j_n \leqslant N$, ибо иначе у нас получится минор с нулевым столбцом из 0_{nn}, и он будет равен нулю. Поэтому мы должны вычеркивать столбцы

$j_1, j_2, \ldots, j_n$ лишь в матрице (1.3.13), так что из (1.3.12) получится квадратная матрица типа $N \times N$. Она будет иметь вид:

$$Y_2 = \| A_{N,\, N-n} \vdots X_{Nn} \|.$$

При этом в матрице $A_{N,N-n}$ строки с номерами $j_1, j_2, \ldots, j_n$ будут состоять сплошь из нулей, а остальные строки будут иметь по одной единице и $(N-n-1)$ нулей. Нас будет интересовать $\det(Y_2)$ как минор, дополнительный к нашему минору. Для его вычисления будем умножать столбцы матрицы $A_{N,N-n}$, в каждой из которых стоит одна единица и $(N-1)$ нулей, на соответствующие элементы столбцов X_{Nn} и вычитать эти столбцы из X_{Nn}. Тогда вместо X_{Nn} получится матрица X'_{Nn}, которая будет иметь строки из сплошных нулей, причем по построению это будут строки с номерами, отличными от $j_1, j_2, \ldots, j_n$. Получим новую матрицу

$$Y_3 = \left\| A_{N,\, N-n} \vdots X'_{Nn} \right\|,$$

где нулевые строки $A_{N,N-n}$ будут иметь номера $j_1, j_2, \ldots, j_n$ а нулевые строки X'_{Nn} — все отличные от них номера. При этом $\det(Y_3) = \det(Y_2)$.

Для вычисления $\det(Y_3)$ снова применим теорему Лапласа, разлагая по минорам последних n столбцов матрицы Y_3. Но из этих миноров отличаться от нуля будет лишь один — со строками $j_1, j_2, \ldots, j_n$. При вычеркивании этих строк в матрице Y_3, матрица $A_{N,N-n}$ превращается в единичную матрицу $E_{N-n,\,N-n}$. Алгебраическое дополнение нашего минора равно

$$(-1)^{j_1+j_2+\ldots+j_n+N-n+1+N-n+2+\ldots+N} \det(E_{N-n,\,N-n}) = \\ = (-1)^{j_1+\ldots+j_n+N-n+1+\ldots+N}.$$

Сам же минор, очевидно, равен исходному минору порядка n матрицы $(X_{Nn})^T$, который мы обозначим δ_k. Произведение δ_k на его алгебраическое дополнение будет равно

$$\delta_k(-1)^{N+1+\ldots+N+n+j_1+\ldots+j_n+j_1+\ldots+j_n+N-n+1+\ldots+N} = \\ = \delta_k(-1)^{n^2} = \delta_k(-1)^n.$$

Поэтому

$$\det(Y) = (-1)^n \sum_k \delta_k^2. \qquad (1.3.15)$$

Сравнивая с (1.3.14), получаем утверждение теоремы 1.3.1.

Нам понадобится также небольшое обобщение этой теоремы. Пусть

$$P_{NN} = P = \begin{Vmatrix} p_1 & 0 & \ldots & 0 \\ 0 & p_2 & \ldots & 0 \\ . & . & . & . \\ 0 & 0 & \ldots & p_N \end{Vmatrix} \qquad (1.3.16)$$

диагональная матрица, причем $p_i > 0 \quad (i = 1, 2, \ldots, N)$.

Обозначим через $\delta_{j_1 \ldots j_n}$ минор, полученный из n столбцов и n строк с номерами $j_1, j_2, \ldots, j_n$ матрицы X_{Nn}.

Теорема 1.3.2. *Имеем*

$$\det(X^T P X) = \sum p_{j_1} p_{j_2} \ldots p_{j_n} \delta^2_{j_1 \ldots j_n}, \tag{1.3.17}$$

где сумма распространяется на все возможные миноры указанного типа.

Доказательство. Введем матрицу $P^{\frac{1}{2}}$ [см. формулу (1.2.9)] и положим $Y = P^{\frac{1}{2}} X$. Тогда $Y^T = X^T \left(P^{\frac{1}{2}}\right)^T = X^T P^{\frac{1}{2}}$, так как $\left(P^{\frac{1}{2}}\right)^T = P^{\frac{1}{2}}$. Далее,

$$X^T P X = Y^T Y.$$

Поэтому применима теорема 1.3.1 с заменой X на Y. Но матрица Y отличается от X тем, что у нее r-я строка помножается на $p_r^{\frac{1}{2}}$, так что (1.3.17) непосредственно следует из (1.3.11).

Нам нужны еще некоторые сведения об инвариантах квадратных матриц. Для теории метода наименьших квадратов существенно лишь понятие следа $\mathrm{Sp}(A)$ квадратной матрицы A *).

Следом матрицы $A_{nn} = A = \|a_{rs}\|$ называют сумму ее диагональных элементов:

$$\mathrm{Sp}(A) = \sum_{r=1}^{n} a_{rr}.$$

Имеет место теорема об инвариантности $\mathrm{Sp}(A)$: если $F = F_{nn}$ неособенная матрица и $A' = FAF^{-1}$, то

$$\mathrm{Sp}(A') = \mathrm{Sp}(A). \tag{1.3.18}$$

Эта теорема просто доказывается рассмотрением „λ-матрицы" $A - \lambda E$.

§ 4. Симметрические матрицы. Квадратичные формы. Ортогональные матрицы

Квадратная матрица называется *симметрической*, если она симметрична относительно главной диагонали, т. е. если $A^T = A$, или $\{A\}_{rs} = \{A\}_{sr}$ для всех возможных r и s.

Если $A = A_{mn}$ — любая матрица, то матрицы $C_1 = AA^T$ и $C_2 = A^T A$ всегда будут симметрическими.

В самом деле,

$$C_1^T = (AA^T)^T = AA^T = C_1;$$

*) Обозначение $\mathrm{Sp}(A)$ происходит от немецкого слова Spur — след.

точно так же $C_2^T = C_2$, что и доказывает симметричность матриц C_1 и C_2.

Квадратичной формой называется функция вектора $X_{n1} = X = \begin{Vmatrix} x_1 \\ \vdots \\ x_n \end{Vmatrix}$, имеющая вид

$$Q = \sum_{r=1}^{n} \sum_{s=1}^{n} a_{rs} x_r x_s,$$

где $a_{rs} = a_{sr}$.

Вводя симметрическую матрицу $A = \| a_{rs} \|$, можно записать Q, как нетрудно проверить, следующим образом:

$$Q = X^T A X. \tag{1.4.1}$$

A называется матрицей квадратичной формы Q.

Запись (1.4.1) весьма удобна в теории метода наименьших квадратов, где квадратичные формы играют основную роль.

Если A — диагональная матрица, то соответствующую ей квадратичную форму будем называть *диагональной*. В частном случае диагональной матрицы, когда $A = E = E_{nn}$, имеем

$$Q = X^T E X = X^T X = x_1^2 + \ldots + x_n^2. \tag{1.4.2}$$

Если вместо вектора $X = X_{n1}$ ввести новый вектор $Y = Y_{n1}$, соответственно линейному преобразованию

$$X_{n1} = G_{nn} Y_{n1}, \tag{1.4.3}$$

где $G_{nn} = G$ — какая-либо квадратная матрица, то из (1.4.1) получим

$$Q = X^T A X = Y^T G^T A G Y = Y^T B Y, \tag{1.4.4}$$

где

$$B = G^T A G. \tag{1.4.5}$$

Очевидно, B тоже симметрическая матрица: $B^T = (G^T A G)^T = G^T A^T G = G^T A G$, так что Q будет квадратичной формой от Y, и соотношения (1.4.4), (1.4.5) доставляют правила преобразования квадратичной формы при линейных подстановках.

В методе наименьших квадратов встречаются в основном положительно определенные и полуопределенные квадратичные формы.

Форма $Q = X^T A X$ называется *полуопределенной*, или *неотрицательной*, если $Q \geqslant 0$ при всех X. Форма Q называется *положительно определенной*, если $Q > 0$ для любого не нулевого вектора X. Если форма Q может принимать значения разных знаков, она называется *неопределенной*.

Весьма важной определенной формой является форма $Q = X^T X = x_1^2 + \ldots + x_n^2$. Геометрически ее можно истолковать, как расстояние точки $(x_1, \ldots, x_n)$ от начала координат в n-мерном пространстве. Если над X сделать подстановку (1.4.3), то Q перейдет в форму $Q = Y^T BY$, где $B = G^T G$. Если матрица G такова, что

$$B = G^T G = E, \tag{1.4.6}$$

то Q переходит в $Y^T Y = y_1^2 + y_2^2 + \ldots + y_n^2$. Такой вид преобразований (1.4.3) описывает вращения и отражения в n-мерном пространстве, или ортогональные преобразования пространства. Если $G = \|g_{rs}\|$, то подробная запись условия (1.4.6) будет

$$\sum_{s=1}^{n} g_{rs}^2 = 1 \quad (r = 1, 2, \ldots, n); \qquad \sum_{s=1}^{n} g_{rs} g_{r_1 s} = 0 \qquad (r \neq r_1).$$

Из $G^T G = E$ следует, что $G^T = G^{-1}$; $GG^T = E$. Матрицы G с условием $GG^T = G^T G = E$ называются *ортогональными*.

Пример. Матрица

$$G = \left\| \begin{matrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{matrix} \right\|$$

в преобразовании $Y = Y_{21} = GX_{21}$ описывает вращение на угол α. Легко подсчитать, что

$$GG^T = G^T G = \left\| \begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix} \right\|.$$

Весьма важным свойством ортогональных преобразований является следующее: всякую квадратичную форму $Q = X^T AX$ путем ортогонального преобразования вектора X можно привести к диагональному виду. Точнее, для всякой симметрической матрицы $A_{nn} = A$ найдется ортогональная матрица F такая, что $B = F^T AF$ есть диагональная матрица.

Если симметрическая матрица A — неособенная ($\det(A) \neq 0$), то A имеет обратную матрицу A^{-1}, также симметрическую в силу соотношения $(A^{-1})^T = (A^T)^{-1} = A^{-1}$. Можно составить квадратичную форму $\tilde{Q} = X^T A^{-1} X$, которую называют *обратной* к форме $Q = X^T AX$. Если симметрическая матрица A особенная ($\det(A) = 0$), то соответствующая ей форма $Q = X^T AX$ называется *вырожденной*. Вырожденную квадратичную форму посредством некоторого ортогонального преобразования (1.4.3) можно привести к диагональному виду $Q = \varkappa_1 y_1^2 + \ldots + \varkappa_r y_r^2$, где $r < n$ и притом r таково, что неособенные линейные преобразования $Y = FX$ не могут привести Q к диагональному виду с меньшим числом слагаемых $\varkappa_i y_i^2$. Оказывается, такое число равно рангу матрицы A; ранг $(A) = r$. Оно называется также *рангом квадратичной формы* Q.

Пусть Q — квадратичная форма с матрицей $A_{nn} = A = \|a_{rs}\|$. Условие, необходимое и достаточное для полуопределенности квадратичной формы Q, состоит в неотрицательности всех главных миноров ее матрицы коэффициентов:

$$\begin{vmatrix} a_{i_1 i_1} & \dots & a_{i_1 i_k} \\ \cdot & \cdot & \cdot \\ a_{i_k i_1} & \dots & a_{i_k i_k} \end{vmatrix} \geqslant 0 \quad (1 \leqslant i_1 < i_2 < \dots < i_k \leqslant n, \quad k = 1, 2, \dots, n).$$

Если форма $Q = X^T A X$ полуопределенная, так что всегда $Q \geqslant 0$, то будем записывать символически: $A \geqslant 0$.

Условие, необходимое и достаточное для положительной определенности формы $Q = X^T A X$, таково:

$$\begin{vmatrix} a_{11} & \dots & a_{1k} \\ \cdot & \cdot & \cdot \\ a_{k1} & \dots & a_{kk} \end{vmatrix} > 0 \qquad (k = 1, 2, \dots, n). \tag{1.4.7}$$

В этом случае будем записывать символически: $A > 0$.

Если $Q \geqslant 0$, то $a_{ii} \geqslant 0$ $(i = 1, 2, \dots, n)$. В самом деле, если бы было $a_{i_0 i_0} < 0$, то, полагая $x_{i_0} = 1$; $x_i = 0$ $(i \neq i_0)$, мы имели бы $Q = X^T A X = a_{i_0 i_0} < 0$, что невозможно.

Далее отметим еще одно нужное нам свойство: если $Q = X^T A X \geqslant 0$, т. е. если Q — полуопределенная форма, и $\det(A) > 0$, то Q — положительно определенная форма.

В самом деле, положим $X = FY$, где F — ортогональная матрица, такая, что

$$F^T A F = \begin{Vmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & d_n \end{Vmatrix}.$$

Имеем, очевидно, $d_i \geqslant 0$. Но $\det(A) = \det(F^T A F) = d_1 d_2 \dots d_n > 0$, так что $d_i > 0$ $(i = 1, 2, \dots, n)$. Но тогда, очевидно, Q должна быть положительно определенной формой.

ГЛАВА II

НЕОБХОДИМЫЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ВЕРОЯТНОСТЕЙ

§ 1. Случайные величины

Предполагаются известными теоретические понятия случайного события и случайной величины одномерной и n-мерной (случайного вектора) и отношение этих понятий к действительности, а также понятие статистической независимости и элементарные теоремы теории вероятностей.

Здесь будут вкратце приведены дальнейшие нужные сведения из теории вероятностей. Большинство из них хорошо известно; доказательства можно найти, например, в книге Б. В. Гнеденко [11]. Другие, известные, но более специальные факты будут приведены с доказательствами.

Общее понятие случайной величины требует введения интеграла Стилтьеса, и мы не будем им пользоваться. Случайные величины, которые будут нас интересовать в основном, — погрешности измерений или округлений; поэтому мы будем рассматривать случайные величины лишь двух типов: дискретные случайные величины и величины, обладающие непрерывной плотностью распределения (преимущественно последние).

Дискретная случайная величина X задается рядом своих значений x_k $(k=1, 2, \ldots)$ и соответствующими ему вероятностями $p_k=P(X=x_k)$ [в дальнейшем запись $P(\ldots)$ будет означать вероятность случайного события, стоящего в скобках]. Приведем некоторые примеры дискретных величин.

Пример 1. Отмечается последняя цифра при округлении длинного ряда наблюдений одним и тем же наблюдателем. В первом приближении ей можно сопоставить величину X, принимающую значения 0, 1, ..., 9, с вероятностями $p_i=\frac{1}{10}$ $(i=1, 2, \ldots, 10)$.

Более подробные сведения показывают, что имеется ряд обстоятельств, заставляющих приписывать различные вероятности разным цифрам, например, ввиду бессознательной склонности наблюдателей к некоторым цифрам (0,5).

В табл. 5. приведены данные для 4 наблюдателей, сделавших около 1000 наблюдений каждый (Кендалл [20], т. 1, стр. 188).

Таблица 5

Распределение последней цифры при измерениях 4 наблюдателями A, B, C, D

Последняя цифра	Число появлений в 1000 измерений			
	A	B	C	D
0	158	122	251	358
1	97	98	37	49
2	125	98	80	90
3	79	90	72	63
4	76	100	55	37
5	71	112	222	211
6	90	98	71	62
7	56	99	75	70
8	126	101	72	44
9	129	81	65	16
Сумма	1001	999	1000	1000

Пример 2. (*Мультиномиальное распределение*). Пусть производятся n независимых испытаний, в каждом из которых некоторая частица (или иной объект) может попасть в любую из имеющихся k зон $Z_1, Z_2, \ldots, Z_k$. При этом в каждом испытании вероятность попадания частицы в зону Z_i есть p_i и $p_1+p_2+\ldots+p_k=1$. После n испытаний в зонах $Z_1, Z_2, \ldots, Z_k$ окажутся некоторые числа $m_1, m_2, \ldots, m_k$ частиц, в сумме составляющие число n. Случайный вектор $X=\|m_1, m_2, \ldots, m_k\|$ описывает такое распределение частиц по зонам. Если при этом $n_1, n_2, \ldots, n_k$ — заданные числа, в сумме составляющие n, то

$$P(X=\|n_1, n_2, \ldots, n_k\|)=\frac{n!}{n_1!\,n_2!\ldots n_k!}\,p_1^{n_1}\ldots p_k^{n_k}.$$

В частном случае $k=2$ предположим, например, что Z_1 — определенный круг стрелковой мишени, а Z_2 — все пространство вне него; $p_1=p$ — вероятность попадания в круг Z_1, $p_2=1-p$. Результат n выстрелов вместо вектора $X=\|m_1, m_2\|$ достаточно описывать одномерной величиной X, задающей m_1 $(m_2=n-m_1)$. При этом

$$P(X=n_1)=\frac{n!}{n_1!\,n_2!}\,p^{n_1}(1-p)^{n-n_1}.$$

Перейдем к примерам случайных величин и векторов с непрерывной плотностью распределения.

Рассмотрим n-мерный случайный вектор $X=\|x_1, \ldots, x_n\|$, заданный во всем n-мерном пространстве. Если имеется n-мерный параллелепипед $I_\Delta=[\alpha_i, \alpha_i+\Delta_i]$ $(i=1, 2, \ldots, n)$, где $\|\alpha_1, \ldots, \alpha_n\|$ —

любая фиксированная точка, а Δ_i — любые положительные числа то будем считать, что определена вероятность $P(X \in I_\Delta)$ *) и притом существует предел

$$\lim_{\Delta_1, \ldots \Delta_n \to 0} \frac{P(X \in I_\Delta)}{\Delta_1 \Delta_2 \ldots \Delta_n} = f(\alpha_1, \alpha_2, \ldots, \alpha_n),$$

где $f(\alpha_1, \alpha_2, \ldots, \alpha_n)$ — непрерывная плотность распределения X.

Если Ω — какая-либо область пространства, допускающая интегрирование по ней, то

$$P(X \in \Omega) = \int \ldots \int_{\Omega} f(x_1, x_2, \ldots, x_n)\, dx_1 \ldots dx_n. \qquad (2.1.1)$$

При Ω, совпадающем со всем n-мерным пространством R_n, интеграл (2.1.1) равен единице.

Пример 3. *Нормальная* случайная величина X задается плотностью распределения

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{(x-a)^2}{2\sigma^2}, \qquad (2.1.2)$$

(Запись $\exp x$ означает e^x; она удобна в случае, когда в экспоненте стоит длинное выражение). Совокупность нормальных законов (2.1.2) будем обозначать $N(a, \sigma)$ и писать: $X \in N(a, \sigma)$.

Отметим некоторые количественные факты, касающиеся нормального закона (2.1.2). Для нормальной величины $X \in N(a, \sigma)$ имеем:

$$P\{|X-a| \leqslant 1{,}96\sigma\} \approx 0{,}95, \qquad (2.1.3)$$

$$P\{|X-a| \geqslant 3\sigma\} \approx 0{,}0027. \qquad (2.1.4)$$

Последнее неравенство часто называется „правилом 3σ“.

Весьма многие количественные показатели, встречающиеся в практике, приближенно подчиняются нормальному распределению, т. е. при достаточно большом числе независимых наблюдений над этими показателями относительная частота их попадания в заданный интервал приближенно равна вероятности попадания случайной нормальной величины $X \in N(a, \sigma)$ в тот же интервал при надлежащим образом подобранных параметрах a и σ. Такими показателями могут быть погрешности наблюдений одного наблюдателя, линейные размеры каких-либо деталей при массовом производстве, проекции на какую-либо ось отрезков от центра стрелковой мишени до пробоины при повторной стрельбе одного и того же стрелка при постоянных установках прицела.

*) Знак $\in$ означает „принадлежит“.

Применимость нормального закона к описанию поведения погрешностей наблюдений хорошо видна из табл. 6 (Н. И. Идельсон [18]).

Здесь N_i — наблюденное число уклонений λ значений прямого восхождения Сириуса от их среднего арифметического, не превосходящих по абсолютной величине числа λ_i. Общее число наблюдений $N=470$; N'_i — выравненное по нормальному закону с параметрами (0; 0,391) число таких уклонений, вычисленное по формуле

$$N'_i = NF(\lambda_i),$$

где

$$F(\lambda_i) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\lambda_i}^{\lambda_i} e^{-\frac{t^2}{2\sigma^2}}\, dt,$$

$$\sigma = 0{,}391.$$

Таблица 6

λ_i	N_i	N'_i
0″,1	94	95
0,2	182	184
0,3	260	262
0,4	318	327
0,5	369	376
0,6	405	411
0,7	431	436
0,8	445	451
0,9	455	460
1,0	462	465
1,4	470	470

Из табл. 6 видно, что нормальный закон N(0; 0,391) довольно хорошо описывает погрешности наблюдений; становится понятной фундаментальная роль этого закона в теории метода наименьших квадратов. В дальнейшем (в § 7 этой главы и в гл. III) будут приведены некоторые теоретические обоснования роли нормального закона в теории погрешностей наблюдений.

В дальнейшем нам понадобится также понятие случайной матрицы. *Случайная матрица* $X_{mn} = X = \|x_{rs}\|$ есть случайный вектор с mn компонентами, собранными в таблицу в виде матрицы X. Сложение и умножение случайных матриц определяется так же, как и для обычных матриц; разумеется, вопрос о распределении результирующей матрицы может быть решен только при знании совокупного распределения матриц, из которых она построена.

Применение таких матриц оказывается иногда полезным и сокращает расчеты и их запись.

Напомним свойства моментов случайных величин.

Математическое ожидание (называемое иногда также центром или средним значением) случайной величины X будем обозначать $E(X)$, а также М. О. X. Если X имеет плотность вероятности $\varphi(x)$ $(x \in (-\infty, \infty))$, то

$$m = E(X) = \int_{-\infty}^{\infty} x\varphi(x)\, dx. \tag{2.1.5}$$

При этом предполагается, что интеграл (2.1.5) абсолютно сходится.

Дисперсию X, если она существует, будем обозначать $D(X)$. По определению,

$$D(X)=E(X-m)^2=\int_{-\infty}^{\infty}(x-m)^2\varphi(x)\,dx. \tag{2.1.6}$$

Здесь $m=E(X)$, как в (2.1.5). *Штандартом* $\sigma(X)$ называется величина

$$\sigma(X)=\sqrt{D(X)}. \tag{2.1.7}$$

Для нормальной плотности $f(x)$ при $X\in N(a,\sigma)$ результаты подстановки $f(x)$ вместо $\varphi(x)$ в (2.1.5), (2.1.6), (2.1.7) дают:

$$m=E(X)=a;$$

$$D(X)=\sigma^2;\quad \sigma(X)=\sigma. \tag{2.1.8}$$

Для n-мерного случайного вектора $X=(X_1, X_2, \ldots, X_n)$, имеющего плотность распределения $f(x_1, x_2, \ldots, x_n)$ [см. (2.1.1)], также вводится понятие моментов. *Первыми моментами* будут математические ожидания компонент X:

$$a_i=E(X_i)=\int_{R_n}\!\!\ldots\int x_i f(x_1, \ldots, x_n)\,dx_1\ldots dx_n \tag{2.1.9}$$

(R_n — все n-мерное пространство точек $(x_1, \ldots, x_n)$).

Вторые моменты b_{rs} определяются следующим образом:

$$b_{rs}=\int_{R_n}\!\!\ldots\int (x_r-a_r)(x_s-a_s)f(x_1, \ldots x_n)\,dx_1\ldots dx_n$$

$$(r, s=1, 2, \ldots, n) \tag{2.1.10}$$

(если написанные интегралы абсолютно сходятся). Мы видим, что $b_{rs}=b_{sr}$ и что $b_{rr}\geqslant 0$.

Симметрическая матрица

$$B_{nn}=B_X=\|b_{rs}\| \tag{2.1.11}$$

называется *корреляционной матрицей* случайного вектора X. Она имеет удобное выражение с помощью случайных матриц. Условимся называть *математическим ожиданием случайной матрицы* $Y=\|y_{rs}\|$ матрицу $E(Y)=\|E(y_{rs})\|$, составленную из математических ожиданий элементов y_{rs}. Полагая

$$X=X_{n1}=\left\|\begin{matrix}x_1\\ \vdots\\ x_n\end{matrix}\right\|, \tag{2.1.12}$$

можем написать, что

$$E(X)=\begin{Vmatrix} a_1 \\ \vdots \\ a_n \end{Vmatrix}=A \tag{2.1.13}$$

и

$$B=B_X=E\left((X-A)(X-A)^T\right). \tag{2.1.14}$$

В самом деле, имеем

$$(X-A)(X-A)^T=\|(x_r-a_r)(x_s-a_s)\|. \tag{2.1.15}$$

Введем вспомогательный вектор $Z_{n1}=Z=\begin{Vmatrix} z_1 \\ \vdots \\ z_n \end{Vmatrix}$ и составим „корреляционную квадратичную форму“

$$\sum_{r=1}^{n}\sum_{s=1}^{n} b_{rs}z_rz_s=Z^TBZ.$$

Эта форма будет положительно определенной или полуопределенной, так как $Z^TBZ\geqslant 0$ для всех Z. В самом деле, для любого Z в силу (2.1.10) имеем

$$Z^TBZ=\sum_{r=1}^{n}\sum_{s=1}^{n} b_{rs}z_rz_s=$$

$$=\int_{R_n}\cdots\int f(x_1,\ldots,x_n)\left[\sum_{r=1}^{n}\sum_{s=1}^{n} z_r(x_r-a_r)z_s(x_s-a_s)\right]dx_1\ldots dx_n=$$

$$=\int_{R_n}\cdots\int f(x_1,\ldots,x_n)\left(\sum_{i=1}^{n} z_i(x_i-a_i)\right)^2 dx_1\,dx_2\ldots dx_n\geqslant 0$$

в силу того, что $f(x_1,\ldots,x_n)\geqslant 0$.

Если B неособенная, так что $\det(B)\neq 0$, то корреляционная квадратичная форма Z^TBZ будет положительно определенной; если же $\det(B)=0$, то она будет полуопределенной. Это следует из замечаний в конце § 4 гл. I.

Приведем еще важное неравенство, по существу являющееся видоизменением неравенства Буняковского — Коши.

Если $X=\begin{Vmatrix} x_1 \\ x_2 \end{Vmatrix}$ — двумерный случайный вектор, то

$$\left(E(x_1x_2)\right)^2\leqslant E\,|x_1x_2|^2\leqslant\left(Ex_1^2\right)\left(Ex_2^2\right). \tag{2.1.16}$$

Важными свойствами моментов $E(X)$ и $D(X)$ являются следующие свойства. Для независимых величин X и Y

$$E(XY)=E(X)E(Y), \tag{2.1.17}$$

если все написанные выражения существуют. Для попарно независимых величин X_1, X_2, ..., X_n

$$D(X_1+\ldots+X_n)=D(X_1)+\ldots+D(X_n), \tag{2.1.18}$$

если величины в правой части существуют.

Весьма важно также следующее неравенство П. Л. Чебышева, связывающее $D(X)$ и $E(X)$: для любого $t>0$

$$P\{|X-E(X)|\geqslant t\}\leqslant\frac{D(X)}{t^2}. \tag{2.1.19}$$

§ 2. Нормальный случайный вектор

Векторы, приближенно нормальные, часто встречаются в практике. Таковы, например, векторы, соединяющие центр стрелковой мишени с пробоинами, результаты многочисленных наблюдений средней скорости ветра в данной точке над поверхностью земли, совокупность погрешностей наблюдений повторных измерений системы n физических величин X_1, X_2, ..., X_n у одного и того же наблюдателя.

Случайный n-мерный вектор $X=(X_1, \ldots, X_n)$ является непосредственным обобщением понятия нормальной случайной величины.

Пусть $Q=Z^T\Lambda Z=\sum\limits_{r=1}^{n}\sum\limits_{s=1}^{n}\lambda_{rs}z_rz_s$ — положительно определенная квадратичная форма от n переменных z_1, ..., z_n с матрицей $\Lambda_{nn}=\Lambda=\|\lambda_{rs}\|$; $A_{n1}=A=\begin{Vmatrix} a_1 \\ \vdots \\ a_n \end{Vmatrix}$ — вектор.

Нормальный n-мерный вектор $X=X_{n1}$, отвечающий заданному вектору средних значений A и положительно определенной матрице Λ, определяется как вектор, имеющий плотность распределения *)

$$f(x_1, \ldots, x_n)=C_0\exp-\frac{1}{2}(X-A)^T\Lambda(X-A), \tag{2.2.1}$$

где C_0 — константа, определяемая из равенства

$$C_0\cdot\int\limits_{R_n}\ldots\int\exp-\frac{1}{2}\left[(X-A)^T\Lambda(X-A)\right]dx_1\ldots dx_n=1. \tag{2.2.2}$$

*) Здесь и во многих других местах мы обозначаем сам случайный вектор X и принимаемые им векторные значения одной буквой, что не должно вести к недоразумениям.

Для того чтобы изучить действие матричного аппарата в вычислениях, посмотрим, как выразить константу C_0 через Λ, а также найдем выражение корреляционной матрицы вектора X через Λ и докажем, что $E(X)=A$ (см., например, С. Рао [42], стр. 51—52).

Теорема 2.2.1. *Имеют место равенства*:

$$C_0=\frac{[\det(\Lambda)]^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}}, \tag{2.2.3}$$

$$E(X)=A, \tag{2.2.4}$$

$$B_X=\Lambda^{-1}, \tag{2.2.5}$$

где B_X — корреляционная матрица вектора X. Таким образом, неособенная матрица B_X обратна Λ.

Доказательство. Сначала докажем (2.2.3). Положим $Y_{n1}=Y=X-A=\begin{Vmatrix} y_1 \\ \vdots \\ y_n \end{Vmatrix}$. Из (2.2.2) находим

$$C_0\int\limits_{R_n}\ldots\int \exp -\frac{1}{2}\,Y^T\Lambda Y\,dy_1\ldots dy_n=1. \tag{2.2.6}$$

Как известно из § 4 гл. I, квадратичную форму $Y^T\Lambda Y$ можно привести к диагональному виду посредством ортогонального преобразования

$$Y=FZ;\quad Z=Z_{n1}=\begin{Vmatrix} z_1 \\ \vdots \\ z_n \end{Vmatrix},\quad F=F_{nn},\quad F^T=F.$$

Получим

$$Y^T\Lambda Y=Z^TF^T\Lambda FZ=Z^TDZ,$$

где

$$D=\begin{Vmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & d_n \end{Vmatrix}$$

диагональная матрица. При этом

$$\det(D)=\det(F^T)\det(\Lambda)\det(F)=\det(\Lambda)>0,$$

так что $d_i>0$ $(i=1,\,2,\,\ldots,\,n)$. Если введем еще преобразование $Z=D^{-\frac{1}{2}}W$, $W=\begin{Vmatrix} w_1 \\ \vdots \\ w_n \end{Vmatrix}$, то в силу $\left(D^{-\frac{1}{2}}\right)^T=D^{-\frac{1}{2}}$ получим

$$Y^T\Lambda Y=Z^T DZ=W^TD^{-\frac{1}{2}}DD^{-\frac{1}{2}}W=W^TW=\sum_{i=1}^{n}w_i^2. \tag{2.2.7}$$

Имеем

$$Y = FZ = FD^{-\frac{1}{2}}W \tag{2.2.8}$$

и

$$\det\left(FD^{-\frac{1}{2}}\right) = \det\left(D^{-\frac{1}{2}}\right) = d_1^{-\frac{1}{2}} \ldots d_n^{-\frac{1}{2}} = [\det(D)]^{-\frac{1}{2}} =$$
$$= [\det(\Lambda)]^{-\frac{1}{2}}. \tag{2.2.9}$$

При вычислении интеграла в (2.2.6) перейдем к переменным $w_1, \ldots, w_n$. Областью интегрирования будет снова пространство $R_n (-\infty < w_i < \infty)$. При переходе к новым переменным, как известно из теории кратных интегралов, в (2.2.6) нужно $dy_1 \ldots dy_n$ заменить на $dw_1 \ldots dw_n$, умноженное на абсолютную величину якобиана $|\det(J)|$, где $J = \left\| \frac{\partial y_r}{\partial w_s} \right\|$. В данном случае имеем

$$Y = FD^{-\frac{1}{2}}W,$$

так что

$$y_r = \sum_{m=1}^{n} \left\{ FD^{-\frac{1}{2}} \right\}_{rm} w_m, \quad \text{и} \quad \frac{\partial y_r}{\partial w_s} = \left\{ FD^{-\frac{1}{2}} \right\}_{rs}.$$

Значит, $J = FD^{-\frac{1}{2}}$ и

$$|\det(J)| = \left|\det\left(FD^{-\frac{1}{2}}\right)\right| = [\det(\Lambda)]^{-\frac{1}{2}}. \tag{2.2.10}$$

Делая замену переменных в интеграле (2.2.6), получим

$$C_0 [\det(\Lambda)]^{-\frac{1}{2}} \int_{R_n} \ldots \int \exp -\frac{1}{2}(W^T W)\, dw_1 \ldots dw_n = 1. \tag{2.2.11}$$

Но $W^T W = w_1^2 + \ldots + w_n^2$, так что

$$\int_{R_n} \ldots \int \exp -\frac{1}{2}(W^T W)\, dw_1 \ldots dw_n =$$
$$= \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} w_1^2\right) dw_1 \ldots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} w_n^2\right) dw_n = (2\pi)^{\frac{n}{2}},$$

так как $\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x^2\right) dx = \sqrt{2\pi}$.

Подставляя это в (2.2.6), находим

$$C_0[\det(\Lambda)]^{-\frac{1}{2}}(2\pi)^{\frac{n}{2}}=1; \quad C_0=[\det(\Lambda)]^{\frac{1}{2}}(2\pi)^{-\frac{n}{2}}, \tag{2.2.12}$$

что доказывает (2.2.3).

Найдем вероятностный смысл вектора A в выражении нормальной плотности (2.2.1). Докажем, что $a_i=E(X_i)$ $(i=1, 2, \ldots, n)$.

Имеем

$$E(X_i)=C_0\int\limits_{R_n}\ldots\int x_i \exp-\frac{1}{2}(X-A)^T\Lambda(X-A)\,dx_1\ldots dx_n.$$

Полагая $X-A=Y$, т. е. $x_j=y_j+a_j$ $(j=1, 2, \ldots, n)$, находим

$$\begin{aligned} E(X_i)&=C_0\int\limits_{R_n}\ldots\int(y_i+a_i)\exp-\frac{1}{2}(Y^T\Lambda Y)\,dy_1\ldots dy_n=\\ &=C_0\int\limits_{R_n}\ldots\int y_i\exp-\frac{1}{2}(Y^T\Lambda Y)\,dy_1\ldots dy_n+\\ &+a_iC_0\int\limits_{R_n}\ldots\int\exp-\frac{1}{2}(Y^T\Lambda Y)\,dy_1\ldots dy_n. \end{aligned} \tag{2.2.13}$$

Далее докажем, что

$$C_0\int\limits_{R_n}\ldots\int y_i\exp-\frac{1}{2}(Y^T\Lambda Y)\,dy_1\ldots dy_n=0. \tag{2.2.14}$$

В самом деле, квадратичная форма $Y^T\Lambda Y=\sum\lambda_{rs}y_ry_s$ есть четная функция y_i, а выражение $y_i\exp-\frac{1}{2}(Y^T\Lambda Y)$ — нечетная функция y_i при любых фиксированных значениях других переменных. Так как наш интеграл берется по y_i от $-\infty$ до ∞, то он равен нулю.

Далее, в силу (2.2.6) второй член (2.2.13) равен a_i. Используя (2.2.14), получаем нужное равенство

$$a_i=E(X_i), \tag{2.2.15}$$

так что имеем $A=E(X)$, что и доказывает (2.2.4).

Теперь найдем связь между корреляционной матрицей B_X вектора X и соответствующей матрицей Λ в выражении (2.2.1) для вероятностной плотности. Нам надо вычислить матрицу

$$B_X=E(YY^T)=C_0\int\limits_{R_n}\ldots\int YY^T\exp-\frac{1}{2}(Y^T\Lambda Y)\,dy_1\ldots dy_n, \tag{2.2.16}$$

притом надлежит понимать написанное выражение так: каждый элемент матрицы YY^T умножается на $C_0\exp-\frac{1}{2}(Y^T\Lambda Y)$ и интегрируется по пространству R_n; полученные результаты собираются в новую матрицу B_X. Для вычисления матрицы (2.2.16) вводим, как и ранее, преобразование

$$Y = FD^{-\frac{1}{2}}W;$$

для упрощения записи положим $FD^{-\frac{1}{2}} = G$. Тогда имеем

$$Y = GW; \quad G^T \Lambda G = E; \quad Y^T \Lambda Y = W^T W = \sum_{i=1}^{n} w_i^2.$$

Далее,

$$YY^T = GW(GW)^T = GWW^TG^T. \tag{2.2.17}$$

Подставляя это в (2.2.16), находим

$$B_X = C_0 \det(G) \int_{R_n} \ldots \int GWW^TG^T \exp - \frac{1}{2}(W^TW)\, dw_1 \ldots dw_n.$$

Или, на основании (2.2.12) и (2.2.9),

$$B_X = (2\pi)^{-\frac{n}{2}} \int_{R_n} \ldots \int GWW^TG^T \exp - \frac{1}{2}(W^TW)\, dw_1 \ldots dw_n =$$

$$= G(2\pi)^{-\frac{n}{2}} \int_{R_n} \ldots \int WW^T \exp - \frac{1}{2}(W^TW)\, dw_1 \ldots dw_n G^T. \tag{2.2.18}$$

Докажем, что

$$(2\pi)^{-\frac{n}{2}} \int_{R_n} \ldots \int WW^T \exp - \frac{1}{2}(W^TW)\, dw_1 \ldots dw_n = E_{nn},$$

где E_{nn} — единичная матрица. Наш интеграл является матрицей $U = \|u_{rs}\|$ с элементами

$$u_{rs} = (2\pi)^{-\frac{n}{2}} \int_{R_n} \ldots \int w_r w_s \exp - \frac{1}{2}\left(w_1^2 + \ldots + w_n^2\right) dw_1 \ldots dw_n.$$

Это выражение, очевидно, равно нулю при $r \neq s$. При $r = s$ имеем

$$u_{rs} = u_{sr} = (2\pi)^{-\frac{n}{2}} \int_{R_n} \ldots \int w_r^2 \exp - \frac{1}{2}\left(w_1^2 + \ldots + w_n^2\right) dw_1 \ldots dw_n =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} w_r^2 \exp - \frac{1}{2} w_r^2\, dw_r = 1$$

на основании свойств нормальной случайной величины $X \in N(0, 1)$ (см. (2.1.8)). Значит,

$$U = \|u_{rs}\| = E,$$

что и доказывает наше утверждение. Из (2.2.18) находим $B_X = GG^T$. Но мы имели ранее $G^T \Lambda G = E$, так что $G^T = G^{-1}\Lambda^{-1}$ и

$$B_X = GG^T = GG^{-1}\Lambda^{-1} = \Lambda^{-1}.$$

Таким образом,

$$B_X = \Lambda^{-1}; \quad \Lambda = B_X^{-1},$$

что доказывает (2.2.5) и завершает доказательство теоремы 2.2.1.

Таким образом, с помощью корреляционной матрицы $B_X = B$ нормальная плотность может быть записана в виде

$$f(x_1, \ldots, x_n) = (2\pi)^{-\frac{n}{2}} [\det(B)]^{-\frac{1}{2}} \exp -\frac{1}{2}\left[(X-A)^T B^{-1}(X-A)\right]. \tag{2.2.19}$$

Для примера возьмем двумерный нормальный вектор $X = \begin{Vmatrix} x_1 \\ x_2 \end{Vmatrix}$. Положим

$$E(x_i - a_i)^2 = \sigma_i^2 \quad (i = 1, 2), \qquad E(x_1 - a_1)(x_2 - a_2) = \rho\sigma_1\sigma_2; \qquad |\rho| < 1.$$

(как известно из элементарного курса теории вероятностей (см. Б. В. Гнеденко [11]), ρ называется *коэффициентом корреляции* между x_1 и x_2).

Как легко подсчитать,

$$B = \begin{Vmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{Vmatrix}; \quad B^{-1} = \frac{1}{1-\rho^2} \begin{Vmatrix} \dfrac{1}{\sigma_1^2} & -\dfrac{\rho}{\sigma_1\sigma_2} \\ -\dfrac{\rho}{\sigma_1\sigma_2} & \dfrac{1}{\sigma_2^2} \end{Vmatrix};$$

$$\det(B) = \sigma_1^2\sigma_2^2(1-\rho^2) > 0;$$

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \exp -\frac{1}{2(1-\rho^2)}\left(\frac{1}{\sigma_1^2}(x_1-a_1)^2 - \right.$$
$$\left. - \frac{2\rho}{\sigma_1\sigma_2}(x_1-a_1)(x_2-a_2) + \frac{1}{\sigma_2^2}(x_2-a_2)^2\right). \tag{2.2.20}$$

При $n = 1$ получаем введенное ранее понятие нормальной случайной величины.

Из формулы (2.2.19) видно, как можно построить нормальную n-мерную плотность непосредственно по вектору средних значений A и корреляционной матрице B с $\det(B) > 0$. Заметим, что в силу доказанного, любая симметрическая матрица $B = B_{nn}$ отвечающая положительно определенной квадратичной форме $Y^T B Y$, будет корреляционной матрицей n-мерного случайного вектора. Именно, можно взять, например, $A = \begin{Vmatrix} 0 \\ \vdots \\ 0 \end{Vmatrix}$, затем положить $\Lambda = B^{-1}$, тогда искомая плотность определяется равенством (2.2.1) с $C_0 = (2\pi)^{-\frac{n}{2}} [\det(B)]^{-\frac{1}{2}}$.

§ 3. Линейные функции n-мерного нормального вектора

В применениях метода наименьших квадратов нам часто придется иметь дело с n-мерным нормальным вектором $X = \begin{Vmatrix} x_1 \\ \vdots \\ x_n \end{Vmatrix}$, компоненты которого будут представлять собой случайные погрешности наблюдений. При этом часто надо будет составлять линейные комбинации наблюдений и их погрешностей, вида

$$u_i = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n \qquad (i = 1,\ 2,\ \ldots,\ k). \tag{2.3.1}$$

Вводя вектор $U = U_{k1} = \begin{Vmatrix} u_1 \\ \vdots \\ u_k \end{Vmatrix}$ и матрицу $A = A_{kn} = \|a_{rs}\|$, можем записать (2.3.1) в виде

$$U = AX, \tag{2.3.2}$$

где U будет линейной векторной функцией случайного вектора X. Если $k = 1$, то мы имеем скалярную функцию вектора X.

Геометрически, как видно из (2.3.1), компоненты $u_i = [A_i \cdot X]$ являются скалярными произведениями вектора X на k фиксированных векторов $A_i = (a_{i1}, \ldots, a_{in})$ $(i = 1, 2, \ldots, k)$, т. е. пропорциональны проекциям X на k фиксированных направлений.

Мы будем различать два случая. Если линейные формы (2.3.1) линейно независимы (см. § 2 гл. I), т. е. если ранг матрицы $A = A_{kn}$ равен k, будем называть линейную векторную функцию U *невырожденным вектором*; в противном случае (ранг $(A) < k$) — *вырожденным вектором*.

Сперва изучим невырожденные векторы U. Начнем с частного случая $k = n$, так что $A = A_{nn}$, ранг $(A) = n$. Докажем, что в этом случае вектор U будет, как и X, n-мерным нормальным случайным вектором.

Пусть заданы числа $v_1, \ldots, v_n$ и положительные приращения $dv_1, \ldots, dv_n$; мы интересуемся событием

$$\mathfrak{A} : \{v_1 \leqslant u_1 < v_1 + dv_1,\ \ldots;\ v_n \leqslant u_n < v_n + dv_n\} \tag{2.3.3}$$

и его вероятностью

$$P\{\mathfrak{A}\} = P\{v_1 \leqslant u_1 < v_1 + dv_1,\ \ldots;\ v_n \leqslant u_n < v_n + dv_n\} \tag{2.3.4}$$

при достаточно малых dv_i $(i = 1, 2, \ldots, n)$.

Рассмотрим сначала сходный вопрос для случая $n=1$ и общей (нелинейной) функциональной зависимости случайных величин.

Пусть X — случайная величина, распределенная на всей оси или на ее отрезке с плотностью вероятности $f(x)$, а $Y=g(X)$ — функция от нее, такая, что производная $g'(x)=\frac{dg(x)}{dx}$ существует и

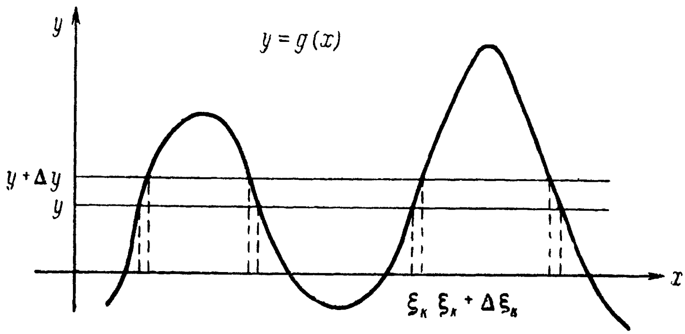

Рис. 3*а*.

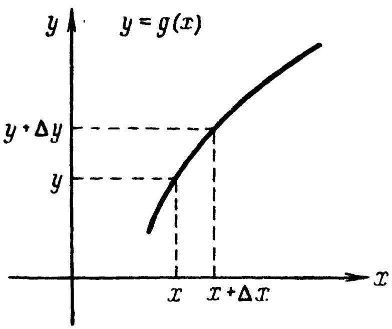

Рис. 3*б*.

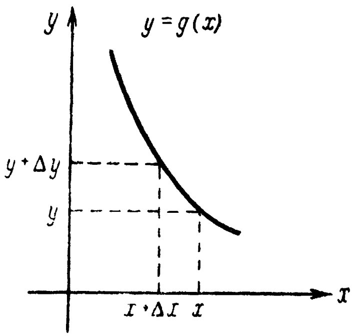

Рис. 3*в*.

непрерывна. Пусть x — какое-либо число, dx — малое приращение, $dy=g'(x)\,dx$. Мы хотим найти плотность $\varphi(y)$ вероятности распределения Y (если она существует). Для этого нужно изучить отношение

$$\frac{P\{y \leqslant Y < y+\Delta y\}}{\Delta y} \quad \text{при} \quad \Delta y \to 0.$$

Из рис. 3*а* видно, что в общем случае, когда функция $y=g(x)$ не монотонна, такое изучение сложно, ибо событие $\{y \leqslant Y < y+\Delta y\}$, вообще говоря, равносильно объединению многих событий вида $P\{\xi_k \leqslant X < \xi_k+\Delta\xi_k\}$. Поэтому мы ограничимся случаем монотонности $y=g(x)$ на отрезке, где распределена величина X. В этом случае будем иметь: $g'(x)>0$ (рис. 3б), либо $g'(x)<0$ (рис. 3в) на этом отрезке, и (для случая на рис. 3б), как легко убедиться,

$$P\{y\leqslant Y<y+\Delta y\}=P\left\{x\leqslant X<x+\frac{dy}{g'(x)}\right\}$$

с точностью до величин порядка малости выше, чем dx и dy. Но для малых δ мы имеем $P\{x\leqslant X<x+\delta\}=f(x)\cdot\delta$ с точностью до величин порядка малости выше δ, что будем записывать так:

$$P\{x\leqslant X<x+\delta\}\approx f(x)\delta.$$

Отсюда

$$P\left\{x\leqslant X<x+\frac{dy}{g'(x)}\right\}\approx f(x)\frac{dy}{g'(x)}.$$

Таким образом,

$$P\{y\leqslant Y<y+\Delta y\}\approx\frac{f(x)}{g'(x)}\Delta y.$$

Стало быть, Y имеет плотность вероятности

$$\varphi(y)=\frac{f(x)}{g'(x)},$$

где x надо выразить через y по формуле $y=g(x)$. Для случая рис. 3в имеем $\varphi(y)=-\frac{f(x)}{g'(x)}$ (тогда $g'(x)<0$), так что в общем случае получаем

$$\varphi(y)=\frac{f(x)}{|g'(x)|}. \tag{2.3.5}$$

Случай монотонности отличается от случая рис. 3а тем, что $g'(x)\neq 0$ на всем отрезке.

Перейдем теперь к случаю векторной функции U случайного вектора X:

$$U=g(X)=\begin{Vmatrix} g_1(x_1,\ldots,x_n)\\ \cdots\cdots\\ g_n(x_1,\ldots,x_n)\end{Vmatrix}.$$

Пусть вектор X имеет плотность вероятности $f(x_1,\ldots,x_n)$. Известно, что в теории преобразований переменных и подстановок в кратных интегралах роль $y'=g'(x)$ играет якобиан $J(x_1,\ldots,x_n)=\det(I)$, где $I=\left\|\frac{\partial g_r}{\partial x_s}\right\|$. Поэтому по аналогии с (2.3.5) мы можем ожидать, что если в области, где распределен вектор X, $J(x_1,\ldots x_n)\neq 0$, то для плотности вероятности $\varphi(u_1,\ldots,u_n)$ вектора U, получим

$$\varphi(u_1,u_2,\ldots,u_n)=\frac{f(x_1,\ldots,x_n)}{|J(x_1,\ldots,x_n)|}, \tag{2.3.6}$$

где x_i надо заменить их выражениями через u_j из формул $u_j=g_j(x_1,\ldots,x_n)$ $(j=1,2,\ldots,n)$.

Это действительно верно, и мы вкратце докажем это для введенной ранее линейной зависимости $U=AX$, где $A=A_{nn}$ — неособенная матрица. При обратном линейном преобразовании $X=A^{-1}U$ параллелепипеду, отвечающему событию $\mathfrak{A}$ из (2.3.3), будет соответствовать некоторый косоугольный параллелепипед объема

$$\frac{du_1 \ldots du_n}{|\det(A)|} \tag{2.3.7}$$

(см. § 3 гл. I, в частности (1.3.3)). Отсюда $P(\mathfrak{A})$ равна вероятности попадания вектора X в указанный косоугольный параллелепипед объема (2.3.7), т. е.

$$P(\mathfrak{A})=\frac{f(x_1, \ldots, x_n)}{|\det(A)|}\, du_1 \ldots du_n. \tag{2.3.8}$$

Таким образом, вектор U является случайным вектором, имеющим плотность вероятности

$$\varphi(u_1, \ldots, u_n)=\frac{f(x_1, \ldots, x_n)}{|\det(A)|}, \tag{2.3.9}$$

где x_i надо выразить через u_j. Но X нормальный вектор, так что, согласно (2.2.19),

$$f(x_1, \ldots, x_n)=\frac{1}{(2\pi)^{\frac{n}{2}}[\det(B)]^{\frac{1}{2}}} \exp -\frac{1}{2}\left(Y^T B^{-1} Y\right), \tag{2.3.10}$$

где $Y=X-M$, M — вектор средних значений и B — корреляционная матрица.

Имеем $U=AX=A(Y+M)$, откуда

$$Y=A^{-1}U-M=A^{-1}(U-AM); \quad Y^T=(U-AM)^T(A^T)^{-1},$$

так что, полагая $U-AM=V$, $Y=A^{-1}V$, найдем

$$\varphi(u_1, \ldots, u_n)=\frac{1}{(2\pi)^{\frac{n}{2}}|\det(A)|(\det(B))^{\frac{1}{2}}} \exp -\frac{1}{2}\left(V^T(A^T)^{-1}B^{-1}A^{-1}V\right).$$

Отсюда, согласно (2.3.10) и теореме 2.2.1, заключаем, что U — n-мерный нормальный случайный вектор с вектором средних значений AM и корреляционной матрицей

$$B_U=AB_XA^T, \tag{2.3.11}$$

при этом $\det(B_U)=(\det(A))^2\det(B)$ и $[\det(B_U)]^{\frac{1}{2}}=|\det(A)|(\det B)^{\frac{1}{2}}$.

Сделаем еще непосредственный расчет корреляционной матрицы вектора $U=AX$ без каких-либо предположений о ранге матрицы

$A = A_{kn}$. Согласно § 1 этой главы [см. формулы (2.1.13) и (2.1.14)], имеем

$$E(U) = E(AX) = AE(X) = AM;$$
$$B_U = E(U - AM)(U - AM)^T = E\left(A(X - M)(A(X - M))^T\right) =$$
$$= E\left(A(X - M)(X - M)^T A^T\right) = AE(X - M)(X - M)^T A^T = AB_X A^T.$$

Здесь A и A^T — постоянные матрицы, которые при вычислении математических ожиданий матриц выносятся за знак математического ожидания. Мы снова пришли к (2.3.11).

В более общем случае, когда $k < n$, $A = A_{kn}$, ранг $(A) = k$, имеет место теорема:

Теорема 2.3.1. *Вектор $U = AX$ есть k-мерный случайный вектор с вектором средних значений AM и корреляционной матрицей $B_U = AB_X A^T$.*

Мы не будем проводить здесь доказательства, требующего несколько кропотливого матричного подсчета, и приведем лишь его идею *).

Вектор

$$U = U_{k1} = \begin{Vmatrix} u_1 \\ \cdot \\ \cdot \\ u_k \end{Vmatrix}$$

дополняется до вектора

$$U'_{n1} = U' = \begin{Vmatrix} u_1 \\ \cdot \\ \cdot \\ u_k \\ u_{k+1} \\ \cdot \\ \cdot \\ u_n \end{Vmatrix}$$

за счет соответствующего дополнения матрицы A до матрицы $A' = A'_{nn}$ с рангом n: ранг $(A') = n$. На основании предыдущего получается нормальный n-мерный вектор U', для которого можно написать плотность вероятности $\varphi(y_1, \ldots, y_n)$. После этого событие $\mathfrak{A}$:

$$(y_1 \leqslant u_1 < y_1 + dy_1, \ \ldots, \ y_k \leqslant u_k < y_k + dy_k)$$

заменяется равносильным событием $\mathfrak{A}_1$:

$$(y_1 \leqslant u_1 < y_1 + dy_1, \ \ldots, \ y_k \leqslant u_k < y_k + dy_k;$$
$$-\infty < u_{k+1} < \infty, \ \ldots, \ -\infty < u_n < \infty).$$

*) Подробное матричное доказательство см., например, в книге С. Рао [42], стр. 53—54.

Тогда

$$P(\mathfrak{A}) \approx dy_1 \ldots dy_k \int_{R_{n-k}} \ldots \int \varphi(y_1, \ldots, y_n)\, dy_{k+1} \ldots dy_n,$$

где через R_{n-k} обозначено $(n-k)$-мерное пространство. Отсюда легко найти плотность вероятности исходного вектора U, задающую его как k-мерный нормальный вектор. В силу сказанного выше вектор средних значений должен быть равен AM, а корреляционная матрица есть $B_U = AB_X A^T$. Перейдем к случаю вырождения линейной векторной функции, когда в выражении $U = A \cdot X$ ранг матрицы $A = A_{kn}$ меньше k. Так будет, например, всегда при $k > n$. В этом случае (см. § 2 гл. I) между компонентами вектора $U = \begin{Vmatrix} u_1 \\ \vdots \\ u_k \end{Vmatrix}$ будут линейные зависимости, которые можно записать в виде $\sum_{j=1}^{k} c_{ij} u_j = 0$, или $CU = 0$, где $C = C_{mk}$ — некоторая матрица.

Геометрически условие $CU = 0$ означает, что вектор U лежит в некотором подпространстве R_k, а с вероятностной точки зрения это условие означает, что распределение вектора сосредоточено в этом подпространстве. В этом случае мы можем описать поведение линейной векторной функции $U = AX$ следующим образом. Среди компонент $u_1, \ldots, u_k$ вектора U отбираем максимальное число линейно независимых, скажем, $u_1, \ldots, u_l$; $l < k$ и составляем вектор $U' = \begin{Vmatrix} u_1 \\ \vdots \\ u_l \end{Vmatrix}$. Это будет l-мерный вектор, а компоненты $u_{l+1}, \ldots, u_k$ будут известными линейными функциями компонент $u_1, \ldots, u_l$.

Все компоненты будут нормальными величинами. Мы не будем доказывать этого (см., например, С. Рао [42], Г. Крамер [25]).

Докажем еще некоторые теоремы о нормальных векторах, полезные для дальнейшего.

Теорема 2.3.2 (теорема Р. А. Фишера*)). *Если X — нормальный n-мерный вектор с нулевым вектором средних и независимыми одинаково распределенными компонентами, а $F = F_{nn}$ — какое-либо ортогональное преобразование, то n-мерный нормальный вектор $Y = FX$ будет иметь то же распределение, что и вектор X.*

Доказательство. Имеем $E(X) = 0$, $B_X = \sigma^2 E$, где σ^2 — дисперсия каждой из компонент X. По теореме 2.3.1 $Y = F \cdot X$ будет

---

*) Р. А. Фишер (род. 1891) — английский статистик.

n-мерным нормальным вектором. При этом

$$E(Y)=E(FX)=FE(X)=F0=0;$$
$$B_Y=FB_XF^T=F\sigma^2E\cdot F^T=\sigma^2FF^T=\sigma^2E=B_X,$$

ибо $FF^T=E$ в силу ортогональности F. Таким образом, $E(Y)=E(X)$; $B_Y=B_X$; значит вектор Y так же распределен, как X, что и требовалось доказать.

Пусть заданы две линейные векторные функции n-мерного нормального вектора X:

$$Y_1=A_1X;\quad Y_2=A_2X, \tag{2.3.12}$$

где A_1 — матрица типа $(m_1\times n)$, а A_2—матрица типа $(m_2\times n)$. Имеет место следующая теорема.

Теорема 2.3.3. *Для того чтобы векторы $Y_1=A_1X$ и $Y_2=A_2X$ были статистически независимыми, необходимо и достаточно, чтобы*

$$A_1B_XA_2^T=0. \tag{2.3.13}$$

Доказательство. Не нарушая общности, можем считать $E(X)=M=0$. В самом деле, если это не так, введем вектор $Z=X-M$. Тогда $B_X=B_Z$, так что условие (2.3.13) имеет прежний вид, а векторы $Y_1'=A_1(X-M)$ и $Y_2'=A_2(X-M)$ будут, очевидно, статистически независимы тогда и только тогда, когда независимы Y_1 и Y_2. Теперь, считая $E(X)=0$, будем доказывать необходимость условия (2.3.13). Так как $B_X=E(XX^T)$, то (2.3.13) можно переписать так:

$$A_1E(XX^T)A_2^T=E\left((A_1X)(A_2X)^T\right)=0,$$

или, в силу (2.3.12),

$$E\left(Y_1Y_2^T\right)=0. \tag{2.3.14}$$

Согласно предыдущему, Y_1 и Y_2 будут нормальными векторами, причем:

$$E(Y_i)=A_iE(X)=0;\qquad i=1,\ 2. \tag{2.3.15}$$

Пусть

$$Y_1=\begin{Vmatrix} y_{11}\\ y_{12}\\ \vdots\\ y_{1m_1}\end{Vmatrix};\qquad Y_2=\begin{Vmatrix} y_{21}\\ y_{22}\\ \vdots\\ y_{2m_2}\end{Vmatrix}. \tag{2.3.16}$$

Если Y_1 и Y_2 независимы, то любая пара компонент y_{1i} и y_{2j} будет парой независимых нормальных случайных величин, так что

$$E(y_{1i}y_{2j})=E(y_{1i})E(y_{2j})=0 \tag{2.3.17}$$

в силу (2.3.15) и (2.3.17). Но (2.3.17) как раз и означает, что

$$E\left(Y_1Y_2^T\right)=0.$$

Это доказывает равенство (2.3.14), которое равносильно (2.3.13). Достаточность (2.3.13) доказать сложнее. Прежде всего, подберем неособенную матрицу $G_{nn}=G$ такую, что

$$GB_XG^T=E. \tag{2.3.18}$$

Мы видели ранее, что G можно выбрать в виде $G=D_0\cdot F_0$, где D_0 — дополнительная неособенная матрица, а F_0 — ортогональная матрица. В самом деле, беря ортогональную матрицу F_0 такую, что

$$F_0B_XF_0^T=D_1,$$

замечаем, что

$$D_1=\begin{Vmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & d_n \end{Vmatrix},$$

где $d_i>0$ $(i=1, 2, \dots, n)$, ибо B_X — неособенная матрица. Беря $G=D_1^{-\frac{1}{2}}F_0$; $G^T=F_0^TD_1^{-\frac{1}{2}}$, получаем требуемое. Пусть теперь выполнено условие (2.3.13). Имеем

$$A_1G^{-1}GBG^T\left(G^T\right)^{-1}A_2^T=A_1G^{-1}\left(A_2G^{-1}\right)^T=0. \tag{2.3.19}$$

Далее,

$$Y_1=A_1G^{-1}GX;\quad Y_2=A_2G^{-1}GX. \tag{2.3.20}$$

Введем обозначения: $GX=Z$; $A_1G^{-1}=A_3$; $A_2G^{-1}=A_4$. Тогда

$$Y_1=A_3Z;\quad Y_2=A_4Z;\quad A_3A_4^T=0, \tag{2.3.21}$$

и Z есть n-мерный случайный вектор с $E(Z)=0$; $B_Z=E$. Нам надлежит доказать независимость Y_1 и Y_2. Пусть

$$A_3=\begin{Vmatrix} a_{11} & \dots & a_{1n} \\ \cdot & \cdot & \cdot \\ a_{m_11} & \dots & a_{m_1n} \end{Vmatrix};\quad A_4=\begin{Vmatrix} b_{11} & \dots & b_{1n} \\ \cdot & \cdot & \cdot \\ b_{m_21} & \dots & b_{m_2n} \end{Vmatrix}. \tag{2.3.22}$$

Строку $(a_{i1}, \dots, a_{in})$ будем рассматривать как вектор и обозначать a_i; строку $(b_{j1}, b_{j2}, \dots, b_{jn})$ — как вектор b_j $(i=1, 2, \dots, m_1;$ $j=1, 2, \dots, m_2)$. Условие $A_3A_4^T=0$ означает попарную ортогональность векторов:

$$[a_ib_j]=0\quad (i=1, 2, \dots, m_1;\ j=1, 2, \dots, m_2). \tag{2.3.23}$$

Среди векторов a_i и среди векторов b_j могут быть линейно зависимые. Пусть ранг $(A_3)=\mu$, ранг $(A_4)=\nu$. (Заметим, что ранг $(A_3)=$ $=$ранг (A_1) и ранг $(A_4)=$ранг (A_2). Это следует из соотношения (1.2.18), ибо $A_3=A_1G$, $A_1=A_3G^{-1}$; ранг $(G)=n$, ранг $(A_1)\leqslant n$; то же касается матриц A_2 и A_4.)

Выделим в A_3 систему μ линейно независимых строк, а в A_4 систему ν линейно независимых строк; пусть им отвечают векторы

$a_{\alpha_1}, a_{\alpha_2}, \ldots, a_{\alpha_\mu}$ и, соответственно, $b_{\beta_1}, \ldots, b_{\beta_\nu}$, так что

$$[a_{\alpha_i} \cdot b_{\beta_j}] = 0 \quad (i = 1, 2, \ldots, \mu;\ j = 1, 2, \ldots, \nu). \qquad (2.3.24)$$

Очевидно, соотношения (2.3.23) будут следствиями соотношений (2.3.24), так как векторы a_i будут линейными комбинациями векторов a_{α_i}, а векторы b_j — линейными комбинациями векторов b_{β_j}. Имеем, соответственно (2.3.16), при $Z = (z_1, \ldots, z_n)$ *)

$$y_{1i} = [a_i \cdot z]; \quad y_{2j} = [b_j \cdot z]. \qquad (2.3.25)$$

Ввиду этого, очевидно, все компоненты y_{1i} будут линейными комбинациями компонент $y_{1\alpha_i}$; все компоненты y_{2j} будут линейными комбинациями компонент $y_{2\beta_j}$. Введем новые векторы

$$Y_1' = \left\| \begin{matrix} y_{1\alpha_1} \\ \vdots \\ y_{1\alpha_\mu} \end{matrix} \right\|, \quad Y_2' = \left\| \begin{matrix} y_{2\beta_1} \\ \vdots \\ y_{2\beta_\nu} \end{matrix} \right\|. \qquad (2.3.26)$$

Если мы докажем, что Y_1' и Y_2' статистически независимы, то, очевидно, этим будет доказано, что Y_1 и Y_2 независимы, так как Y_1 и Y_2 суть векторные функции Y_1' и Y_2'. Мы можем написать

$$Y_1' = A_5 Z; \qquad Y_2' = A_6 Z, \qquad (2.3.27$$

где A_5 составлена из строк $\alpha_1, \ldots, \alpha_\mu$ матрицы A_3, а A_6 — из строк $\beta_1, \ldots, \beta_\nu$ матрицы A_4. Очевидно,

$$A_5 A_6^T = 0, \qquad (2.3.28)$$

ибо это равенство равносильно (2.3.24). По теореме 2.3.1, Y_1' есть μ-мерный, а Y_2' есть ν-мерный нормальный вектор. Составим теперь новый случайный вектор из всех компонент Y_1' и Y_2':

$$Y = \left\| \begin{matrix} y_{1\alpha_1} \\ \vdots \\ y_{1\alpha_\mu} \\ y_{2\beta_1} \\ \vdots \\ y_{2\beta_\nu} \end{matrix} \right\|. \qquad (2.3.29)$$

---

*) Здесь и иногда в дальнейшем мы будем таким образом записывать вектор

$$Z = \left\| \begin{matrix} z_1 \\ \vdots \\ z_n \end{matrix} \right\|.$$

Имеем

$$Y = AZ; \qquad A = \left\| \begin{array}{c} A_5 \\ \hdashline A_6 \end{array} \right\|. \tag{2.3.30}$$

Докажем, что Y есть $(\mu+\nu)$-мерный нормальный вектор (так что $\mu+\nu \leqslant n$). Ясно, что Y — нормальный вектор (теорема 2.3.1); число его компонент равно $\mu+\nu$. Чтобы доказать, что Y есть $(\mu+\nu)$-мерный нормальный вектор, достаточно убедиться, что его корреляционная матрица B_Y неособенная, тогда при $E(Y)=0$ плотность вероятности для него задается формулой (2.2.10), показывающей его $(\mu+\nu)$-мерность.

По теореме 2.3.1 имеем $B_Y = AB_ZA^T = AA^T$. В силу (2.3.30) и правил умножения разделенных матриц (§ 2 гл. I)

$$AA^T = \left\| \begin{array}{c} A_5 \\ \hdashline A_6 \end{array} \right\| \cdot \left\| A_5^T \,\vdots\, A_6^T \right\| = \left\| \begin{array}{c:c} A_5A_5^T & A_5A_6^T \\ \hdashline A_6A_5^T & A_6A_6^T \end{array} \right\|.$$

Из (2.3.28) имеем $A_5A_6^T = 0$; транспонируя это соотношение, находим $A_6A_5^T = 0$, откуда

$$AA^T = \left\| \begin{array}{c:c} A_5A_5^T & 0 \\ \hdashline 0 & A_6A_6^T \end{array} \right\|. \tag{2.3.31}$$

Здесь $A_5A_5^T$ и $A_6A_6^T$ — симметрические матрицы типов $(\mu \times \mu)$ и $(\nu \times \nu)$. Докажем, что они неособенные. На основании теоремы 1.3.1, ввиду того, что $\mu \leqslant n$ и ранг $(A_5) = \mu$, получаем, что $\det\left(A_5A_5^T\right) > 0$. Аналогично $\det\left(A_6A_6^T\right) > 0$. Таким образом,

$$\det\left(AA^T\right) = \det\left(A_5A_5^T\right)\det\left(A_6A_6^T\right) > 0,$$

так что $B_Y = AA^T$ — неособенная, и Y есть $(\mu+\nu)$-мерный нормальный вектор. Для того, чтобы доказать независимость Y_1' и Y_2', составим выражение для плотности Y соответственно формуле (2.2.19).

Имеем

$$B_Y^{-1} = \left(AA^T\right)^{-1} = \left\| \begin{array}{c:c} \left(A_5A_5^T\right)^{-1} & 0 \\ \hdashline 0 & \left(A_6A_6^T\right)^{-1} \end{array} \right\|. \tag{2.3.32}$$

Далее, имеем:

$$B_{Y_1'} = A_5A_5^T; \quad B_{Y_2'} = A_6A_6^T, \tag{2.3.33}$$

так что

$$B_Y^{-1} = \left\| \begin{array}{c:c} B_{Y_1'}^{-1} & 0 \\ \hdashline 0 & B_{Y_2'}^{-1} \end{array} \right\|. \tag{2.3.34}$$

Составляя плотность для Y по формуле (2.2.19) и учитывая, что

$$Y = Y_{\mu+\nu,\,1} = \left\| \begin{array}{c} Y_1' \\ \hdashline Y_2' \end{array} \right\|$$

и

$$Y^T B_Y^{-1} Y = \left\| (Y_1')^T \;\vdots\; (Y_2')^T \right\| \cdot \left\| \begin{array}{c:c} B_{Y_1'}^{-1} & 0 \\ \hdashline 0 & B_{Y_2'}^{-1} \end{array} \right\| \cdot \left\| \begin{array}{c} Y_1' \\ \hdashline Y_2' \end{array} \right\| =$$
$$= (Y_1')^T B_{Y_1'}^{-1} Y_1' + (Y_2')^T B_{Y_2'}^{-1} Y_2',$$

найдем для плотности вероятности Y выражение

$$\frac{1}{(2\pi)^{\frac{\mu}{2}} \left(\det\left(B_{Y_1'}\right)\right)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(Y_1')^T B_{Y_1'} Y_1'\right] \frac{1}{(2\pi)^{\frac{\nu}{2}} \left(\det\left(B_{Y_2'}\right)\right)^{\frac{1}{2}}} \times$$
$$\times \exp -\frac{1}{2}(Y_2')^T B_{Y_2'} Y_2'.$$

Мы видим, что плотность вероятности для Y распадается на произведение плотностей вероятности для Y_1' и для Y_2', что и доказывает их независимость. Этим, наконец, теорема 2.3.3 доказана.

§ 4. Приведение нормального вектора к простейшему виду. Корреляционный эллипсоид и эллипсоид постоянной дисперсии

Рассмотрим n-мерный нормальный вектор $X_{n1} = X$ с корреляционной матрицей B_X. Вектор средних $E(X)$ будем считать нулевым, так как замена X на $X - E(X)$ приводит к той же корреляционной матрице B_X. Мы знаем, что можно подобрать ортогональную матрицу $F_{nn} = F$, такую, что, полагая $Y = FX$, получим

$$B_Y = F B_X F^T = D, \tag{2.4.1}$$

где D — диагональная неособенная матрица с положительными диагональными элементами:

$$D = \left\| \begin{array}{cccc} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & d_n \end{array} \right\|.$$

Поэтому плотность вероятности нормального вектора Y имеет вид

$$f(y_1, \ldots, y_n) = \frac{1}{(2\pi)^{\frac{n}{2}} (d_1 \ldots d_n)^{\frac{1}{2}}} \exp - \frac{1}{2}\left(\frac{y_1^2}{d_1} + \ldots + \frac{y_n^2}{d_n}\right). \quad (2.4.2)$$

Мы видим, что вектор Y имеет независимые компоненты y_i, причем

$$E(y_i) = 0; \quad D(y_i) = d_i. \quad (2.4.3)$$

Если введем еще преобразование $Z = D^{-\frac{1}{2}} Y$, то получим

$$E(Z) = 0; \quad B_Z = D^{-\frac{1}{2}} D D^{-\frac{1}{2}} = E. \quad (2.4.4)$$

Плотность вероятности для вектора Z будет иметь вид (2.4.2), где все $d_i = 1$. Z будет нормальным вектором с независимо распределенными по $N(0, 1)$ компонентами. Такой вектор мы можем назвать *простейшим нормальным вектором.*

Рассмотрим нормальный n-мерный вектор $X = \begin{Vmatrix} x_1 \\ \vdots \\ x_n \end{Vmatrix}$; x_i будем считать прямоугольными координатами. Пусть

$$E(X) = M = \begin{Vmatrix} m_1 \\ \vdots \\ m_n \end{Vmatrix}; \quad B_X = \| b_{rs} \|.$$

Введем нормальный вектор

$$X - M = Z = \begin{Vmatrix} z_1 \\ \vdots \\ z_n \end{Vmatrix}$$

(вектор уклонений от среднего).

Пусть $a = A_{1n} = \| a_1, \ldots, a_n \|$ — единичный вектор $(a_1^2 + \ldots + a_n^2 = 1)$; составим скалярное произведение

$$[aZ] = \sum_{i=1}^{n} a_i Z_i = AZ.$$

Это будет нормальная величина с нулевым средним и дисперсией $D(AZ) = E(AZZ^T A^T) = AB_X A^T$ (см. теорему 2.3.1). Итак, в направлении, характеризуемом единичным вектором $(a_1, \ldots, a_n)$, получается случайная проекция с дисперсией $AB_X A^T$. Рассмотрим теперь поверхность, которая получится, если в направлении $(a_1, \ldots, a_n)$ откладывать соответствующую дисперсию $AB_X A^T$. Ее мы назовем *корреляционной поверхностью* n-мерного нормального вектора X.

Теорема 2.4.1. *Корреляционная поверхность нормального n-мерного вектора X представляет собой эллипсоид с уравнением*

$$(X-M)^T B_X^{-1}(X-M)=1. \tag{2.4.5}$$

Таким образом, как видим из (2.2.19), плотность вероятности постоянна на этой поверхности.

Для доказательства введем квадратичную форму $U^T B_X U$, где $U=U_{n1}=\begin{Vmatrix} u_1 \\ \vdots \\ u_n \end{Vmatrix}$. Если $u_i=a_i$ $(i=1, 2, \ldots, n)$ и $A=A_{1n}=\|a_1, \ldots, a_n\|$, то $U^T B_X U=AB_XA^T$. Сделаем ортогональное преобразование $U=FV$, $FF^T=E$, такое, что

$$F^T B_X F=\begin{Vmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & \sigma_n^2 \end{Vmatrix}=D,$$

где D — диагональная матрица.

Единичному вектору $\|a_1, \ldots, a_n\|$ в старой системе координат будет отвечать также единичный вектор $\|b_1, \ldots, b_n\|$ в новой системе координат, в направлении которого будет откладываться величина $\sum_{i=1}^{n}\sigma_i^2 b_i^2$. Таким образом, получится корреляционная поверхность.

Преобразуем эту поверхность, полагая $w_i=\frac{b_i}{\sigma_i}$ $(i=1, 2, \ldots, n)$. В новой поверхности в направлении $(b_1, \ldots, b_n)$ будет откладываться расстояние $\sum_{i=1}^{n} b_i^2=1$, так что получится сфера единичного радиуса. Значит, прежняя поверхность была эллипсоидом с полуосями $\sigma_1, \ldots, \sigma_n$ и уравнением

$$V^T D^{-1} V=\sum_{i=1}^{n}\frac{v_i^2}{\sigma_i^2}=1.$$

В старых координатах ее уравнение ввиду $V=F^T U$ имеет вид

$$U^T F D^{-1} F^T U=U^T B_X^{-1} U=1,$$

так как $F^T B_X F=D$; $B_X=FDF^{-1}$; $B_X^{-1}=FD^{-1}F^{-1}=FD^{-1}F^T$. Итак, для корреляционной поверхности имеем уравнение

$$U^T B_X^{-1} U=1, \tag{2.4.6}$$

что, при замене $U = X - M$ (где X понимается как обычный вектор $\left\| \begin{matrix} x_1 \\ \vdots \\ x_n \end{matrix} \right\|$), приводит к (2.4.5).

Рассмотрим частные случаи $n = 1$ и $n = 2$. При $n = 1$ имеем случайную величину; $M = m$; $B_X = \sigma^2$, и (2.4.5) превращается в

$$|x - m| = \sigma; \quad x = m \pm \sigma \tag{2.4.7}$$

(корреляционный отрезок).

При $n = 2$ из (2.2.20) имеем

$$\frac{1}{\sigma_1^2}(x_1 - m_1)^2 - \frac{2\rho}{\sigma_1\sigma_2}(x_1 - m_1)(x_2 - m_2) + \frac{1}{\sigma_2^2}(x_2 - m_2)^2 = 1 \tag{2.4.8}$$

(корреляционный эллипс).

Если корреляционный эллипсоид (2.4.5) растянуть в $\sqrt{n+2}$ раз, то получается эллипсоид

$$(X - M)^T B_X^{-1} (X - M) = n + 2, \tag{2.4.9}$$

который называется *эллипсоидом рассеяния*, или *эллипсоидом концентрации* (см. Г. Крамер [25], стр. 331). Плотность вероятности вектора X, очевидно, также постоянна на этом эллипсоиде.

Вероятностный смысл эллипсоида концентрации следующий: если по данному нормальному вектору X разыскивать случайный вектор R, который был бы равномерно распределен по некоторому эллипсоиду, имея внутри него постоянную плотность вероятности и нулевую плотность вероятности вне его, и для которого было бы $E(X) = E(R)$, $B_X = B_R$ (т. е. у которого вектор средних и корреляционная матрица совпадали бы с данными для X), то этот эллипсоид определялся бы однозначно уравнением (2.4.9).

Доказательство этого утверждения для $n = 2$ можно найти в книге Г. Крамера ([25], стр. 312—314); оно легко может быть проведено в общем случае матричным методом, но мы не будем на нем останавливаться, так как для теории метода наименьших квадратов такая интерпретация не понадобится.

Рассмотрим тот же n-мерный нормальный вектор X с $E(X) = M$ и данной корреляционной матрицей B_X. Пусть дан вектор

$$Z = Z_{n1} = \left\| \begin{matrix} z_1 \\ \vdots \\ z_n \end{matrix} \right\|.$$

Составим скалярное произведение $\sum (z_i - m_i)(x_i - m_i) = [Z - M, \quad X - M]$. Дисперсия этого произведения будет

$(Z-M)^T B_X (Z-M)$. Совокупность точек $(z_1, \ldots, z_n)$, где эта дисперсия равна единице, будет задаваться уравнением

$$(Z-M)^T B_X (Z-M) = 1. \tag{2.4.10}$$

Для каждого вектора в направлении $Z-M$ эллипсоид (2.4.10) задает длину, при которой дисперсия соответствующей проекции будет единичной.

Мы видим, что в левых частях уравнений (2.4.5) и (2.4.10) стоят взаимно обратные квадратичные формы.

Теорема 2.4.2 *). *Если для двух n-мерных нормальных векторов X и Y с общим вектором средних $M = E(X) = E(Y)$ корреляционный эллипсоид Y лежит внутри корреляционного эллипсоида X, то эллипсоид постоянной дисперсии X лежит внутри эллипсоида постоянной дисперсии Y, и обратно.*

Мы не будем приводить здесь доказательства этой теоремы.

§ 5. Сопоставление различных нормальных распределений

Допустим, что мы производим измерение некоторой физической величины m, причем наблюдения сопровождаются случайными погрешностями ξ без систематической погрешности, так что результат наблюдений может быть выражен случайной величиной $X = m + \xi$. Мы будем считать, что X нормальна с параметрами m, σ:

$$X \in N(m, \sigma).$$

Если другой способ измерения той же величины (или другой наблюдатель) дает результаты, выражаемые случайной величиной $X_1 = m + \xi_1 \in N(m, \sigma_1)$, причем $\sigma_1 < \sigma$, то естественно считать эти результаты точнее прежних, так как величина ξ_1 „более сосредоточена“ вблизи нуля, чем ξ. Точный смысл этого высказывания следующий: если задать какое-либо число $\varepsilon > 0$, то уклонения наблюдений от m более чем на ε будут более вероятны в первом случае, чем во втором, т. е.

$$P\{|X_1 - m| > \varepsilon\} < P\{|X - m| > \varepsilon\}. \tag{2.5.1}$$

В самом деле, имеем

$$P\{|X_1 - m| > \varepsilon\} = \frac{2}{\sqrt{2\pi}\,\sigma_1} \int_{\varepsilon}^{\infty} e^{-\frac{t^2}{2\sigma_1^2}}\, dt = \frac{2}{\sqrt{2\pi}} \int_{\varepsilon/\sigma_1}^{\infty} e^{-u^2/2}\, du$$

и аналогично

$$P\{|X - m| > \varepsilon\} = \frac{2}{\sqrt{2\pi}} \int_{\varepsilon/\sigma}^{\infty} e^{-u^2/2}\, du.$$

*) Автору неизвестно, нова ли эта теорема.

Если $\sigma_1 < \sigma$, то $\frac{\varepsilon}{\sigma_1} > \frac{\varepsilon}{\sigma}$, и первый интеграл меньше второго, что и доказывает (2.5.1).

Мы можем сформулировать это же следующим образом: вторые измерения точнее первых, если корреляционный отрезок вторых измерений $[m-\sigma_1,\ m+\sigma_1]$ лежит внутри корреляционного отрезка первых измерений $[m-\sigma,\ m+\sigma]$, или нормальная величина ξ_1 „лучше" чем ξ, если $E(\xi_1)=E(\xi)$ и корреляционный отрезок ξ_1 лежит внутри корреляционного отрезка ξ. Попробуем перенести эти соображения на случай n-мерного нормального вектора. Пусть измеряется без систематической ошибки n физических величин $m_1, \ldots, m_n$ с погрешностями $\xi_1, \ldots, \xi_n$, так что результат измерений выражается случайным вектором

$$X=\begin{Vmatrix} x_1 \\ \vdots \\ x_n \end{Vmatrix}=\begin{Vmatrix} m_1+\xi_1 \\ \vdots \\ m_n+\xi_n \end{Vmatrix}.$$

Во многих задачах случайные погрешности ξ_i (а стало быть, и сами величины x_i) независимы, но иногда приходится иметь дело со статистически зависимыми (связанными) погрешностями. В дальнейшем такие случаи нам встретятся, например, в теории непрямых (косвенных) наблюдений.

Таким образом, в общем случае X будет n-мерным нормальным вектором общего вида. Пусть

$$E(X)=\begin{Vmatrix} m_1 \\ \vdots \\ m_n \end{Vmatrix}; \quad B_X=E\,[(X-M)(X-M)^T]=\|b_{rs}\|.$$

Пусть другой способ измерений (или другой наблюдатель) приводит к результатам, выражаемым n-мерным нормальным вектором $Y=\begin{Vmatrix} y_1 \\ \vdots \\ y_n \end{Vmatrix}$ с тем же вектором средних $E(Y)=\begin{Vmatrix} m_1 \\ \vdots \\ m_n \end{Vmatrix}$, но с другой корреляционной матрицей $B_Y=\|c_{rs}\|$.

Составим уравнения корреляционных эллипсоидов в одной и той же системе координат $(x_1, \ldots, x_n)$:

$$(X-M)^T B_X^{-1}(X-M)=1, \tag{2.5.2}$$

$$(X-M)^T B_Y^{-1}(X-M)=1. \tag{2.5.3}$$

Пусть эллипсоид (2.5.3) расположен целиком внутри эллипсоида (2.5.2). Можно ли считать, как в одномерном случае, что нормальный вектор Y в каком-либо смысле „лучше" нормального вектора X? В этом направлении нетрудно доказать следующее утверждение.

Теорема 2.5.1. *Пусть* $\alpha = (\alpha_1, \ldots, \alpha_n)$ — *какой-либо вектор. Составим скалярные произведения векторов* $X - M$ *и* $Y - M$ *на вектор* α:

$$\xi = \sum_{i=1}^{n} \alpha_i (x_i - m_i); \quad \eta = \sum_{i=1}^{n} \alpha_i (y_i - m_i). \tag{2.5.4}$$

Величины ξ *и* η *будут нормальными, и* $D(\eta) < D(\xi)$, *так что*, как пояснено выше, *для любого заданного* $\varepsilon > 0$ *будем иметь*

$$P(|\eta| > \varepsilon) < P(|\xi| > \varepsilon). \tag{2.5.5}$$

Доказательство. Имеем $E(\xi) = E(\eta) = 0$. На основании сказанного выше, полагая $A = \|\alpha_1, \ldots, \alpha_n\|$, найдем, что

$$D(\xi) = A^T B_X A > A^T B_Y A = D(\eta),$$

что требовалось доказать.

В частности, полагая $\alpha_i = 1$; 0, имеем $D(y_i - m_i) < D(x_i - m_i)$ $(i = 1, 2, \ldots, n)$.

Можно несколько перефразировать полученную теорему. Условие $\left|\sum \alpha_i (x_i - m_i)\right| \leqslant \varepsilon$ определяет плоский слой в n-мерном пространстве с центром симметрии $(m_1, \ldots, m_n)$. При любом $\varepsilon > 0$ вероятность попадания в такой слой для вектора Y больше, чем для вектора X.

Указанные выше утверждения помогут нам выяснить реальный смысл метода наименьших квадратов.

§ 6. Распределения случайных величин, связанных с нормальным распределением, встречающиеся в математической статистике

χ^2-распределение. Во многих отраслях математической статистики играют большую роль квадраты нормальных случайных величин. В методе наименьших квадратов часто приходится рассматривать сумму квадратов независимых случайных погрешностей (без систематических ошибок), которые хорошо описываются нормальными случайными величинами с нулевыми математическими ожиданиями. Ввиду этого большое значение имеет случайная величина

$$\chi_n^2 = x_1^2 + \ldots + x_n^2, \tag{2.6.1}$$

где $x_1, \ldots, x_n$ — независимые нормальные простейшие случайные величины ($x_i \in N(0, 1)$). Эта величина называется „хи-квадрат с n степенями свободы“; n может быть равно 1, 2, 3, ... При $n = 2$ и $n = 3$ величина χ_n^2 имеет наглядный геометрический смысл: это квадрат расстояния от начала координат точки с нормальными случайными компонентами $x_i \in N(0, 1)$.

Вычислим плотность вероятности $k_n(x)$ величины χ_n^2. Имеем

$$P\{x \leqslant \chi_n^2 < x + dx\} \approx k_n(x)\,dx. \tag{2.6.2}$$

Ввиду неотрицательности χ_n^2 имеем $k_n(x) = 0$ при $x < 0$. При $n = 1$, $k_1(x)$ вычисляется непосредственно. Имеем при $x > 0$

$$P\{x \leqslant \chi_1^2 < x + dx\} = P\left\{x^{\frac{1}{2}} \leqslant x_1 < (x+dx)^{\frac{1}{2}}\right\} + \\ + P\left\{-(x+dx)^{\frac{1}{2}} \leqslant x_1 < -x^{\frac{1}{2}}\right\} = 2P\left\{x^{\frac{1}{2}} \leqslant x_1 < (x+dx)^{\frac{1}{2}}\right\}.$$

Далее,

$$(x + dx)^{\frac{1}{2}} = x^{\frac{1}{2}} + \frac{dx}{2x^{\frac{1}{2}}} + \ldots,$$

так что

$$2P\left\{x^{\frac{1}{2}} \leqslant x_1 < (x+dx)^{\frac{1}{2}}\right\} \approx 2P\left\{x^{\frac{1}{2}} \leqslant x_1 < x^{\frac{1}{2}} + \frac{dx}{2x^{\frac{1}{2}}}\right\} \approx \\ \approx \frac{2}{\sqrt{2\pi}} e^{-\frac{x}{2}} \frac{dx}{2x^{\frac{1}{2}}} = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} dx,$$

откуда

$$k_1(x) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} \qquad (x > 0).$$

Докажем теперь общий результат:

$$k_n(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \text{ при } x \geqslant 0; \; k_n(x) = 0 \text{ при } x < 0. \tag{2.6.3}$$

Здесь $\Gamma\left(\frac{n}{2}\right)$ — гамма-функция Эйлера, определяемая при $s > 0$ интегралом

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x}\,dx.$$

Мы доказали, что $k_1(x) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}$. Это совпадает с формулой (2.6.3) при $n = 1$ ввиду известного равенства $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, которое легко вывести из очевидного тождества

$$1=\int_0^\infty k_1(x)\,dx=\frac{1}{\sqrt{2\pi}}\int_0^\infty x^{-\frac{1}{2}}e^{-\frac{x}{2}}\,dx=$$
$$=\frac{1}{\sqrt{\pi}}\int_0^\infty y^{\frac{1}{2}-1}e^{-y}\,dy=\frac{1}{\sqrt{\pi}}\Gamma\left(\frac{1}{2}\right).$$

Будем доказывать (2.6.3) по индукции. Допустим, что формула верна при индексах 1, 2, 3, ..., n и докажем ее для индекса $n+1$.

Функция $k_{n+1}(x)$ есть плотность вероятности случайной величины $\chi^2_{n+1}=\chi^2_n+\chi^2_1$, где слагаемые χ^2_n и χ^2_1 — независимые случайные величины с плотностями вероятностей $k_n(x)$ и $k_1(x)$. Поэтому для плотности вероятности $k_{n+1}(x)$ имеем

$$k_{n+1}(x)=\int_0^x k_n(y)\,k_1(x-y)\,dy=C_n\int_0^x y^{\frac{n}{2}-1}(x-y)^{-\frac{1}{2}}e^{-\frac{x}{2}}\,dy$$

согласно предположению индукции; C_n — некоторая константа, зависящая только от n. Отсюда имеем

$$k_{n+1}(x)=C_n e^{-\frac{x}{2}}\int_0^x y^{\frac{n}{2}-1}(x-y)^{-\frac{1}{2}}\,dy=$$
$$=C_n e^{-\frac{x}{2}}x^{\frac{n-1}{2}}\int_0^x\left(\frac{y}{x}\right)^{\frac{n}{2}-1}\left(1-\frac{y}{x}\right)^{-\frac{1}{2}}\frac{dy}{x}=$$
$$=C_n e^{-\frac{x}{2}}x^{\frac{n-1}{2}}\int_0^1 z^{\frac{n}{2}-1}(1-z)^{-\frac{1}{2}}\,dz=C'_n x^{\frac{n-1}{2}}e^{-\frac{x}{2}},$$

где $C'_n=C_n\int_0^1 z^{\frac{n}{2}-1}(1-z)^{-\frac{1}{2}}\,dz$ — новая константа, зависящая только от n. Для вычисления C'_n заметим, что должно быть

$$\int_0^\infty k_{n+1}(x)\,dx=C'_n\int_0^\infty x^{\frac{n-1}{2}}e^{-\frac{x}{2}}\,dx=1.$$

Полагая $\frac{x}{2}=y$, получаем

$$C'_n 2^{\frac{n+1}{2}}\int_0^\infty y^{\frac{n+1}{2}-1}e^{-y}\,dy=1,$$

или, по определению Г-функции, $C'_n \cdot 2^{\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right) = 1$, так что

$$C'_n = \frac{1}{2^{\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right)}$$

и

$$k_{n+1}(x) = \frac{1}{2^{\frac{n+1}{2}} \Gamma\left(\frac{n+1}{2}\right)} x^{\frac{n+1}{2}-1} e^{-\frac{x}{2}} \quad (x > 0),$$

что и требовалось доказать. Таким образом, формула (2.6.3) верна для всех целых положительных n.

В статистических вычислениях приходится применять не $k_n(x)$, а

$$K_n(x) = \int_0^x k_n(t)\, dt = P\{\chi_n^2 < x\};$$

эта функция $\left(\text{по удалении множителя } \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}\right)$ называется иногда неполной Г-функцией. Для нее существуют таблицы с двумя входами: по степеням свободы n и аргументу x.

Отметим некоторые свойства случайной величины χ_n^2. По определению, $\chi_n^2 = x_1^2 + \ldots + x_n^2$, где x_i независимы. Ввиду этого, имеем: $\chi_n^2 = \chi_p^2 + \chi_q^2$, если p и q — любые натуральные числа, $p + q = n$, и χ_p^2 и χ_q^2 независимы. Поэтому мы должны иметь

$$k_n(x) = \int_0^x k_p(y)\, k_q(x - y)\, dy; \quad p + q = n. \tag{2.6.4}$$

Это равенство называется законом композиции плотностей.

Часто встречаются распределения, непосредственно связанные с χ_n^2. Если $X = y_1^2 + \ldots + y_n^2$, где случайные величины $\frac{y_i}{\sigma} \in N(0, 1)$ и независимы, то $\frac{X}{\sigma^2} = \left(\frac{y_1}{\sigma}\right)^2 + \ldots + \left(\frac{y_n}{\sigma}\right)^2$, где $\frac{y_i}{\sigma} \in N(0, 1)$ и независимы. Отсюда $X/\sigma^2 = \chi_n^2$, или $X = \sigma^2 \chi_n^2$.

Встречаются также величины вида $\frac{1}{n} X = \frac{1}{n}(y_1^2 + \ldots + y_n^2)$ (среднее арифметическое y_i^2),

$$\sqrt{X} = \sqrt{y_1^2 + \ldots + y_n^2}; \quad \sqrt{\frac{X}{n}} = \sqrt{\frac{1}{n}(y_1^2 + \ldots + y_n^2)}.$$

Они имеют плотности вероятности, просто связанные с $k_n(x)$ и легко рассчитываемые по формуле (2.4.5) из (2.6.3). Приведем сводку выражений таких плотностей (см. Г. Крамер [25]).

Величина	Плотность вероятности
$X=\sum_{i=1}^{n} y_i^2$	$\frac{1}{\sigma^2} k_n\left(\frac{y}{\sigma^2}\right)=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)\sigma^n} y^{\frac{n}{2}-1} e^{-\frac{y}{2\sigma^2}}$
$\frac{1}{n} X=\frac{1}{n}\sum_{i=1}^{n} y_i^2$	$\frac{n}{\sigma^2} k_n\left(\frac{ny}{\sigma^2}\right)=\frac{\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)\sigma^n} y^{\frac{n}{2}-1} e^{-\frac{ny}{2\sigma^2}}$
$\sqrt{X}=\sqrt{\sum_{i=1}^{n} y_i^2}$	$\frac{2y}{\sigma^2} k_n\left(\frac{y^2}{\sigma^2}\right)=\frac{2}{2^{\frac{n}{2}}\sigma^n\Gamma\left(\frac{n}{2}\right)} y^{n-1} e^{-\frac{y^2}{2\sigma^2}}$
$\sqrt{\frac{X}{n}}=\sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2}$	$\frac{2ny}{\sigma^2} k_n\left(\frac{ny^2}{\sigma^2}\right)=\frac{2\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\sigma^n\Gamma\left(\frac{n}{2}\right)} y^{n-1} e^{-\frac{n}{2\sigma^2}y^2}$

Наглядные примеры получаются при $n=2$ и $n=3$. Если горизонтальное и вертикальное отклонения пробоины от центра мишени — y_1 и y_2 независимы и $y_i \in N(0, \sigma)$, то расстояние $r=\sqrt{y_1^2+y_2^2}$ пробоины от центра мишени имеет плотность вероятности

$$\frac{2y}{\sigma^2} k_2\left(\frac{y^2}{\sigma^2}\right)=\frac{y}{\sigma^2} e^{-\frac{y^2}{2\sigma^2}}.$$

Если $y_i \in N(0, \sigma)$ — статистически независимые компоненты скорости движения молекулы в прямоугольной системе координат, то скорость $r=\sqrt{y_1^2+y_2^2+y_3^2}$ имеет плотность вероятности

$$\frac{2y}{\sigma^2} k_3\left(\frac{y^2}{\sigma^2}\right)=\sqrt{\frac{2}{\pi}}\frac{y^2}{\sigma^3} e^{-\frac{y^2}{2\sigma^2}}.$$

Распределение Стьюдента. Это распределение имеет важное значение при статистических вычислениях, связанных с нормальным законом $N(a, \sigma)$, где σ неизвестно и подлежит определению на основе опытных данных, например при статистической обработке наблюдений с неизвестной точностью. Пусть y; $y_1, y_2, \ldots, y_n$—независимые нормальные величины с нулевыми средними и одинаковой дисперсией σ^2. Безразмерная величина

$$t=\frac{y}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2}} \tag{2.6.5}$$

называется дробью Стьюдента *). Мы видим, что ее распределение не зависит от σ в силу безразмерности t, так что можно, не нарушая общности, считать $\sigma=1$; y, $y_i \in N(0, 1)$. Число n называется числом степеней свободы t. Плотность вероятности t имеет выражение

$$s_n(x)=\frac{1}{\sqrt{\pi n}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}. \tag{2.6.6}$$

Докажем эту формулу. Сперва вычислим

$$S_n(x)=\int_{-\infty}^{x} s_n(t)\,dt=P(t<x)$$

(именно эта функция и нужна в расчетах и затабулирована с двумя входами: по n и x).

Имеем $t=\frac{y}{\sqrt{\frac{1}{n}X}}$ в обозначениях таблицы на стр. 71; полагаем $\sigma=1$. Совместная плотность вероятности y и $\sqrt{\frac{X}{n}}$ в точке (u, v) равна

$$c_n v^{n-1}\exp-\left(\frac{u^2+nv^2}{2}\right), \tag{2.6.7}$$

где

$$c_n=\sqrt{\frac{2}{\pi}}\,\frac{n^{\frac{n}{2}}2^{-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}. \tag{2.6.8}$$

Имеем

$$S_n(x)=P\left(\frac{y}{\sqrt{\frac{1}{n}X}}<x\right)=c_n\iint_{\Omega} v^{n-1}\exp-\frac{u^2+nv^2}{2}\,du\,dv,$$

где Ω — область $\{v>0,\ u\leqslant vx\}$. Для вычисления двойного интеграла введем новые переменные $\frac{u}{v}=\alpha$, $v=\beta$. Тогда получим

$$J=\frac{\partial(u, v)}{\partial(\alpha, \beta)}=\begin{vmatrix} v & 0 \\ 0 & 1 \end{vmatrix}=v=\beta>0,$$

*) Стьюдент (Student) — псевдоним английского статистика В. Госсета (1876—1937).

откуда

$$S_n(x)=c_n\int_0^x d\alpha\int_0^\infty \beta^n \exp-\left(\frac{n+\alpha^2}{2}\beta^2\right)d\beta.$$

Далее, сделаем замену $\beta=z^{\frac{1}{2}}$, $d\beta=\frac{1}{2}z^{-\frac{1}{2}}dz$, так что

$$\int_0^\infty \beta^n \exp-\left(\frac{n+\alpha^2}{2}\beta^2\right)d\beta=\frac{1}{2}\int_0^\infty z^{\frac{n+1}{2}-1}\exp-\left(\frac{n+\alpha^2}{2}z\right)dz=$$

$$=2^{\frac{n-1}{2}}\Gamma\left(\frac{n+1}{2}\right)\frac{1}{(n+\alpha^2)^{\frac{n+1}{2}}},$$

откуда

$$S_n(x)=2^{\frac{n-1}{2}}\Gamma\left(\frac{n+1}{2}\right)c_n\int_{-\infty}^x\frac{d\alpha}{(n+\alpha^2)^{\frac{n+1}{2}}}.$$

Подставляя значение c_n из (2.6.8), находим

$$S_n(x)=\frac{1}{\sqrt{\pi n}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}\int_{-\infty}^x\frac{d\alpha}{\left(1+\frac{\alpha^2}{n}\right)^{\frac{n+1}{2}}},$$

откуда формула (2.6.7) получается дифференцированием по x. При больших значениях n и ограниченных x величина $\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$ в формуле (2.6.6) близка к $e^{-\frac{x^2}{2}}$, а множитель $\frac{1}{\sqrt{\pi n}}\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$ близок к $\frac{1}{\sqrt{2\pi}}$, поэтому распределение Стьюдента близко к нормальному $N(0, 1)$, чем и пользуются при табулировании $S_n(x)$ для большого числа степеней свободы. Это свойство $S_n(x)$ нетрудно было бы усмотреть и из самого определения дроби t.

Таблица распределения Стьюдента $S_n(x)$ (табл. I в приложении) имеет два входа — число степеней свободы n и вероятность (надежность) p_0. По этим двум входам табл. I дает число γ такое, что

$$P\{|t_n|\leqslant\gamma\}=p_0,$$

где t_n — дробь Стьюдента с n степенями свободы, Распределение Фишера. Это распределение встречается в вопросах о сравнении и оценке неизвестных дисперсий нормальных распределений на основе опытных данных о них. Пусть величины $x_1, \ldots, x_m$; $y_1, \ldots, y_n$ независимы и нормальны $N(0, \sigma)$. Составим безразмерную дробь

$$\varkappa = \frac{\sum_{i=1}^{m} x_i^2}{\sum_{i=1}^{n} y_i^2}. \tag{2.6.9}$$

Мы, видим, что распределение дроби $\varkappa$ в силу ее безразмерности не зависит от σ, и можно считать $\sigma = 1$. Пусть $F_{mn}(x) = P(\varkappa < x)$ и $f_{mn}(x) = F'_{mn}(x)$. Очевидно, $f_{mn}(x) = 0$ при $x < 0$. Докажем, что

$$f_{mn}(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{x^{\frac{m}{2}-1}}{(x+1)^{\frac{m+n}{2}}} \quad (x > 0). \tag{2.6.10}$$

Мы можем написать

$$\varkappa = \frac{\chi_m^2}{\chi_n^2}, \tag{2.6.11}$$

где χ_m^2 и χ_n^2 независимы. Дальнейшие выкладки проводятся так же, как в случае распределения Стьюдента. Совместная плотность вероятности величин χ_m^2 и χ_n^2 (в точке (u, v)) имеет вид

$$c_{mn} u^{\frac{m}{2}-1} v^{\frac{n}{2}-1} e^{-\frac{u+v}{2}}, \tag{2.6.12}$$

где

$$c_{mn} = \frac{1}{2^{\frac{m+n}{2}} \Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}.$$

Сперва вычисляется $P(\varkappa < x)$, так что областью интегрирования (2.6.12) будет область Ω: $\{v > 0,\ u \leqslant vx\}$ $(x > 0)$. Введя такую же замену u, v на α, β, как и ранее, находим:

$$F_{mn}(x) = P(\varkappa < x) = c_{mn} \int_0^x \alpha^{\frac{m}{2}-1} d\alpha \cdot \int_0^1 \beta^{\frac{m+n}{2}-1} e^{-\frac{\alpha+1}{2}\beta} d\beta.$$

Подсчет, вполне аналогичный предыдущему, дает

$$F_{mn}(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \int_0^x \frac{\alpha^{\frac{m}{2}-1}}{(\alpha+1)^{\frac{m+n}{2}}} d\alpha. \tag{2.6.13}$$

Дифференцирование по x приводит к нужной формуле (2.6.10). Распределение Фишера $F_{mn}(x)$ затабулировано в виде таблиц *), приложенных в конце книги. Там же объяснены правила пользования таблицами.

§ 7. Приближенно нормальные распределения, их роль в теории вероятностей

В § 1 этой главы мы приводили уже некоторые примеры, в которых было видно, что эмпирическое распределение ошибок наблюдения приближенно нормально (см. табл. 6). Подобных примеров приближенной нормальности эмпирических распределений статистиками собрано очень много; частое появление приближенной нормальности и объясняет само название „нормального закона". Важным классом случаев возникновения приближенно нормального распределения является такой, когда наблюдаемый количественный показатель может быть описан как суммарный эффект большого числа независимых причин. Так обстоит дело с ошибками физических или астрономических измерений, по „гипотезе элементарных ошибок" Хагена и Бесселя и в действительности. Приближенную нормальность распределения этих ошибок можно принять в очень многих случаях, и почти всегда в тех условиях, когда обычно применяется метод наименьших квадратов. Подобная же ситуация встречается в антропометрии при изучении распределения роста и размеров отдельных органов людей одного пола и возраста. Это распределение с хорошим приближением описывается нормальным законом. Много примеров приближенной нормальности распределения показателей можно найти в народном хозяйстве; таковы линейные размеры однотипных деталей в массовом производстве, общее потребление электроэнергии данной электростанции как сумма количеств энергии, потребляемых отдельными абонентами, и т. д.

Разумеется, речь идет лишь о приближенной нормальности; например, многие из перечисленных показателей существенно положительны, тогда как нормальная случайная величина распределена на всей оси $(-\infty, \infty)$. Далее, часты случаи эмпирических распределений и не по нормальному закону. Так, например, происходит при качественной неоднородности рассматриваемой совокупности; скажем, если рассматривать распределение по росту большой группы людей разного пола или возраста, или распределение по размерам деталей различного типа и т. п., то получается „многовершинное", не нормальное распределение (которое, однако, часто можно изобразить как „смесь" нормальных).

Часто также встречаются распределения, тесно связанные с нормальными. Например, логарифмически-нормальное распределение

*) Как следует из исследований П. Л. Чебышева о дифференциальном биноме [53], $F_{mn}(x)$ всегда выражается через элементарные функции.

положительной величины X, т. е. такое, что $\log X$ нормально, хорошо приближает такие эмпирические распределения, как веса песчинок в песке однородного состава, веса крупинок и самородков золота, веса мелких камней, прошедших камнедробилку (см. по этому поводу статью А. Н. Колмогорова [21]).

В. Н. Зимовнов в брошюре [17] рассматривает композицию нормального закона с законом равномерного на некотором отрезке распределения, как один из возможных законов распределения погрешностей в геодезии.

Теоретическим основанием для объяснения появления приближенной нормальности в схеме „суммарного эффекта большого числа независимых причин“ служит центральная предельная теорема А. М. Ляпунова (см. Б. В. Гнеденко [11]). Мы приведем ее здесь без доказательства.

Теорема 2.7.1 (А. М. Ляпунов). *Пусть* $S_n = Y_1 + \ldots + Y_n$ — *сумма* n *независимых случайных величин, имеющих математические ожидания* $E(Y_i) = a_i$, *дисперсии* $D(Y_i) = b_i$ *и абсолютные третьи моменты* $\gamma_i = E\,|\,Y_i - a_i\,|^3$.

Рассмотрим нормированное уклонение суммы S_n, т. е. *величину*

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{D(S_n)}}\,; \textit{ пусть } F_n(z) = P(Z_n < z).$$

Составим „дробь Ляпунова“ $L_n = \dfrac{\sum\limits_{i=1}^{n} \gamma_i}{[D(S_n)]^{\frac{3}{2}}}$. *Если* $L_n \to 0$ *при* $n \to \infty$, *то для любого заданного числа* z

$$F_n(z) \to \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{z} e^{-\frac{t^2}{2}} dt \tag{2.7.1}$$

равномерно по z.

Мы будем соотношение (2.7.1) выражать словами: Z_n *асимптотически нормальна* $N(0, 1)$, а S_n асимптотически нормальна $N\left(E(S_n), \sqrt{D(S_n)}\right)$.

Условие $L_n \to 0$ количественно выражает требование того, чтобы ни одно из слагаемых не доминировало в S_n, т. е. требование „предельной пренебрегаемости“ слагаемых.

В частном случае, когда все слагаемые Y_i одинаково распределены, достаточно одного лишь существования дисперсии для выполнения (2.7.1).

Центральная предельная теорема хорошо объясняет появление приближенной нормальности в схеме суммы независимых, предельно пренебрегаемых факторов. Часто, однако, встречаются ситуации, когда отдельные факторы не являются независимыми.

Для отдельных типов „слабой зависимости" и здесь удается обосновать приближенную нормальность (см. по этому поводу С. Н. Бернштейн [4], М. Лёв [31]).

Отметим также важную для нас ситуацию, встречающуюся при обработке наблюдений, когда составляются функции $\Phi(x_1, \ldots, x_n)$ от „равноточных" погрешностей наблюдений (см. об этом подробнее в гл. III и IV), которые должны приближенно выражать некоторые измеряемые факторы, — такие функции, как, например, $\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2$ и др. Эти функции в большинстве случаев бывают симметрическими (что выражает равноправие равноточных измерений). Оказывается, для широкого класса таких симметрических функций можно теоретически обосновать приближенную нормальность. Приведем (без доказательства) соответствующую теорему В. Хёфдинга [49].

Теорема 2.7.2 (В. Хёфдинг*)). *Пусть $x_1, \ldots, x_n$ — независимые в совокупности случайные величины, $\Phi(x_1, \ldots, x_m)$ (m фиксировано) — интегрируемая функция от них, такая, что существует $E\Phi^2(x_1, \ldots, x_n)$ (а стало быть, и $E\Phi(x_1, \ldots, x_n) = \theta$). Составим симметрическую функцию*

$$U = \frac{1}{n(n-1)\ldots(n-m+1)} \sum_{(\alpha_1, \ldots, \alpha_m)} \Phi(x_{\alpha_1}, \ldots, x_{\alpha_m}),$$

где суммирование идет по всем выборкам $(\alpha_1, \ldots, \alpha_m)$ из n различных чисел. Тогда при $n \to \infty$ распределение случайной величины $\sqrt{n}(U - \theta)$ стремится к нормальному распределению.

Закон больших чисел.

Для осмысливания реального значения вероятностных вычислений часто достаточно применение закона больших чисел в форме П. Л. Чебышева. Эта теорема менее точна, чем центральная предельная теорема, но доказывается при менее жестких условиях.

Теорема 2.7.3. (Закон больших чисел в форме П. Л. Чебышева.) *Пусть $Y_1, \ldots, Y_n$ — попарно независимые случайные величины, имеющие математические ожидания $E(Y_i) = a_i$ и дисперсии $D(Y_i) \leqslant L$, где L — одно и то же для $i = 1, 2, \ldots, n$. Пусть $S_n = Y_1 + \ldots + Y_n$; $A_n = a_1 + \ldots + a_n$. Тогда для любого $\varepsilon > 0$*

$$P\left\{\left|\frac{S_n}{n} - \frac{A_n}{n}\right| \leqslant \varepsilon\right\} > 1 - \frac{L}{n\varepsilon^2}. \tag{2.7.2}$$

При больших n и заданных ε дробь $\frac{S_n}{n}$ будет, таким образом, с вероятностью, сколь угодно близкой к единице, лежать не более чем на расстоянии ε от дроби $\frac{A_n}{n}$.

---

*) Для простоты изложения формулировка В. Хёфдинга несколько сужается.

ГЛАВА III

НЕОБХОДИМЫЕ СВЕДЕНИЯ ИЗ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

§ 1. Выборка. Статистика

Основным понятием математической статистики является *выборка*, или совокупность, наблюдений $X = (x_1, \ldots, x_n)$ какого-либо количественного показателя. Если этот показатель является случайной величиной, то выборка $X = (x_1, \ldots, x_n)$ является случайным вектором. Число n называется *объемом выборки*. Выборка называется *повторной*, если компоненты вектора X независимы и одинаково распределены, так что $P(X_i < x) = F_0(x)$, $i = 1, 2, \ldots, n$ ($F_0(x)$ — функция распределения каждой компоненты), и *бесповторной*, если это условие не выполнено. Всякая функция $g(x_1, \ldots, x_n)$ наблюдений $(x_1, \ldots, x_n)$ называется *статистикой*; статистика есть некоторая сводная характеристика наблюдений. Мы будем рассматривать статистики лишь двух типов: непрерывные статистики и целочисленные статистики (в основном, первые).

Примеры непрерывных статистик и их общепринятые обозначения:

выборочное среднее $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$;

выборочная дисперсия $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$;

выборочный третий центральный момент $m_3 = \frac{1}{n}\sum (x_i - \bar{x})^3$;

выборочный коэффициент изменчивости $v = \frac{s}{\bar{x}}$;

наибольший элемент выборки $x_{\max}$;
наименьший элемент выборки $x_{\min}$.

Пример целочисленных статистик (*статистики В. Н. Гостева*). Пусть выборка $x_1, x_2 \ldots, x_n$ представляет собой диаметры валиков, выбранных из текущей продукции токарного станка, и имеются две скобы ширины u_1, u_2; $u_1 < u_2$. Пусть n^- —

число валиков, проходящих в первую скобу, n^0 — число валиков, не проходящих в первую скобу и проходящих во вторую, а n^+ — число остальных валиков ($n^- + n^0 + n^+ = n$). Величины n^- и n^+ будут целочисленными статистиками, характеризующими распределение выборки по размерам валиков.

§ 2. Оценивание параметров

К вопросам оценивания параметров для наших целей естественнее всего подойти с точки зрения теории ошибок наблюдений. Рассмотрим простейший случай n независимых измерений физической величины m, в результате которых получается повторная выборка $(x_1, x_2, \ldots, x_n)$ наблюдений $x_i = m + \Delta_i$, где Δ_i — независимые нормальные погрешности с нулевым средним и дисперсией σ_0^2, т. е. $\Delta_i \in N(0, \sigma_0)$. (Такие погрешности называют несмещенными погрешностями с известной одинаковой точностью.) Сами наблюдения $x_i \in N(m, \sigma_0)$, так что плотность вероятности наблюдений есть

$$f(x, m) = \frac{1}{\sqrt{2\pi}\,\sigma_0} \exp - \frac{(x-m)^2}{2\sigma_0^2}. \tag{3.2.1}$$

В данном случае мы знаем величину σ_0 и не знаем m.

Плотность вероятности случайного вектора $X = (x_1, \ldots, x_n)$ (повторной выборки) имеет вид

$$L(x_1, \ldots, x_n, m) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_0^n} \exp - \frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - m)^2}{\sigma_0^2}. \tag{3.2.2}$$

По отношению к этой плотности вероятности неизвестная величина m является параметром распределения. Мы хотим получить представление о значении неизвестного параметра m на основании наблюдений $x_1, \ldots, x_n$ (выборки). Такая задача называется *задачей оценивания параметра m на основании выборки.* Мы видим, что в данном случае это и есть задача измерения физической величины при наличии случайных погрешностей наблюдений.

Общая постановка этой задачи такова. Пусть имеется выборка $x_1, \ldots, x_n$, повторная или бесповторная, с плотностью вероятности

$$L(x_1, \ldots, x_n;\ \alpha_1, \alpha_2, \ldots, \alpha_s), \tag{3.2.3}$$

зависящей от s параметров $\alpha_1, \ldots, \alpha_s$. Плотность вероятности выборки в математической статистике носит особое название — *функция правдоподобия выборки.* На основании выборки $(x_1, \ldots, x_n)$ мы должны составить приближения к значениям параметров $\alpha_1, \ldots, \alpha_s$ (оценить параметры $\alpha_1, \ldots, \alpha_s$). Для этого составляем и вычисляем s соответствующим образом выбранных функций наблюдений: $g_1(x_1, \ldots, x_n)$, $g_2(x_1, \ldots, x_n)$, $\ldots$, $g_s(x_1, \ldots, x_n)$ (s статистик).

Эти статистики называются *оценками* соответственно для $\alpha_1, \ldots, \alpha_s$; оценка $g_i(x_1, \ldots, x_n)$ для параметра α_i есть случайная величина, дающая в некотором смысле „хорошее“ приближение к параметру. (В дальнейшем мы будем употреблять более совершенный способ оценивания с помощью так называемых доверительных интервалов, но сделаем это непосредственно для метода наименьших квадратов.) Для случая измерения одной физической величины, которому отвечает функция правдоподобия (3.2.2), „хорошей“ и в некотором смысле „наилучшей“ оценкой является статистика $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Заметим, что она обладает следующими свойствами:

$$E(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = m; \quad D(\bar{x}) = \frac{1}{n^2}\sum_{i=1}^{n} D(x_i) = \frac{\sigma^2}{n}. \qquad (3.2.4)$$

В случае повторной выборки функция правдоподобия имеет вид

$$L(x_1, \ldots, x_n; \alpha_1, \ldots, \alpha_s) = \prod_{i=1}^{n} f(x_i; \alpha_1, \ldots, \alpha_s). \qquad (3.2.5)$$

где $f(x; \alpha_1, \ldots, \alpha_s)$ — плотность вероятности x_i.

В случае повторной выборки наблюдения $x_1, x_2, \ldots, x_n$ являются вполне „равноправными“, и поэтому естественно в качестве статистик $g_1 = g_1(x_1, \ldots, x_n), \ldots, g_s = g_s(x_1, \ldots, x_n)$ для оценок параметров $\alpha_1, \ldots, \alpha_s$ употреблять симметрические функции наблюдений. При большом объеме выборки n эти функции, вообще говоря, будут приближенно нормальными (см. теорему В. Хёфдинга 2.7.2). Случайный вектор $G = \begin{Vmatrix} g_1 \\ \vdots \\ g_n \end{Vmatrix}$, составленный из этих статистик, будет также в довольно общем случае приближенно (асимптотически) нормальным. Это весьма повышает роль асимптотически нормальных статистик в теории оценивания параметров. Кроме того, наиболее часто используемые методы оценивания: метод максимального правдоподобия и метод моментов, о которых будет сказано позже, приводят к асимптотически нормальным оценкам, что также увеличивает значение этих оценок. Таким образом, в большинстве случаев оценивания параметров при больших выборках мы будем иметь дело с асимптотически нормальными в совокупности оценками.

Введем еще важное понятие несмещенных и асимптотически несмещенных оценок параметров.

В задаче об оценке физической величины при помощи выборочного среднего $\bar{x}$, мы имели в рассмотренной там ситуации соотношение

$$E(\bar{x}) = a,$$

где a — измеряемая физическая величина; $\bar{x}$ можно назвать *несмещенной оценкой* a, или оценкой без систематической погрешности. В общем случае, вектор $G = \left\| \begin{matrix} g_1 \\ \cdot \\ \cdot \\ \cdot \\ g_s \end{matrix} \right\|$; $g_i = g_i(x_1, \ldots, x_n)$, назовем *вектором несмещенных оценок* для параметров $\alpha_1, \ldots, \alpha_s$, или вектором оценок без постоянных ошибок, если

$$E(g_i) = \alpha_i \quad (i = 1, 2, \ldots, s). \tag{3.2.6}$$

Почему нужно стараться получить несмещенные оценки для измеряемых параметров $\alpha_1, \alpha_2, \ldots; \alpha_s$? Мы уже видели, что, вообще говоря, при больших выборках будем иметь дело с асимптотически нормальными оценками; согласно формуле (2.1.3), мы будем при этом иметь

$$P\{E(g_i) - 1{,}96\sigma(g_i) \leqslant g_i \leqslant E(g_i) + 1{,}96\sigma(g_i)\} \approx 0{,}95$$

при больших выборках. При точных измерениях [т. е. при малом $\sigma(g_i)$], если $E(g_i) \neq \alpha_i$, то g_i часто будет попадать в окрестность $E(g_i)$, а не α_i, и тем самым измерять параметр с систематической ошибкой, равной $E(g_i) - \alpha_i$, что недопустимо.

Часто, однако, наряду с несмещенными оценками g_i для параметров α_i применяются *асимптотически несмещенные* оценки, т. е. такие, что $E(g_i) - \alpha_i \to 0$ при увеличении выборки и выборе последовательности измеряющих α_i статистик g_i. Приведем пример. Пусть для повторной выборки $(x_1, x_2, \ldots, x_n)$ из нормальных наблюдений $x_i \in N(a, \sigma)$ надо дать оценку для σ^2. Если составим выборочную дисперсию

$$s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2,$$

то получаем

$$E(s^2) = \frac{1}{n} \sum_{i=1}^{n} E(x_i^2) - E(\bar{x}^2) = \left(1 - \frac{1}{n}\right)\sigma^2, \tag{3.2.7}$$

как показывает непосредственный подсчет. Таким образом, эта оценка для σ^2 смещена в меньшую сторону. Ясно, далее, что несмещенной оценкой для σ^2 будет оценка

$$s_1^2 = \frac{n}{n-1} s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

так как

$$E(s_1^2) = \frac{n}{n-1} E(s^2) = \sigma^2.$$

Однако при большом объеме выборки s^2 будет асимптотически несмещенной оценкой σ^2, так как $E(s^2)-\sigma^2=-\frac{\sigma^2}{n}\to 0$ при $n\to\infty$.

§ 3. Как точно можно оценивать параметры при заданном числе наблюдений

Неравенства Рао — Крамера*). Имея дело с несмещенными или асимптотически несмещенными и приближенно нормальными оценками g_i параметров α_i, мы будем, естественно, считать дисперсии оценок $D(g_i)$ характеристиками точности оценок. Если имеем две оценки указанного типа g и h, причем $D(g)=\sigma^2$, $D(h)=\rho^2$, $\rho>\sigma$, то в силу (2.5.1) и (2.1.3) получим, что для данного $\varepsilon>0$ при больших выборках имеют место соотношения

$$P\{|g-\alpha|<\varepsilon\}>P\{|h-\alpha|<\varepsilon\}, \tag{3.3.1}$$

$$P\{|g-\alpha|\leqslant 1{,}96\sigma(g)\}\approx 0{,}95;\ P\{|h-\alpha|\leqslant 1{,}96\sigma(h)\}\approx 0{,}95. \tag{3.3.2}$$

Из этих соотношений видно, что естественной мерой точности будет $\frac{1}{\sigma(g_i)}=\frac{1}{\sqrt{D(g_i)}}$ — величина, которую можно назвать *весом оценки*.

Если оценка g_i не является приближенно нормальной, то соотношения (3.3.1) и (3.3.2), вообще говоря, не будут иметь места. Однако, в силу неравенства Чебышева (2.1.9), для несмещенных оценок g_i получим

$$P\{|g_i-\alpha|\leqslant t\sigma_i\}\geqslant 1-\frac{1}{t^2} \tag{3.3.3}$$

для любых $t>0$, что также оправдывает до известной степени понимание $\frac{1}{\sigma_i}=\frac{1}{\sqrt{D(g_i)}}$ как меры точности и в общем случае.

Встает вопрос, насколько точно в указанном смысле можно оценить параметры при данном числе наблюдений (есть ли граница снизу для дисперсии оценки параметра при данном объеме выборки?).

Эта задача поставлена и решена независимо друг от друга С. Рао [43] и Г. Крамером [24]. Здесь будут изложены теоремы Рао — Крамера сперва для оценки одного, а затем — нескольких параметров. Мы будем следовать изящному изложению Рао [42] **) и ограничимся случаем несмещенных оценок.

Обратимся к случаю одного параметра.

---

*) С. Радхакришна Рао (род. 1921) — индийский статистик-математик. Г. Крамер (род. 1893) — шведский математик.

**) Заполняя некоторые небольшие пробелы в этом изложении.

Пусть функция правдоподобия выборки зависит от параметра α, так что имеет вид $L(x_1, \ldots, x_n, \alpha)$, и пусть $t(x_1, \ldots, x_n)$ — несмещенная оценка для α. Имеем тогда

$$\int \ldots \int_R L(x_1, \ldots, x_n, \alpha)\, dx_1, \ldots, dx_n = 1. \tag{3.3.4}$$

Здесь R — все „выборочное пространство", т. е. набор всех точек $(x_1, \ldots, x_n)$. Будем считать в дальнейшем законными употребляемые нами дифференцирования под знаком интеграла по параметрам. Имеем

$$\int \ldots \int_R \frac{\partial L}{\partial \alpha} dx_1 \ldots dx_n = 0. \tag{3.3.5}$$

Далее, в силу того, что $Et(x_1, \ldots, x_n) = \alpha$, имеем

$$\int \ldots \int_R tL\, dx_1 \ldots dx_n = \alpha,$$

откуда

$$\int \ldots \int_R t \frac{\partial L}{\partial \alpha} dx_1 \ldots dx_n = 1, \tag{3.3.6}$$

или, в силу (3.3.5),

$$\int \ldots \int_R [t(x_1, \ldots, x_n) - \alpha] \frac{\partial}{\partial \alpha} L(x_1, \ldots, x_n, \alpha)\, dx_1 \ldots dx_n = 1. \tag{3.3.7}$$

Для сокращения записи будем опускать аргументы, и перепишем (3.3.7) в виде

$$\int \ldots \int_R (t - \alpha)\left(\frac{1}{L}\frac{\partial L}{\partial \alpha}\right) \cdot L\, dx_1 \ldots dx_n = 1. \tag{3.3.8}$$

Если $L = L(x_1, \ldots, x_n, \alpha)$ в некоторых точках может обращаться в нуль, то мы попросту считаем $\frac{\partial}{\partial \alpha} L = \frac{1}{L}\frac{\partial L}{\partial \alpha} \cdot L$, так что интеграл будет иметь смысл. Теперь (3.3.8) можно переписать так:

$$E(t - \alpha)\left(\frac{1}{L}\frac{\partial L}{\partial \alpha}\right) = 1. \tag{3.3.9}$$

Важно заметить, что здесь $L = L(x_1, \ldots, x_n, \alpha)$ считается случайной величиной — функцией случайного вектора $(x_1, \ldots, x_n)$.

Полагая $t - \alpha = X$; $\frac{1}{L}\frac{\partial L}{\partial \alpha} = Y$, находим

$$E(XY) = 1. \tag{3.3.10}$$

Далее, из неравенства (2.1.16) имеем

$$EX^2 \cdot EY^2 \geqslant (E(XY))^2.$$

Отсюда и из (3.3.10) получаем

$$E(t-\alpha)^2 \cdot E\left(\frac{1}{L}\frac{\partial L}{\partial \alpha}\right)^2 \geqslant 1, \tag{3.3.11}$$

если только величины в левой части существуют как несобственные интегралы.

Сделаем еще небольшое преобразование выражения $E\left(\frac{1}{L}\frac{\partial L}{\partial \alpha}\right)^2$. Имеем (при $L \neq 0$)

$$\frac{\partial \ln L}{\partial \alpha} = \frac{1}{L}\frac{\partial L}{\partial \alpha};$$

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = -\frac{1}{L^2}\left(\frac{\partial L}{\partial \alpha}\right)^2 + \frac{1}{L}\frac{\partial^2 L}{\partial \alpha^2}. \tag{3.3.12}$$

Если умножим обе части (3.3.12) на L и проинтегрируем по всему пространству R, получим математические ожидания входящих туда величин. Далее,

$$E\left(\frac{1}{L}\frac{\partial^2 L}{\partial \alpha^2}\right) = \int \underset{R}{\ldots} \int \frac{\partial^2 L}{\partial \alpha^2} dx_1 \ldots dx_n = 0,$$

поэтому

$$E\left(\frac{1}{L^2}\frac{\partial^2 L}{\partial \alpha^2}\right) = -E\frac{\partial^2 \ln L}{\partial \alpha^2}.$$

Следовательно, (3.3.11) можно переписать так:

$$D(t) = E(t-\alpha)^2 \geqslant \frac{1}{E\left(-\frac{\partial^2 \ln L}{\partial \alpha^2}\right)}. \tag{3.3.13}$$

Величина

$$E\left(-\frac{\partial^2 \ln L}{\partial \alpha^2}\right) = -\int \underset{R}{\ldots} \int \frac{\partial^2 \ln L}{\partial \alpha^2} dx_1 \ldots dx_n$$

(разумеется, предполагается сходимость выражающего ее интеграла) называется *информационным количеством Фишера*, по имени впервые рассматривавшего ее ученого. Эта величина не зависит от способа оценки α, т. е. от статистики t, и представляет собой нижнюю границу точности любой оценки.

Мы пришли к следующему выводу: при данном объеме выборки точность несмещенной оценки параметра будет ограничена снизу. Разберем некоторые важные частные случаи. Пусть выборка из наблюдений повторная, так что

$$L(x_1, \ldots, x_n, \alpha) = f(x_1, \alpha) \ldots f(x_n, \alpha) \quad \text{и} \quad f(x_i, \alpha) \neq 0.$$

Тогда

$$\frac{\partial^2 \ln L}{\partial \alpha^2} = \sum_{i=1}^{n} \frac{\partial^2 \ln f(x_i, \alpha)}{\partial \alpha^2}$$

и

$$E\frac{\partial^2 \ln L}{\partial \alpha^2} = \sum_{i=1}^{n} E\frac{\partial^2 \ln f(x_i, \alpha)}{\partial \alpha^2} = nE\frac{\partial^2 \ln f(x_1, \alpha)}{\partial \alpha^2},$$

ибо, очевидно, величины

$$E\frac{\partial^2 \ln f(x_i, \alpha)}{\partial \alpha^2} \qquad (i = 1, 2, \ldots n)$$

все равны между собой. Отсюда

$$D(t) \geqslant \frac{1}{nE\left(-\frac{\partial^2 \ln f(x_1, \alpha)}{\partial \alpha^2}\right)}, \tag{3.3.14}$$

так что

$$\frac{1}{\sqrt{D(t)}} \leqslant \sqrt{nE\left(-\frac{\partial^2 \ln f(x_1, \alpha)}{\partial \alpha^2}\right)}.$$

Мы приходим к выводу: в весьма широком классе несмещенных оценок вес оценки при повторной выборке не может быть больше величины, пропорциональной квадратному корню из числа наблюдений.

Оценка t, для которой в неравенстве (3.3.14) достигается знак равенства, называется *эффективной*.

Вернемся к случаю измерений физической величины $\alpha = m$ с нормальным законом ошибок, рассмотренному в § 2 этой главы.

Здесь имеем:

$$L = \prod_{i=1}^{n} f(x_i, m); \quad f(x_i, m) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp - \frac{(x_i - m)^2}{2\sigma^2};$$

$$\ln f(x_1, m) = \ln\left(\sqrt{2\pi}\,\sigma\right) - \frac{(x_1 - m)^2}{2\sigma^2};$$

$$\frac{\partial^2 \ln f(x_1, m)}{\partial m^2} = -\frac{1}{\sigma^2};$$

$$E\left(-\frac{\partial^2 \ln f(x_1, m)}{\partial m^2}\right) = E\left(\frac{1}{\sigma^2}\right) = \frac{1}{\sigma^2},$$

откуда

$$D(t) \geqslant \frac{\sigma^2}{n}. \tag{3.3.15}$$

Для специальной оценки $t = \bar{x}$, имеем

$$D(\bar{x}) = \frac{\sigma^2}{n}, \tag{3.3.16}$$

т. е. эта оценка имеет максимальную возможную точность. Помимо этого, реальный смысл этого свойства оценки в весьма широком классе асимптотически несмещенных и асимптотически нормальных оценок t можно пояснить на основании § 6 гл. II [см. формулу (2.5.1)]:

$$P\{|\bar{x}-m|>\varepsilon\} \leqslant P\{|t-m|>\varepsilon\} \tag{3.3.17}$$

при больших выборках и заданном ε.

Перейдем к случаю нескольких оцениваемых параметров $\alpha_1, \ldots, \alpha_s$. Здесь функция правдоподобия выборки имеет вид

$$L=L(x_1, \ldots, x_n; \ \alpha_1, \ldots, \alpha_s).$$

Пусть

$$t_1=t_1(x_1, \ldots, x_n), \ \ldots, \ t_s=t_s(x_1, \ldots, x_n) \tag{3.3.18}$$

— несмещенные оценки параметров $\alpha_1, \ldots, \alpha_s$. Составим количества

$$L_{ij}=-\frac{\partial^2 \ln L}{\partial\alpha_i\,\partial\alpha_j}, \tag{3.3.19}$$

которые будут случайными величинами, так как зависят от случайной выборки. Затем рассмотрим математические ожидания этих величин

$$I_{ij}=E(L_{ij}); \qquad i, j=1, 2, \ldots, s, \tag{3.3.20}$$

и из этих количеств составим матрицу

$$I=I_{(ss)}=\|I_{ij}\|, \tag{3.3.21}$$

которая называется *информационной матрицей Фишера*. Заметим, что при $X=\begin{Vmatrix} x_1 \\ \cdot \\ \vdots \\ \cdot \\ x_n \end{Vmatrix}$ квадратичная форма $X^T I X$ является положительно определенной или полуопределенной. Причина этого заключается в том, что

$$E\left(-\frac{\partial^2 \ln L}{\partial\alpha_i\,\partial\alpha_j}\right)=E\left(\frac{1}{L}\frac{\partial L}{\partial\alpha_i}\cdot\frac{1}{L}\frac{\partial L}{\partial\alpha_j}\right), \tag{3.3.22}$$

так что I есть корреляционная матрица для случайного вектора

$$\left(\frac{1}{L}\frac{\partial L}{\partial\alpha_1}, \ \ldots, \ \frac{1}{L}\frac{\partial L}{\partial\alpha_s}\right) \tag{3.3.23}$$

с нулевым вектором средних.

Докажем (3.3.22). Имеем

$$\frac{\partial \ln L}{\partial\alpha_i}=\frac{1}{L}\frac{\partial L}{\partial\alpha_i}; \qquad \frac{\partial^2 \ln L}{\partial\alpha_i\,\partial\alpha_j}=\frac{1}{L}\frac{\partial^2 L}{\partial\alpha_i\partial\alpha_j}-\frac{1}{L^2}\frac{\partial L}{\partial\alpha_i}\frac{\partial L}{\partial\alpha_j},$$

откуда, как и ранее, находим

$$E\left(-\frac{\partial^2 \ln L}{\partial\alpha_i\partial\alpha_j}\right)=E\left(\frac{1}{L}\frac{\partial L}{\partial\alpha_i}\cdot\frac{1}{L}\frac{\partial L}{\partial\alpha_j}\right),$$

что и требовалось доказать.

Будем теперь считать информационную матрицу Фишера I неособенной: $\det(I) \neq 0$, так что будет существовать I^{-1}. Корреляционный эллипсоид случайного вектора (3.3.23) будет иметь вид

$$X^T I^{-1} X = 1 \tag{3.3.24}$$

[см. формулу (2.4.5)].

Рассмотрим теперь набор несмещенных оценок $t_1, \ldots t_s$ из (3.3.18) и составим случайный вектор уклонений

$$(t_1 - \alpha_1,\ t_2 - \alpha_2,\ \ldots,\ t_s - \alpha_s), \tag{3.3.25}$$

который имеет нулевой вектор средних, корреляционную матрицу

$$B_t = \|b_{ij}\| = \|E(t_i - \alpha_i)(t_j - \alpha_j)\|,\quad i, j = 1, 2, \ldots, s \tag{3.3.26}$$

и корреляционную квадратичную форму

$$Z^T B_t Z;\quad Z = \begin{Vmatrix} z_1 \\ \vdots \\ z_s \end{Vmatrix}. \tag{3.3.27}$$

Имеет место теорема, обобщающая неравенство (3.3.13).

***Теорема 3.3.1* (*Рао — Крамер*).** *При условии существования и неособенности информационной матрицы Фишера* $I = \left\| E\left(-\frac{\partial^2 \ln L}{\partial \alpha_i \partial \alpha_j}\right)\right\|$ *и величин* $E\left(t_j \frac{1}{L} \frac{\partial L}{\partial \alpha_i}\right)$ *при любом* $Z = \begin{Vmatrix} z_1 \\ \vdots \\ z_s \end{Vmatrix}$ *имеем*

$$Z^T B_t Z \geqslant Z^T I^{-1} Z, \tag{3.3.28}$$

где B_t — *корреляционная матрица вектора статистик* $t_1, \ldots$ $\ldots, t_s$ *).

Доказательство. Мы должны доказать, что для всех Z

$$Q = Z^T (B_t - I^{-1}) Z \geqslant 0 \tag{3.3.29}$$

т. е. что квадратичная форма Q положительно определенная или полуопределенная.

Воспользуемся признаком полуопределенности квадратичной формы, указанным в § 4 гл. I. Если $A = A_{ss}$ — какая-либо квадратная матрица, то через $(A)_{kk}$ $(k \leqslant s)$ будем обозначать квадратную матрицу, вырезанную из левого верхнего угла данной матрицы и имеющую k строк и k столбцов („главную диагональную подматрицу“).

*) При доказательстве используется еще допустимость некоторых перестановок операций интегрирования и дифференцирования по параметру.

Пусть $F_{ss}=F$ — любое ортогональное преобразование. Очевидно (см. § 4 гл. I), квадратичная форма Q будет неотрицательна тогда и только тогда, когда неотрицательна форма с матрицей

$$F\left(B_t-I^{-1}\right)F^T=FB_tF^{-1}-FI^{-1}F^{-1}=B'-H^{-1}, \quad (3.3.30)$$

где

$$B'=FB_tF^{-1}; \quad H^{-1}=FI^{-1}F^{-1}. \quad (3.3.31)$$

Очевидно, при этом

$$H=FIF^{-1}. \quad (3.3.32)$$

Составим теперь корреляционную матрицу случайного вектора

$$\left(t_1, \ldots, t_s, \quad \frac{1}{L}\frac{\partial L}{\partial \alpha_1}, \ldots, \frac{1}{L}\frac{\partial L}{\partial \alpha_s}\right). \quad (3.3.33)$$

Имеем $E(t_i)=\alpha_i$, и

$$E\left(\frac{1}{L}\frac{\partial L}{\partial \alpha_i}\right)=\int\underset{R}{\cdots}\int\frac{\partial L}{\partial \alpha_i}dx_1\ldots dx_n=$$

$$=\frac{\partial}{\partial \alpha_i}\int\underset{R}{\cdots}\int L dx_1\ldots dx_n=\frac{\partial}{\partial \alpha_i}1=0.$$

В силу (3.3.22) имеем

$$E\left(\frac{1}{L}\frac{\partial L}{\partial \alpha_i}\cdot\frac{1}{L}\frac{\partial L}{\partial \alpha_j}\right)=\{I\}_{ij}. \quad (3.3.34)$$

Далее,

$$E\left[(t_i-\alpha_i)\left(\frac{1}{L}\frac{\partial L}{\partial \alpha_j}\right)\right]=E\left(t_i\frac{1}{L}\frac{\partial L}{\partial \alpha_j}\right)-\alpha_iE\left(\frac{1}{L}\frac{\partial L}{\partial \alpha_j}\right)=E\left(t_i\frac{1}{L}\frac{\partial L}{\partial \alpha_j}\right). \quad (3.3.35)$$

$$E\left(t_i\frac{1}{L}\frac{\partial L}{\partial \alpha_j}\right)=\frac{\partial}{\partial \alpha_j}\int\underset{R}{\cdots}\int t_iL\,dx_1\ldots dx_n=\frac{\partial}{\partial \alpha_j}E(t_i)=\frac{\partial \alpha_i}{\partial \alpha_j}. \quad (3.3.36)$$

Последнее выражение, очевидно, равно нулю, если $i\neq j$, и равно единице, если $i=j$.

Соотношения (3.3.34)—(3.3.36) позволяют записать корреляционную матрицу вектора (3.3.33) в виде

$$B_{2s,\,2s}=\left\|\begin{array}{c|c} B_{ss} & E_{ss} \\ \hline E_{ss} & I_{ss}\end{array}\right\| \quad (3.3.37)$$

Здесь матрица B_{ss} есть корреляционная матрица вектора $t=(t_1, \ldots, t_s)$, I_{ss} — корреляционная матрица вектора $l=\left(\frac{1}{L}\frac{\partial L}{\partial \alpha_1}, \ldots, \frac{1}{L}\frac{\partial L}{\partial \alpha_s}\right)$, а единичная матрица E_{ss} совпадает с матрицей $E\left(tl^T\right)$. Пусть теперь векторы t и l подвергнуты ортогональному преобразованию F, так что получились векторы Ft и Fl. Тогда B_{ss} переходит в B', I_{ss} — в H и $E(tl^T)$ — в $E(Ftl^TF^T) = E_{ss}$, т. е. в себя. Возьмем теперь первые k компонент ($k \leqslant s$) векторов Ft и Fl и составим корреляционную матрицу получившегося вектора из $2k$ компонент. Легко заметить, что мы получим матрицу

$$M_{2k,\,2k} = \left\| \begin{array}{c|c} (B')_{kk} & E_{kk} \\ \hline E_{kk} & (H)_{kk} \end{array} \right\|. \tag{3.3.38}$$

В самом деле, полагая $F = \|f_{ij}\|$, находим

$$E\left(\sum_{j=1}^{s} f_{ij}t_j \sum_{n=1}^{s} f_{mn}l_n\right) = 0$$

при $i \neq m$ и единице при $i = m$ в силу (3.3.36). Таким образом, матрица (3.3.38) является корреляционной и потому неотрицательной (порождает определенную или полуопределенную квадратичную форму).

Теперь выберем F в (3.3.31) так, чтобы H стала диагональной матрицей. Тогда $(H)_{kk}$ будет тоже диагональной матрицей, и имеет место простое, но важное для нас свойство:

$$(H^{-1})_{kk} = (H)_{kk}^{-1}. \tag{3.3.39}$$

Рассмотрим теперь вспомогательную матрицу

$$N_{2k,\,2k} = \left\| \begin{array}{c|c} E_{kk} & -(H)_{kk}^{-1} \\ \hline 0 & (H)_{kk}^{-1} \end{array} \right\|. \tag{3.3.40}$$

По теореме Лапласа (§ 2 гл. 1) имеем

$$\det(N_{2k,\,2k}) = [\det(H)_{kk}]^{-1} > 0 \qquad (k = 1,\ 2,\ \ldots,\ s). \tag{3.3.41}$$

Имея в виду, что

$$\det(M_{2k,\,2k}) \geqslant 0; \qquad (k = 1,\ 2,\ \ldots,\ s), \tag{3.3.42}$$

находим

$$\det(N_{2k,\,2k} \cdot M_{2k,\,2k}) \geqslant 0; \quad (k = 1,\ 2,\ \ldots,\ s). \tag{3.3.43}$$

Но мы имеем

$$N_{2k,\,2k} \cdot M_{2k,\,2k} = \left\| \begin{array}{c|c} E_{kk} & -(H)_{kk}^{-1} \\ \hline 0_{kk} & (H)_{kk}^{-1} \end{array} \right\| \cdot \left\| \begin{array}{c|c} (B')_{kk} & E_{kk} \\ \hline E_{kk} & (H)_{kk} \end{array} \right\| =$$

$$= \left\| \begin{array}{c|c} (B')_{kk} - (H)_{kk}^{-1} & 0_{kk} \\ \hline 0_{kk} & E_{kk} \end{array} \right\| = \left\| \begin{array}{c|c} (B')_{kk} - (H^{-1})_{kk} & 0_{kk} \\ \hline (H^{-1})_{kk} & E_{kk} \end{array} \right\|$$

по правилу умножения разделенных матриц (§ 2 гл. 1) с применением (3.3.39). Отсюда

$$\det[(B')_{kk}-(H^{-1})_{kk}]\geqslant 0; \quad k=1,\,2,\,\ldots,\,s. \qquad (3.3.44)$$

Совершенно аналогично вместо первых k компонент ($k\leqslant s$) векторов Ft и Fl можно было взять компоненты с индексами $i_1,\,i_2,\,\ldots,\,i_k$. Тогда получается соотношение, аналогичное (3.3.44), где вместо угловых главных миноров стоят главные миноры с номерами столбцов и строк $i_1,\,i_2,\,\ldots,\,i_k$. Отсюда следует, что матрица $B'-H^{-1}$ неотрицательная. Тогда, в силу сказанного ранее, матрица B_t-I^{-1} также неотрицательна, что и требовалось доказать.

Пусть $Y=\begin{Vmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_s \end{Vmatrix}$ — единичный вектор $\left(\sum\limits_{i=1}^{s} y_i^2=1\right)$; спроектируем на него случайный вектор $(t_1,\,\ldots,\,t_s)$; получим (см. § 4 гл. II) случайную величину с дисперсией Y^TB_tY, причем

$$Y^TB_tY\geqslant Y^TI^{-1}Y. \qquad (3.3.45)$$

Это означает, что корреляционный эллипсоид $Y^TB_t^{-1}Y=1$ случайного вектора $(t_1,\,\ldots,\,t_s)$ охватывает фиксированный эллипсоид $Y^TIY=1$ (см. теоремы § 4 гл. II). В том случае, когда корреляционный эллипсоид системы несмещенных оценок $t_1^{(0)},\,\ldots,\,t_s^{(0)}$ имеет вид $Y^TB_0^{-1}Y=1$ и совпадает с эллипсоидом $Y^TIY=1$, все другие системы несмещенных оценок будут иметь корреляционные эллипсоиды, охватывающие данный. Система оценок $t_1^{(0)},\,\ldots,\,t_s^{(0)}$ называется в этом случае системой *совместно эффективных* оценок. Подобный случай нам будет встречаться в теории метода наименьших квадратов, однако гораздо чаще встречается система асимптотически несмещенных и совместно асимптотически эффективных оценок. Под этим понимаются последовательности систем оценок для параметров $\alpha_1,\,\ldots,\,\alpha_s$: $(t_{1\nu},\,\ldots\,t_{s\nu})$, зависящих от объема выборки n, асимптотически несмещенных и имеющих корреляционные эллипсоиды $Y^T(B^{(\nu)})^{-1}Y=1$, при $n\to\infty$ асимптотически приближающихся к даваемому информационной матрицей Фишера эллипсоиду $Y^TIY=1$.

§ 4. Дополнительные сведения об оценивании параметров. Основные методы оценивания

Займемся вопросом о реальном смысле совместно эффективных и асимптотически совместно эффективных оценок. Пусть $t_1,\,\ldots,\,t_s$ — некоторая система несмещенных оценок параметров $\alpha_1,\,\ldots\alpha_s$, а $t_1^{(0)},\,\ldots$ $\ldots t_s^{(0)}$ — такая же система совместно эффективных оценок. Составим случайные векторы уклонений $(\Delta_1^{(0)}, \ldots, \Delta_s^{(0)})$, где $\Delta_i^{(0)} = t_i^{(0)} - \alpha_i$, и $(\Delta_1, \Delta_2, \ldots, \Delta_s)$, где $\Delta_i = t_i - \alpha_i$. Корреляционный эллипсоид вектора $(\Delta_1^{(0)}, \ldots, \Delta_s^{(0)})$ будет лежать внутри корреляционного эллипсоида вектора $(\Delta_1, \ldots, \Delta_s)$. Мы можем теперь применить теорему 2.5.1. Пусть

$$\xi = \sum \beta_i \Delta_i; \quad \eta = \sum \beta_i \Delta_i^{(0)},$$

где β_i $(i = 1, 2, \ldots, s)$ — любые фиксированные числа. Тогда при любом фиксированном $\varepsilon > 0$

$$P\{|\eta| > \varepsilon\} < P\{|\xi| > \varepsilon\}, \tag{3.4.1}$$

т. е. для любой заданной линейной комбинации уклонений оценок от истинных значений параметров вероятность уклонения более заданного от нуля будет меньше для системы совместно эффективных оценок, чем для иных оценок в весьма широком классе несмещенных оценок. В частности, можно в (3.4.1) взять $\xi = \Delta_i$, $\eta = \Delta_i^{(0)}$ $(i = 1, 2, \ldots, s)$.

Для случая последовательности асимптотически несмещенных и асимптотически совместно эффективных оценок утверждение (3.4.1) будет иметь место при достаточно больших выборках и с точностью до слагаемого, стремящегося к нулю.

Перейдем к основным способам оценивания параметров. Основными способами оценивания параметров в математической статистике являются способ максимального правдоподобия и способ моментов. Способ моментов введен К. Пирсоном*). Пусть имеем повторную выборку $x_1, \ldots, x_n$ с функцией правдоподобия

$$L(x_1, \ldots, x_n; \alpha_1, \ldots, \alpha_s) = \prod_{i=1}^{n} f(x_i; \alpha_1, \ldots, \alpha_s). \tag{3.4.2}$$

Составим s выборочных моментов

$$a_\nu = \frac{x_1^\nu + \ldots + x_n^\nu}{n}; \qquad \nu = 1, 2, \ldots, s. \tag{3.4.3}$$

Если существует $Ex_i^{2\nu} = \alpha_{2\nu}$, то при $n \to \infty$ будет действовать закон больших чисел (см. § 7 гл. II), так что будет

$$P\{|a_\nu - \alpha_\nu| > \varepsilon\} \to 0 \quad \text{при} \quad n \to \infty,$$

т. е. при большом объеме выборки a_ν с большой вероятностью будет близко к α_ν. Поэтому естественно составить уравнения

$$\alpha_\nu = \int_{-\infty}^{\infty} x^\nu f(x_1, \alpha_1, \ldots, \alpha_s)\, dx = a_\nu \qquad (\nu = 1, 2, \ldots, s), \tag{3.4.4}$$

*) К. Пирсон (1857—1936) — английский статистик.

откуда находятся оценки для параметров $\alpha_1, \ldots, \alpha_s$. При $n \to \infty$ эти оценки оказываются асимптотически несмещенными и асимптотически нормальными с дисперсиями порядка $\frac{1}{n}$. Мы не будем останавливаться на примерах, так как этот способ, хотя иногда и применяется в силу сравнительной простоты, теоретически не наилучший. Для нас важно, что он приводит к асимптотически нормальным оценкам.

Способ максимального правдоподобия восходит еще к Даниилу Бернулли*) (1776 г.) и К. Ф. Гауссу, но систематически разработан Р. А. Фишером в 1912 г. При этом способе выбираются такие оценки параметров для $\alpha_1, \ldots, \alpha_s$, которые дают максимальное значение функции правдоподобия $L(x_1, \ldots, x_n; \alpha_1, \ldots, \alpha_s)$, если подставить их на место $\alpha_1, \ldots, \alpha_s$.

Заметим, что таким образом достигается максимум плотности вероятности выборки, но не „наивероятнейшее значение", о котором говорится в старых учебниках. При распределениях, имеющих плотность вероятности, „наивероятнейших значений" вообще нет, так как вероятность любого значения равна нулю.

В случае повторной выборки, когда функция правдоподобия имеет вид (3.2.5) и допустимы соответствующие дифференцирования, уравнения правдоподобия приобретают вид

$$\frac{\partial \ln L}{\partial \alpha_i} = \sum_{j=1}^{n} \frac{\partial \ln f(x_j; \alpha_1, \ldots, \alpha_s)}{\partial \alpha_i} = 0; \qquad i = 1, 2, \ldots, s. \tag{3.4.5}$$

Оценки, получаемые по методу максимального правдоподобия, обозначаются обычно $\hat{\alpha}_1, \ldots, \hat{\alpha}_s$. Они обладают рядом замечательных свойств, которые будут здесь изложены без доказательств для случая одного параметра α.

Пусть $f(x, \alpha)$ — плотность вероятности x_1; $L = \prod_{i=1}^{n} f(x_i, \alpha)$ — функция правдоподобия повторной выборки, и

1) $\frac{\partial \ln L}{\partial \alpha}$ и $\frac{\partial^2 \ln L}{\partial \alpha^2}$ существуют и непрерывны для всех $(x_1, \ldots, x_n)$ и для всех α в интервале, заключающем оцениваемое значение α_0 параметра α. Далее, должны быть выполнены неравенства

$$\left|\frac{\partial \ln L}{\partial \alpha}\right| < F_1(x_1, \ldots, x_n); \quad \left|\frac{\partial^2 \ln L}{\partial \alpha^2}\right| < F_2(x_1, \ldots x_n),$$

где интегралы $\int_R \ldots \int F_1\, dx_1 \ldots dx_n$ и $\int_R \ldots \int F_2\, dx_1 \ldots dx_n$ сходятся;

*) На это указал акад. В. И. Смирнов.

2) $\frac{\partial^3 \ln L}{\partial \alpha^3}$ существует; $\left|\frac{\partial^3 \ln L}{\partial \alpha^3}\right| < M(x_1, \ldots, x_n)$,

$$|E\{M(x_1, \ldots x_n)\}| < \text{const};$$

3) $E\left(-\frac{\partial^2 \ln L}{\partial \alpha^2}\right) = I(\alpha)$ существует и $\neq 0$.

Тогда имеют место теоремы:

Теорема 1. При $n \to \infty$ с вероятностью, стремящейся к единице, уравнение правдоподобия $\frac{\partial \ln L}{\partial \alpha} = 0$ имеет решение $\hat{\alpha}$, причем

$$P\{|\hat{\alpha} - \alpha_0| > \varepsilon\} \to 0 \tag{3.4.6}$$

при $n \to \infty$ и заданном $\varepsilon > 0$ (Дюге, 1937, [16]).

Теорема 2. Всякое решение $\hat{\alpha}$ уравнения правдоподобия со свойством (3.4.6) дает максимум функции правдоподобия L с вероятностью, стремящейся к единице при $n \to \infty$ (Гузурбазар, 1948 [13]).

Теорема 3. Решение $\hat{\alpha}$ уравнения правдоподобия со свойством (3.4.6) будет асимптотически нормальным со средним α_0 (Г. Крамер, 1948, [25]).

Теорема 4. Решение $\hat{\alpha}$, о котором идет речь в предыдущих теоремах, является асимптотически эффективным (Г. Крамер, 1948, [25]).

Наиболее важными для нас здесь являются свойства асимптотической несмещенности, асимптотической нормальности и асимптотической эффективности оценок $\hat{\alpha}$ метода максимального правдоподобия. Эти свойства выделяют указанный метод среди других методов (как, например, метод моментов).

Укажем еще одну теорему, важную для дальнейшего (см. Г. Крамер, [25], стр. 543).

Теорема 5. Если для параметров $\alpha_1, \alpha_2, \ldots \alpha_s$ существуют совместно эффективные оценки $a_1, \ldots, a_s$ и соблюдены условия, аналогичные 1), 2) и 3), то уравнения максимального правдоподобия имеют единственное решение $(a_1, \ldots, a_s)$.

Примеры по применению метода максимального правдоподобия будут даны непосредственно при изложении метода наименьших квадратов.

ГЛАВА IV

ПРЯМЫЕ РАВНОТОЧНЫЕ ИЗМЕРЕНИЯ

§ 1. Точечная оценка измеряемой величины

Пусть измерения физической величины a дают значения x_1, $x_2, \ldots, x_n$. Будем считать, что эти значения представляют собой сумму a и случайных погрешностей

$$\Delta_i = x_i - a \qquad (i = 1, \ldots, n), \tag{4.1.1}$$

причем погрешности Δ_i независимы в совокупности и нормальны. Далее, будем считать, что

$$E(\Delta_i) = 0; \quad E\left(\Delta_i^2\right) = \sigma^2 \tag{4.1.2}$$

(т. е. $\Delta_i \in N(0, \sigma)$).

Равенство $E(\Delta_i) = 0$ означает *несмещенность измерений* (отсутствие систематической ошибки). То, что число σ, неизвестное нам, одно и то же для всех значений i, означает *равноточность измерений*.

Требуется найти приближенное значение физической величины a на основании наблюдений $x_1, \ldots, x_n$. Как мы видели, это — задача оценивания параметра a при повторной выборке. Применим для ее решения способ максимального правдоподобия (§ 4 гл. III). Функция правдоподобия имеет вид

$$L(x_1, \ldots, x_n, a) = \prod_{i=1}^{n} f(x_i, a),$$

где

$$f(x_i, a) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp - \frac{(x_i - a)^2}{2\sigma^2}.$$

Таким образом,

$$L(x_1, \ldots, x_n, a) = (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - a)^2. \tag{4.1.3}$$

Предписание максимального правдоподобия требует при заданном σ выбирать $a = \hat{a}$ так, чтобы было

$$L(x_1, \ldots, x_n, a) = \max. \tag{4.1.4}$$

Из (4.1.3) очевидно, что это равносильно условию

$$Q=\sum_{i=1}^{n}(x_i-a)^2=\min, \tag{4.1.5}$$

которое и есть предписание наименьших квадратов.

Введя среднее выборочное

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i, \tag{4.1.6}$$

находим

$$Q=\sum_{i=1}^{n}(x_i-a)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(a-\bar{x})^2, \tag{4.1.7}$$

откуда видно, что $\min Q$ получается тогда и только тогда, если заменим a на $\bar{x}$, т. е.

$$\hat{a}=\bar{x}. \tag{4.1.8}$$

Итак, оценка по методу наименьших квадратов в данном случае дает $\hat{a}=\bar{x}$ в качестве приближенного значения a. Выясним реальный смысл такого приближения.

Имеем

$$E(\bar{x})=a; \quad D(\bar{x})=\frac{\sigma^2}{n}.$$

Поскольку $\bar{x}$ — нормальная случайная величина, то

$$\bar{x}\in N\left(a, \frac{\sigma}{\sqrt{n}}\right).$$

Значит, $\bar{x}$ будет нормальной величиной со средним, равным измеряемой физической величине a, и штандартом, равным $\frac{\sigma}{\sqrt{n}}$. В частности, будем иметь

$$P\left\{|\bar{x}-a|>\frac{1{,}96\sigma}{\sqrt{n}}\right\}=0{,}05. \tag{4.1.9}$$

При большом числе наблюдений n мы будем иметь, при работе с оценкой $\bar{x}$, почти наверное приближение порядка $\frac{2\sigma}{\sqrt{n}}$ к измеряемой величине a. Однако изредка (с вероятностью $<0{,}05$) возможны и бóльшие отклонения. Отклонение более чем на $\frac{3\sigma}{\sqrt{n}}$ имеет уже вероятность около 0,003. Далее, $\bar{x}$ является эффективной оценкой для a, ибо из (3.3.13) имеем для весьма широкого класса других несмещенных оценок t параметра a неравенство

$$D(t)\geqslant\frac{\sigma^2}{n}=D(\bar{x}).$$

Таким образом, на основании сказанного в § 5 гл. II, реальный смысл оценки $\bar{x}$ по методу наименьших квадратов таков: в весьма широком классе асимптотически несмещенных и асимптотически нормальных оценок t имеем при заданном сколь угодно малом $\varepsilon > 0$

$$P\{|\bar{x}-a|>\varepsilon\} \leqslant P\{|t-a|>\varepsilon\}, \tag{4.1.10}$$

с точностью до возможных добавочных слагаемых, стремящихся к нулю при увеличении объема выборки. Это служит оправданием выбора предписания наименьших квадратов вместо какого-либо иного.

§ 2. Оценивание с помощью доверительных интервалов

Хотя мы знаем, что точечная оценка $\bar{x}$, полученная в § 1, выгодна в широком классе иных точечных оценок, она имеет тот недостаток, что наверное не совпадает с измеряемой величиной a; кроме того, для определения ее точности нужно знать дисперсию единичного наблюдения. Последний вопрос требует отдельного решения. Теперь же мы обратимся к новому, более совершенному способу оценивания — способу доверительных интервалов, принадлежащему известному американскому статистику Ю. Нейману [38]. Мы будем рассматривать этот способ не во всей полноте его применений, а лишь в приложении к методу наименьших квадратов.

Наряду с выборочным средним $\bar{x}$ введем еще выборочную дисперсию $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$. Докажем следующую теорему.

Теорема 4.2.1. *s^2 есть случайная величина, статистически независимая от $\bar{x}$ и распределенная как величина*

$$\frac{\sigma^2}{n}\chi^2_{n-1}. \tag{4.2.1}$$

Для доказательства независимости s^2 и $\bar{x}$ докажем, что $\bar{x}$ независимо от нормального случайного вектора $(x_1-\bar{x}, \ldots, x_n-\bar{x})$ (вырожденного, так как сумма его компонент равна нулю).

Рассмотрим n-мерный случайный нормальный вектор

$$(x_1-\bar{x}, \ldots, x_{n-1}-\bar{x}, \bar{x}). \tag{4.2.2}$$

Компоненты $\bar{x}$ некоррелированы с остальными компонентами. В самом деле,

$$E(x_i-\bar{x}) = 0,$$

$$E(x_i-\bar{x})(\bar{x}-a) = E[(x_i-a)-(\bar{x}-a)](\bar{x}-a) = \\ = \frac{1}{n}E(x_i-a)^2 - E(\bar{x}-a)^2 = \frac{\sigma^2}{n}-\frac{\sigma^2}{n} = 0.$$

Ввиду нормальности плотности распределения (4.2.2), это означает, что $\bar{x}$ не зависит от $x_i - \bar{x}$ ($i = 1, 2, \ldots, n-1$) в совокупности, стало быть и от любой их непрерывной функции $f(x_1 - \bar{x}, \ldots, x_{n-1} - \bar{x})$. Следовательно, $\bar{x}$ не зависит от $x_n - \bar{x} = -\sum_{i=1}^{n-1}(x_i - \bar{x})$, от совокупности $(x_1 - \bar{x}, \ldots, x_n - \bar{x})$ и от любой непрерывной функции $f(x_1 - \bar{x}, \ldots, x_n - \bar{x})$, в частности, от $s^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$. Остается доказать, что s^2 распределено как $\frac{\sigma^2}{n}\chi^2_{n-1}$. Рассмотрим величину $\frac{ns^2}{\sigma^2}$ и положим $y_i = \frac{x_i - a}{\sigma}$, так что $y_i \in N(0, 1)$.

Тогда

$$\frac{ns^2}{\sigma^2} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2.$$

Рассмотрим ортогональное преобразование

$$\begin{aligned} z_1 &= \sqrt{n}\,\bar{y} = \frac{1}{\sqrt{n}}y_1 + \ldots + \frac{1}{\sqrt{n}}y_n \\ z_2 &= c_{21}y_1 + \ldots + c_{2n}y_n \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ z_n &= c_{n1}y_1 + \ldots + c_{nn}y_n \end{aligned}$$

с первой строкой, отвечающей $\sqrt{n}\,\bar{y}$, и другими соответствующим образом подобранными строками.

Согласно теореме Фишера (теорема 2.3.2), $(z_1, \ldots, z_n)$ будет таким же нормальным вектором, как $(y_1, \ldots, y_n)$, т. е. $z_i \in N(0, 1)$ и z_i независимы в совокупности ($i = 1, 2, \ldots, n$). Далее,

$$\sum y_i^2 = \sum z_i^2, \quad \frac{ns^2}{\sigma^2} = \sum y_i^2 - n\bar{y}^2 = \sum_{i=1}^{n} z_i^2 - z_1^2 = \sum_{i=2}^{n} z_i^2.$$

Таким образом, величина $\frac{ns^2}{\sigma^2}$ распределена по закону χ^2_{n-1}. Отсюда следует важная теорема:

Теорема 4.2.2. *Дробь*

$$t_{n-1} = \frac{\bar{x} - a}{s}\sqrt{n-1} \tag{4.2.3}$$

распределена по закону Стьюдента с $n-1$ степенями свободы.

Для доказательства замечаем, что $t_{n-1} = \frac{\xi}{\sqrt{(n-1)^{-1}(\eta_1^2 + \ldots + \eta_{n-1}^2)}}$, где $\xi, \eta_1, \ldots, \eta_{n-1}$ независимы в совокупности и нормальны $N(0, 1)$.

Заметим, что на основании предыдущего величина $\sqrt{\frac{ns^2}{\sigma^2(n-1)}}$ распределена, как $\sqrt{\frac{\chi^2_{n-1}}{n-1}}$, а величина $\frac{(\bar{x}-a)\sqrt{n}}{\sigma}$ от нее независима и нормальна $N(0, 1)$. Поэтому величина

$$t_{n-1}=\frac{\bar{x}-a}{s}\sqrt{n-1}=\frac{\bar{x}-a}{\sigma}\sqrt{n}\left(\frac{s\sqrt{n}}{\sigma\sqrt{n-1}}\right)^{-1}$$

действительно распределена по закону Стьюдента с плотностью вероятности $s_{n-1}(x)$ [см. (2.6.6)], т. е. с $n-1$ степенями свободы.

Доказанная теорема дает путь к оцениванию a по методу доверительных интервалов. По заданному числу степеней свободы $n-1$ выберем γ по табл. I приложений так, чтобы было

$$P\{|t_{n-1}|\leqslant\gamma\}=0{,}95. \tag{4.2.4}$$

Это означает, что

$$P\left\{\left|\frac{\bar{x}-a}{s}\right|\sqrt{n-1}\leqslant\gamma\right\}=0{,}95, \tag{4.2.5}$$

или

$$P\left\{\bar{x}-\frac{\gamma s}{\sqrt{n-1}}\leqslant a\leqslant\bar{x}+\frac{\gamma s}{\sqrt{n-1}}\right\}=0{,}95. \tag{4.2.6}$$

Иначе говоря, вероятность того, что интервал со случайными концами

$$I=\left[\bar{x}-\frac{\gamma s}{\sqrt{n-1}},\quad \bar{x}+\frac{\gamma s}{\sqrt{n-1}}\right] \tag{4.2.7}$$

накрывает неизвестное значение a, есть 0,95.

Такой интервал I со случайными концами называется *доверительным интервалом*, а изложенные соображения — *методом оценивания с помощью доверительного интервала*. Этот метод требует мало дополнительных вычислений по сравнению с методом точечной оценки.

Величина 0,95 в данном случае называется *надежностью* доверительного интервала, а математическое ожидание длины интервала, т. е. величина

$$E\left(\frac{2\gamma s}{\sqrt{n-1}}\right), \tag{4.2.8}$$

— *точностью* оценивания.

Надежность можно задавать равной не только 0,95, но и 0,99; 0,999, и вообще любой, употребляя при необходимости интерполяцию в таблицах. При данной надежности точность будет обратно пропорциональна $\sqrt{n-1}$, а при заданном числе наблюдений и увеличении надежности число γ возрастает, так что точность уменьшается.

Остановимся на вопросе о реальном смысле оценивания по методу доверительных интервалов. Допустим, что наблюдатель производит большое число N серий равноточных прямых измерений различных величин $a_1, a_2, \ldots, a_N$, измеряя их по методу доверительных интервалов с одной и той же надежностью, скажем, равной 0,95. Мы можем тогда утверждать, что при большом числе N серий измерений, около $0{,}95N$ построенных наблюдателем доверительных интервалов I_i будут накрывать измеряемые величины a_i.

В самом деле, введем случайную величину ξ_i, принимающую значение единица, если доверительный интервал I_i накрывает a_i, и нуль, если он ее не накрывает. Тогда $\Xi_N = \xi_1 + \ldots + \xi_N$ будет числом успешных измерений наблюдателя. Здесь числа $\xi_1, \ldots, \xi_N$ независимы, и при этом

$$E(\xi_i) = 1 \cdot 0{,}95 + 0 \cdot 0{,}05 = 0{,}95,$$
$$D(\xi_i) = 0{,}95 \cdot 0{,}05 < 0{,}05.$$

Отсюда, по закону больших чисел (теорема 2.7.3)

$$P\left\{\left|\frac{\Xi_N}{N} - 0{,}95\right| > \varepsilon\right\} \to 0$$

при $N \to \infty$ и любом заданном $\varepsilon > 0$, что доказывает наше утверждение.

Встает вопрос о том, почему мы употребляем именно доверительный интервал (4.2.7), а не какие-либо иные. Мы замечаем, что центр доверительного интервала (4.2.7) — оценка $\bar{x}$ — найден по методу наименьших квадратов и представляет собой точечную оценку для a, оптимальную в определенном смысле [см. (4.1.10)]. Оказывается, что и доверительный интервал (4.2.7) в некотором смысле оптимален, но мы не будем на этом останавливаться, ибо это требует углубления в теорию так называемых достаточных статистик.

Мы видели, что $D(\bar{x}) = \frac{\sigma^2}{n}$. Очевидно, несмещенной оценкой для $\frac{\sigma^2}{n}$ будет оценка

$$\frac{s_1^2}{n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2. \tag{4.2.9}$$

При обработке наблюдений указанного типа обычно употребляются терминология и обозначения, восходящие еще к Гауссу. Величины $\Delta_i = x_i - a$ (вообще говоря, неизвестные нам) называются *истинными ошибками*.

Величины $v_i = x_i - \bar{x}$ называются *кажущимися ошибками*. Суммы вида $\sum\limits_{i=1}^{n} a_i$ обозначаются символом $[a]$, а суммы вида $\sum\limits_{i=1}^{n} b_i^2$ — символом $[bb]$.

Правила обработки прямых равноточных наблюдений

а) Точечное оценивание

Находим $\bar{x}=\frac{1}{n}[x]$; $s_1^2=\frac{1}{n-1}[vv]$; $\frac{s_1^2}{n}=\frac{1}{n(n-1)}[vv]$.

Вычисление удобно делать с помощью подходящим образом выбранного числа β и очевидных равенств

$$\bar{x}=\frac{1}{n}[x-\beta]+\beta, \tag{4.2.10}$$

$$[vv]=\frac{1}{n}[x-\beta,\ x-\beta]-(\bar{x}-\beta)^2. \tag{4.2.11}$$

Для контроля вычислений $\bar{x}$ пользуемся равенством $[v]=0$.

б) Оценивание по методу доверительных интервалов

Вычисляем величины $\bar{x}$ и $s=\sqrt{\frac{1}{n}[vv]}$; выбираем подходящую надежность p_0; она будет первым входом табл. I приложений. Вторым входом будет число степеней свободы $n-1=k$. Эти два входа дадут по таблице число γ. Доверительный интервал

$$I=\left[\bar{x}-\frac{\gamma s}{\sqrt{n-1}},\ \bar{x}+\frac{\gamma s}{\sqrt{n-1}}\right]$$

будет накрывать измеряемую величину a с заданной надежностью p_0.

Если пользоваться несмещенной оценкой s_1^2 для σ^2, то доверительный интервал для a имеет вид

$$I=\left[\bar{x}-\frac{\gamma s_1}{\sqrt{n}},\ \bar{x}+\frac{\gamma s_1}{\sqrt{n}}\right].$$

Оценивание точности измерений с помощью доверительных интервалов излагается ниже.

§ 3. Оценивание точности равноточных измерений

Метод доверительных интервалов позволял нам делать выводы об измеряемой величине на основании повторной выборки наблюдений $x_1, x_2, \ldots, x_n$, распределенных с плотностью вероятности

$$\varphi(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp-\frac{(x-a)^2}{2\sigma^2}. \tag{4.3.1}$$

При этом знание дисперсии наблюдений $\sigma^2=E(x_i-a)^2$, характеризующей их точность, не было нужно для данного способа оценивания. Однако часто бывает нужно изучать и сам по себе штандарт наблюдений σ. Такой вопрос встает при изучении точности работы измерительных приборов, механизмов на производстве, рассеяния снарядов, даваемого артиллерийскими системами, и т. д.

С точки зрения математической статистики (см. § 2 гл. III), данная задача есть задача оценивания параметра σ на основании повторной выборки наблюдений $(x_1, \ldots, x_n)$.

Как мы знаем (см. § 2 этой главы), статистика

$$s_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{4.3.2}$$

будет несмещенной оценкой для σ^2, т. е.

$$E\left(s_1^2\right) = \sigma^2. \tag{4.3.3}$$

Исследуя статистику (4.3.2) (например, используя теорему В. Хёфдинга 2.7.2), можно показать, что при данном $\varepsilon > 0$ и $n \to \infty$

$$P\left\{\left|s_1^2 - \sigma^2\right| > \varepsilon\right\} \to 0. \tag{4.3.4}$$

Можно, кроме того, охарактеризовать быстроту такого стремления. Поэтому для больших выборок можно брать $\sigma^2 \approx s_1^2$; $\sigma \approx s_1$ ($\approx$ знак приближенного равенства). Таким образом, получаются приближенные точечные оценки для σ^2 и σ.

Мы, однако, применим более точный способ оценивания с помощью доверительных интервалов. Он будет основан на утверждении, доказанном в § 2 (теорема 4.2.1) о том, что величина $u = \frac{ns^2}{\sigma^2}$ распределена по закону χ^2_{n-1}. Составим доверительный интервал

$$I_s = [\gamma_1 s^2, \ \gamma_2 s^2]. \tag{4.3.5}$$

Вероятность того, что этот интервал I_s накрывает σ^2, равна

$$P\{\gamma_1 s^2 \leqslant \sigma^2 \leqslant \gamma_2 s^2\} = P\left\{\frac{n}{\gamma_2} \leqslant \frac{ns^2}{\sigma^2} \leqslant \frac{n}{\gamma_1}\right\} = P\left\{\frac{n}{\gamma_2} \leqslant \chi^2_{n-1} \leqslant \frac{n}{\gamma_1}\right\}.$$

Обозначим $k_{n-1}(x)$ плотность вероятности для χ^2_{n-1} (см. § 6 гл. II).

Надежность накрытия σ^2 интервалом I_s равна

$$\int_{\frac{n}{\gamma_2}}^{\frac{n}{\gamma_1}} k_{n-1}(x)\, dx = K_{n-1}\left(\frac{n}{\gamma_1}\right) - K_{n-1}\left(\frac{n}{\gamma_2}\right).$$

Таким образом, чтобы при заданном $\frac{\gamma_2}{\gamma_1}$ получить нужную надежность, нужно иметь таблицы неполной гамма-функции $K_{n-1}(x)$. Можно пользоваться таблицами Е. Е. Слуцкого [45]. Мы не останавливаемся на том, оптимален ли в каком-либо смысле такой метод оценивания.

Для упрощения дела можно выбрать (см. А. Н. Колмогоров [22]) γ_1 и γ_2 так, что

$$K_{n-1}\left(\frac{n}{\gamma_2}\right)=\frac{1-\omega}{2}=P_1;\quad K_{n-1}\left(\frac{n}{\gamma_1}\right)=\frac{1+\omega}{2}=P_2,$$

где ω — заданная надежность. Соответствующие таблицы имеются в статье А. Н. Колмогорова [22].

§ 4. Примеры

Пример 1. При определении величины заряда электрона, равной $e_0\cdot 10^{-10}$ единиц $CGSE$, Милликен получил 58 значений величины e_0 (в табл. 7 они обозначены через x'_i).

Таблица 7

x'_i	$x'_i-\beta$	$(x'_i-\beta)^2$	x'_i	$x'_i-\beta$	$(x'_i-\beta)^2$
4,781	0,081	0,00656	4,771	0,071	0,00504
4,795	0,095	0,00903	4,809	0,109	0,01188
4,769	0,069	0,00476	4,790	0,090	0,00810
4,792	0,092	0,00846	4,779	0,079	0,00624
4,779	0,079	0,00624	4,788	0,088	0,00774
4,775	0,075	0,00563	4,772	0,072	0,00518
4,772	0,072	0,00518	4,791	0,091	0,00828
4,791	0,091	0,00828	4,788	0,088	0,00774
4,782	0,082	0,00672	4,783	0,083	0,00689
4,767	0,067	0,00449	4,740	0,040	0,00160
4,764	0,064	0,00410	4,775	0,075	0,00563
4,776	0,076	0,00578	4,761	0,061	0,00372
4,771	0,071	0,00504	4,792	0,092	0,00846
4,789	0,089	0,00792	4,758	0,058	0,00336
4,772	0,072	0,00518	4,764	0,064	0,00410
4,789	0,089	0,00792	4,810	0,110	0,01210
4,764	0,064	0,00410	4,799	0,099	0,00980
4,774	0,074	0,00548	4,799	0,099	0,00980
4,778	0,078	0,00608	4,797	0,097	0,00941
4,791	0,091	0,00828	4,790	0,090	0,00810
4,777	0,077	0,00593	4,747	0,047	0,00221
4,765	0,065	0,00423	4,769	0,069	0,00476
4,785	0,085	0,00723	4,806	0,106	0,01124
4,805	0,105	0,01103	4,779	0,079	0,00624
4,768	0,068	0,00462	4,785	0,085	0,00723
4,801	0,101	0,01020	4,790	0,090	0,00810
4,785	0,085	0,00723	4,777	0,077	0,00593
4,783	0,083	0,00689	4,749	0,049	0,00240
4,808	0,108	0,01166	4,781	0,081	0,00656
			Сумма	4,687	0,39209

Оценим истинное значение a величины заряда и штандарт σ погрешностей наблюдений.

Точечное оценивание параметров. Оценим сперва истинное значение и штандарт погрешностей измерения величины e_0. Тогда оценки $\bar{x}$ и s_1 для a и σ получатся умножением оценок для e_0 на 10^{-10}. Оценки для e_0 обозначим соответствующими буквами со штрихами. Тогда

$$\overline{x'}=\frac{1}{n}[x'-\beta]+\beta.$$

Принимаем $\beta=4,7$. В таблице приведены значения $x'_i-\beta$ и $(x'_i-\beta)^2$ *). Имеем $[x'-\beta]=4,687$ и $\overline{x'}=\frac{4,687}{58}+4,7=4,78081$ **). Значит, оценка величины заряда равна $\bar{x}=4,78081\cdot 10^{-10}$ ед.

Оценка s'_1 для штандарта погрешностей измерения величины e_0 получается по формуле

$$s'_1=\sqrt{\frac{1}{n-1}[vv]},$$

где

$$[vv]=[x'-\beta,\ x'-\beta]-n\left(\overline{x'}-\beta\right)^2.$$

Вычисляем $[vv]=0,39209-0,37876=0,01333$. Отсюда $s'_1=0,0153$. Следовательно, оценкой для σ будет $s_1=0,0153\cdot 10^{-10}$ ед.

Оценивание параметров с помощью доверительных интервалов. Оценим значения a и σ с помощью доверительных интервалов, т. е. найдем интервал, о котором с определенной вероятностью p_0 (доверительной вероятностью, или надежностью) можно утверждать, что он накрывает оцениваемое значение параметра.

Доверительным интервалом для значения a будет интервал

$$I=\left[\bar{x}-\frac{\gamma s_1}{\sqrt{n}},\ \bar{x}+\frac{\gamma s_1}{\sqrt{n}}\right];$$

здесь γ определяется по таблицам t-распределения Стьюдента из условия $P\{|t_{n-1}|\leqslant\gamma\}=p_0$, где p_0 — заданная надежность, а число степеней свободы равно $n-1=k$.

*) Столбцы промежуточных вычислений могут быть опущены, если вычисления производятся на счетных машинах с накапливающим устройством.

**) При вычислении среднего значения делением на целое число мы удерживаем в результате большее число значащих цифр, учитывая возможность уничтожения значащих цифр в последующих вычислениях. (Например, ниже при вычислении $[vv]$ уничтожаются три значащие цифры.)

Возьмем $p_0=0,95$. По таблице t-распределения *) находим для $k=n-1=57$ и $p_0=0,95$ значение $\gamma=2,02$. Значит, $\frac{\gamma s_1}{\sqrt{n}}=$ $=0,00406\cdot 10^{-10}$, и имеем доверительный интервал

$$4,776\cdot 10^{-10}\leqslant a\leqslant 4,785\cdot 10^{-10}.$$

Как показано в § 3 настоящей главы, доверительным интервалом для σ^2, построенным с помощью χ^2-распределения (с $n-1$ степенями свободы), служит интервал $I'=\left[\gamma_1'^2 s^2,\ \gamma_2'^2 s^2\right]$. Он соответствует вероятности

$$P\left\{\gamma_1'^2 s^2\leqslant\sigma^2\leqslant\gamma_2'^2 s^2\right\}=P_2-P_1=p.$$

Доверительным интервалом для штандарта σ является тогда интервал

$$I''=\left[\sqrt{\frac{n}{\chi^2_{P_1}}}\,s,\ \sqrt{\frac{n}{\chi^2_{P_2}}}\,s\right].$$

Мы будем пользоваться несмещенной оценкой s_1^2 для σ^2, а потому, используя равенство $ns^2=(n-1)s_1^2$, запишем доверительный интервал для σ в виде

$$I=\left[\sqrt{\frac{n-1}{\chi^2_{P_1}}}\,s_1,\ \sqrt{\frac{n-1}{\chi^2_{P_2}}}\,s_1\right].$$

Коэффициенты $\sqrt{\frac{n-1}{\chi^2_{P_1}}}=\gamma_1$ и $\sqrt{\frac{n-1}{\chi^2_{P_2}}}=\gamma_2$ для ряда значений n и p протабулированы **).

Возьмем $p_1=0,90$ ($P_1=0,05$; $P_2=0,95$). Тогда для $k=n-1=57$ находим по табл. III значения $\gamma_1\approx 0,868$, $\gamma_2\approx 1,186$; значит, $P\{0,868\,s_1\leqslant\sigma\leqslant 1,186\,s_1\}=0,90$. Доверительным интервалом для σ будет

$$0,0132\cdot 10^{-10}\leqslant\sigma\leqslant 0,0182\cdot 10^{-10}.$$

Возьмем для этого же примера $p_1=0,95$ ($P_1=0,025$; $P_2=0,975$). Тогда $\gamma_1\approx 0,845$, $\gamma_2\approx 1,225$, и при надежности 0,95 получается неравенство

$$0,0129\cdot 10^{-10}\leqslant\sigma\leqslant 0,0188\cdot 10^{-10}.$$

Таким образом, при значительной надежности для данного числа наблюдений получаем, что оценка точности наблюдений довольно груба, что отвечает существу дела. В известном смысле ее нельзя улучшить.

---

*) См. табл. I приложений.
**) См. табл. III приложений.

Пример 2. Английский физик Кэвендиш в 1789 г. опубликовал результаты 29 измерений плотности земли (см. первый и четвертый столбцы табл. 8).

Таблица 8

x_i	$x_i-\beta$	$(x_i-\beta)^2$	x_i	$x_i-\beta$	$(x_i-\beta)^2$
5,50	0,50	0,250	5,34	0,34	0,116
5,61	0,61	0,372	5,79	0,79	0,624
5,88	0,88	0,774	5,10	0,10	0,010
5,07	0,07	0,005	5,27	0,27	0,073
5,26	0,26	0,068	5,39	0,39	0,152
5,55	0,55	0,303	5,42	0,42	0,176
5,36	0,36	0,130	5,47	0,47	0,221
5,29	0,29	0,084	5,63	0,63	0,397
5,58	0,58	0,336	5,34	0,34	0,116
5,65	0,65	0,423	5,46	0,46	0,212
5,57	0,57	0,325	5,30	0,30	0,090
5,53	0,53	0,281	5,75	0,75	0,563
5,62	0,62	0,384	5,68	0,68	0,462
5,29	0,29	0,084	5,85	0,85	0,723
5,44	0,44	0,194			
			Сумма . .	13,99	7,948

Требуется оценить измеряемую величину a и штандарт погрешностей σ.

Произведем сначала точечное оценивание параметров. Положим $\beta=5$. Оценка a есть $\bar{x}=\frac{13,99}{29}+5=5,482$. Оценим σ. Вычислим $[vv]=7,948-6,749=1,199$. Оценка s_1 для σ равна $s_1=$ $=\sqrt{\frac{1}{28}1,199}=0,207$.

Найдем доверительный интервал для a. Возьмем $p_0=0,90$. Число степеней свободы $k=n-1=28$. По таблице t-распределения находим для $k=28$ и $p_0=0,90$ значение $\gamma=1,701$. Тогда $\frac{\gamma s_1}{\sqrt{n}}=\frac{1,701\cdot 0,207}{5,385}=0,065$. Значит, при надежности 0,90 получаем доверительный интервал

$$5,41<a<5,55.$$

Оценим точность наблюдений при той же надежности $p_0=0,90$. Для этой надежности и $k=n-1=28$ находим по табл. III значения $\gamma_1=0,823$, $\gamma_2=1,286$. Следовательно, при надежности 0,90 имеем неравенство $0,823\,s_1<\sigma<1,286\,s_1$, или

$$0,170<\sigma<0,267.$$

Пример 3. Приводим данные по наблюдению азимута на Хвалынском базисе Елшанка — Благодатное *) (табл. 9) за 1926 г.

Таблица 9

Дата	Прием	Азимут x_i	$x_i-\beta$	$(x_i-\beta)^2$
16/IX	1	62°01′11″,46	1″,46	2,132
	2	11,01	1,01	1,020
	3	12,29	2,29	5,244
	4	13,47	3,47	12,041
	5	14,07	4,07	16,565
	6	10,60	0,60	0,360
17/IX	7	10,76	0,76	0,578
	8	14,13	4,13	17,057
	9	12,98	2,98	8,880
	10	10,46	0,46	0,212
	11	11,25	1,25	1,563
18/IX	12	12,02	2,02	4,080
		Сумма . . .	24,50	69,732

Найдем точечные оценки для измеряемого значения a и стандарта σ погрешностей; построим для a и σ доверительные интервалы.

Полагаем $\beta=62°01'10''$. Оценка измеряемого азимута a равна

$$\bar{x}=\frac{24'',50}{12}+62°01'10''=62°01'12'',04.$$

Оценим σ. Вычисляем

$$[vv]=[x-\beta,\ x-\beta]-n(\bar{x}-\beta)^2=69,73-49,94=19,79.$$

Следовательно, оценка s_1 для σ равна

$$s_1=\sqrt{\frac{1}{11}\,19,79}=1'',34.$$

Найдем доверительный интервал для a при надежности $p_0=0,90$. Число степеней свободы равно $k=11$. Для таких p_0 и k имеем $\gamma=1,796$. Значит,

$$\frac{\gamma s_1}{\sqrt{n}}=\frac{1,796\cdot 1,34}{3,464}=0,69.$$

Получаем доверительный интервал

$$62°01'11'',35<a<62°01'12'',73.$$

*) Данные заимствованы из книги А. С. Чеботарёва [51].

Оценим σ с помощью доверительного интервала. Для $k = n-1 = 11$ и $p = 0{,}90$, имеем, по табл. III, $\gamma_1 = 0{,}748$, $\gamma_2 = 1{,}550$. Значит, при надежности 0,90 получим доверительный интервал $0{,}748\, s_1 < \sigma < 1{,}550\, s_1$, или $1'',00 < \sigma < 2'',08$.

Пример 4. Измерение угла a привело к данным табл. 10.

Таблица 10

Номер измерения	Измеренное значение угла x_i	$x_i - \beta$	$(x_i - \beta)^2$
1	$72^\circ 13' 38'',3$	$8'',3$	68,89
2	44,6	14,6	213,16
3	33,7	3,7	13,69
4	41,1	11,1	123,21
5	43,0	13,0	169,00
6	36,2	6,2	38,44
7	39,6	9,6	92,16
8	37,8	7,8	60,84
9	40,3	10,3	106,09
	Сумма	84,6	885,48

Найдем точечные оценки для измеряемой величины a, стандарта σ измерений и построим доверительные интервалы для них.

Полагаем $\beta = 72^\circ 13' 30''$. Тогда оценка среднего значения равна

$$\bar{x} = \frac{84'',6}{9} + 72^\circ 13' 30'' = 72^\circ 13' 39'',4.$$

Оценим σ. Вычислим

$$[vv] = 885{,}48 - 795{,}24 = 90{,}24.$$

Значит, оценка для σ равна

$$s_1 = \sqrt{\frac{1}{8}\, 90{,}24} = 3'',36.$$

Найдем доверительные границы для a. Пусть надежность $p_0 = 0{,}90$. Число степеней свободы $k = n - 1 = 8$. По таблице t-распределения находим $\gamma = 1{,}860$, значит

$$\frac{\gamma s_1}{\sqrt{n}} = \frac{1{,}860 \cdot 3{,}36}{3} = 2'',08.$$

При надежности 0,90 имеем доверительный интервал

$$72^\circ 13' 37'',3 < a < 72^\circ 13' 41'',5,$$

Оценим σ с помощью доверительных интервалов. Пусть $p_0=0,90$. Тогда, по табл. III, для $k=n-1=8$ находим $\gamma_1=0,718$, $\gamma_2=1,711$. Значит, доверительным интервалом для σ будет $0,718\,s_1<\sigma<1,711\,s_1$, или $2'',41<\sigma<5'',75$.

§ 5. Резко выделяющиеся наблюдения

Предыдущие методы относились к ситуации, когда наблюдения $x_1, x_2, \ldots, x_n$ представляют повторную выборку из нормальной совокупности: $x_i \in N(a, \sigma)$. Иногда, однако, в рядах наблюдений попадаются данные, которые на глаз резко выделяются среди других. Это может быть случайным обстоятельством, так как в длинном ряду наблюдений теоретически могут быть данные x_i, сильно уклоняющиеся в обе стороны от a, но может быть приписано также особым обстоятельствам, нарушившим обычный режим образования случайных погрешностей и вызвавшим грубую ошибку наблюдателя.

Рассмотрим, например, в ряду наблюдений $x_1, x_2, \ldots, x_n$ максимальное $x_{\max}$. Следуя Ф. Грэббсу [12], будем рассматривать безразмерную дробь „типа Стьюдента"

$$v=\frac{x_{\max}-\overline{x}}{s}. \tag{4.5.1}$$

Если считать, что в ряду наблюдений $x_1, \ldots, x_n$ всякое из них принадлежит к одной и той же нормальной совокупности $N(a, \sigma)$, то распределение дроби v будет зависеть не от параметров a и σ, а только от объема выборки n. В этом случае распределение v может быть вычислено и затабулировано со входами n и p, что и сделано Ф. Грэббсом [12].

По заданному объему выборки n и заданной вероятности p можно найти такое число ξ_n, что

$$P\{v \geqslant \xi_n\}=p. \tag{4.5.2}$$

Такие числа ξ_n для $p=0,10$; $0,05$; $0,025$; $0,01$ даны в табл. IV приложений*). Мы можем пользоваться следующим правилом (критерием).

П р а в и л о. Для того чтобы решить, нужно ли отбросить наблюдение $x_{\max}$ как грубую ошибку, надлежит выбрать p (скажем, 0,05), затем по данному числу n наблюдений и вероятности p найти по табл. IV число ξ_n и составить дробь

$$v=\frac{x_{\max}-\overline{x}}{s}.$$

Если окажется $v>\xi_n$, то следует отбросить $x_{\max}$ как наблюдение, содержащее грубую ошибку, а если $v \leqslant \xi_n$, то $x_{\max}$ можно оставить.

---

*) Таблица заимствована из статьи [12], стр. 29.

Если подозрение вызывает уклонившееся в левую сторону наблюдение $x_{\min}$, то можно заменить наблюдения $x_1, \ldots, x_n$ на $(-x_1), \ldots, (-x_n)$; тогда $x_{\min} = (-x)_{\max}$. В этом случае можно применить то же правило с заменой v на

$$v_1 = -\frac{x_{\min} - \bar{x}}{s} = \frac{\bar{x} - x_{\min}}{s}.$$

Мы не будем разбирать здесь, насколько этот критерий оптимален.

П р и м е р. В книге А. Р. Кларка „Geodesy" приводятся результаты 20 микроизмерений штриха на масштабе:

3,68	5,08	2,81	4,43
3,11	2,95	4,65	3,43
4,76	6,35	3,27	3,26
2,75	3,78	4,08	2,48
4,15	4,49	4,51	4,84

Предполагая, что измерения равноточны, независимы и принадлежат одной и той же нормальной совокупности с неизвестными a и σ, оценим, следует ли отвергнуть значение $x_{\max} = 6{,}35$ как грубую ошибку измерения или нет.

Составим дробь

$$v = \frac{x_{\max} - \bar{x}}{s}.$$

У нас

$$\bar{x} = \frac{\sum\limits_{i=1}^{20} x_i}{20} = \frac{78{,}86}{20} = 3{,}943,$$

$$s = \sqrt{\frac{1}{n}\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]} = \sqrt{0{,}8892} = 0{,}943.$$

Значит,

$$v = \frac{6{,}35 - 3{,}943}{0{,}943} = 2{,}552.$$

По табл. IV приложений находим, что при $n = 20$ значение ξ_n для 5%-го уровня значимости равно 2,623. Поскольку значение $v = 2{,}552$ меньше ξ_n, то $x_{\max} = 6{,}35$ следует оставить как допустимое в данном ряду наблюдений.

§ 6. Уточнение критерия Аббе*)

Критерий Аббе имеет целью проверку гипотезы об отсутствии систематического сдвига в наблюдениях $x_1, x_2, \ldots, x_n$, т. е. гипотезы о том, что

$$E(x_i)=a \qquad (i=1, 2, \ldots, n). \tag{4.6.1}$$

Наиболее естественно было бы трактовать общий вопрос проверки гипотезы о том, что все x_i принадлежат нормальной совокупности $N(a, \sigma)$. Проверка такой гипотезы — вопрос математической статистики. Однако поскольку вопрос о критерии Аббе (Э. Аббе [1], 1863) затрагивается в некоторых старых учебниках (см., например, Ф. Гельмерт [10]),—мы рассмотрим его. Критерий Аббе состоял в сравнении суммы квадратов погрешностей наблюдений с суммой квадратов их последовательных разностей, менее чувствительной к систематическому сдвигу математического ожидания.

Сформулируем его в уточненной форме. Вводим статистики

$$q^2=\frac{1}{2(n-1)}\sum_{i=1}^{n-1}(x_{i+1}-x_i)^2,$$

$$s_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

Составляем частное

$$r=\frac{q^2}{s_1^2}. \tag{4.6.2}$$

Можно найти (А. Хальд [48]), что при условии $x_i \in N(a, \sigma)$ имеют место равенства

$$E(r)=1, \quad D(r)=\frac{1}{n+1}\left(1-\frac{1}{n+1}\right).$$

Далее известно, что при $n \geqslant 20$ величина r распределена приблизительно нормально, с параметрами $\left(1;\ \left(\frac{1}{n+1}\left(1-\frac{1}{n-1}\right)\right)^{\frac{1}{2}}\right)$.

Если имеем систематический сдвиг $E(x_i)$, то следует ожидать, что s_1^2 будет много больше, чем q^2. В табл. V приложений даны при разных n числа r_p такие, что вероятности $P\{r \leqslant r_p\}=p$ ($p=0{,}001$; $0{,}01$; $0{,}05$). Если обнаружится, что для наблюдений $x_1, \ldots, x_n$ будет $r<r_p$, то такое событие в случае верности гипотезы (4.6.1) имело бы вероятность $<p$, и потому рассматриваемая гипотеза отвергается.

*) Э. Аббе (1840—1905) — немецкий оптик.

Получаем следующее правило: чтобы проверить отсутствие систематического сдвига в погрешностях наблюдений $x_1, x_2, \ldots, x_n$, вычисляем величины

$$q^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2,$$

$$s_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

и их частное $r = \frac{q^2}{s_1^2}$.

Далее, по табл. V находим соответствующее число r_p.

Если $r \geqslant r_p$, то можно допустить, что наши наблюдения не содержат систематического сдвига математических ожиданий; если же $r < r_p$, то считаем, что существует такой сдвиг.

Пример. В первом столбце табл. 11 приведены значения величины сопротивления сжатию для контрольных кубов в *кг/см*^{2} (из которых вычтено 400)*). Значения получены из проб, взятых на производстве в течение 45 дней.

Требуется определить, можно ли считать среднее сопротивление сжатию не подверженным систематическим изменениям во времени, или, напротив, в процессе производства имело место постепенное изменение генерального среднего.

Проверим гипотезу об отсутствии систематического смещения среднего с помощью критерия среднего квадрата последовательных разностей. Именно, по данным выборки вычисляем величины

$$q^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2,$$

$$s_1^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

и их отношение

$$r = \frac{q^2}{s_1^2}.$$

Малые значения r являются значимыми, т. е. при малых r гипотеза отвергается. В табл. V приложений приведены значения квантилей распределения величины r для $P = 0{,}001$; $0{,}01$; $0{,}05$ и n от 4 до 60.

---

*) Данные взяты из книги А. Хальда [48], стр. 307—309.

Таблица 11

x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$x_{i+1} - x_i$	$(x_{i+1} - x_i)^2$
40	—26,778	717,1	—7	49
33	—33,778	1141,0	42	1764
75	8,222	67,6	—57	3249
18	—48,778	2379,3	44	1936
62	—4,778	22,8	—29	841
33	—33,778	1141,0	5	25
38	—28,778	828,2	31	961
69	2,222	4,9	—4	16
65	—1,778	3,2	35	1225
100	33,222	1103,7	24	576
124	57,222	3274,4	—33	1089
91	24,222	586,7	—12	144
79	12,222	149,4	—37	1369
42	—24,778	613,9	21	441
63	—3,778	14,3	—40	1600
23	—43,778	1916,5	24	576
47	—19,778	391,2	5	25
52	—14,778	218,4	46	2116
98	31,222	974,8	—1	1
97	30,222	913,4	—24	576
73	6,222	38,7	12	144
85	18,222	332,0	3	9
88	21,222	450,4	—48	2304
40	—26,778	717,1	2	4
42	—24,778	613,9	9	81
51	—15,778	248,9	—28	784
23	—43,778	1916,5	52	2704
75	8,222	67,6	—23	529
52	—14,778	218,4	74	5476
126	59,222	3507,2	—36	1296
90	23,222	539,3	21	441
111	44,222	1955,6	—19	361
92	25,222	636,1	17	289
109	42,222	1782,7	—37	1369
72	5,222	27,3	—44	1936
28	—38,778	1503,7	28	784
56	—10,778	116,2	—39	1521
17	—49,778	2477,8	35	1225
52	—14,778	218,4	16	256
68	1,222	1,5	7	49
75	8,222	67,6	27	729
102	35,222	1240,6	5	25
107	40,222	1617,8	—30	900
77	10,222	104,5	—32	1024
45	—21,778	474,3		
Сумма . . . 3005		37335,9		42819

Проведем вычисления для нашего примера. Значения соответствующих разностей, квадратов разностей, а также суммы квадратов даны в табл. 11 $\left(\text{среднее } \bar{x} = \frac{3005}{45} = 66{,}778\right)$. Значит,

$$q^2 = \frac{1}{2 \cdot 44} \cdot 42\,819 = 486{,}6, \quad s_1^2 = \frac{1}{44} \cdot 37\,336 = 848{,}5.$$

Найдем отношение

$$r = \frac{q^2}{s_1^2} = \frac{486{,}6}{848{,}5} = 0{,}573.$$

Примем для нашего критерия 5-процентный уровень значимости, т. е. будем наблюденное r считать значимым (отвергать гипотезу), если r не превосходит квантили r_P при $P = 0{,}05$. По табл. V находим для $n = 45$ и $P = 0{,}05$ значение $r_P = 0{,}760$. Мы видим, что наблюденное r (равное 0,573) меньше 5-процентной квантили (и даже 1-процентной квантили, равной 0,666). Значит, гипотеза об отсутствии систематического смещения среднего тем самым отвергается.

Заметим, что величина r при больших n распределена приближенно нормально со средним, равным 1, и дисперсией $\sigma^2 = \frac{n-2}{(n-1)(n+1)}$. Поэтому квантили r_P распределения r при $n > 60$ можно вычислять по формуле

$$r_P \approx 1 + u_P \sqrt{\frac{n-2}{(n-1)(n+1)}}$$

(или с достаточной для практики точностью по формуле

$$r_P \approx 1 + \frac{u_P}{\sqrt{n+1}},$$

где u_P есть P-процентная квантиль нормального распределения с параметрами (0, 1))

§ 7. Групповые прямые равноточные измерения

Иногда встречаются случаи равноточных измерений целой группы физических величин $a_1, a_2, \ldots, a_n$. Величине a_i отвечают n_i измерений

$$x_{i1}, x_{i2}, \ldots, x_{in_i} \qquad (i = 1, 2, \ldots, r), \tag{4.7.1}$$

причем x_{ij} нормальны, т. е. $x_{ij} \in N(a_i, \sigma)$, и независимы в совокупности. Здесь величина $D(x_{ij}) = \sigma^2$ — одна и та же для всех наблюдений x_{ij}.

Воспользовавшись равноточностью измерений и используя сразу все наблюдения x_{ij} ($i = 1, 2, \ldots, r$; $j = 1, 2, \ldots, n_i$), можно получить для величин $a_1, a_2, \ldots, a_r$ более „точные" доверительные интервалы, чем в том случае, если для каждой величины a_i пользоваться лишь измерениями x_{ij} $(j=1, 2, \ldots, n_i)$.

Рассмотрим для каждой системы измерений (4.7.1) выборочные средние $\overline{x}_i$ и выборочные дисперсии s_i^2. Имеем

$$\overline{x}_i=\frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}; \quad s_i^2=\frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij}-\overline{x}_i)^2. \tag{4.7.2}$$

Как известно из ранее изложенного (см. теорему 4.2.1), величина $\frac{n_i s_i^2}{\sigma^2}$ распределена, как $\chi^2_{n_i-1}$, так что для суммы независимых величин $\frac{n_i s_i^2}{\sigma^2}$ получаем

$$\frac{1}{\sigma^2}\sum_{i=1}^{r} n_i s_i^2=\chi^2_{n-r}, \tag{4.7.3}$$

где

$$n=n_1+n_2+\ldots+n_r. \tag{4.7.4}$$

Положим еще $S^2=\sum_{i=1}^{r} n_i s_i^2$.

Случайная величина $\frac{(\overline{x}_i-a_i)\sqrt{n_i}}{\sigma}$ будет нормальной $N(0, 1)$ и независимой от S^2. Поэтому величина

$$t_i=\frac{\overline{x}_i-a_i}{S}\sqrt{n_i(n-r)}=\frac{\frac{(\overline{x}_i-a_i)}{\sigma}\sqrt{n_i}}{\sqrt{\frac{S^2}{n-r}}}$$

распределена по закону Стьюдента с $n-r$ степенями свободы.

Для построения доверительного интервала с заданной надежностью p находим по табл. I при $n-r$ степенях свободы абсциссу γ_p такую, что

$$P\{|t_i|\leqslant\gamma_p\}=p.$$

Тогда доверительный интервал

$$I_i=\left[\overline{x}_i-\frac{\gamma_p S}{\sqrt{n_i(n-r)}},\ \overline{x}_i+\frac{\gamma_p S}{\sqrt{n_i(n-r)}}\right] \tag{4.7.5}$$

будет накрывать измеряемую величину a_i с вероятностью p. Такое утверждение верно порознь для $i=1, 2, \ldots, r$, но вероятность одновременного накрытия величин a_i интервалами I_i не равна p^r, так как отдельные события, состоящие в накрытии a_i интервалами I_i, зависимы.

Рассмотрим частный случай $n_1 = n_2 = \ldots = n_r = 2$, встречающийся на практике. Здесь $n = 2r$, $n - r = r$, $\bar{x}_i = \frac{1}{2}(x_{i1} + x_{i2})$,

$$s_i^2 = \frac{1}{4}(x_{i1} - x_{i2})^2, \quad S^2 = \frac{1}{2}\sum_{i=1}^{r}(x_{i1} - x_{i2})^2.$$

Доверительные интервалы примут вид

$$I_i = \left[\bar{x}_i - \frac{1}{2}\gamma_p\sqrt{\frac{1}{r}\sum_{i=1}^{r}(x_{i1} - x_{i2})^2},\ \bar{x}_i + \frac{1}{2}\gamma_p\sqrt{\frac{1}{r}\sum_{i=1}^{r}(x_{i1} - x_{i2})^2}\right], \tag{4.7.6}$$

где γ_p надо находить при r степенях свободы. Сравним полученные доверительные интервалы с теми, которые получились бы, если бы для оценивания a_i применялись только наблюдения x_{i1} и x_{i2}. Доверительный интервал I_i', построенный по правилу § 4 этой главы, имел бы вид

$$I_i' = \left[\bar{x}_i - \frac{1}{2}\gamma_p'\,|x_{i1} - x_{i2}|,\ \bar{x}_i + \frac{1}{2}\gamma_p'\,|x_{i1} - x_{i2}|\right]. \tag{4.7.7}$$

Здесь γ_p' берется при $2 - 1 = 1$ степени свободы; из табл. I видно, что γ_p' убывает при увеличении числа степеней свободы, и небольшое исследование показывает, что вследствие этого I_i в среднем будет короче I_i' при больших n.

Мы приходим к правилу:

Правило обработки групповых равноточных в совокупности измерений по методу доверительных интервалов

Составим

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}; \quad s_i^2 = \frac{1}{n_i}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2;$$

$$n = \sum_{i=1}^{r} n_i; \qquad S^2 = \sum_{i=1}^{r} n_i s_i^2.$$

Выберем надежность p, в табл. I распределения Стьюдента с $(n - r)$ степенями свободы возьмем число γ_p, и построим доверительный интервал

$$I_i = \left[\bar{x}_i - \frac{\gamma_p S}{\sqrt{n_i(n-r)}},\quad \bar{x}_i + \frac{\gamma_p S}{\sqrt{n_i(n-r)}}\right].$$

Остановимся на вопросе об оценке точности при группов наблюдениях.

Произведем сперва точечное оценивание σ^2 по методу максимального подобия. Функция правдоподобия выборки (4.7.1) имеет вид

$$L(x_{ij}, a_i, \sigma) = (2\pi)^{-\frac{n}{2}} \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - a_i)^2\right\}.$$

Имеем

$$\ln L = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2} \sum_{i=1}^{r} \frac{1}{\sigma^2} \sum_{j=1}^{n_i} (x_{ij} - a_i)^2.$$

Уравнения максимального правдоподобия таковы:

$$\frac{\partial \ln L}{\partial \sigma^2} = 0; \tag{4.7.8}$$

$$\frac{\partial \ln L}{\partial a_i} = 0, \qquad i = 1, 2, \ldots, r. \tag{4.7.9}$$

Последние r уравнений сводятся к предписанию наименьших квадратов

$$Q = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - a_i)^2 = \min$$

и имеют простое решение

$$\hat{a}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} = \bar{x}_i.$$

После этого первое уравнение дает

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = 0.$$

Отсюда

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = \frac{1}{n} S^2 = \frac{1}{n} \sum n_i s_i^2.$$

Заметим, что полученная оценка для σ^2 является смещенной, так как

$$E\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{r} (n_i - 1)\, \sigma^2 = \left(1 - \frac{r}{n}\right) \sigma^2.$$

Несмещенная оценка σ^2 получается из нее по формуле

$$\tilde{\sigma}^2 = \frac{1}{n-r} S^2.$$

Можно построить и доверительный интервал для σ^2, пользуясь тем, что

$$\frac{1}{\sigma^2} S^2 = \chi^2_{n-r}.$$

Действуя так же, как в § 3 этой главы, получим следующий результат. Для данных γ_1', γ_2' и доверительного интервала

$$I_s = [S^2\gamma_1', S^2\gamma_2'],$$

имеем

$$P\{S^2\gamma_1' \leqslant \sigma^2 \leqslant S^2\gamma_2'\} = P\left\{\frac{n}{\gamma_2'} \leqslant \chi^2_{n-r} \leqslant \frac{n}{\gamma_1'}\right\} = \\ = K_{n-r}\left(\frac{n}{\gamma_2'}\right) - K_{n-r}\left(\frac{n}{\gamma_1'}\right), \qquad (4.7.10)$$

где $K_{n-r}(x)$ — интегральное распределение χ^2_{n-r}. Как и ранее, для упрощения дела будем выбирать (см. А. Н. Колмогоров [22]) γ_1' и γ_2' так, что

$$K_{n-r}\left(\frac{n}{\gamma_2'}\right) = \frac{1-p_0}{2}, \quad K_{n-r}\left(\frac{n}{\gamma_1'}\right) = \frac{1+p_0}{2},$$

где p_0 — предписанная надежность.

§ 8. Пример *)

В табл. 12 приведен ряд двойных измерений ($x_{i1} = \beta_i + x_{i1}'$, $x_{i2} = \beta_i + x_{i2}'$) угла между береговыми предметами, произведенных секстаном. После каждого измерения отсчет сбивался и заново делалось сведение изображений.

Пользуясь всей совокупностью наблюдений, оценим значения m_i для каждой группы из двух измерений.

Найдем доверительные интервалы для m_i. В случае

$$n_1 = n_2 = \ldots = n_r = 2, \quad n = \sum_{i=1} n_i = 2r, \quad n - r = r,$$

*) Данные взяты из книги А. П. Ющенко [53], стр. 42.

Таблица 12

β_i	x'_{i1}	x'_{i2}	$x'_{i1}-x'_{i2}$	$(x'_{i1}-x'_{i2})^2$	$\overline{x}'_i = \frac{1}{2}(x'_{i1}+x'_{i2})$	$m_i^H=\beta_i+\overline{x}'_i-6{,}87$	$m_i^B = \beta_i+\overline{x}'_i+6{,}87$
60°18′	20″	30″	—10	100	25	60° 18′ 18″,13	60° 18′ 31″,87
112 16	10	10	0	0	10	112 16 3,13	112 16 16,87
47 2	20	15	5	25	17,5	47 2 10,63	47 2 24,37
83 37	30	40	—10	100	35	83 37 28,13	83 37 41,87
17 12	40	40	0	0	40	17 12 33,13	17 12 46,87
32 12	35	40	—5	25	37,5	32 12 30,63	32 12 44,37
8 10	20	30	—10	100	25	8 10 18,13	8 10 31,87
51 19	30	20	10	100	25	51 19 18,13	51 19 31,87
73 51	40	30	10	100	35	73 51 28,13	73 51 41,87
90 3	55	50	5	25	52,5	90 3 45,63	90 3 59,37
Сумма . .				575			

формулы упрощаются. Именно,

$$\bar{x}_i = \frac{1}{2}(x_{i1}+x_{i2}), \quad s_i^2 = \frac{1}{4}(x_{i1}-x_{i2})^2$$

и

$$S^2 = \frac{1}{2}\sum_{i=1}^{r}(x_{i1}-x_{i2})^2.$$

Следовательно, доверительным интервалом для m_i будет интервал

$$I_i = \left[\bar{x}_i - \frac{\gamma_p S}{\sqrt{2r}}, \quad \bar{x}_i + \frac{\gamma_p S}{\sqrt{2r}}\right].$$

(У нас $\bar{x}_i = \beta_i + \bar{x}'_i$.) Вычисления приведены в табл. 12. Очевидно,

$$S^2 = \frac{1}{2}\cdot 575 \doteq 287{,}5; \qquad S = 16{,}96$$

Возьмем надежность $p=0{,}90$. По таблице распределения Стьюдента для $p=0{,}90$ и числа степеней свободы $n-r=r=10$ находим $\gamma_p=1{,}812$. Значит

$$\gamma_p = \frac{1{,}812\cdot 16{,}96}{4{,}472} = 6{,}87.$$

В двух последних столбцах таблицы выписаны значения нижней и верхней границ доверительных интервалов для m_i.

ГЛАВА V

ПРЯМЫЕ НЕРАВНОТОЧНЫЕ НАБЛЮДЕНИЯ

§ 1. Постановка задачи

В практических применениях теории обработки наблюдений, например в астрономии и геодезии, часто встречается ситуация, когда при измерении одной и той же величины получаются системы по несколько равноточных наблюдений, объединяемые в группы, которые трактуются как исходные наблюдения.

В качестве примера возьмем приведенные в книге А. С. Чеботарёва [51] данные по определению широты города Щегловска по нескольким наведениям на полярную звезду и на звезды из созвездий Андромеды, Треугольника, Лебедя и Цефея (табл. 13).

Таблица 13

№	Дата наблюдения	Название звезды	Наблюденная широта города φ	Число наблюдений
1	21/IX	α Ursae minoris	55°21′13″,5	8
2		β Andromedae	12,2	4
3		α Trianguli	11,5	2
4	4/XI	α Ursae minoris	12,4	4
5		γ Cigni	11,3	8
6		η Cephei	12,4	4

При этом для отдельных систем равноточных наблюдений (№ 1, 2, 3, ...) указаны выборочные средние $\bar{x}_1 = 55°21'13'',5$; $\bar{x}_2 = 55°21'12'',2$ и т. д. и количество наблюдений $n_i = 8, 4, 2, \ldots$. В дальнейшем (§ 6 гл. VI) мы увидим, насколько целесообразно сведéние наблюдений к выборочным средним в таком случае.

Если $\sigma^2 = D(x_i)$ — дисперсия единичного наблюдения, то $D(\bar{x}_i) = \frac{\sigma^2}{n_i}$. Таким образом, величины $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n$ нормальны; $\bar{x}_i \in N\left(a, \frac{\sigma}{\sqrt{n_i}}\right)$ $(i = 1, 2, \ldots, k)$. Можно считать, что мы должны произвести оценивание измеряемой величины a на основании наблюдений $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ с дисперсиями $\frac{\sigma^2}{n_i}$; здесь n_i — целые числа, представляющие собой количества соответствующих равноточных наблюдений. Рассмотрим более общую ситуацию. Даны неравноточные прямые наблюдения $x_1, x_2, \ldots, x_n$, независимые и нормальные, причем

$$x_i \in N\left(a, \frac{\sigma}{\sqrt{p_i}}\right),$$

где σ неизвестно, но величины p_i, обратно пропорциональные дисперсиям наблюдений $D(x_i) = \frac{\sigma^2}{p_i}$ $(i = 1, 2, \ldots, n)$, известны. Эти величины называются *весами* наблюдений. Их естественно считать определенными с точностью до коэффициента пропорциональности λ, так как

$$D(x_i) = \frac{\lambda\sigma^2}{\lambda p_i} = \frac{\mu^2}{\lambda p_i} \qquad (i = 1, 2, \ldots, n);$$

в нужные расчеты будут входить не сами веса, а отношение

$$\frac{p_i}{\sum\limits_{i=1}^{n} p_i} = \frac{p_i}{[p]}.$$

В предыдущем примере веса были пропорциональны числам наблюдений n_i. Во многих случаях веса имеют иную структуру; мы встретимся с примерами, когда веса, естественно, приписываются наблюдениям, связанным с некоторыми исходными линейной функциональной зависимостью.

Рассмотрим задачу оценивания измеряемой величины a по n неравноточным прямым независимым наблюдениям $x_1, \ldots, x_n$ с известными весами $p_1 : p_2 : \ldots : p_n$, так что

$$D(x_i) = \frac{\sigma^2}{p_i},$$

где σ^2 — неизвестный коэффициент пропорциональности, подлежащий оцениванию наряду с измеряемой величиной a.

Мы начнем с точечного оценивания a и σ^2 по методу максимального правдоподобия, убедимся, что при оценивании a получается предписание метода наименьших квадратов, с учетом весов p_i, и рассмотрим „качество“ полученных оценок. Затем перейдем к построению доверительных интервалов.

§ 2. Точечное оценивание a и σ^2

Наблюдения x_i имеют плотность вероятности

$$f(x_i, a, \sigma^2) = \frac{\sqrt{p_i}}{\sqrt{2\pi}\,\sigma} \exp -\frac{p_i}{2\sigma^2}(x_i - a)^2. \tag{5.2.1}$$

Ввиду этого и в силу независимости наблюдений, функция правдоподобия выборки $x_1, \ldots, x_n$ наших наблюдений имеет вид

$$L(x_1, \ldots, x_n, a, \sigma^2) = \\ = (p_1 p_2 \ldots p_n)^{\frac{1}{2}} (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp - \frac{1}{2\sigma^2} \sum_{i=1}^{n} p_i (x_i - a)^2.$$

Каково бы ни было $\sigma^2 > 0$, максимум $L(x_1, \ldots, x_n, a, \sigma^2)$ достигается при условии

$$Q = \sum_{i=1}^{n} p_i (x_i - a)^2 = \min, \tag{5.2.2}$$

т. е. при выборе a по предписанию наименьших квадратов с учетом весов p_i.

Для нахождения минимума Q имеем

$$\frac{\partial Q}{\partial a} = \sum_{i=1}^{n} 2p_i (x_i - a) = 0,$$

откуда находим единственное решение

$$\hat{a} = \frac{p_1 x_1 + \ldots + p_n x_n}{p_1 + \ldots + p_n} = \frac{[px]}{[p]}. \tag{5.2.3}$$

Легко обнаружить, что $\hat{a}$ доставляет минимум Q. Имеем

$$\sum p_i (x_i - \hat{a}) = \sum p_i x_i - \hat{a} \sum p_i = [px] - \hat{a}[p] = 0. \tag{5.2.4}$$

Поэтому при любом a получаем

$$Q = \sum p_i (x_i - a)^2 = \sum p_i (x_i - \hat{a} + \hat{a} - a)^2 = \\ = \sum p_i (x_i - \hat{a})^2 + (\hat{a} - a) \sum p_i (x_i - \hat{a}) + \sum p_i (\hat{a} - a)^2 = \\ = \sum p_i (x_i - \hat{a})^2 + (\hat{a} - a)^2 \sum p_i, \tag{5.2.5}$$

откуда вытекает наше утверждение.

Полученную для a оценку $\hat{a} = \frac{[px]}{[p]}$ будем обозначать $\tilde{x}$ и называть *взвешенным средним*.

Теорема 5.2.1. *Взвешенное среднее $\tilde{x}$ является несмещенной оценкой для a, т. е.*

$$E(\tilde{x}) = a, \tag{5.2.6}$$

и имеет дисперсию

$$D(\tilde{x}) = \frac{\sigma^2}{\sum_{i=1}^{n} p_i}, \tag{5.2.7}$$

так что вес взвешенного среднего равен сумме весов наблюдений.

Таким образом, теорема утверждает, что $\tilde{x} \in N\left(a, \frac{\sigma}{\sqrt{[p]}}\right)$. Для доказательства заметим, что $E(x_i) = a$ $(i = 1, 2, \ldots, n)$, поэтому

$$E\,\frac{[px]}{[p]} = \frac{1}{[p]} \sum p_i E(x_i) = \frac{a[p]}{[p]} = a,$$

далее,

$$D(\tilde{x}) = D\,\frac{1}{[p]}\left(\sum p_i x_i\right) = \frac{1}{[p]^2} \sum p_i^2 D(x_i) =$$
$$= \frac{1}{[p]^2} \sum p_i^2 \frac{\sigma^2}{p_i} = \sigma^2 \frac{[p]}{[p]^2} = \frac{\sigma^2}{[p]},$$

что доказывает (5.2.7).

Покажем, что оценка $\tilde{x}$ эффективна. На основании (3.3.13) имеем для весьма широкого класса несмещенных оценок t:

$$D(t) = E(t-a)^2 \geqslant \frac{1}{E\left(-\frac{\partial^2 \ln L}{\partial a^2}\right)}. \tag{5.2.8}$$

Здесь

$$\ln L = \ln (p_1 p_2 \ldots p_n)^{\frac{1}{2}} - \frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} p_i (x_i - a)^2;$$

$$\frac{\partial^2 \ln L}{\partial a^2} = -\frac{\sum p_i}{\sigma^2} = -\frac{[p]}{\sigma^2};$$

$$E\left(-\frac{\partial^2 \ln L}{\partial a^2}\right) = \frac{[p]}{\sigma^2}.$$

Таким образом,

$$D(t) \geqslant \frac{\sigma^2}{[p]} = D(\tilde{x}),$$

что и доказывает эффективность оценки $\tilde{x}$.

Реальный смысл оценивания с помощью $\tilde{x}$ таков: при заданном $\varepsilon > 0$ в весьма широком классе асимптотически нормальных и асимптотически несмещенных оценок t имеем

$$P\{|\tilde{x} - a| \leqslant \varepsilon\} \geqslant P\{|t - a| \leqslant \varepsilon\} \tag{5.2.9}$$

при достаточно большом объеме выборки n.

Для оценивания σ^2 составляем уравнение

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} p_i (x_i - a)^2 = 0.$$

Отсюда

$$\widehat{\sigma^2} = \frac{1}{n} \sum p_i (x_i - \hat{a})^2 = \frac{1}{n} \sum p_i (x_i - \tilde{x})^2. \tag{5.2.10}$$

Найдем $E(\widehat{\sigma^2})$, чтобы установить, насколько смещена оценка $\hat{\sigma}^2$. Полагаем $y_i = x_i - a$; тогда $\tilde{y} = \frac{[py]}{[p]} = \tilde{x} - a$, и

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} p_i (x_i - \tilde{x})^2 = \frac{1}{n}\sum_{i=1}^{n} p_i (y_i - \tilde{y})^2.$$

Поэтому

$$E(\widehat{\sigma^2}) = \frac{1}{n}\left[\sum_i p_i E y_i^2 - 2\sum_i p_i E y_i \tilde{y} + \sum_i p_i E y_i^2\right]. \quad (5.2.11)$$

Далее, имеем

$$E(y_i^2) = \frac{\sigma^2}{p_i},$$

$$E(y_i \tilde{y}) = \frac{1}{[p]}\sum_j p_j E(y_i y_j).$$

Заметим, что $E(y_i y_j) = 0$ при $i \neq j$, так что

$$E(y_i \tilde{y}) = \frac{1}{[p]} p_i \frac{\sigma^2}{p_i} = \frac{\sigma^2}{[p]};$$

$$E(\tilde{y}^2) = \frac{\sigma^2}{[p]}.$$

Отсюда, подставляя в (5.2.11), получаем

$$E(\widehat{\sigma^2}) = \frac{1}{n}[n\sigma^2 - 2\sigma^2 + \sigma^2] = \frac{n-1}{n}\sigma^2.$$

Таким образом, $\widehat{\sigma^2}$ является смещенной оценкой для σ^2, но асимптотически несмещенной при $n \to \infty$. Несмещенную оценку для σ^2, очевидно, дает величина:

$$\tilde{q}^2 = \frac{n}{n-1}\sigma^2 = \frac{1}{n-1}\sum p_i (x_i - \tilde{x})^2 = \frac{Q}{n-1}.$$

Далее, найдем несмещенную оценку для дисперсии $\tilde{x}$. Имеем

$$D(\tilde{x}) = \frac{\sigma^2}{[p]},$$

так что несмещенная оценка для $D(\tilde{x})$ получится из $\tilde{q}^2$, если возьмем:

$$u^2 = \frac{\tilde{q}^2}{[p]} = \frac{1}{[p](n-1)}\sum_{i=1}^{n} p_i (x_i - \tilde{x})^2. \quad (5.2.12)$$

Полученные формулы дают решение задачи точечного оценивания в случае неравноточных наблюдений с известными весами; реальный смысл такого оценивания поясняется формулой (5.2.9). Частный случай при $p_1 = p_2 = \ldots = p_n = 1$ приводит к ранее изученным (гл. IV) равноточным измерениям.

Остановимся еще на упрощении вычислений при расчете $\tilde{x}$, u^2 и $\tilde{q}^2$. Если β — любое число, то, как легко проверить непосредственно,

$$\tilde{x} = \frac{1}{[p]}[px] = \frac{1}{[p]}[p(x-\beta)] + \beta. \tag{5.2.13}$$

Подставляя сюда вместо β число $\tilde{x}$, находим:

$$\tilde{x} = \frac{1}{[p]}[p(x-\tilde{x})] + \tilde{x},$$

откуда

$$\frac{1}{[p]}[p(x-\tilde{x})] = 0. \tag{5.2.14}$$

Это равенство может служить для контроля вычисления $\tilde{x}$. Далее, согласно (5.2.5), с заменой $\hat{a}$ на $\tilde{x}$ и a на β, имеем

$$\sum p_i(x_i-\tilde{x})^2 = \sum p_i(x_i-\beta)^2 - (\tilde{x}-\beta)^2\cdot[p], \tag{5.2.15}$$

что также можно использовать для упрощения вычислений.

§ 3. Оценивание a и σ^2 с помощью доверительных интервалов

Для оценивания a и σ^2 с помощью доверительных интервалов докажем сначала следующую теорему.

Теорема 5.3.1. *Величины $\tilde{x}$ и*

$$q^2 = \frac{1}{n-1}Q = \frac{1}{n-1}\sum_i p_i(x_i-\tilde{x})^2$$

статистически независимы, и величина Q распределена, как $\sigma^2\chi^2_{n-1}$, для рассматриваемого нами случая $x_i \in N\left(a, \frac{\sigma}{\sqrt{p_i}}\right)$.

Сперва докажем, подобно тому, как это делалось в § 2 гл. IV, что нормальный вектор $(x_1-\tilde{x}, x_2-\tilde{x}, \ldots, x_{n-1}-\tilde{x})$ не зависит от $\tilde{x}$. Для этого достаточно проверить, что

$$E[(x_i-\tilde{x})\tilde{x}] = 0.$$

Полагая, как и ранее, $y_i = x_i - a$, $\tilde{y} = \tilde{x} - a$, имеем

$$E[(x_i-\tilde{x})\tilde{x}] = E[(y_i-\tilde{y})(\tilde{y}+a)] = E(y_i-\tilde{y})\tilde{y} =$$
$$= Ey_i\tilde{y} - E\tilde{y}^2 = \frac{\sigma^2}{[p]} - \frac{\sigma^2}{[p]} = 0. \text{ (см. § 2 этой главы).}$$

Значит, случайный вектор $(x_1-\tilde{x}, \ldots, x_{n-1}-\tilde{x})$ не зависит от $\tilde{x}$. Далее, $x_n - \tilde{x} = -\sum_{i=1}^{n-1}(x_i-\tilde{x})$, так что любые непрерывные функции $\varphi(x_1 - \tilde{x}, \ldots, x_n - \tilde{x})$ статистически не зависят от $\tilde{x}$; в частности, $Q = \sum_{i=1}^{n} p_i (x_i - \tilde{x})^2$ не зависит от $\tilde{x}$. Докажем теперь, что величина

$$Q = \sum_{i=1}^{n} p_i (x_i - \tilde{x})^2 \tag{5.3.1}$$

распределена, как $\sigma^2\chi^2_{n-1}$. Имеем, полагая в (5.2.13) $\beta = 0$,

$$Q = \sum_{i=1}^{n} p_i (x_i - \tilde{x})^2 = \sum_{i=1}^{n} p_i (y_i - \tilde{y})^2 = \sum_{i=1}^{n} p_i y_i^2 - [p]\,\tilde{y}^2.$$

Здесь y_i независимы и нормальны; $y_i \in N\left(0, \frac{\sigma}{\sqrt{p_i}}\right)$. Поэтому величины $u_i = y_i \sqrt{p_i}$ также независимы и нормальны $N(0, \sigma)$. Имеем

$$Q = \sum_{i=1}^{n} u_i^2 - (\tilde{y}\sqrt{[p]})^2;$$

$$\tilde{y}\sqrt{[p]} = \sqrt{[p]}\,\frac{\sum p_i y_i}{[p]} = \frac{1}{\sqrt{[p]}} \sum_{i=1}^{n} u_i \sqrt{p_i}.$$

Введем ортогональное преобразование

$$Z = FU, \quad \text{где} \quad U = \begin{Vmatrix} y_1 \\ \cdot \\ \cdot \\ y_n \end{Vmatrix}, \quad Z = \begin{Vmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ z_n \end{Vmatrix},$$

матрица $F = \|f_{rs}\|$ ортогональна: $FF^T = E$, и первая строка F отвечает равенству

$$z_1 = \sum_{i=1}^{n} u_i \frac{\sqrt{p_i}}{\sqrt{[p]}} = \tilde{y}\sqrt{[p]} \qquad \left(f_{1s} = \frac{\sqrt{p_s}}{\sqrt{[p]}}\right).$$

По теореме Фишера (теорема 2.3.2), случайный вектор $(z_1, \ldots, z_n)$ будет нормальным с независимыми компонентами $z_i \in N(0, \sigma)$.

Имеем

$$Q = \sum_{i=1}^{n} u_i^2 - u_1^2 = \sum_{i=2}^{n} u_i^2.$$

Поэтому величина $\frac{Q}{\sigma^2}$ будет распределена как χ^2_{n-1}. Этим доказательство теоремы завершается.

Составим теперь дробь, аналогичную дроби $\frac{\bar{x}-a}{s}$ в случае равноточных наблюдений и распределенную по закону Стьюдента.

Величина $\bar{x}-a$ будет нормальной $N\left(0, \frac{\sigma}{\sqrt{[p]}}\right)$ и независимой от $\sqrt{Q}$; таким образом, величина

$$t_{n-1}=\frac{(\bar{x}-a)\sqrt{[p]}\,\sigma^{-1}}{\sqrt{\frac{Q}{[p]}}\,\sigma^{-1}}=\frac{(\bar{x}-a)\sqrt{n-1}}{\tilde{s}}, \tag{5:3.2}$$

где

$$\tilde{s}=\sqrt{\frac{Q}{[p]}}=\left(\frac{1}{[p]}\sum p_i(x_i-\tilde{x})^2\right)^{\frac{1}{2}} \tag{5.3.3}$$

будет распределена по закону Стьюдента с $n-1$ степенями свободы. Поэтому построение доверительных интервалов проводится так же, как в § 2 гл. IV для равноточных наблюдений: выбираем надежность p_0, и по числу степеней свободы $n-1$ находим число γ такое, что

$$P\{|t_{n-1}|\leqslant\gamma\}=p_0.$$

Из (5.3.2) видим, что доверительный интервал

$$I_\gamma=\left[\tilde{x}-\frac{\tilde{s}}{\sqrt{n-1}},\quad \tilde{x}+\frac{\tilde{s}}{\sqrt{n-1}}\right] \tag{5.3.4}$$

будет накрывать измеряемую величину a с вероятностью p_0.

Перейдем к оцениванию точности по методу доверительных интервалов. Имеем, как и в § 2 гл. IV,

$$\frac{Q}{\sigma^2}=\frac{[p]\,\tilde{s}^2}{\sigma^2}=\chi^2_{n-1}.$$

Составим доверительный интервал

$$I=[q^2\gamma_1^2,\ q^2\gamma_2^2].$$

Известно, что величина $\frac{(n-1)q^2}{\sigma^2}$ распределена как χ^2_{n-1}. Таким образом,

$$P\{q^2\gamma_1^2\leqslant\sigma^2\leqslant q^2\gamma_2^2\}=$$
$$=P\left\{\frac{n-1}{\gamma_2^2}\leqslant\frac{(n-1)q^2}{\sigma^2}\leqslant\frac{n-1}{\gamma_1^2}\right\}=P\left\{\frac{n-1}{\gamma_2^2}\leqslant\chi^2_{n-1}\leqslant\frac{n-1}{\gamma_1^2}\right\}=$$
$$=K_{n-1}\left(\frac{n-1}{\gamma_1^2}\right)-K_{n-1}\left(\frac{n-1}{\gamma_2^2}\right).$$

Ввиду этого надежность устанавливается по заданным γ_1, γ_2 (или наоборот) с помощью таблиц.

Мы можем сформулировать теперь следующие правила.

Правила оценивания измеряемой величины и точности независимых наблюдений с известными весами

а) Точечное оценивание

Находим взвешенное среднее $\tilde{x}$, q^2, u^2 по формулам:

$\tilde{x} = \dfrac{\sum_{i=1}^{n} p_i x_i}{\sum p_i} = \dfrac{[px]}{[p]}$ — несмещенная оценка для измеряемой величины a,

$q^2 = \dfrac{1}{n-1} \sum p_i (x_i - \tilde{x})^2 = \dfrac{[p\tilde{v}\tilde{v}]}{n-1}$ — несмещенная оценка для σ^2,

где $v = x_i - \tilde{x}$ — „кажущиеся погрешности“,

$u^2 = \dfrac{q^2}{[p]} = \dfrac{[p\tilde{v}\tilde{v}]}{(n-1)[p]}$ — несмещенная оценка для дисперсии $\tilde{x}$, $D(\tilde{x}) = \dfrac{\sigma^2}{[p]}$.

Для облегчения вычислений выберем вспомогательное число β наиболее удобным образом. Имеем

$$\tilde{x} = \frac{1}{[p]} [p(x - \beta)] + \beta;$$

контрольное соотношение: $[v] = 0$.

$$(n-1) q^2 = \sum_{i=1}^{n} p_i (x_i - \beta)^2 - (\tilde{x} - \beta)^2 [p],$$

$$u^2 = \frac{q^2}{[p]}.$$

В частном случае, когда веса p_i пропорциональны числам наблюдений n_i в сериях равноточных наблюдений, можно взять $p_i = n_i$; вместо x_i берется $\bar{x}_i$ — выборочное среднее в серии n_i наблюдений.

б) Оценивание с помощью доверительных интервалов

Находим $\tilde{x}$ и $\tilde{s} = \sqrt{\dfrac{[p\tilde{v}\tilde{v}]}{[p]}}$. Назначаем надежность p_0 и по таблице распределения Стьюдента (табл. I) по заданной надежности p_0 и числу степеней свободы $n-1$ находим число γ. Доверительный интервал

$$I_\gamma = \left[\tilde{x} - \gamma \frac{\tilde{s}}{\sqrt{n-1}},\ \tilde{x} + \gamma \frac{\tilde{s}}{\sqrt{n-1}}\right]$$

будет накрывать измеряемую величину a с вероятностью p_0. Для оценивания σ^2 назначаем надежность p_0 и по табл. III при $n-1$ степенях свободы находим γ_1, γ_2. Доверительный интервал

$$I_{\gamma_1\gamma_2}=[q^2\gamma_1^2,\ q^2\gamma_2^2]$$

накрывает σ^2 с вероятностью p_0.

Логически естественно далее рассмотреть случай неравноточных наблюдений с неизвестными весами, т. е. задачу оценивания измеряемой величины a по выборке $x_1, x_2, \ldots, x_n$ независимых наблюдений $x_i \in N(a, \sigma_i)$; σ_i различны и неизвестны нам. Эта задача теоретически может быть решена, но пока ее решение не приведено в удобную для практического использования форму. Поэтому мы не будем ею заниматься.

§ 4. Примеры

Пример 1 *). В первом столбце табл. 14 приведены значения угла α, полученные при измерении его различным числом приемов.

Таблица 14

α_i	Число приемов	p_i	$\alpha_i'=\alpha_i-\beta$	$p_i\alpha_i'$	$\alpha_i-\tilde{\alpha}$	$(\alpha_i-\tilde{\alpha})^2$	$p_i(\alpha_i-\tilde{\alpha})^2$
46°10′12″	8	2	2″	4	−1″,75	3,063	6,126
14	4	1	4	4	0,25	0,063	0,063
12	12	3	2	6	−1,75	3,063	9,189
18	8	2	8	16	4,25	18,063	36,126
Сумма . .		8		30			51,504

Оценим значение угла α и точность σ его определения. При этом веса отдельных результатов измерений принимаем пропорциональными числу приемов.

Точечной оценкой для α будет

$$\tilde{\alpha}=\frac{[p\alpha]}{[p]}=\beta+\frac{[p\alpha']}{[p]}.$$

Вычислим

$$\tilde{s}=\sqrt{\frac{[p\tilde{v}\tilde{v}]}{[p]}},\quad \text{где}\quad \tilde{v}_i=\alpha_i-\tilde{\alpha};\quad [p\tilde{v}\tilde{v}]=\sum_{i=1}^{n}p_i(\alpha_i-\tilde{\alpha})^2.$$

Оценить σ можно с помощью величины

$$q=\sqrt{\frac{[p\tilde{v}\tilde{v}]}{n-1}}.$$

*) Данные заимствованы из книги А. П. Ющенко [56], стр. 51.

Вычисления приведены в табл. 14. Здесь принято $\beta = 46°10'10''$. Значит,

$$\tilde{\alpha} = 46°10'10'' + \frac{30''}{8} = 46°10'13'',75.$$

Очевидно,

$$\tilde{s} = \sqrt{\frac{51,504}{8}} = 2,537, \quad q = \sqrt{\frac{51,504}{3}} = 4,144.$$

Оценим α и σ с помощью доверительных интервалов. Доверительным интервалом для α является интервал

$$I = \left[\tilde{x} - \frac{\gamma_p \tilde{s}}{\sqrt{n-1}}, \quad \tilde{x} + \frac{\gamma_p \tilde{s}}{\sqrt{n-1}}\right],$$

где γ_p определяется по таблице Стьюдента для надежности p и числа степеней свободы $k = n - 1$. Возьмем надежность $p = 0,90$. По таблице Стьюдента для $n - 1 = 3$ степеней свободы находим $\gamma_p = 2,353$. Значит,

$$\frac{\gamma_p \cdot \tilde{s}}{\sqrt{n-1}} = \frac{2,353 \cdot 2,537}{1,732} = 3'',45,$$

и мы имеем для α доверительный интервал

$$46°10'10'',30 < \alpha < 46°10'17'',20.$$

Доверительный интервал для σ находится с помощью таблиц χ^2-распределения из условия

$$P\left\{\sqrt{\frac{[p]}{\chi_1^2}}\,\tilde{s} < \sigma < \sqrt{\frac{[p]}{\chi_2^2}}\,\tilde{s}\right\} = P_2 - P_1 = p,$$

или из условия

$$P\left\{\sqrt{\frac{n-1}{\chi_1^2}}\,q < \sigma < \sqrt{\frac{n-1}{\chi_2^2}}\,q\right\} = P_2 - P_1 = p.$$

Здесь χ_2^2 и χ_1^2 соответствуют вероятностям P_2, P_1 при $n - 1$ степенях свободы. Удобнее пользоваться вторым условием, поскольку коэффициенты $\sqrt{\frac{n-1}{\chi_1^2}} = \gamma_1$ и $\sqrt{\frac{n-1}{\chi_2^2}} = \gamma_2$ протабулированы для различных p и $k = n - 1$ (см. табл. III приложений).

Возьмем $p = 0,90$ ($P_1 = 0,05$; $P_2 = 0,95$). По табл. III для $k = n - 1 = 3$ и $p = 0,90$ находим $\gamma_1 = 0,620$, $\gamma_2 = 2,92$. Значит, доверительным интервалом для σ при надежности $p = 0,90$ является интервал $0,620\,q < \sigma < 2,92\,q$ или $2,56 < \sigma < 12,10$.

Пример 2*). В первом столбце табл. 15 даны 7 значений постоянной солнечного параллакса, полученных разными наблюдателями и разными методами. Во втором столбце даны веса.

Таблица 15

x_i	p_i	$x'_i = x_i - \beta$	$p_i x'_i$	$(x_i - \tilde{x})$	$(x_i - \tilde{x})^2$	$p_i (x_i - \tilde{x})^2$
8″,780	1,0	0,080	0,0800	—0,0259	0,000671	0,000671
8,794	0,8	0,094	0,0752	—0,0119	0,000142	0,000114
8,857	0,8	0,157	0,1256	0,0511	0,002611	0,002089
8,802	8,2	0,102	0,8364	—0,0039	0,000015	0,000123
8,806	0,2	0,106	0,0212	0,0001	0,000000	0,000000
8,806	11,1	0,106	1,1766	0,0001	0,000000	0,000000
8,807	25,0	0,107	2,6750	0,0011	0,000001	0,000025
Сумма . .	47,1		4,9900			0,003022

Оценим значение a постоянной солнечного параллакса, а также точность σ измерений.

Находим точечные оценки для a и σ. Вычисления приведены в табл. 15. Полагаем $\beta = 8'',7$, тогда

$$\tilde{x} = \frac{[px]}{[p]} = \beta + \frac{[px']}{[p]} = 8'',7 + \frac{4'',99}{47,1} = 8'',8059.$$

Вычисляем

$$\tilde{s} = \sqrt{\frac{0,003022}{47,1}} = 0,00801, \qquad q = \sqrt{\frac{0,003022}{6}} = 0,0224.$$

Ищем доверительный интервал для a. Принимаем $p = 0,90$. Тогда для $k = n - 1 = 6$ степеней свободы и $p = 0,90$ находим по таблице распределения Стьюдента значение $\gamma_p = 1,943$. Имеем

$$\frac{\gamma_p \tilde{s}}{\sqrt{n-1}} = \frac{1,943 \cdot 0,00801}{2,45} = 0'',0064.$$

Следовательно, доверительным интервалом для a является интервал

$$8'',7995 < a < 8'',8123.$$

Оценим точность наблюдений при той же надежности $p = 0,90$. Для $k = n - 1 = 6$ и $p = 0,90$ находим по табл. III значения $\gamma_1 = 0,690$, $\gamma_2 = 1,92$. Следовательно, доверительным интервалом для σ является интервал

$$0,690\, q < \sigma < 1,92\, q \quad \text{или} \quad 0,015 < \sigma < 0,043.$$

*) Данные взяты из книги Н. И. Идельсона [18], стр. 99.

Пример 3*). Даны результаты 5 измерений высоты h некоторого пункта. Эти результаты измерений h_i в миллиметрах, а также веса приведены в первых двух столбцах табл. 16.

Таблица 16

h_i	p_i	$h'_i = h_i - \beta$	$p_i h'_i$	$(h_i - \tilde{h})$	$(h_i - \tilde{h})^2$	$p_i(h_i - \tilde{h})^2$
12356	1,0	6	6,0	−0,64	0,410	0,410
12361	0,4	11	4,4	4,36	19,010	7,604
12357	3,0	7	21,0	0,36	0,130	0,390
12355	0,2	5	1,0	−1 64	2,690	0,538
12352	0,4	2	0,8	−4,64	21,530	8,612
Сумма . .	5,0		33,2			17,554

Оценим значение h, а также точность наблюдений σ.

Найдем сначала точечные оценки для h и σ. Вычисления приведены в табл. 16. Полагаем $\beta = 12\,350$ *мм*. Тогда

$$\tilde{x} = \beta + \frac{[ph']}{[p]} = 12\,350 + \frac{33{,}2}{5} = 12\,356{,}64 \text{ мм}.$$

Вычисляем $\tilde{s} = \sqrt{\frac{17{,}554}{5{,}0}} = 1{,}87$, $q = \sqrt{\frac{17{,}554}{4}} = 2{,}09$.

Ищем доверительный интервал для h. Возьмем надежность $p = 0{,}90$. Тогда для числа степеней свободы $k = n - 1 = 4$ и вероятности $p = 0{,}90$ находим по таблице распределения Стьюдента значение $\gamma_p = 2{,}132$. Стало быть,

$$\frac{\gamma_p \tilde{s}}{\sqrt{n-1}} = \frac{2{,}132 \cdot 1{,}87}{2} = 1{,}99.$$

Доверительный интервал для h таков:

$$12\,354{,}6 < h < 12\,358{,}7.$$

Оценим точность наблюдений при той же надежности $p = 0{,}90$. По табл. III для $k = n - 1 = 4$ находим $\gamma_1 = 0{,}649$, $\gamma_2 = 2{,}37$. Значит, доверительным интервалом для σ будет интервал

$$0{,}649q < \sigma < 2{,}37q \quad \text{или} \quad 1{,}35 < \sigma < 4{,}96.$$

Пример 4**). В табл. 17 приводятся данные по измерению широты г. Щегловска, определенной по разным звездам и с разным числом n_i наведений. В таблице даны средние значения x_i из n_i

*) Данные взяты из книги Н. И. Идельсона [19], стр. 84.
**) Данные заимствованы из книги А. С. Чеботарёва [51].

наведений на эти звезды. Измерения в каждой группе считаем одинаково точными.

Таблица 17

x_i	$p_i\,(n_i)$	$x'_i = x_i - \beta$	$p_i x'_i$	$x_i - \tilde{x}$	$(x_i - \tilde{x})^2$	$p_i(x_i - \tilde{x})^2$
55°21′13″,5	8	3″,5	28,0	1,187	1,409	11,272
12,2	4	2,2	8,8	—0,113	0,013	0,052
11,5	2	1,5	3,0	—0,813	0.661	1,322
12,4	4	2,4	9,6	0,087	0,008	0,032
11,3	8	1,3	10,4	—1,013	1,026	8,208
12,4	4	2,4	9,6	0,087	0,008	0,032
Сумма . .	30		69,4			20,918

Оценим значение a широты г. Щегловска и штандарт σ наблюдений.

Найдем точечные оценки для a и σ. Вычисления приведены в табл. 17. Полагаем $\beta = 55°21'10''$. Тогда оценка $\tilde{x}$ для a равна

$$\tilde{x} = \beta + \frac{[px']}{[p]} = 55°21'10'' + \frac{69'',4}{30} = 55°21'12'',313.$$

Вычислим величины

$$\tilde{s} = \sqrt{\frac{20{,}918}{30}} = 0{,}835 \quad \text{и} \quad q = \sqrt{\frac{20{,}918}{5}} = 2{,}045.$$

Строим доверительный интервал для a. Возьмем $p = 0{,}90$. Тогда для числа степеней свободы $k = n - 1 = 5$ и вероятности $p = 0{,}90$ находим по таблице t-распределения значение $\gamma_p = 2{,}015$. Значит,

$$\frac{\gamma_p \tilde{s}}{\sqrt{n-1}} = \frac{2{,}015 \cdot 0{,}835}{2{,}236} \doteq 0'',752.$$

Итак, с надежностью 0,90 можем утверждать, что имеет место неравенство

$$55°21'11'',56 < a < 55°21'13'',07.$$

Оценим σ при той же надежности $p = 0{,}90$. Для $p = 0{,}90$ и $k = n - 1 = 5$ находим по табл. III приложений $\gamma_1 = 0{,}67$; $\gamma_2 = 2{,}09$. Следовательно, доверительным интервалом для σ будет интервал

$$0{,}67q < \sigma < 2{,}09q \quad \text{или} \quad 1{,}37 < \sigma < 4{,}28.$$

ГЛАВА VI

НЕПРЯМЫЕ (КОСВЕННЫЕ) БЕЗУСЛОВНЫЕ ИЗМЕРЕНИЯ

§ 1. Постановка задачи

Во введении [§ 1, формулы (0.1.1), (0.1.21) и др.] мы видели, что многие задачи, в которых применяется метод наименьших квадратов, касаются определения коэффициентов $a_1, a_2, \ldots, a_k$ в избыточной системе уравнений вида

$$y_r = \sum_{j=1}^{n} a_j x_{rj} \qquad (r = 1, 2, \ldots, N, \quad N > n). \tag{6.1.1}$$

Эта избыточная система уравнений в § 1 введения естественно появлялась в ряде различных примеров.

В дальнейшем у нас будут еще примеры подобной ситуации, теперь же поставим задачу более точно.

Пусть нам желательно найти n величин $a_1, a_2, \ldots, a_n$ на основе уравнений (6.1.1), связывающих y_r, a_j, x_{jr}. Здесь нам точно известны величины x_{jr}, а величины $y_1, y_2, \ldots, y_N$ измерены с погрешностями $\Delta_1, \Delta_2, \ldots, \Delta_N$, так что мы наблюдаем искаженные погрешностями величины

$$l_r = y_r + \Delta_r \qquad (r = 1, 2, \ldots, N). \tag{6.1.2}$$

Далее, мы предположим, что в случае точного определения $y_1, y_2, \ldots, y_N$ [при котором уравнения (6.1.1) должны оказаться совместными] мы можем найти из этих уравнений $a_j (j = 1, 2, \ldots, n)$. Это значит, что ранг матрицы

$$X = X_{Nn} = \| x_{rj} \| \tag{6.1.3}$$

равен n и, стало быть, $N \geqslant n$ (фактически нам понадобится, чтобы N было больше n). При данной постановке задачи можно применять предписание наименьших квадратов, и, соответственно ему, однозначно вычислить $a_1, \ldots, a_n$. При этом вычислительный аппарат будет состоять в применении хорошо изученных приемов решения линейных уравнений, облегчаемых специальным видом (симметрией) получаемых уравнений (называемых нормальными).

Мы, однако, введем дополнительные условия о характере погрешностей Δ_r. Эти условия будут хорошо схематизировать то, что встречается на практике и, вместе с тем, с теоретической стороны дадут возможность составить понятие о точном математико-статистическом смысле получаемых результатов.

Допустим, что погрешности $\Delta_1, \Delta_2, \ldots, \Delta_N$ независимы, несмещены и нормальны; $\Delta_r \in N(0, \sigma_r)$. При этом

$$\sigma_r^2 = \frac{\sigma^2}{p_r} \qquad (r = 1, 2, \ldots, N), \tag{6.1.4}$$

где веса p_r будут считаться известными, а σ^2 неизвестным. На основании наблюдений $l_1, l_2, \ldots, l_N$ мы должны произвести оценивание параметров $a_1, a_2, \ldots, a_n$ и σ. Заметим, что рассматриваемый нами случай неравноточных погрешностей легко сводится к случаю равноточных. Именно, если умножим r-е уравнение (6.1.1) на $\sqrt{p_r}$ и заменим y_r на $\sqrt{p_r}\, y_r$ и, соответственно, l_r на $l_r\sqrt{p_r}$, Δ_r на $\Delta_r' = \sqrt{p_r}\,\Delta_r$, x_{rj} на $\sqrt{p_r}\, x_{rj}$, то получим системы

$$y_r\sqrt{p_r} = \sum_{j=1}^{n} a_j x_{rj}\sqrt{p_r}\,, \tag{6.1.5}$$

$$l_r\sqrt{p_r} = y_r\sqrt{p_r} + \Delta_r', \tag{6.1.6}$$

где

$$E(\Delta_r') = 0; \quad D(\Delta_r') = \sigma^2; \quad \Delta_r' \in N(0, \sigma), \tag{6.1.7}$$

и Δ_r' независимы в совокупности, так что получается случай равноточных измерений. В дальнейшем, однако, мы будем трактовать непосредственно случай неравноточных измерений, сводя его к случаю равноточных измерений лишь тогда, когда это сильно упрощает выводы.

§ 2. Применение метода наименьших квадратов

Как и ранее, начнем с применения метода максимального правдоподобия. Составим функцию правдоподобия выборки из независимых наблюдений $l_1, l_2, \ldots, l_N$. Имеем

$$l_r = y_r + \Delta_r = \Delta_r + \sum_{j=1}^{n} a_j x_{rj}. \tag{6.2.1}$$

Здесь $\Delta_r \in N\left(0, \frac{\sigma}{\sqrt{p_r}}\right)$ случайно, а второе слагаемое постоянно и содержит неизвестные параметры $a_1, \ldots, a_n$.

Имеем

$$E(\Delta_r) = 0; \quad E(l_r) = \sum_{j=1}^{n} a_j x_{rj}. \tag{6.2.2}$$

Ввиду этого, l_r нормально и имеет плотность вероятности

$$f_r(l_r)=\frac{p_r^{\frac{1}{2}}(2\pi)^{-\frac{1}{2}}}{\sigma}\exp\left\{-\frac{p_r}{2\sigma^2}\left(l_r-\sum_{j=1}^{n}a_jx_{rj}\right)^2\right\}. \tag{6.2.3}$$

Таким образом, функция правдоподобия выборки $l_1, l_2, \ldots, l_N$ примет вид

$$L(l_1, \ldots, l_N)=(2\pi)^{-\frac{N}{2}}(\sigma^2)^{-\frac{N}{2}}(p_1p_2\ldots p_N)^{\frac{1}{2}}\times$$
$$\times\exp\left\{-\frac{1}{2\sigma^2}\sum_{r=1}^{N}p_r\left(l_r-\sum_{j=1}^{n}a_jx_{rj}\right)^2\right\}.$$

Далее, выражение

$$\ln L(l_1, \ldots, l_N)=-\frac{N}{2}\ln 2\pi-\frac{N}{2}\ln\sigma^2+\frac{1}{2}\ln(p_1p_2\ldots p_N)-$$
$$-\frac{1}{2\sigma^2}\sum_{r=1}^{N}p_r\left(l_r-\sum_{j=1}^{n}a_jx_{rj}\right)^2$$

должно быть максимальным.

Мы видим, что при всяком значении σ^2 максимум функции правдоподобия достигается при выборе $a_1, \ldots, a_n$, не зависящем от σ^2, таком, что

$$\sum_{r=1}^{N}p_r\left(l_r-\sum_{j=1}^{n}a_jx_{rj}\right)^2=\min, \tag{6.2.4}$$

т. е. при выборе a_j по предписанию наименьших квадратов. Для того чтобы получить представление о наглядном смысле этого предписания, изложим расчет для случая равноточных измерений ($p_i=1$) по методу наименьших квадратов на языке многомерной геометрии (см. А. Н. Колмогоров [22]). Введем N-мерные векторы

$$l=(l_1, \ldots, l_N);\quad y=(y_1, \ldots, y_N);\quad \Delta=(\Delta_1, \Delta_2, \ldots, \Delta_N),$$
$$x_j=(x_{1j}, x_{2j}, \ldots, x_{Nj}).$$

Предписание наименьших квадратов для данного случая будет иметь вид

$$\sum_{r=1}^{N}\left(l_r-\sum_{j=1}^{n}a_jx_{rj}\right)^2=\min. \tag{6.2.5}$$

Значения $a_1, a_2, \ldots, a_n$, получаемые из условия (6.2.5), обозначим $\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n$, и введем вектор

$$\tilde{l}=\sum_{j=1}^{n}\tilde{a}_jx_j. \tag{6.2.6}$$

Положим

$$l-\tilde{l}=\varepsilon;\quad l=\tilde{l}+\varepsilon. \tag{6.2.7}$$

Здесь $\varepsilon=(l_1-\tilde{l}_1, l_2-\tilde{l}_2, \ldots, l_n-\tilde{l}_n)$ — вектор разностей.

Если $a_1, a_2, \ldots, a_n$ могут принимать любые значения, то совокупность векторов

$$m=\sum_{j=1}^{n} a_j x_j$$

образует линейное пространство, построенное на $x_1, x_2, \ldots, x_n$, $l=(l_1, \ldots, l_N)$ — фиксированный вектор. Условие (6.2.5) означает, что $\sum\limits_{i=1}^{n}(l_i-\tilde{l}_i)^2=\min$, т. е. что скалярный квадрат вектора ε, равный $[\varepsilon\varepsilon]=\sum\limits_{i=1}^{n}\varepsilon_i^2$, должен быть минимальным.

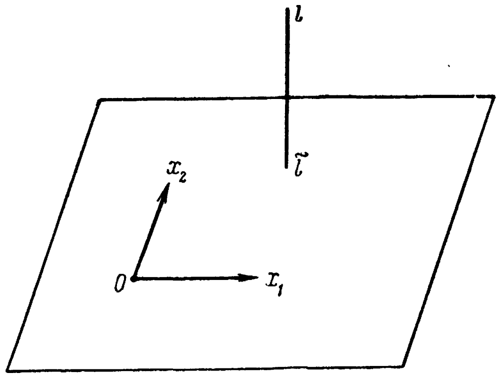

Рис. 4.

С точки зрения многомерной геометрии $[\varepsilon\varepsilon]$ есть квадрат расстояния от конца вектора l до конца переменного вектора $\tilde{l}$, лежащего в линейном пространстве, образованном векторами $x_1, x_2, \ldots, x_n$. Такое расстояние будет минимальным тогда и только тогда, когда вектор $l-\tilde{l}$ будет перпендикулярен к этому линейному пространству, т. е. $l-\tilde{l}\perp x_i$ $(i=1, 2, \ldots, n)$, или $\varepsilon\perp x_i$, $(i=1, 2, \ldots, n)$. (См. рис. 4, где взято $N=3$, $n=2$).

Это означает, что

$$[\varepsilon x_i]=\sum_{r=1}^{n}\varepsilon_r x_{ri}=0 \qquad (i=1, 2, \ldots, n). \tag{6.2.8}$$

Мы имеем из (6.2.6) и (6.2.7)

$$\sum_{j=1}^{n}\tilde{a}_j x_j=l-\varepsilon. \tag{6.2.9}$$

Умножая это равенство скалярно на $x_i (i=1, 2, \ldots, n)$ и учитывая (6.2.8), найдем:

$$\sum_{j=1}^{n}\tilde{a}_j[x_i x_j]=[x_i l],\quad i=1, 2, \ldots, n. \tag{6.2.10}$$

Система линейных уравнений (6.2.10) для нахождения коэффициентов $\tilde{a}_j$ называется *нормальной* системой. Ее матрица $X = X_{nn} = \|[x_i x_j]\|$ неособенная. Ее определитель $\det(X) = \det\|[x_i x_j]\|$ есть определитель Грама системы векторов $x_1, x_2, \ldots, x_n$ (см. § 3 гл. I). Так как ранг $(X) = n$, то $\det(X) \neq 0$, и система уравнений (6.2.10) разрешается однозначно. Геометрический вывод А. Н. Колмогорова в статье [22] соответственно сказанному в § 1 этой главы легко распространяется и на случай неодинаковых весов. Однако в более сложных случаях условных измерений и оценивания линейных комбинаций параметров геометрические рассуждения становятся громоздкими. В этих случаях предпочтительнее матричные выводы, которые к тому же позволяют удобную и короткую запись. Мы будем их проводить для всех случаев систематически.

§ 3. Матричный вывод

Перепишем введенные ранее векторы в виде одностолбцовых матриц:

$$Y = Y_{N1} = \begin{Vmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{Vmatrix}; \quad L = L_{N1} = \begin{Vmatrix} l_1 \\ l_2 \\ \vdots \\ l_N \end{Vmatrix}, \quad \tilde{L} = \tilde{L}_{N1} = \begin{Vmatrix} \tilde{l}_1 \\ \vdots \\ \tilde{l}_N \end{Vmatrix}, \quad \Delta = \begin{Vmatrix} \Delta_1 \\ \vdots \\ \Delta_N \end{Vmatrix},$$

$$A = \tilde{A}_{n1} = \begin{Vmatrix} \tilde{a}_1 \\ \tilde{a}_2 \\ \vdots \\ \tilde{a}_n \end{Vmatrix}.$$

Матрица A есть матрица оценок, найденных по методу наименьших квадратов.

Матрицу известных величин $\|x_{rj}\|$ $(r = 1, 2, \ldots, N;\ j = 1, 2, \ldots, n)$ обозначим $X = X_{Nn} = \|x_{rj}\|$, ранг $(X) = n$. Основные уравнения (6.1.1) запишутся в виде

$$Y = XA, \tag{6.3.1}$$

$$L = Y + \Delta,$$

$$XA - L = -\Delta. \tag{6.3.2}$$

Если L_{N1} — фиксированные наблюдения, то Δ_{N1} — фиксированные (неизвестные нам) поправки; если же Δ_{N1} — случайные погрешности, то, очевидно, и левая часть (6.3.2) будет записью случайного вектора. Предписание наименьших квадратов (6.2.5) требует выбора матрицы

$$\tilde{A} = \tilde{A}_{n1} = \begin{Vmatrix} \tilde{a}_1 \\ \tilde{a}_2 \\ \cdot \\ \cdot \\ \tilde{a}_n \end{Vmatrix}$$

таким образом, что если обозначим

$$\tilde{V} = X\tilde{A} - L, \tag{6.3.3}$$

то должно быть

$$[p\tilde{v}\tilde{v}] = \sum_{r=1}^{N} p_r \tilde{v}_r^2 = \min. \tag{6.3.4}$$

Введем теперь диагональную матрицу весов

$$P = P_{NN} = \begin{Vmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & p_N \end{Vmatrix}. \tag{6.3.5}$$

Тогда (см. § 4 гл. I), получим

$$[p\tilde{v}\tilde{v}] = \sum_{r=1}^{N} p_r \tilde{v}_r^2 = \tilde{V}^T P \tilde{V}, \tag{6.3.6}$$

так что матричная запись условия минимальности такова:

$$\tilde{V}^T P \tilde{V} = \min. \tag{6.3.7}$$

Необходимые условия для минимума имеют вид

$$\frac{\partial}{\partial \tilde{a}_i}[pvv] = 2\left[p\frac{\partial v}{\partial \tilde{a}_i} \cdot v\right] = 0 \qquad (i = 1, 2, \dots, n). \tag{6.3.8}$$

Из (6.3.3) видим, что

$$v_r = x_{r1}\tilde{a}_1 + x_{r2}\tilde{a}_2 + \dots + x_{rn}\tilde{a}_n - l_r, \tag{6.3.9}$$

$$\frac{\partial v_r}{\partial \tilde{a}_i} = x_{ri}. \tag{6.3.10}$$

Покажем, что совокупность условий (6.3.8) равносильна условию

$$X^T P V = 0, \tag{6,3.11}$$

Имеем

$$X = X_{Nn} = \begin{Vmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ x_{N1} & x_{N2} & \ldots & x_{Nn} \end{Vmatrix}; \quad (X_{Nn})^T = \begin{Vmatrix} x_{11} & x_{21} & \ldots & x_{N1} \\ x_{12} & x_{22} & \ldots & x_{N2} \\ \cdot & \cdot & \cdot & \cdot \\ x_{1n} & x_{2n} & \ldots & x_{Nn} \end{Vmatrix},$$

$$\left[p \frac{\partial v}{\partial \tilde{a}_i} v\right] = \sum_{r=1}^{N} x_{ri} p_r v_r; \tag{6.3.12}$$

$$\begin{Vmatrix} p_1 v_1 \\ \vdots \\ p_N v_N \end{Vmatrix} = PV. \tag{6.3.13}$$

Рассматривая (6.3.12) и (6.3.13), видим, что условие (6.3.8) можно записать так:

$$X^T PV = 0_{N1}, \tag{6.3.14}$$

что и требовалось доказать.

Теорема 6.3.1. *Вектор оценок $\tilde{A}_{n1}$ по методу наименьших квадратов однозначно находится из уравнения*

$$C\tilde{A} = X^T PL, \tag{6.3.15}$$

где

$$C = X^T PX.$$

В более подробной записи:

$$C_{nn}\tilde{A}_{n1} = (X_{Nn})^T L_{N1}, \tag{6.3.16}$$

$$C_{nn} = (X_{Nn})^T P_{NN} X_{Nn}. \tag{6.3.17}$$

Доказательство. Сопоставим (6.3.14) с (6.3.3). Из (6.3.3) имеем $\tilde{V} = X\tilde{A} - L$; умножая слева на $X^T P$ и учитывая (6.3.14), находим

$$0 = X^T PX\tilde{A} - X^T PL,$$

откуда

$$X^T PX\tilde{A} = X^T PL. \tag{6.3.18}$$

Обозначая $C = X^T PX$, приходим к доказательству теоремы.

Рассмотрим более внимательно матрицу

$$C = C_{nn} = X^T PX.$$

***Теорема 6.3.2.** Матрица $C = C_{nn}$ неособенная, и*

$$\det(C) = p_1 p_2 \ldots p_N \det\left(X^T X\right) = \det(P)\det\left(X^T X\right). \quad (6.3.19)$$

Доказательство непосредственно вытекает из определения матрицы C и теоремы 1.3.1.

Из теоремы 6.3.2 мы видим, что система уравнений (6.3.15) однозначно разрешима относительно $\tilde{A}_{n1}$ и

$$\tilde{A} = C^{-1} X^T P L,$$

или, подробнее,

$$\tilde{A}_{n1} = (C_{nn})^{-1} (X_{Nn})^T P_{NN} L_{N1}. \quad (6.3.20)$$

Таким образом, оценки $\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n$ для параметров $a_1, a_2, \ldots, a_n$ получаются по методу наименьших квадратов однозначно. Проверим еще, что $\tilde{A}_{n1}$ доставляет минимум $[p\tilde{v}\tilde{v}]$, так что необходимые условия минимизации $[p\tilde{v}\tilde{v}]$ являются и достаточными.

Пусть $A' = A'_{n1} = \left\| \begin{matrix} a'_1 \\ \vdots \\ a'_n \end{matrix} \right\|$ — какой-либо набор чисел;

$$\begin{aligned} V' &= XA' - L, \\ \tilde{V} &= X\tilde{A} - L. \end{aligned} \quad (6.3.21)$$

Докажем, что

$$[pv'v'] \geqslant [pvv]. \quad (6.3.22)$$

Из (6.3.11) находим

$$V' - \tilde{V} = X\left(A' - \tilde{A}\right); \quad V' = \tilde{V} + X\left(A' - \tilde{A}\right). \quad (6.3.23)$$

Используя последнее соотношение, получаем

$$\begin{aligned} [pv'v'] = (V')^T P V' &= \left(\tilde{V}^T + \left(A' - \tilde{A}\right)^T X^T\right) P \left(V + X\left(A' - \tilde{A}\right)\right) = \\ &= \tilde{V}^T P \tilde{V} + \left(A' - \tilde{A}\right) X^T P V + \tilde{V}^T P X\left(A' - \tilde{A}\right) + \\ &\quad + \left(A' - \tilde{A}\right)^T X^T P X \left(A' - \tilde{A}\right). \end{aligned} \quad (6.3.24)$$

Но из (6.3.14): $X^T P \tilde{V} = 0$. Транспонируя это соотношение, имеем

$$\tilde{V}^T P X = 0 \quad (6.3.25)$$

в силу того, что $P^T = P$. Таким образом, второй и третий члены (6.3.24) исчезают. Далее, из (6.3.24), $X\left(A' - \tilde{A}\right) = V' - \tilde{V}$, так что четвертый член в (6.3.24) записывается в виде

$$\left(V' - \tilde{V}\right)^T P \left(V' - \tilde{V}\right) = \left[p\left(v' - \tilde{v}\right)\left(v' - \tilde{v}\right)\right],$$

и мы имеем

$$[pv'v'] = [p\tilde{v}\tilde{v}] + [p(v' - \tilde{v})(v' - \tilde{v})]. \tag{6.3.26}$$

Величина

$$[p(v' - v)(v' - v)] = \sum_{i=1}^{N} p_r (v'_r - \tilde{v}_r)^2 \geqslant 0,$$

так что $\tilde{A}_{n1}$ действительно дает минимум $[p\tilde{v}\tilde{v}]$.

§ 4. Нормальные уравнения, статистические свойства их решений

Вернемся к системе уравнений (6.3.15), полученной нами для определения параметров $\tilde{A}_{n1}$. Имеем

$$C\tilde{A} = X^T PL,$$

или

$$C_{nn}\tilde{A}_{n1} = (X_{Nn})^T PL_{N1}, \tag{6.4.1}$$

где

$$C = X^T PX. \tag{6.4.2}$$

Эта система называется *системой нормальных уравнений*. Ее матрица C является симметрической; действительно, $P^T = P$, поэтому

$$C^T = (X^T PX)^T = X^T PX = C.$$

Далее, из подробной записи матриц X_{Nn}, X^T_{Nn} видим, что

$$C = C_{nn} = \begin{Vmatrix} [px_1x_1] & [px_1x_2] & \dots & [px_1x_n] \\ \dots & \dots & \dots & \dots \\ [px_nx_1] & [px_nx_2] & \dots & [px_nx_n] \end{Vmatrix}, \tag{6.4.3}$$

т. е. что C есть некоторая обобщенная матрица Грама векторов $x_1, x_2, \dots, x_n$ $(x_i = (x_{1i}, x_{2i}, \dots, x_{Ni}))$; соответствующие скалярные произведения берутся с весами по координатам $p_1, p_2, \dots, p_N$.

Далее, $X^T PL = (X_{Nn})^T P_{NN} L_{N1}$ есть одностолбцовая матрица вида

$$\begin{Vmatrix} [px_1l] \\ [px_2l] \\ \vdots \\ [px_nl] \end{Vmatrix}. \tag{6.4.4}$$

Так что система нормальных уравнений (6.4.1) записывается так:

$$\sum_{\nu=1}^{n} [px_ix_\nu]\,\tilde{a}_\nu = [px_il], \quad i = 1, 2, \dots, n \tag{6.4.5}$$

$$x_i = (x_{i1}, x_{i2}, \dots, x_{iN}), \quad l = (l_1, l_2, \dots, l_N).$$

Из (6.4.5) также легко усмотреть симметрию матрицы C системы нормальных уравнений. На симметрии матрицы C основаны многие вычислительные приемы решения нормальных уравнений (6.4.5), хорошо описанные в ряде старых учебников (см., например, Н. И. Идельсон [19]). Мы остановимся на этих приемах в § 12, где будут даны и примеры, теперь же систематически изучим статистические свойства $\tilde{A}=(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n)$.

Имеем

$$\tilde{A}=C^{-1}X^TPL, \tag{6.4.6}$$

$$L=L_{N1}=Y_{N1}+\Delta_{N1},$$

где

$$\Delta=\Delta_{N1}=\begin{Vmatrix} \Delta_1 \\ \cdot \\ \vdots \\ \Delta_N \end{Vmatrix} \tag{6.4.7}$$

матрица „истинных погрешностей при измерении Y“. В силу постановки задачи имеем

$$E\Delta=0; \quad EL=Y. \tag{6.4.8}$$

Теорема 6.4.1. *Оценки $\tilde{A}_{n1}$ для параметров A_{n1}, полученные по методу наименьших квадратов, являются несмещенными*, т. е.

$$E(\tilde{A}_{n1})=A_{n1}. \tag{6.4.9}$$

Для доказательства воспользуемся (6.4.6) и (6.4.7). Заметим, что в равенстве (6.4.6) лишь $\tilde{A}$ и L являются случайными матрицами, а матрица $C^{-1}X^TP$ — фиксированная матрица. Ввиду этого,

$$E(\tilde{A})=E(C^{-1}X^TPL)=C^{-1}X^TPE(L)=C^{-1}X^TPY.$$

Но в силу постановки задачи,

$$Y=XA, \tag{6.4.10}$$

откуда

$$E(\tilde{A})=C^{-1}X^TPXA=C^{-1}CA=A,$$

что и требовалось доказать.

Теорема 6.4.2. *Несмещенные оценки $\tilde{A}_{n1}$ представляют собой случайный, нормальный n-мерный вектор с корреляционной матрицей*

$$B_{\tilde{A}}=\sigma^2C^{-1}. \tag{6.4.11}$$

Доказательство. Положим

$$D = \dot{D}_{nN} = C^{-1}X^T P, \tag{6.4.12}$$

так что

$$\tilde{A} = DL. \tag{6.4.13}$$

Далее, имеем тождество

$$A = C^{-1}CA = C^{-1}X^T PXA, \tag{6.4.14}$$

откуда

$$\tilde{A} - A = C^{-1}X^T P(L - XA) = C^{-1}X^T P(\Delta + Y - XA) = C^{-1}X^T P\Delta,$$

ибо $Y - XA = 0$.

В силу (6.4.12)

$$\tilde{A} - A = D\Delta \tag{6.4.15}$$

(подробная запись: $\tilde{A}_{n1} - A_{n1} = D_{nN}\Delta_{N1}$).

Далее замечаем, что $\Delta = \begin{Vmatrix} \Delta_1 \\ \vdots \\ \Delta_n \end{Vmatrix}$ есть случайный нормальный вектор с нулевым вектором средних и независимыми компонентами, имеющими дисперсии

$$\sigma_r^2 = \frac{\sigma^2}{p_r} (r = 1, 2, \ldots, N), \tag{6.4.16}$$

так что его корреляционная матрица, очевидно, имеет вид

$$B_\Delta = \sigma^2 P^{-1} = \sigma^2 P_{NN}^{-1}. \tag{6.4.17}$$

По формуле (2.3.11) из (6.4.15) и (6.4.12) заключаем, что корреляционная матрица вектора $\tilde{A}$ имеет вид

$$B = B_{nn} = D\sigma^2 P^{-1} D^T = C^{-1}X^T P\sigma^2 P^{-1} PX(C^{-1})^T = \sigma^2 C^{-1} X^T PX(C^{-1})^T.$$

Но $C = X^T PX$ симметрическая матрица, так что (см. § 4 гл. 1)

$$(C^{-1})^T = C^{-1}, \quad \text{и} \quad B_{\tilde{A}} = \sigma^2 C^{-1} X^T PXC^{-1} = \sigma^2 C^{-1},$$

что и требовалось доказать.

Мы можем теперь записать совместную плотность вероятностей $(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n)$; она будет иметь вид [см. (2.2.10)]

$$f(x_1, x_2, \ldots, x_n) = \\ = (2\pi)^{-\frac{n}{2}} \sigma^{-n} (\det(C))^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[(X - A)^T C(X - A)\right]\right\}, \tag{6.4.18}$$

где

$$X = \begin{Vmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{Vmatrix} = X_{n1}; \qquad A = A_{n1} = \begin{Vmatrix} a_1 \\ \vdots \\ a_n \end{Vmatrix}; \quad C = X^T P X.$$

Рассмотрим теперь вопрос о совместной эффективности оценок $\tilde{A}_{n1}$ и постараемся составить понятие о реальном теоретико-вероятностном смысле способа точечного оценивания по методу наименьших квадратов (см. § 3 гл. III). Составим для функции правдоподобия $L(l_1, \ldots, l_n)$ (см. § 2 этой главы) информационную матрицу Фишера I [см. (3.3.21)] при заданном σ для параметров $a_1, a_2, \ldots, a_n$

$$I = I_{nn} = \left\| E\left(-\frac{\partial^2 \ln L}{\partial a_i \, \partial a_j} \right) \right\|; \qquad i, j = 1, 2, \ldots, n. \tag{6.4.19}$$

Из § 2 непосредственно находим

$$-\frac{\partial^2 \ln L}{\partial a_i \, \partial a_j} = \frac{1}{\sigma^2} \sum_{r=1}^{N} x_{ri} p_r x_{rj} = \frac{1}{\sigma^2} [p x_i x_j] = \frac{1}{\sigma^2} \left(X^T P X \right)_{ij}$$

откуда

$$I = I_{nn} = \frac{1}{\sigma^2} X^T P X = \frac{C}{\sigma^2} \tag{6.4.20}$$

и, согласно (6.4.11),

$$I^{-1} = \sigma^2 C^{-1} = B_{\tilde{A}}. \tag{6.4.21}$$

По теореме 3.3.1, в весьма широком классе несмещенных оценок $(a_1', a_2', \ldots, a_n')$ с корреляционной матрицей B_1 корреляционный эллипсоид оценок $(\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n)$ будет содержаться внутри корреляционного эллипсоида оценок $(a_1', a_2', \ldots, a_n')$.

Соответственно результатам § 5 гл. II, это позволяет установить некоторый реальный теоретико-вероятностный смысл оценивания по методу наименьших квадратов.

§ 5. Реальный смысл точечного оценивания по методу наименьших квадратов

Мы можем убедиться, что оценки $\tilde{a}_1, \ldots, \tilde{a}_n$ в весьма широком классе несмещенных и асимптотически нормальных оценок $a_1', \ldots, a_n'$ имеют в некотором смысле экстремальный характер. Пусть

$$\tilde{\Delta}_i = \tilde{a}_i - a_i; \qquad \Delta_i' = a_i' - a_i; \tag{6.5.1}$$

$\beta_1, \beta_2, \ldots, \beta_n$ — любые фиксированные числа;

$$\xi = \sum_{i=1}^{n} \beta_i \tilde{\Delta}_i; \qquad \eta = \sum_{i=1}^{n} \beta_i \Delta_i'. \tag{6.5.2}$$

Тогда, при любом фиксированном $\varepsilon > 0$, заданном n, достаточно большом N:

$$P\{|\xi| \leqslant \varepsilon\} \geqslant P\{|\eta| \leqslant \varepsilon\}. \tag{6.5.3}$$

Доказательство непосредственно явствует из (6.4.21) и теоремы 2.5.1. Заметим, в частности, что, беря $\beta_{i_0} = 1$; $\beta_i = 0$ $(i \neq i_0)$, получим

$$P\{|\tilde{\Delta}_{i_0}| \leqslant \varepsilon\} \geqslant P\{|\Delta'_{i_0}| \leqslant \varepsilon\} \tag{6.5.4}$$

при достаточно большой выборке. Таким образом, каждая погрешность порознь лучше всего оценивается по методу наименьших квадратов среди весьма широкого класса методов, приводящих к несмещенным и асимптотически нормальным оценкам. Но нужно заметить, что мы не утверждаем этого относительно некоторых погрешностей одновременно. Мы можем сделать еще некоторые добавления об асимптотическом поведении оценок по методу наименьших квадратов.

Пусть $g(a_1, \ldots, a_n)$ — некоторая функция измеряемых параметров, разложимая в ряд Тейлора в окрестности a точки $(a_1, \ldots, a_n)$, а $a'_1, \ldots, a'_n$ — набор асимптотически нормальных и несмещенных оценок для $a_1, \ldots, a_n$. Сравним поведение $g(\tilde{a}_1, \ldots, \tilde{a}_n)$ и $g(a'_1, \ldots, a'_n)$. Если положим $\tilde{a}_i = a_i + \tilde{\delta}_i$, $a'_i = a_i + \delta'_i$, где $\tilde{\delta}_i$, δ'_i — погрешности, которые малы с вероятностью, весьма близкой к единице (при заданном n и большом N), то можно написать приближенные равенства

$$g(\tilde{a}_1, \ldots, \tilde{a}_n) \approx g(a_1, \ldots, a_n) + \sum_{i=1}^{n} \tilde{\delta}_i \frac{\partial g}{\partial a_i},$$

$$g(a'_1, \ldots, a'_n) \approx g(a_1, \ldots, a_n) + \sum_{i=1}^{n} \delta'_i \frac{\partial g}{\partial a_i}.$$

Поэтому $g(\tilde{a}_1, \ldots, \tilde{a}_n)$ и $g(a'_1, \ldots, a'_n)$ приближенно линейны и к ним применимы предыдущие рассуждения о линейных функциях от погрешностей, так что

$$\begin{gathered} P\{|g(\tilde{a}_1, \ldots, \tilde{a}_n) - g(a_1, \ldots, a_n)| \leqslant \varepsilon\} \geqslant \\ \geqslant P\{|g(a'_1, \ldots, a'_n) - g(a^1, \ldots, a_n)| \leqslant \varepsilon\} \end{gathered} \tag{6.5.5}$$

при заданном ε, фиксированном n и больших N.

§ 6. Статистическое поведение уклонений $\tilde{V}$

Для оценивания точности полученных результатов и построения доверительных интервалов большое значение имеет одностолбцовая матрица уклонений $\tilde{V} = \tilde{V}_{N1}$ и функции от нее, как, например,

$[p\tilde{v}\tilde{v}] = \tilde{V}^T P \tilde{V}$. Изучим статистическое поведение $\tilde{V}$. По определению,

$$\tilde{V} = X\tilde{A} - L = \tilde{L} - L; \quad \tilde{L} = X\tilde{A}. \tag{6.6.1}$$

Теорема 6.6.1. *Имеем*

$$\tilde{V} = \tilde{V}_{N1} = (U_{NN} - E_{NN})\Delta_N, \tag{6.6.2}$$

где

$$U = U_{NN} = XC^{-1}X^T P = X(X^T P X)^{-1} X^T P. \tag{6.6.3}$$

Доказательство. Имеем

$$\tilde{V} = X\tilde{A} - L; \quad \tilde{A} = C^{-1}X^T PL,$$

так что

$$\tilde{V} = XC^{-1}X^T PL - L = (XC^{-1}X^T P - E)L = (U - E)L;$$

$$U = XC^{-1}X^T P.$$

Далее,

$$L = Y + \Delta, \quad Y = XA.$$

Таким образом,

$$\tilde{V} = (U - E)\Delta + (U - E)XA;$$

далее,

$$(U - E)XA = X(X^T P X)^{-1} X^T P X A - XA = XA - XA = 0$$

и

$$\tilde{V} = (U - E)\Delta,$$

что и требовалось доказать.

Мы знаем, что корреляционная матрица Δ имеет вид (см. 6.4.17)

$$B_\Delta = \sigma^2 P^{-1}.$$

Для дальнейшего важна следующая теорема.

Теорема 6.6.2. *Корреляционная матрица вектора уклонений $\tilde{V} = \tilde{V}_{n1}$ имеет вид*

$$B_{\tilde{V}} = \sigma^2 P^{-1}(E - U^T) = \sigma^2 P^{-1}\left(E - PX(X^T P X)^{-1} X^T\right). \tag{6.6.4}$$

Для доказательства уже здесь удобно перейти к случаю равноточных наблюдений: $P = E = E_{NN}$, а затем, соответственно сказанному в § 1 этой главы [формулы (6.1.5) и (6.1.6)], заменить в найденных для равноточных измерений формулах $\tilde{V}$ на $P^{-\frac{1}{2}}\tilde{V}$ и X на $P^{\frac{1}{2}}X$.

При этом матрица

$$U = X(X^T P X)^{-1} X^T P$$

для равноточных измерений получит вид

$$U = X(X^T X)^{-1} X^T. \text{ *)}$$

Из такого определения U имеем непосредственно

$$U^T = U, \quad U^2 = U. \tag{6.6.5}$$

В самом деле,

$$U^T = X(X^T X)^{-1} X^T = U;$$
$$U^2 = X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T = X(X^T X)^{-1} X^T = U.$$

Из равенства

$$\tilde{V} = (U - E)\Delta \quad \text{и} \quad B_L = \sigma^2 E$$

(равноточность) по теореме 2.3.1 находим

$$B_{\tilde{V}} = (U - E) B_L (U - E)^T = \sigma^2 (U - E)^2 = \sigma^2 (E - U), \tag{6.6.6}$$

ибо $(U - E)^2 = U^2 - 2U + E = U - 2U + E = E - U$ на основании (6.6.5).

Теперь остается лишь перейти от равноточных наблюдений к неравноточным. Переписывая (6.6.6) в виде

$$B_{\tilde{V}} = \sigma^2 (E - X(X^T X)^{-1} X^T),$$

получим, согласно теореме 2.3.1 и равенству $\left(P^{-\frac{1}{2}}\right)^T = P^{-\frac{1}{2}}$,

$$B_{\tilde{V}} = \sigma^2 P^{-\frac{1}{2}} \left(E - P^{\frac{1}{2}} X (X^T P X)^{-1} X^T P^{\frac{1}{2}}\right) P^{-\frac{1}{2}} =$$
$$= \sigma^2 P^{-1} \left(E - P X (X^T P X) X^T\right) = \sigma^2 P^{-1} \left(E - U^T\right),$$

что и требовалось доказать.

Докажем теперь весьма важную теорему.

Теорема 6.6.3. *Случайные векторы $\tilde{A} = \tilde{A}_{n1}$ и $\tilde{V} = \tilde{V}_{N1}$ нормальны и независимы.*

Доказательство будет непосредственно опираться на теорему 2.3.3. Имеем

$$\tilde{V} = (U - E)\Delta; \quad U = X(X^T P X)^{-1} X^T P,$$
$$\tilde{A} - A = (X^T P X)^{-1} X^T P \Delta = W\Delta$$

[см. (6.4.14)]. Здесь $W = (X^T P X)^{-1} X^T P$. Отсюда явствует, что $\tilde{V}$ и $\tilde{A}$ — нормальные векторы (см. теорему 2.3.3). Далее,

$$B_\Delta = \sigma^2 P^{-1} E. \tag{6.6.7}$$

---

*) В этом случае на языке многомерной геометрии матрица U получает сравнительно простой смысл — это матрица проектирования (см., например, В. И. Смирнов [46], стр. 158—163).

Из теоремы 2.3.3 следует, что достаточно лишь проверить матричное равенство

$$(U-E)B_\Delta W^T=0. \tag{6.6.8}$$

Полагая, как и ранее, $C=X^TPX$, имеем

$$\begin{aligned}(U-E)B_\Delta W^T&=\sigma^2\left(XC^{-1}X^TP-E\right)P^{-1}PXC^{-1}=\\&=\sigma^2\left(XC^{-1}X^TPXC^{-1}-XC^{-1}\right)=\\&=\sigma^2\left(XC^{-1}CC^{-1}-XC^{-1}\right)=\sigma^2\left(XC^{-1}-XC^{-1}\right)=0,\end{aligned}$$

что и доказывает (6.6.8) и тем самым всю теорему. Заметим, что сама теорема 2.3.3 доказывается сравнительно сложно; можно было бы дать другой вывод этой теоремы, на основании соображений, изложенных в статье А. Н. Колмогорова [22]. Мы не будем останавливаться на этом и перейдем к следующей важной теореме.

Теорема 6.6.4. *Случайный вектор $\tilde{A}$ является n-мерным нормальным вектором, а случайный вектор $\tilde{V}$ является $(N-n)$-мерным нормальным вектором. Эти случайные векторы независимы. Случайная величина*

$$\frac{1}{\sigma^2}\left[p\tilde{v}\tilde{v}\right] \tag{6.6.9}$$

распределена как χ^2_{N-n} и независима от $\tilde{A}$.

Доказательство. Как и ранее, рассмотрим сначала случай равноточных измерений, считая $P=E$, Получим

$$\tilde{A}-A=W\Delta;\ \tilde{V}=(U-E)\Delta;\quad W=(X^TX)^{-1}X^T;\quad U=X(X^TX)^{-1}X^T. \tag{6.6.10}$$

Имеем, далее,

$$\text{ранг }(W)=n;\quad \text{ранг }(U)=n. \tag{6.6.11}$$

В самом деле, ранг $(X)=$ ранг $\left(X^T\right)=n$; X^TX — неособенная матрица. Ввиду этого, ранг $(W)=$ ранг$\left(X^T\right)=n$, Далее, ранг $(U)\leqslant$ ранг$(X)=n$ в силу неравенства (1.2.18). С другой стороны,

$$UX=X(X^TX)^{-1}X^TX=X \tag{6.6.12}$$

(что отвечает тому, что U — матрица проектирования). Поэтому в силу того же неравенства (1.2.18) имеем

$$n=\text{ранг}(X)\leqslant\min(\text{ранг}(U),\ \text{ранг}(X))=\min(\text{ранг}(U),\ n),$$

откуда ранг$(U)\geqslant n$. Сопоставляя это с ранее найденным, находим, что ранг $(U)=n$. Таким образом, $\tilde{A}-A$ и $\tilde{A}$ суть n-мерные векторы.

Покажем теперь, что ранг $(E-U)=N-n$.

Имеем $(E-U)^T = E-U$, т. е. $E-U$ — симметрическая матрица. Далее, $U^2 = U$, откуда

$$(E-U)^2 = E - 2U + U^2 = E - U. \tag{6.6.13}$$

Существует ортогональная матрица $F = F_{NN}$, которая приводит U к диагональной форме:

$$F^T U F = D = \begin{Vmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & d_N \end{Vmatrix}. \tag{6.6.14}$$

При этом $D^2 = F^{-1} U^2 F = F^{-1} U F = D$, так что $d_i^2 = d_i$ и $d_i = 0; 1$. Очевидно, число таких d_i, что $d_i = 1$, равно рангу U, т. е. числу n. Далее,

$$F^{-1}(E-U)F = E - F^{-1}UF = E - D = \begin{Vmatrix} 1-d_1 & 0 & \dots & 0 \\ 0 & 1-d_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 1-d_N \end{Vmatrix}, \tag{6.6.15}$$

откуда ясно, что ранг $(E-U) =$ ранг $(E-D) = N-n$. Таким образом, ранг $(U-E) = N-n$, и из (6.6.10) и теоремы 2.3.1 заключаем, что $\tilde{V}$ есть $(N-n)$-мерный нормальный вектор. В ходе этих рассуждений можно заметить, что матрицы U и $(E-U)$ приводятся к диагональному виду одной и той же диагональной матрицей F.

При этом, если введем одностолбцовую матрицу $Z = Z_{N1} = \begin{Vmatrix} z_1 \\ \vdots \\ z_N \end{Vmatrix}$,

и составим квадратичные формы

$$Z^T U Z; \quad Z^T (E-U) Z, \tag{6.6.16}$$

то при подстановке $Z = FY$, $Y = Y_{N1}$ получим новые квадратичные формы

$$Y^T D Y = \sum_{i=1}^{n} y_{\alpha_i}^2; \quad Y^T (E-D) Y = \sum_{j=1}^{N-n} y_{\beta_j}^2. \tag{6.6.17}$$

Как видно из (6.6.14) и (6.6.15), набор переменных y_{α_i} и y_{β_j} будет пробегать все переменные $y_1, \dots, y_N$, причем $\alpha_i \neq \beta_j$ ($i = 1, \dots, n$; $j = 1, \dots, N-n$), т. е. это будут наборы различных переменных y_i.

Перейдем теперь к случайной величине (6.6.9). Полагая пока $P=E$; $p_i=1$ $(i=1, \ldots, N)$, составим

$$[\tilde{v}\tilde{v}]=\tilde{V}^T\tilde{V}.$$

Имеем

$$\tilde{V}^T\tilde{V}=\Delta^T(U-E)^T(U-E)\Delta=\Delta^T(U-E)^2\Delta=\Delta^T(E-U)\Delta$$

в силу (6.6.13). Далее, Δ для равноточных измерений есть нормальный случайный вектор с $E(\Delta)=0$ и $B_\Delta=\sigma^2E$. По теореме Фишера (теорема 2.3.2), для любой ортогональной матрицы F случайный вектор $F\cdot\Delta$ будет иметь то же распределение, что и вектор Δ. Возьмем ортогональную матрицу F, о которой говорилось выше. Имеем

$$[\tilde{v}\tilde{v}]=\Delta^T(E-U)\Delta=\Delta^TFF^T(E-U)FF^T\Delta=$$
$$=(\Delta^TF)(E-D)(F^T\Delta)=\sum_{j=1}^{N-n} y_{\beta_j}^2.$$

[см. (6.6.17)]. Здесь мы положили $Y=Y_{N1}=\begin{Vmatrix} y_1 \\ \cdot \\ \vdots \\ \cdot \\ y_N \end{Vmatrix}=F^T\Delta$. На основании сказанного ранее, $Y=Y_{N1}$ есть случайный нормальный вектор с независимыми компонентами типа $N(0, \sigma)$. Отсюда по определению χ^2_{N-n} имеем

$$\frac{1}{\sigma^2}[\tilde{v}\tilde{v}]=\chi^2_{N-n}.$$

Теорема доказана. Ясно, что $[\tilde{v}\tilde{v}]$ как функция случайного вектора $\tilde{V}$, статистически независимого от $\tilde{A}$, не будет зависеть от $\tilde{A}$.

Теперь остается вернуться к случаю неравноточных измерений. При таком переходе вектор $\tilde{A}$ имеет прежнее выражение; вместо X надо будет брать $P^{\frac{1}{2}}X$, причем ранг $\left(P^{\frac{1}{2}}X\right)=$ ранг (X), так что $\tilde{A}$ по-прежнему будет n-мерным случайным вектором. Далее, $\tilde{V}$ заменяется на $P^{-\frac{1}{2}}\tilde{V}$ при неравноточных измерениях; стало быть, для случая неравноточных измерений величина $\frac{1}{\sigma^2}[p\tilde{v}\tilde{v}]=\chi^2_{N-n}$ статистически независима от $\tilde{A}$.

Выведем еще некоторые формулы, относящиеся к $[p\tilde{v}\tilde{v}]$.

Теорема 6.6.5. *Имеем*

$$[p\tilde{v}\tilde{v}]=L^TPL-\tilde{A}^TC\tilde{A}=[pll]-\tilde{A}^TC\tilde{A}. \tag{6.6.18}$$

Доказательство. Согласно формулам (6.4.1) и (6.1.1),

$$C\tilde{A}=X^TPL;\quad C=X^TPX;\quad \tilde{V}=X\tilde{A}-L.$$

Отсюда $\tilde{A}^TC=L^TPX$. Далее,

$$\begin{aligned}[p\tilde{v}\tilde{v}] &= \tilde{V}^TP\tilde{V}=(\tilde{A}^TX^T-L^T)P(X\tilde{A}-L)=\\ &=\tilde{A}^TX^TPX\tilde{A}+L^TPL-L^TPX\tilde{A}-\tilde{A}^TX^TPL=\\ &=\tilde{A}^TC\tilde{A}+L^TPL-\tilde{A}^TC\tilde{A}-\tilde{A}^TC\tilde{A}=\\ &=L^TPL-\tilde{A}^TC\tilde{A},\end{aligned}$$

что и требовалось доказать.

Следующая теорема может быть полезна для вычисления или контроля вычислений $[p\tilde{v}\tilde{v}]$.

Теорема 6.6.6. *Имеем*

$$\begin{aligned}[p\tilde{v}\tilde{v}] &= L^TPL-L^TPX\tilde{A}=\\ &=[pll]-[px_1l]\tilde{a}_1-[px_2l]\tilde{a}_2-\ \ldots\ -[px_nl]\tilde{a}_n.\end{aligned}\tag{6.6.19}$$

Для доказательства вернемся к равенству (6.6.18); из него находим

$$[p\tilde{v}\tilde{v}]=L^TPL-L^TPX\tilde{A}-\tilde{A}^TX^TPX\tilde{A}-\tilde{A}^TX^TPL.$$

Подставляя вместо X^TPL равную ей матрицу $C\tilde{A}$ и учитывая, что $X^TPX=C$, находим, что два последних члена обращаются в нуль, так как $\tilde{A}^TC\tilde{A}-\tilde{A}^TC\tilde{A}=0$. Это доказывает первое из равенств (6.6.19). Докажем второе из равенств (6.6.19). Имеем

$$L^TPX=(X^TPL)^T=\|[px_1l],\ [px_2l],\ \ldots,\ [px_nl]\|.$$

Отсюда

$$L^TPX\tilde{A}=[px_1l]\tilde{a}_1+[px_2l]\tilde{a}_2+\ \ldots\ +[px_nl]\tilde{a}_n,$$

что дает нам второе из равенств (6.6.19).

Докажем еще формулу

$$[p\tilde{v}\tilde{v}]=L^TPL-\tilde{L}^TP\tilde{L}=[pll]-[p\tilde{l}\tilde{l}],\tag{6.6.20}$$

где

$$\tilde{L}=X\tilde{A}.$$

Из (6.6.18) имеем

$$\begin{aligned}[p\tilde{v}\tilde{v}] &= L^TPL-\tilde{A}^TC\tilde{A}=L^TPL-\tilde{A}^TX^TPX\tilde{A}=\\ &=L^TPL-\tilde{L}^TP\tilde{L}=[pll]-[p\tilde{l}\tilde{l}],\end{aligned}$$

что и требовалось доказать.

Дадим еще одно выражение для $[p\tilde{v}\tilde{v}]$, принадлежащее Ю. Нейману и его ученице Ф. Дэвид [15].

Положим

$$H_0 = \sum_{r=1}^{N} p_r l_r^2, \tag{6.6.21}$$

$$H_h = \sum_{r=1}^{N} p_r l_r x_{rh} \qquad (h = 1, \ldots, n), \tag{6.6.22}$$

$$C_{hh} = \sum_{r=1}^{N} p_r x_{rh} x_{rk} \qquad (h,\ k = 1,\ 2,\ \ldots,\ n). \tag{6.6.23}$$

Тогда, как легко подсчитать,

$$C = \|c_{hh}\| = \begin{Vmatrix} c_{11} & c_{12} \ldots c_{1n} \\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot\ \cdot \\ c_{n1} & c_{n2} \ldots c_{nn} \end{Vmatrix}, \tag{6.6.24}$$

Построим окаймление матрицы C при помощи чисел $H_0, H_1, \ldots, H_n$, т. е. построим матрицу

$$C_{HH} = \begin{Vmatrix} H_0 & H_1 & H_2 \ldots H_n \\ H_1 & c_{11} & c_{12} \ldots c_{1n} \\ H_2 & c_{21} & c_{22} \ldots c_{2n} \\ \cdot\ \cdot & \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot\ \cdot \\ H_n & c_{n1} & c_{n2} \ldots c_{nn} \end{Vmatrix}. \tag{6.6.25}$$

Тогда имеет место теорема:

Теорема 6.6.7. *Справедливо равенство*

$$[p\tilde{v}\tilde{v}] = \frac{\det(C_{HH})}{\det(C)}. \tag{6.6.26}$$

Для доказательства запишем (6.6.21) и (6.6.22) в матричном виде

$$H_0 = L^T P L, \tag{6.6.27}$$

$$H_h = \{X^T P L\}_{h1}. \tag{6.6.28}$$

Рассмотрим теперь матрицу C_{HH} и ее определитель $\det(C_{HH})$. Для вычисления $\det(C_{HH})$ будем разлагать этот определитель сперва по элементам первой строки, потом по элементам первого столбца с тем, чтобы выделить коэффициенты при H_0 и $H_h H_k$. Получим

$$\det(C_{HH}) = H_0 \det(C) - \sum_{h,\,k=1}^{n} H_h H_k M_{hk},$$

где M_{hk} — алгебраическое дополнение элемента c_{hk} в матрице C. Учитывая симметричность C $(C^T = C)$, имеем

$$\|M_{hk}\| = C^{-1} \det(C);$$

$$\frac{\det(C_{HH})}{\det(C)} = H_0 - \sum_{h,\,k=1}^{n} H_h H_k \{C^{-1}\}_{hk} = H_0 - H^T C^{-1} H,$$

где через H обозначена матрица

$$H = H_{n1} = \left\| \begin{matrix} H_1 \\ \vdots \\ H_n \end{matrix} \right\| = X^T PL,$$

согласно (6.6.28). Подставляя сюда еще (6.6.21), находим

$$\frac{\det(C_{HH})}{\det(C)} = L^T PL - L^T PXC^{-1} X^T PL. \qquad (6.6.29)$$

Но, согласно (6.4.6), $C^{-1} X^T PL = \tilde{A}$, так что

$$\frac{\det(C_{HH})}{\det(C)} = L^T PL - L^T PX\tilde{A} = [p\tilde{v}\tilde{v}]$$

в силу (6.6.19), что и требовалось доказать.

§ 7. Точечное оценивание величин y_i ($i = 1, 2, \ldots, N$)

Величины y_i связаны с величинами x_j следующим точным соотношением:

$$Y = Y_{N1} = XA = \left\| \begin{matrix} y_1 \\ \vdots \\ y_N \end{matrix} \right\|. \qquad (6.7.1)$$

В качестве оценок для $(y_1, \ldots, y_n)$ будем брать вектор

$$\tilde{L} = \tilde{L}_{N1} = X\tilde{A}. \qquad (6.7.2)$$

Из (6.7.2) получаем

$$E(\tilde{L}) = XA = Y, \qquad (6.7.3)$$

так что компоненты $\tilde{L}$ образуют несмещенные оценки для компонентов Y. Далее, для корреляционной матрицы $B_{\tilde{L}}$ вектора $\tilde{L}$ имеем

$$B_{\tilde{L}} = \sigma^2 XC^{-1} X^T. \qquad (6.7.4)$$

В самом деле,

$$B_{\tilde{L}} = E\left[(X(\tilde{A} - A))(X(\tilde{A} - A))^T\right] = XB_{\tilde{A}} X^T = \sigma^2 XC^{-1} X^T,$$

что и требовалось доказать.

Таким образом, для случайной величины $\tilde{l}_i = \{\tilde{L}\}_{i1}$ имеем

$$E(\tilde{l}_i) = y_i; \quad D(\tilde{l}_i) = \sigma^2 \{XC^{-1} X^T\}_{ii}. \qquad (6.7.5)$$

Докажем еще теорему о дисперсиях $\tilde{l}_i$.

Теорема 6.7.1. *Имеем*

$$\sum_{i=1}^{N} p_i D(\tilde{l}_i) = [pD(\tilde{l})] = n\sigma^2. \tag{6.7.6}$$

В силу (6.7.5) достаточно доказать, что

$$\sum_{i=1}^{N} p_i \{XC^{-1}X^T\}_{ii} = n. \tag{6.7.7}$$

Введем диагональную матрицу весов P и заметим, что

$$p_i \{XC^{-1}X^T\}_{ii} = \{XC^{-1}X^TP\}_{ii}.$$

Далее, $XC^{-1}X^TP = U$ [см. (6.6.3)], так что достаточно обнаружить, что

$$\sum_{i=1}^{N} \{U\}_{ii} = n, \tag{6.7.8}$$

или $\operatorname{Sp}(U) = n$ (см. § 3 гл. I).

Мы имеем далее

$$U^2 = U \text{ и ранг } (U) = n$$

(см. § 6 этой главы).

Матрица $U_1 = P^{-\frac{1}{2}}UP^{\frac{1}{2}} = P^{-\frac{1}{2}}XC^{-1}X^TP^{\frac{1}{2}}$ симметрическая и $U_1^2 = U_1$. Далее, ранг $(U_1) =$ ранг $(U) = n$.

Имеем по формуле (1.3.18)

$$\operatorname{Sp}(U_1) = \operatorname{Sp}(U). \tag{6.7.9}$$

Как и в § 6, подберем ортогональную матрицу F, такую, что

$$FU_1F^T = D, \tag{6.7.10}$$

где D — диагональная матрица. Так как $F^T = F^{-1}$, то получаем

$$FU_1F^T = FU_1F^{-1},$$

так что

$$\operatorname{Sp}(U_1) = \operatorname{Sp} D \quad \text{и} \quad D^2 = D.$$

Имеем

$$D = \begin{Vmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & d_N \end{Vmatrix}; \quad d_i^2 = d_i,$$

поэтому $d_i = 0;\ 1$.

Далее, ранг (D) = ранг $(U_1) = n$. Поэтому количество чисел d_i, равных единице, равно n. Отсюда, очевидно,

$$\operatorname{Sp}(D) = \sum_{i=1}^{N} d_i = n. \tag{6.7.11}$$

Поскольку

$$\operatorname{Sp}(U) = \operatorname{Sp}(U_1) = \operatorname{Sp}(D),$$

то

$$\operatorname{Sp}(U) = n,$$

что и доказывает теорему.

Теорема 6.7.1 подчеркивает разницу между случайными векторами $\widetilde{L}$ и L. Величины $\widetilde{l}_1, \ldots, \widetilde{l}_N$, вообще говоря, зависимы, а величины $l_1, \ldots, l_N$ независимы. Мы имеем

$$\sum_{r=1}^{N} p_r D(\widetilde{l}_r) = [pD(\widetilde{l})] = n\sigma^2 \quad [\text{см. } (6.7.6)],$$

в то время как, очевидно:

$$\sum_{r=1}^{N} p_r D(l_r) = [pD(l)] = N\sigma^2. \tag{6.7.12}$$

Равенство (6.7.6) дает нам возможность составить представление об общей точности оценок $\widetilde{l}_1, \ldots, \widetilde{l}_N$. Для простоты пусть $p_i = 1$, так что наблюдения равноточны. Тогда

$$\sum_{i=1}^{N} D(\widetilde{l}_i) = n\sigma^2. \tag{6.7.13}$$

Таким образом, среднее дисперсий $\widetilde{l}_i$ равно

$$\frac{1}{N} \sum_{i=1}^{N} D(\widetilde{l}_i) = \frac{n}{N} \sigma^2. \tag{6.7.14}$$

Отсюда видно, что среднее дисперсий $\widetilde{l}_i$ пропорционально $\frac{n}{N}$ при данном σ^2, т. е. оценки будут тем лучше, чем больше число наблюдений по отношению к числу измеряемых параметров.

§ 8. Оценивание параметров с помощью доверительных интервалов

Предыдущие результаты позволяют построить доверительные интервалы для оценивания параметров $a_1, a_2, \ldots, a_n$ порознь или их линейной комбинации *). Мы знаем, что вектор оценок по методу

---

*) В гл. XIV будут построены для более общей ситуации оценивания по элементам и доверительные области для любых наборов параметров $a_1, a_2, \ldots, a_n$.

наименьших квадратов $\tilde{A}$ имеет вектор средних A и корреляционную матрицу $B=\sigma^2 C^{-1}$. Таким образом, для оценки $\tilde{a}_i$ имеем

$$E(\tilde{a}_i)=a_i; \quad D(\tilde{a}_i)=\sigma^2\{C^{-1}\}_{ii}. \tag{6.8.1}$$

Поэтому

$$\frac{\tilde{a}_i-a}{\sigma\sqrt{\{C^{-1}\}_{ii}}}\in N(0,\ 1). \tag{6.8.2}$$

Далее, величина $q=\frac{[p\tilde{v}\tilde{v}]}{N-n}$ не зависит от нормального вектора $\tilde{A}$ и, стало быть, от величин (6.8.2). По теореме 6.6.4

$$\frac{1}{\sigma^2}q=\frac{1}{\sigma^2}\frac{[p\tilde{v}\tilde{v}]}{N-n}=\frac{\chi^2_{N-n}}{N-n}.$$

Отсюда следует (см. § 6 гл. II), что величина

$$t_{N-n}=\frac{\tilde{a}_i-a_i}{\sqrt{\{C^{-1}\}_{ii}q}}=\frac{\tilde{a}_i-a_i}{\sqrt{\{C^{-1}\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}}} \tag{6.8.3}$$

имеет распределение Стьюдента с $(N-n)$ степенями свободы. Это позволяет построить доверительный интервал для оценивания $\tilde{a}$.

Выберем нужную надежность p, имеющуюся в первом входе табл. I распределения Стьюдента, в качестве второго входа возьмем число степеней свободы $k=N-n$. Тогда по таблице определяется число γ такое, что

$$P\{|t_{N-n}|\leqslant\gamma\}=p.$$

Доверительный интервал с надежностью p для оценивания a_i получается так же, как в гл. V для случая оценки одного параметра:

$$I_\gamma^{(i)}=\left[\tilde{a}_i-\gamma\sqrt{\{C^{-1}\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}},\ \tilde{a}_i+\gamma\sqrt{\{C^{-1}\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}}\right]. \tag{6.8.4}$$

Аналогично этому, можно построить и доверительный интервал для оценивания $y_1, y_2, \ldots, y_N$. Для этого используем (6.7.4) и заметим, что $\tilde{L}=X\tilde{A}$ не зависит от $[p\tilde{v}\tilde{v}]$. Таким образом, дробь

$$t_{N-n}=\frac{\tilde{l}_i-y_i}{\sqrt{\{XC^{-1}X^T\}_{ii}q}}=\frac{\tilde{l}_i-y_i}{\sqrt{\{XC^{-1}X^T\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}}} \tag{6.8.5}$$

распределена по закону Стьюдента с $N-n$ степенями свободы. Для нее можно выбрать надежность p_0, определить число γ, как в предыдущем случае, и построить доверительный интервал

$$J_\gamma^{(i)}=\left[\tilde{l}_i-\gamma\sqrt{\{XC^{-1}X^T\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}},\ \tilde{l}_i+\gamma\sqrt{\{XC^{-1}X^T\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}}\right] \tag{6.8.6}$$

с надежностью p_0.

Мы видим, что длина доверительного интервала $J_\gamma^{(i)}$ является случайной величиной:

$$d_\gamma^{(i)}=2\gamma s\sqrt{\{XC^{-1}X^T\}_{ii}}, \tag{6.8.7}$$

где $s=\sqrt{\frac{[p\tilde{v}\tilde{v}]}{N-n}}$.

(6.8.7) позволяет составить представление о поведении средней квадратичной из длин $J_\gamma^{(i)}$

$$E\left\{\frac{1}{N}\sum_{i=1}^{N}p_i\left(d_\gamma^{(i)}\right)^2\right\}=\frac{1}{N}E\left[pd_\gamma^2\right]. \tag{6.8.8}$$

Эта величина равна

$$\frac{1}{N}4\gamma^2\sigma^2\sum_{i=1}^{N}p_i\{XC^{-1}X^T\}_{ii}=\frac{4\gamma^2 n\sigma^2}{N},$$

согласно (6.7.7).

Таким образом, средний взвешенный квадрат длин доверительного интервала при заданном σ^2 пропорционален $\gamma^2\frac{n}{N}$. При этом число γ определяется как число, для которого $P\{|t_{N-n}|\leqslant\gamma\}=p_0$. Для простоты будем считать наблюдения равноточными ($p_i=1$). Если $N-n\geqslant 20$, можно без особой погрешности считать дробь Стьюдента t_{N-n} приближенно нормальной $N(0, 1)$ (см. § 6 гл. II). Поэтому, полагая $p_0=0{,}95$, можно взять $\gamma\approx 1{,}96$ (см. 2.1.3). Средний квадрат длины доверительного интервала $J_\gamma^{(i)}$ будет приближенно равен

$$\frac{4\cdot(1{,}96)^2\sigma^2 n}{N}=3{,}84\sigma^2\frac{n}{N}.$$

Таким образом, для точности работы $\frac{N}{n}$ должно быть большим.

§ 9. Оценивание точности измерений

Для ориентировки в выборе нужной оценки произведем сначала оценивание σ^2 по методу максимального правдоподобия. Логарифм функции правдоподобия дается формулой § 2 этой главы, так что имеем уравнение

$$\frac{\partial\ln L}{\partial\sigma^2}=-\frac{N}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{r=1}^{N}p_r\left(l_r-\sum_{j=1}^{n}a_jx_{rj}\right)^2=0. \tag{6.9.1}$$

Мы видели в § 2 этой главы, что оценки для a_i по методу наименьших квадратов совпадают с оценками максимального правдоподобия. Имеем

$$\sum_r p_r\left(l_r - \sum \tilde{a}_r x_{rj}\right)^2 = [p\tilde{v}\tilde{v}].$$

Поэтому оценивание σ^2 по методу максимального правдоподобия согласно уравнению (6.9.1) дает

$$\hat{\sigma}^2 = \frac{[p\tilde{v}\tilde{v}]}{N}. \tag{6.9.2}$$

Мы знаем, однако, что $\frac{1}{\sigma^2}[p\tilde{v}\tilde{v}] = \chi^2_{N-n}$, так что

$$E\,[p\tilde{v}\tilde{v}] = \sigma^2 E\left(\chi^2_{N-n}\right) = (N-n)\,\sigma^2. \tag{6.9.3}$$

Таким образом, оценка $\hat{\sigma}^2$ получается смещенной в сторону занижения:

$$E\,(\hat{\sigma}^2) = \sigma^2\,\frac{N-n}{N}. \tag{6.9.4}$$

Из $\hat{\sigma}^2$ легко получить несмещенную оценку

$$\hat{\sigma}_1^2 = \frac{N\hat{\sigma}^2}{N-n} = \frac{[p\tilde{v}\tilde{v}]}{N-n}. \tag{6.9.5}$$

Эта оценка асимптотически эквивалентна оценке $\hat{\sigma}^2$ при заданном числе n оцениваемых параметров и увеличивающемся числе наблюдений N. Мы будем употреблять ее в качестве точечной оценки для σ^2; в качестве точечной оценки для σ будем использовать величину

$$\tilde{s} = \sqrt{\frac{[p\tilde{v}\tilde{v}]}{N-n}}. \tag{6.9.6}$$

Заметим, что последняя оценка является уже смещенной, но асимптотически несмещенной при заданном n и увеличивающемся N.

Можно построить доверительный интервал для оценивания σ, пользуясь тем, что $\frac{1}{\sigma^2}[p\tilde{v}\tilde{v}] = \chi^2_{N-n}$. Это делается совершенно так же, как в гл. V, § 3. Пользуясь табл. III приложений при числе степеней свободы $k = N - n$, составим доверительный интервал

$$I = \left[\gamma_1\sqrt{\frac{[p\tilde{v}\tilde{v}]}{N-n}},\quad \gamma_2\sqrt{\frac{[p\tilde{v}\tilde{v}]}{N-n}}\right], \tag{6.9.7}$$

накрывающий σ с вероятностью p_0. Так как $\tilde{s}^2 = \chi^2_{N-n}$, то на основании свойств χ^2_{N-n} можно заметить, что

$$D(s) \propto \frac{\sigma^2}{2(N-n)} \tag{6.9.8}$$

в смысле асимптотического равенства, т. е. $\frac{D(\tilde{s})}{\sigma^2/2(N-n)} \to 1$ при увеличении $N-n$.

§ 10. Обзор прямых измерений с новой точки зрения. О весах

Мы получили все данные, достаточные для оценивания измеряемых параметров по методу наименьших квадратов при косвенных (непрямых) безусловных измерениях. Разобранный в гл. V случай прямых неравноточных измерений с известными весами является частным случаем изложенного. Разберем его в виде примера.

В этом случае имеем один параметр a; матрица X имеет весьма простой вид:

$$X = X_{N1} = \begin{Vmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{Vmatrix}; \quad y_r = a \qquad (r = 1, 2, \ldots, N), \qquad (6.10.1)$$

так что уравнения вида (6.2.1) сводятся к N равенствам:

$$l_r = y_r + \Delta_r = a + \Delta_r \qquad (r = 1, 2, \ldots, N).$$

Имеем, принимая во внимание (6.4.6) и (6.4.11),

$$\tilde{A} = \tilde{A}_{n1} = \tilde{A}_{11} = \tilde{a};$$

$$C = X^T P X = \sum_{i=1}^{n} p_i = [p];$$

$$\tilde{a} = \tilde{A}_{11} = C^{-1} X^T P L = \frac{[pl]}{[p]};$$

$$D(\tilde{a}) = B = B_{11} = \sigma^2 C^{-1} = \frac{\sigma^2}{[p]}.$$

Для оценивания σ^2 берем несмещенную точечную оценку

$$\frac{[p\tilde{v}\tilde{v}]}{N-1} = \frac{[p(l-\tilde{a})(l-\tilde{a})]}{N-1},$$

что отвечает старой формуле (5.2.11).

Далее, для оценивания $D(\tilde{a}) = \frac{\sigma^2}{[p]}$ берем $\frac{[p(l-\tilde{a})(l-\tilde{a})]}{[p](N-1)}$ [соответственно формуле (5.2.11)].

Рассмотрим теперь вопрос о группах равноточных наблюдений в общем процессе комбинирования неравноточных измерений.

Пусть имеем основную систему (6.1.1)

$$y_r = \sum_{j=1}^{n} a_j x_{rj} \qquad (r = 1, 2, \ldots, N, \ N > n),$$

причем для каждого r производится не одно, а n_r наблюдений:

$$l_{rj} = y_r + \Delta_{rj}; \quad j = 1, 2, \ldots, n_r;$$

$$\Delta_{rj} \text{ равноточны } (r = 1, 2, \ldots, N, \ j = 1, 2, \ldots, n_r).$$

Мы имеем, таким образом, равноточные наблюдения, соединенные в группы соответственно наборам абсцисс $(x_{r1}, x_{r2}, \ldots, x_{rn})$. Это отвечает той же основной схеме (6.1.1), где каждый набор абсцисс $(x_{r1}, \ldots, x_{rn})$ повторяется n_r раз. Новая схема будет иметь $N_1 = \sum_{r=1}^{N} n_r$ фундаментальных уравнений.

Если обозначить, как и ранее,

$$x_\mu = (x_{\mu 1}, \ldots, x_{\mu N}) \qquad (\mu = 1, 2, \ldots, n),$$

$$[x_\mu x_\nu] = \sum_{r=1}^{N} x_{\mu r} x_{\nu r}, \tag{6.10.2}$$

то для матрицы C новой системы N_1 фундаментальных уравнений, согласно (6.3.17), получим

$$C = \left\| \begin{matrix} [px_1x_1] & [px_1x_2] & \ldots & [px_1x_n] \\ \cdot & \cdot & \cdot & \cdot \\ [px_nx_1] & [px_nx_2] & \ldots & [px_nx_n] \end{matrix} \right\|, \tag{6.10.3}$$

где $[px_\mu x_\nu] = \sum_{r=1}^{n} n_r x_{\mu r} x_{\nu r}; \quad p_r = n_r.$

Далее, одностолбцовая матрица $X^T L$ примет вид

$$\left\| \begin{matrix} z_1 \\ \cdot \\ \cdot \\ z_n \end{matrix} \right\|,$$

где $z_\mu = \sum_{r=1}^{N} x_{\mu r} (l_{r1} + \ldots + l_{rn_r}) = \sum_{r=1}^{N} n_r x_{\mu r} \bar{l}_r, \quad \bar{l}_r = \dfrac{l_{r1} + \ldots + l_{rn_r}}{n_r}.$

Таким образом, положив $p_r = n_r$, получим

$$X^T L = \left\| \begin{matrix} [px_1\bar{l}] \\ \cdot \\ \cdot \\ [px_n\bar{l}] \end{matrix} \right\| = X^T P \bar{L},$$

где

$$P = \left\| \begin{matrix} p_1 & 0 & \ldots & 0 \\ 0 & p_2 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & p_N \end{matrix} \right\| \text{— диагональная матрица весов,} \quad \bar{L} = \left\| \begin{matrix} \bar{l}_1 \\ \cdot \\ \cdot \\ \bar{l}_N \end{matrix} \right\|.$$

Мы видим, что новая система $N_1=\sum_{r=1}^{N} n_r$ фундаментальных уравнений приводит к той же системе нормальных уравнений, что и старая, при замене l_{rj} на $\bar{l}_r$:

$$\bar{l}_r-\bar{\Delta}_r=\sum_{j=1}^{n} a_j x_{rj}, \tag{6.10.4}$$

где $\bar{\Delta}_r \in N\left(0, \frac{\sigma}{\sqrt{n_r}}\right)$.

Таким образом, замена наблюдений $l_{r1}, \ldots, l_{rn_r}$ на выборочное среднее $\bar{l}_r$ и отдельных равноточных погрешностей Δ_{rj} на $\bar{\Delta}_r=\frac{\Delta_{r1}+\ldots+\Delta_{rn_r}}{n_r}$ целесообразна с точки зрения метода наименьших квадратов и приводит к оптимальному использованию информации от этих наблюдений в том смысле, в каком была пояснена выше оптимальность метода наименьших квадратов при нормальном векторе погрешностей.

Кроме того, при таком комбинировании наблюдений возникает еще одно преимущество и для случая ненормального вектора погрешностей — при замене $l_{r1}, \ldots, l_{rn_r}$ на $\bar{l}_r$ уже при сравнительно малых значениях r величина $\bar{\Delta}_r=\frac{\Delta_{r1}+\ldots+\Delta_{rn_r}}{n_r}$ соответственно теореме Ляпунова (теорема 2.7.1) будет довольно близка к нормальной $N\left(0, \frac{\sigma}{\sqrt{n_r}}\right)$, и, тем самым, использование метода наименьших квадратов при обработке наблюдений будет близко к оптимальному в смысле, указанном выше.

§ 11. Сводка формул и правила оценивания

а) Точечное оценивание

Подробная запись	Матричная запись
Основная система уравнений	
$l_r-\Delta_r=\sum_{j=1}^{n} a_j x_{rj}$ $(r=1, 2, \ldots, N)$	$L-\Delta=XA$, подробнее (6.11.1)
$x_i=(x_{1i}, \ldots, x_{Ni})$	$L_{N1}-\Delta_{N1}=X_{Nn}A_{n1}$
$l=(l_1, \ldots, l_N)$, веса $p_1, \ldots, p_N$.	Матрица весов $P=\begin{Vmatrix} p_1 & 0 & \ldots & 0 \\ 0 & p_2 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & p_N \end{Vmatrix}$.

Нормальные уравнения для определения $\tilde{a}_i$

$$\sum_{\nu=1}^{n} [px_i x_\nu]\, \tilde{a}_\nu = [px_i l]; \quad i = 1, 2, \ldots, n. \qquad \Bigg| \qquad C\tilde{A} = X^T PL, \quad C = X^T PX. \tag{6.11.2}$$

Для оценивания $\tilde{a}_i$ надлежит решить нормальные уравнения (6.11.2). В § 12 даны некоторые часто используемые вычислительные приемы для решения этой системы уравнений, они также подробно описаны во многих учебниках (см., например, Н. И. Идельсон [19]).

Далее надлежит вычислять $[p\tilde{v}\tilde{v}]$.

Подробная запись	Матричная запись
$\tilde{v}_r = l_r + \sum_{i=1}^{n} \tilde{a}_i x_{ri},$	$\tilde{V} = X\tilde{A} - L,$
$[p\tilde{v}\tilde{v}] = \sum_{r=1}^{N} p_r \left(l_r - \sum_{i=1}^{n} \tilde{a}_i x_{ri}\right)^2.$	$\tilde{V}^T P \tilde{V}.$

Для вычисления $[p\tilde{v}\tilde{v}]$ также есть практические приемы, которые мы проиллюстрируем на примерах. После вычисления $\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n$ по нормальным уравнениям можно воспользоваться формулой (6.6.19)

$$[p\tilde{v}\tilde{v}] = [pll] - \sum_{\nu=1}^{n} [px_\nu l]\, \tilde{a}_\nu \qquad \Big| \qquad \tilde{V}^T P \tilde{V} = L^T PL - L^T PX\tilde{A} \tag{6.11.3}$$

для вычисления $[p\tilde{v}\tilde{v}]$ или контроля такого вычисления. Далее составляется $\tilde{s} = \sqrt{\dfrac{[p\tilde{v}\tilde{v}]}{N-n}}$ для оценивания σ.

С помощью величины $\tilde{s}$ можно оценить штандарт оценки $\tilde{a}_i$. Для системы уравнений вида

$$\sum_{\nu=1}^{n} [px_i x_\nu]\, \tilde{a}_\nu = \tilde{b}_i \qquad (i = 1, 2, \ldots, n) \tag{6.11.4}$$

рассматривается обратная система

$$\sum_{\nu=1}^{n} q_{j\mu} \tilde{b}_\mu = \tilde{a}_j, \tag{6.11.5}$$

из которой нам нужны величины $q_{11}, q_{22}, \ldots, q_{nn}$. Величина $\{C^{-1}\}_{ii}$ совпадает с q_{ii}. При решении примеров будут указаны способы нахождения q_{ii} одновременно с решением нормальных уравнений. Теперь штандарт $\sigma(\tilde{a}_i)$ оценки $\tilde{a}_i$ оценивается через

$$\tilde{s}\sqrt{(C)_{ii}^{-1}} = \tilde{s} q_{ii}. \tag{6.11.6}$$

Далее, σ оценивается через

$$\tilde{s}=\sqrt{\frac{[p\tilde{v}\tilde{v}]}{N-n}}, \tag{6.11.7}$$

причем штандарт оценки $\tilde{s}$

$$\sigma(\tilde{s})\approx\frac{\sigma}{\sqrt{2(N-n)}} \tag{6.11.8}$$

оценивается через

$$\frac{\tilde{s}}{\sqrt{2(N-n)}}. \tag{6.11.9}$$

б) Оценивание с помощью доверительных интервалов*)

Для оценивания a_i выбираем нужную надежность p_0. Она служит первым входом в табл. I; вторым входом служит число степеней свободы $k=N-n$. По этим двум входам выбираем γ и составляем доверительный интервал

$$I_\gamma^{(i)}=\left[\tilde{a}_i-\gamma\sqrt{\{C^{-1}\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}},\ \tilde{a}_i+\gamma\sqrt{\{C^{-1}\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}}\right]. \tag{6.11.10}$$

Он будет накрывать измеряемый параметр a_i с вероятностью p_0.

По поводу оценивания $y_r=\sum\limits_{r=1}^{n}a_ix_{ri}$ $(r=1, 2, \ldots, N)$ см. § 7 и § 8 этой главы, а также гл. VII.

Для оценивания точности пользуемся величиной $[p\tilde{v}\tilde{v}]$.

При данной надежности p_0 по табл. III приложений при $k=N-n$ степенях свободы находим числа γ_1 и γ_2 и по ним составляем доверительный интервал $[\gamma_1\tilde{s}, \gamma_2\tilde{s}]$, накрывающий σ с вероятностью p_0. Далее имеет место формула

$$\frac{1}{\sigma^2}[pD(\tilde{l}_i)]=\sum_{r=1}^{N}p_r\{XC^{-1}X^T\}_{rr}=n, \tag{6.11.11}$$

которая может быть употреблена для контроля вычислений. Из этой формулы можно видеть, что точность оценивания зависит, в основном, от отношения $\frac{N}{n}$.

---

*) Как уже говорилось, в гл. XIV будут указаны доверительные области для нескольких параметров.

§ 12. Некоторые вычислительные методы решения нормальных уравнений. Метод Гаусса и Гаусса — Дулиттла

Рассмотрим систему нормальных уравнений

$$\begin{gathered}x_{11}a_1+x_{12}a_2+\ldots+x_{1n}a_n=b_1,\\ x_{21}a_1+x_{22}a_2+\ldots+x_{2n}a_n=b_2,\\ \cdots\cdots\cdots\cdots\cdots\\ x_{n1}a_1+x_{n2}a_2+\ldots+x_{nn}a_n=b_n.\end{gathered}$$

Матрица X из коэффициентов при неизвестных a_i симметрична $(x_{ij}=x_{ji})$ и положительно определена. В матричных обозначениях система запишется в виде $XA=B$.

В настоящее время известен целый ряд различных методов для решения такой системы. Среди них исторически первым и наиболее широко распространенным является метод исключения Гаусса. Он состоит в цепочке последовательных исключений неизвестных из уравнений и переходе, таким образом, от исходной системы к системе с треугольной матрицей, которая решается просто.

В задачах, связанных с обработкой статистических данных, помимо получения решения системы нормальных уравнений часто представляет интерес задача одновременного обращения матрицы X коэффициентов системы. Например, при расчете доверительных интервалов для параметров (см. стр. 164) нужно знать диагональные элементы обратной матрицы X^{-1}. В силу $A=X^{-1}B$ решение системы можно находить, в частности, с помощью обращения матрицы ее коэффициентов. Однако обращение матрицы — задача более трудоемкая, чем решение системы. В самом деле, построение матрицы $X^{-1}=C$ сводится к отысканию n^2 величин c_{mk}, элементов матрицы C, удовлетворяющих условиям

$$\sum_{j=1}^{n}x_{ij}c_{jl}=\begin{cases}1, & \text{если} \quad i=l,\\ 0, & \text{если} \quad i\neq l,\end{cases}$$

$i=1, 2, \ldots, n$, $l=1, 2, \ldots, n$, т. е. к решению n систем уравнений относительно c_{mk}. Поэтому мы приведем здесь вычислительные схемы, приспособленные как для обращения матрицы, так и для решения системы без отыскания обратной матрицы (с соответствующим сокращением количества вычислений и записей).

В методе исключения различают целый ряд вычислительных схем *). Среди них особое место занимают так называемые компактные схемы. В настоящее время при работе на настольных счетных машинах компактные схемы более употребительны. Они основаны на возможности производить ряд вычислений на этих машинах

*) См. В. Н. Фаддеева, Вычислительные методы линейной алгебры, Гостехиздат, 1950; P. S. Dwyer, Linear Computations, 1951.

накоплением, без записи промежуточных результатов. Этим достигается уменьшение ошибок округления и сокращение числа записей.

Не касаясь теоретических основ вопроса, ограничимся простым описанием двух вычислительных схем (развернутой и компактной) для решения системы нормальных уравнений и обращения матрицы ее коэффициентов. При этом пояснения к схемам дадим на конкретном примере для случая четырех неизвестных.

Рассмотрим следующую систему нормальных уравнений:

$$\begin{aligned} a_1+0{,}42a_2+0{,}54a_3+0{,}66a_4&=0{,}3,\\ 0{,}42a_1+a_2+0{,}32a_3+0{,}44a_4&=0{,}5,\\ 0{,}54a_1+0{,}32a_2+a_3+0{,}22a_4&=0{,}7,\\ 0{,}66a_1+0{,}44a_2+0{,}22a_3+a_4&=0{,}9. \end{aligned}$$

Схема А. Вычисления приведены в табл. 18. Поясним последовательность заполнения таблицы.

I. В левом верхнем углу [столбцы (1)—(4)] помещена матрица X системы. Рядом с ней — столбец (5) свободных членов. В столбцах (6)—(9) — единичная матрица, т. е. свободные члены четырех систем уравнений, из которых находятся элементы обратной матрицы $C=X^{-1}$. Если требуется только найти решение исходной системы, то эти столбцы следует опустить. Если же, помимо решения, нужно еще найти обратную матрицу C, то можно опустить столбец (5), так как решение системы легко получить с помощью элементов C. Столбец (10) — контрольный столбец. На первом шаге в этом столбце помещены значения сумм всех элементов строки. На следующих этапах вычислений числа контрольного столбца подвергаются тем же вычислительным операциям, что и остальные числа строки. При отсутствии вычислительных ошибок число в контрольном столбце равно (с точностью до ошибок округления) сумме всех элементов соответствующей строки [столбец (11)].

Для элементов таблицы введены обозначения $x_{p.ij}$, где p — номер шага преобразования (I — IX) и i, j — соответственно номера строки на p-м шаге и столбца.

II. Делим каждый элемент строки на элемент четвертого столбца этой строки: $x_{2.ij}=\dfrac{x_{1.ij}}{x_{1.i4}}$. При делении в частном следует брать большее число значащих цифр, чем требуемое их число в окончательных результатах, для того чтобы на точность результатов не повлияли ошибки округления, а также исчезновение значащих цифр при вычитании близких по величине чисел.

III. Вычитаем поэлементно последнюю строку II из всех остальных. [Столбец (4) из нулей не выписываем.]

Таблица 18

Схема А

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
I	$X_1 (x_{1.ij})$	1,00	0,42	0,54	0,66	0,3	1	0	0	0	3,92	
		0,42	1,00	0,32	0,44	0,5	0	1	0	0	3,68	
		0,54	0,32	1,00	0,22	0,7	0	0	1	0	3,78	
		0,66	0,44	0,22	1,00	0,9	0	0	0	1	4,22	
II	$X_2 (x_{2.ij})$	1,51515	0,63636	0,81818	1	0,45455	1,51515	0	0	0	5,93939	5,93939
		0,95455	2,27273	0,72727	1	1,13636	0	2,27273	0	0	8,36364	8,36364
		2,45455	1,45455	4,54545	1	3,18182	0	0	4,54545	0	17,18182	17,18182
		0,66000	0,44000	0,22000	1	0,90000	0	0	0	1	4,22000	4,22000
III	$X_3 (x_{3.ij})$	0,85515	0,19636	0,59818		—0,44545	1,51515	0	0	—1	1,71939	1,71939
		0,29455	1,83273	0,50727		0,23636	0	2,27273	0	—1	4,14364	4,14364
		1,79455	1,01455	4,32545		2,28182	0	0	4,54545	—1	12,96182	12,96182
IV	$X_4 (x_{4.ij})$	1,42959	0,32826	1		—0,74468	2,53293	0	0	—1,67174	2,87437	2,87436
		0,58066	3,61293	1		0,46595	0	4,48032	0	—1,97134	8,16851	8,16852
		0,41488	0,23455	1		0,52753	0	0	1,05086	—0,23119	2,99664	2,99663
V	$X_5 (x_{5.ij})$	1,01471	0,09371			—1,27221	2,53293	0	— 1,05086	—1,44055	—0,12227	—0,12227
		0,16578	3,37838			—0,06158	0	4,48032	— 1,05086	—1,74015	5,17187	5,17189
VI	$X_6 (x_{6.ij})$	10,82819	1			—13,57603	27,02945	0	—11,21396	—15,37243	—1,30477	—1,30478
		0,04907	1			—0,01823	0	1,32617	—0,31105	—0,51508	1,53087	1,53088
VII	$X_7 (x_{7.ij})$	10,77912				—13,55780	27,02945	—1,32617	—10,90291	—14,85735	—2,83564	—2,83566
VIII	$X_8 (x_{8.ij})$	1	0	0	0		2,50757	—0,12303	—1,01148	—1,37835	0,99473	0,99471
		0	1	0	0		—0,12305	1,33221	0,26142	—0,44744	1,50028	1,50030
		0	0	1	0		—1,01147	—0,26143	1,53182	0,44561	1,70452	1,70452
		0	0	0	1		—1,37832	—0,44746	0,44560	2,00855	1,62836	1,62837
IX	a_i	—1,25778	0,04349	1,03916	1,48238		— 1,25780	0,04350	1,03916	1,48239		

IV. Делим каждое число, полученное на предыдущем шаге III на элемент столбца (3) соответствующей строки: $x_{4.ij}=\frac{x_{3.ij}}{x_{3.i3}}$.

V. Вычитаем поэлементно последнюю строку IV из всех остальных.

VI, VII. Повторяем те же операции, что и выше.

Если требуется только решить систему, то после шага VII неизвестные находятся последовательно по формулам:

$$a_1=\frac{x_{7.15}}{x_{7.11}},$$

$$a_2=x_{6.i5}-x_{6.i1}\cdot a_1, \qquad i=1 \text{ или } 2$$

$$a_3=x_{4.i5}-x_{4.i1}\cdot a_1-x_{4.i2}\cdot a_2, \qquad i=1, \text{ или } 2, \text{ или } 3,$$

$$a_4=x_{2.i5}-x_{2.i1}\cdot a_1-x_{2.i2}\cdot a_2-x_{2.i3}\cdot a_3, \quad i=1, \text{ или } 2, \text{ или } 3, \text{ или } 4.$$

Для отыскания a_2, a_3, a_4 можно использовать разные уравнения, полагая $i=1$, или 2, или 3. Найдем, например, неизвестные с помощью последних строк на соответствующих шагах. (Полагаем $i=2$ для a_2, $i=3$ для a_3, $i=4$ для a_4.) Получим

$$a_1=\frac{-13{,}55780}{10{,}77912}=-1{,}25778,$$

$$a_2=-0{,}01823-0{,}04907\cdot(-1{,}25778)=0{,}04349,$$

$$a_3=0{,}52753-0{,}41488\cdot(-1{,}25778)-0{,}23455\cdot 0{,}04349=1{,}03916,$$

$$a_4=0{,}90-0{,}66\cdot(-1{,}25778)-0{,}44\cdot 0{,}04349-0{,}22\times$$
$$\times 1{,}03916=1{,}48238.$$

Полученные значения неизвестных выписаны в строке IX таблицы.

VIII. Поясним схему получения элементов обратной матрицы, выписанной в столбцах (6)—(9) (VIII). Элементы первой строки матрицы C получены по формуле

$$c_{1k}=\frac{x_{7.1j}}{x_{7.11}}, \qquad j=6,\ 7,\ 8,\ 9 \quad (k=j-5).$$

Например, $c_{11}=\frac{27{,}02945}{10{,}77912}=2{,}50757$, $c_{12}=\frac{-1{,}32617}{10.77912}=-0{,}12303$ и т. д.

Слева выписываем единичную матрицу. Контроль вычислений тот же (по сумме элементов строки). Числа контрольного столбца получаются по тем же формулам, что и c_{mk} при $j=10$. Элементы следующих строк матрицы C получаются по формулам

$$c_{2k}=x_{6.ij}-x_{6.i1}c_{1k}, \qquad i=1 \text{ или } 2,$$

$$c_{3k}=x_{4.ij}-x_{4.i1}c_{1k}-x_{4.i2}c_{2k}, \qquad i=1, \text{ или } 2, \text{ или } 3,$$

$$c_{4k}=x_{2.ij}-x_{2.i1}c_{1k}-x_{2.i2}c_{2k}-x_{2.i3}c_{3k}, \quad i=1, \text{ или } 2, \text{ или } 3, \text{ или } 4,$$

$$j=6,\ 7,\ 8,\ 9, \quad k=j-5.$$

Элементы обратной матрицы также можно находить из разных уравнений, полагая $i=1$, или 2, или 3, т. е. выбирая ту или иную строку таблицы. Найдем, например, c_{2k} с помощью второй строки VI:

$$c_{21}=x_{6.26}-x_{6.21}c_{11}=0-0{,}04907\cdot 2{,}50757=-0{,}12305.$$

Мы видим, что c_{21} получилось отличным от $c_{12}=-0{,}12303$ (матрица C — также симметрична) в пятом десятичном знаке, на котором сказалась степень точности вычислений. Далее,

$$c_{22}=1{,}32617-0{,}04907\cdot(-0{,}12303)=1{,}33221 \text{ и т. д.}$$

Найдем c_{3k} и c_{4k} с помощью первых строк IV и II (полагаем $i=1$). Например,

$$c_{32}=0-1{,}42959\cdot(-0{,}12303)-0{,}32826\cdot 1{,}33221=-0{,}26143,$$

$$c_{43}=0-1{,}51515\cdot(-1{,}01148)-0{,}63636\cdot(-0{,}26142)-0{,}81818\times \times 1{,}53182=0{,}44560.$$

В качестве окончательного контроля вычисления элементов матрицы C можно использовать условие $XC=E$, где E—единичная матрица. Элементы d_{ij} матрицы E на главной диагонали равны единице, а вне нее — нулю. Очевидно, $d_{ij}=\sum_{k=1}^{4}x_{ik}c_{kj}$. Вычислим диагональные элементы d_{ii} матрицы E для нашего случая.

$$d_{11}=x_{11}c_{11}+x_{12}c_{21}+x_{13}c_{31}+x_{14}c_{41}=1\cdot 2{,}50757+0{,}42\times \times(-0{,}12305)+0{,}54\cdot(-1{,}01147)+0{,}66\cdot(-1{,}37832)=0{,}999999,$$

$$d_{22}=0{,}999997,\quad d_{33}=0{,}999998,\quad d_{44}=1{,}000\,000.$$

В случаях, когда точность вычисления элементов обратной матрицы нас не удовлетворяет, можно, не проделывая заново вычислений с бóльшим числом значащих цифр, воспользоваться методами для уточнения значений c_{mk} (например, итеративным процессом Хотеллинга *)).

Значения неизвестных a_i вычисляются по элементам C по формуле

$$a_i=\sum_{j=1}^{4}c_{ij}b_j.$$

В нашем случае они равны

$$a_1=c_{11}b_1+c_{12}b_2+c_{13}b_3+c_{14}b_4=$$
$$=2{,}50757\cdot 0{,}3-0{,}12303\cdot 0{,}5-1{,}01148\cdot 0{,}7-1{,}37835\cdot 0{,}9=$$
$$=-1{,}25780,$$
$$a_2=0{,}04350,\quad a_3=1{,}03916,\quad a_4=1{,}48238.$$

*) См. В. Н. Фаддеева, Вычислительные методы линейной алгебры, стр. 103—106.

После отыскания a_i целесообразно подставить их в исходные уравнения:

$$a_1 + 0{,}42a_2 + 0{,}54a_3 + 0{,}66a_4 = -1{,}25780 + 0{,}42 \cdot 0{,}04350 + \\ + 0{,}54 \cdot 1{,}03916 + 0{,}66 \cdot 1{,}48239 = 0{,}29999,$$
$$0{,}42a_1 + a_2 + 0{,}32a_3 + 0{,}44a_4 = 0{,}50001,$$
$$0{,}54a_1 + 0{,}32a_2 + a_3 + 0{,}22a_4 = 0{,}69999,$$
$$0{,}66a_1 + 0{,}44a_2 + 0{,}22a_3 + a_4 = 0{,}90000.$$

Однако с теоретической точки зрения этот контроль не является достаточным.

Может встретиться случай, когда какой-либо из элементов, на которые в процессе вычислений приходится делить, равен нулю. Тогда, чтобы не усложнять вычислений (соответствующей перестановкой столбцов), можно использовать другую схему расчета, например схему Б, где деление производится только на диагональные элементы, т. е. на положительные числа.

Схема А довольно громоздка, хотя и крайне проста однотипностью вычислительных операций. Приведем здесь одну из компактных схем для решения системы нормальных уравнений и обращения симметричной матрицы X*).

Схема Б. Вычисления приведены в табл. 19. Поясним порядок заполнения таблицы.

Таблица 19

Схема Б

		(1)	(2)	(3)	(4)	(5)	(6)	(7)
I		1,00	0,42 1,00	0,54 0,32 1,00	0,66 0,44 0,22 1,00	0,3 0,5 0,7 0,9	2,92 2,68 2,78 3,22	
II	X_1 Y_1	1,00 1,00	0,42 0,42	0,54 0,54	0,66 0,66	0,3 0,3	2,92 2,92	2,92 2,92
III	X_2 Y_2		0,8236 1	0,0932 0,11316	0,1628 0,19767	0,3740 0,45410	1,4536 1,76493	1,4536 1,76493
IV	X_3 Y_3			0,69785 1	— 0,15482 — 0,22185	0,49568 0,71030	1,03871 1,48844	1,03871 1,48845
V	X_4 Y_4				0,49787 1	0,73804 1,48240	1,23591 2,48240	1,23591 2,48240

*) Схема квадратного корня (см. цитированную на стр. 165 литературу) представляется также очень удобной для задач, связанных с обработкой статистических данных.

$$\text{VI.}\quad \begin{Vmatrix} 2{,}50759 & -0{,}12304 & -1{,}01149 & -1{,}37834 \\ -0{,}12304 & 1{,}33221 & -0{,}26142 & -0{,}44746 \\ -1{,}01149 & -0{,}26142 & 1{,}53183 & 0{,}44560 \\ -1{,}37834 & -0{,}44746 & 0{,}44560 & 2{,}00856 \end{Vmatrix}$$

I. Слева помещаем элементы из верхнего правого угла исходной матрицы X, рядом — столбец (5) свободных членов b_i, справа — контрольный столбец (6). На первом шаге в контрольный столбец выписываются суммы всех элементов строки (при этом рассматривается вся матрица X, а не только ее угол). На дальнейших шагах расчета числа контрольного столбца подвергаются тем же вычислительным операциям, что и остальные числа строки.

II. Строка X_1: переписываем первую строку I, т. е. $X_{1j}=x_{1j}$. Строка Y_1: делим элементы строки X_1 на X_{11}. В нашем примере $X_{11}=1$, и потому строки X_1 и Y_1 совпадают.

Контроль вычислений для всех шагов: сумма всех элементов строки должна равняться (с точностью до ошибок округления) числу, стоящему в контрольном столбце.

III. Элементы строки X_2 вычисляем по формуле

$$X_{2j}=x_{2j}-[X_{12}Y_{1j} \text{ или } X_{1j}Y_{12}],$$

где x_{2j} — элементы второй строки I и (в пределах точности вычислений) имеет место равенство $X_{12}Y_{1j}=X_{1j}Y_{12}$. В нашем случае строки X_1 и Y_1 совпадают и имеет место точное равенство. Находим

$$\begin{aligned} X_{22}&=1\quad\;\; -0{,}42\cdot 0{,}42=0{,}8236,\\ X_{23}&=0{,}32-0.42\cdot 0{,}54=0{,}0932,\\ X_{24}&=0{,}44-0{,}42\cdot 0{,}66=0{,}1628 \text{ и т. д.} \end{aligned}$$

Элементы строки Y_2 равны $Y_{2j}=\dfrac{X_{2j}}{X_{22}}$. Значит $Y_{22}=1$, $Y_{23}=$ $=0{,}11316$ и т. д.

IV. Вычисляем элементы строки X_3 по формуле:

$$X_{3j}=x_{3j}-[(X_{13}Y_{1j}+X_{23}Y_{2j}) \text{ или } (X_{1j}Y_{13}+X_{2j}Y_{23})].$$

В нашем примере для двух сумм в квадратных скобках имеет место точное равенство, а не приближенное. Получим

$$\begin{aligned} X_{33}&=1-(0{,}54\cdot 0{,}54+0{,}0932\cdot 0{,}11316)=0{,}69785,\\ X_{34}&=0{,}22-(0{,}54\cdot 0{,}66+0{,}0932\cdot 0{,}19767)=-0{,}15482 \text{ и т. д.} \end{aligned}$$

Элементы строки Y_3 равны $Y_{3j}=\dfrac{X_{3j}}{X_{33}}$.

V. Элементы строки X_4 вычисляем по формуле

$$X_{4j}=x_{4j}-[(X_{14}Y_{1j}+X_{24}Y_{2j}+X_{34}X_{3j}) \text{ или } (X_{1j}Y_{14}+X_{2j}Y_{24}+X_{3j}Y_{34})].$$

Например,

$$X_{44}=1-(0{,}66\cdot 0{,}66+0{,}1628\cdot 0{,}19767+$$
$$+0{,}15482\cdot 0{,}22185)=0{,}49787.$$

Элементы строки Y_4 вычисляются по формуле $Y_{4j}=\frac{X_{4j}}{X_{44}}$.

Не обращая матрицы коэффициентов, решение системы можно находить по формулам:

$$a_4=Y_{45},$$
$$a_3=Y_{35}-Y_{34}a_4,$$
$$a_2=Y_{25}-Y_{23}a_3-Y_{24}a_4,$$
$$a_1=Y_{15}-Y_{12}a_2-Y_{13}a_3-Y_{14}a_4.$$

В нашем случае применение формул дает

$$a_4=1{,}48240,\quad a_3=1{,}03917,\quad a_2=0{,}04348,\quad a_1=-1{,}25780.$$

VI. Опишем схему получения элементов обратной матрицы. Элементы c_{ij} мы будем последовательно получать, начиная с c_{i4}, а потому матрица C заполняется по столбцам слева направо, причем снизу вверх.

Вычисляем c_{i4} по формулам

$$c_{44}=\frac{1}{X_{44}},$$
$$c_{34}=-c_{44}Y_{34},$$
$$c_{24}=-c_{34}Y_{23}-c_{44}Y_{24},$$
$$c_{14}=-c_{24}Y_{12}-c_{34}Y_{13}-c_{44}Y_{14}.$$

В нашем примере

$$c_{44}=\frac{1}{0{,}49787}=2{,}00856,$$
$$c_{34}=-2{,}00856\cdot(-0{,}22185)=0{,}44560,$$
$$c_{24}=-0{,}44746,$$
$$c_{14}=-1{,}37834.$$

Далее вычисляем c_{i3} по формулам

$$c_{43}=c_{34},$$
$$c_{33}=\frac{1}{X_{33}}-c_{34}Y_{34},$$
$$c_{23}=-c_{33}Y_{23}-c_{34}Y_{24},$$
$$c_{13}=-c_{23}Y_{12}-c_{33}Y_{13}-c_{34}Y_{14}.$$

Например, $c_{33}=\frac{1}{0{,}69785}-0{,}44560\cdot(-0{,}22185)=1{,}53183.$

Находим c_{i2} по формулам

$$c_{42}=c_{24}, \quad c_{32}=c_{23},$$
$$c_{22}=\frac{1}{X_{22}}-c_{23}Y_{23}-c_{24}Y_{24},$$
$$c_{12}=-c_{22}Y_{12}-c_{23}Y_{13}-c_{24}Y_{14}.$$

И, наконец, $c_{41}=c_{14}, \quad c_{31}=c_{13}, \quad c_{21}=c_{12}$,

$$c_{11}=\frac{1}{X_{11}}-c_{12}Y_{12}-c_{13}Y_{13}-c_{14}Y_{14}.$$

Обратная матрица C выписана внизу в схеме Б. Окончательный контроль для c_{ij} сводится к вычислению сумм произведений $d_{ii}=\sum_{k=1}^{4} x_{ik}c_{ki}$ (диагональных элементов единичной матрицы), которые должны быть равны единице. В нашем примере $d_{11}=1{,}000004$, $d_{22}=0{,}999996$, $d_{33}=1{,}000003$, $d_{44}=1{,}000005$.

При решении системы с помощью матрицы C дальнейшие вычисления те же, что и в схеме А. Для уточнения элементов обратной матрицы, в случаях, когда в этом есть надобность, целесообразно применять итерационный метод Хотеллинга (см. ссылку в схеме А).

§ 13. Примеры

Пример 1. При исследовании влияния температуры на ход хронометра получены результаты *), помещенные в первых двух столбцах табл. 20.

Таблица 20

t_i	ω_i	t_i-15	$(t_i-15)^2$	$\left(\frac{t_i-15}{15}\right)=x'_i$	$\left(\frac{t_i-15}{15}\right)^2=x''_i$
5,0	2,60	−10,0	100,00	−0,667	0,445
9,6	2,01	−5,4	29,16	−0,360	0,130
16,0	1,34	1,0	1,00	0,067	0,004
19,6	1,08	4,6	21,16	0,307	0,094
24,4	0,94	9,4	88,36	0,627	0,393
29,8	1,06	14,8	219,04	0,987	0,974
34,4	1,25	19,4	376,36	1,293	1,672

Полагают, что зависимость хода хронометра от температуры может быть выражена уравнением

$$\omega=\omega_{15}+(t-15)\beta+(t-15)^2\gamma.$$

*) Данные взяты из книги А. П. Ющенко [56], стр. 83.

Требуется оценить постоянные ω_{15}, β, γ методом наименьших квадратов, найти доверительные интервалы для этих величин и оценить точность наблюдений.

Прежде чем искать систему нормальных уравнений для ω_{15}, β, γ, целесообразно сделать в исходном уравнении замену переменных, так как множители $(t_1-15)^2$ очень велики, и потому коэффициенты нормальных уравнений будут колебаться в широких пределах, что усложняет вычисления. Положим $a_1=15\beta$, $a_2=225\gamma$, $a_3=\omega_{15}$. Тогда исходное уравнение перепишется в виде

$$\left(\frac{t-15}{15}\right)a_1+\left(\frac{t-15}{15}\right)^2 a_2+a_3=\omega;$$

значения величин $x_i'=\frac{t_i-15}{15}$ и $x_i''=\left(\frac{t_i-15}{15}\right)^2$ приведены в пятом и шестом столбцах табл. 20.

Составим систему нормальных уравнений для определения a_1, a_2, a_3. Коэффициенты x_{ij} нормальных уравнений, равные $x_{11}=[x'x']$, $x_{12}=[x'x'']$ и т. д., получаем накоплением, не выписывая промежуточных результатов. Выпишем их в виде матрицы

$$X=\begin{Vmatrix} 3{,}712 & 3{,}055 & 2{,}254 \\ 3{,}055 & 4{,}122 & 3{,}712 \\ 2{,}254 & 3{,}712 & 7 \end{Vmatrix}.$$

Свободные члены равны: $\omega_1=1{,}215$, $\omega_2=5{,}017$, $\omega_3=10{,}280$. Таким образом, система нормальных уравнений для a_i имеет вид

$$\begin{aligned} 3{,}712a_1+3{,}055a_2+2{,}254a_3&=\ \ 1{,}215,\\ 3{,}055a_1+4{,}122a_2+3{,}712a_3&=\ \ 5{,}017,\\ 2{,}254a_1+3{,}712a_2+7{,}000a_3&=10{,}280. \end{aligned}$$

Решим эту систему и обратим матрицу ее коэффициентов, пользуясь компактной вычислительной схемой (схема Б на стр. 170). В табл. 21 приведены результаты вычислений по этой схеме. Под таблицей помещена обратная матрица X^{-1}. Для контроля элементов матрицы $X^{-1}=C$ вычисляем диагональные элементы единичной матрицы $E=XC$ по формуле $d_{ii}=\sum_{k=1}^{3}x_{ik}c_{ki}$. Получим $d_{11}=1{,}00002$, $d_{22}=0{,}99980$, $d_{33}=0{,}99989$.

Решение системы находим по формуле $\tilde{a}_i=\sum_{j=1}^{3}c_{ij}\omega_j$: $\tilde{a}_1=-1{,}246$, $\tilde{a}_2=0{,}874$, $\tilde{a}_3=1{,}406$. Значит, точечными оценками для ω_{15}, β, γ будут $\tilde{\omega}_{15}=1{,}406$, $\tilde{\beta}=\frac{\tilde{a}_1}{15}=-0{,}0831$, $\tilde{\gamma}=\frac{\tilde{a}_2}{225}=0{,}00388$. Таким

Таблица 21

	(1)	(2)	(3)	(4)	Контроль	Сумма элементов строки
	3,712	3,055 4,122	2,254 3,712 7,000	1,215 5,017 10,280	10,236 15,906 23,246	
X_1 Y_1	3,712 1	3,055 0,8230	2,254 0,6072	1,215 0,3273	10,236 2,7575	10,236 2,7575
X_2 Y_2		1,608 1	1,857 1,155	4,017 2,498	7,482 4,653	7,482 4,653
X_3 Y_3			3,487 1	4,903 1,406	8,390 2,406	8,390 2,406

$$C = X^{-1} = \begin{Vmatrix} 0,72444 & -0,62555 & 0,09847 \\ -0,62555 & 1,00446 & -0,33123 \\ 0,09847 & -0,33123 & 0,28678 \end{Vmatrix}.$$

образом, уравнение, выражающее зависимость хода хронометра от температуры, запишется в виде

$$\omega = 1,406 - 0,0831\,(t - 15) + 0,0039\,(t - 15)^2.$$

Перейдем к отысканию доверительных интервалов для параметров, которые определяются по формуле

$$I_{\gamma_1}^{(i)} = \left[\tilde{a}_i' - \gamma_1 \sqrt{c_{ii}\frac{[\tilde{v}\tilde{v}]}{N-n}},\ \ \tilde{a}_i' + \gamma_1 \sqrt{c_{ii}\frac{[\tilde{v}\tilde{v}]}{N-n}}\right],$$

где $\tilde{a}_i'$ — точечная оценка параметра, N — число наблюдений, n—число параметров, γ_1 определяется по таблице распределения Стьюдента из условия $P\{|t| < \gamma_1\} = p$ при $k = N - n$ степенях свободы, c_{ii} — диагональный элемент матрицы $C = X^{-1}$ и

$$[\tilde{v}\tilde{v}] = [\omega\omega] - \sum_{i=1}^{n} [a_i'\omega]\,\tilde{a}_i'.$$

В нашем случае $[\omega\omega] = 17,3318$, $[a_1'\omega] = [(t - 15)\,\omega] = 18,228$, $[a_2'\omega] = [(t - 15)^2\,\omega] = 1128,50$, $[a_3'\omega] = [1 \cdot \omega] = 10,28$.

Значит,

$$[\tilde{v}\tilde{v}] = 17,3318 - (-18,228 \cdot 0,0831 + 1128,50 \cdot 0,00388 + 10,28 \cdot 1,406) = 0,01429.$$

Возьмем надежность $p=0{,}90$. Тогда по таблице t-распределения для $k \doteq 7-3=4$ находим $\gamma_1=2{,}132$. Диагональные элементы матрицы C равны $c_{11}=0{,}7244$, $c_{22}=1{,}0045$, $c_{33}=0{,}2868$; точечные оценки для параметров равны

$$\tilde{a}'_1=-0{,}0831 \text{ (для } \beta), \quad \tilde{a}'_2=0{,}0039 \text{ (для } \gamma), \quad \tilde{a}'_3=1{,}406 \text{ (для } \omega_{15}).$$

Подставив значения соответствующих величин в написанную выше формулу, получим доверительные интервалы

для β:

$$I^{(1)}=[-0{,}191;\ 0{,}025],$$

для γ:

$$I^{(2)}=[-0{,}124;\ 0{,}132],$$

для ω_{15}:

$$I^{(3)}=[1{,}338;\ 1{,}474].$$

Оценим точность наблюдений с помощью соотношения

$$P\{\gamma_1\tilde{s}<\sigma<\gamma_2\tilde{s}\}=P_2-P_1=p.$$

Вычислим

$$\tilde{s}=\sqrt{\frac{[\tilde{v}\tilde{v}]}{N-n}}=\sqrt{\frac{0{,}01429}{4}}=0{,}0598.$$

Возьмем надежность $p=0{,}90$. Тогда для числа степеней свободы $k=N-n=4$ находим по табл. III приложений значения $\gamma_1=0{,}649$, $\gamma_2=2{,}37$. Значит, доверительный интервал для σ определяется неравенством

$$0{,}038<\sigma<0{,}142.$$

Пример 2. В табл. 22 *) приведены показания барометра — анероида (A) и ртутного барометра (B_0) при различной температуре (t).

Таблица 22

№	t (°C)	A (мм)	B_0 (мм)
1	10,0	749,0	744,4
2	6,2	746,1	741,3
3	6,3	756,6	752,7
4	5,3	758,9	754,7
5	4,8	751,7	747,9
6	3,8	757,5	754,0
7	17,1	752,4	747,8
8	22,2	752,5	748,6
9	20,8	752,2	747,7
10	21,0	759,5	755,6

*) Данные взяты из книги А. С. Чеботарёва [51], стр. 231.

Известна эмпирическая формула

$$B_0 = A + a_1 + a_2 t + a_3 (760 - A). \qquad (6.12.1)$$

Здесь a_1, a_2, a_3 — неизвестные постоянные.

Найдем оценки для a_1, a_2, a_3 по методу наименьших квадратов, доверительные интервалы для этих величин и оценим точность наблюдений.

Положив $x_1 = 1$, $x_2 = t$, $x_3 = 760 - A$, $b = B_0 - A$, перепишем формулу (6.12.1) в виде

$$x_1 a_1 + x_2 a_2 + x_3 a_3 = b.$$

Данные табл. 22 для величин x_i, b представятся тогда табл. 23. (Наблюденные значения величин будем обозначать соответствующими буквами со штрихами.)

Таблица 23

№	x'_1	x'_2	x'_3	b'
1	1	10,0	11,0	—4,6
2	1	6,2	13,9	—4,8
3	1	6,3	3,4	—3,9
4	1	5,3	1,1	—4,2
5	1	4,8	8,3	—3,8
6	1	3,8	2,5	—3,5
7	1	17,1	7,6	—4,6
8	1	22,2	7,5	—3,9
9	1	20,8	7,8	—4,5
10	1	21,0	0,5	—3,9
Сумма . .	10	117,5	63,6	—41,7

Система нормальных уравнений для определения a_1, a_2, a_3 имеет вид

$$x_{11} a_1 + x_{12} a_2 + x_{13} a_3 = b_1,$$
$$x_{21} a_1 + x_{22} a_2 + x_{23} a_3 = b_2,$$
$$x_{31} a_1 + x_{32} a_2 + x_{33} a_3 = b_3,$$

где коэффициенты при неизвестных определяются по формулам

$$x_{11} = [x'_1 x'_1], \quad x_{12} = [x'_1 x'_2], \quad x_{13} = [x'_1 x'_3], \quad b_1 = [x'_1 b'],$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$x_{31} = [x'_3 x'_1], \quad \ldots\ldots\ldots\ldots \quad b_3 = [x'_3 b'],$$

и, очевидно, $x_{ij} = x_{ji}$.

Не выписывая отдельно значений слагаемых (результаты получаются накоплением), приведем значения x_{ij} в виде матрицы X

коэффициентов при неизвестных в нормальных уравнениях и матрицы B свободных членов:

$$X = \begin{Vmatrix} 10 & 117{,}5 & 63{,}6 \\ 117{,}5 & 1902{,}59 & 741{,}97 \\ 63{,}6 & 741{,}97 & 577{,}22 \end{Vmatrix}, \quad B = \begin{Vmatrix} -41{,}7 \\ -494{,}87 \\ -276{,}75 \end{Vmatrix}.$$

Таким образом, система нормальных уравнений для нашего случая имеет вид:

$$\begin{aligned} 10{,}0a_1 + 117{,}5a_2 + 63{,}6a_3 &= -41{,}7, \\ 117{,}5a_1 + 1902{,}59a_2 + 741{,}97a_3 &= -494{,}87, \\ 63{,}6a_1 + 741{,}97a_2 + 577{,}22a_3 &= -276{,}75. \end{aligned}$$

Решим систему, пользуясь компактной вычислительной схемой (схема Б). В табл. 24 приведены результаты вычислений по этой схеме. Полное описание схемы дано на стр. 170—173.

Таблица 24

	(1)	(2)	(3)	(4)	Контроль	Сумма элементов строки
	10	117,5 1902,59	63,6 741,97 577,22	—41,70 —494,87 —276,75	149,40 2267,19 1106,04	
X_1 Y_1	10 1	117,5 11,75	63,6 6,36	—41,70 —4,17	149,40 14,94	149,40 14,94
X_2 Y_2		521,965 1	—5,330 —0,01021141	—4,895 —0,00937802	511,740 0,98041056	511,740 0,98041057
X_3 Y_3			172,6696 1	—11,5880 —0,0671	161,0816 0,9329	161,0816 0,9329

$$X^{-1} = C = \begin{Vmatrix} 0{,}6076815 & -0{,}0228938 & -0{,}0375282 \\ -0{,}0228938 & 0{,}0019164 & 0{,}00005914 \\ -0{,}0375282 & 0{,}00005914 & 0{,}0057914 \end{Vmatrix}.$$

Элементы соответствующих строк табл. 24 получены по формулам

$$X_{1j} = x_{1j}, \qquad Y_{1j} = \frac{X_{1j}}{X_{11}},$$

$$X_{2j} = x_{2j} - X_{12}Y_{1j}, \qquad Y_{2j} = \frac{X_{2j}}{X_{22}},$$

$$X_{3j} = x_{3j} - (X_{13}Y_{1j} + X_{23}Y_{2j}), \qquad Y_{3j} = \frac{X_{3j}}{X_{33}}.$$

Например, $X_{23} = 741{,}97 - 117{,}5 \cdot 6{,}36 = -5{,}330$,

$X_{34} = -276{,}75 - (-63{,}6 \cdot 4{,}17 + 5{,}330 \cdot 0{,}00937802) = -11{,}5880$.

Решение системы, как уже говорилось выше, можно получить по формулам

$$\tilde{a}_3 = Y_{34},$$
$$\tilde{a}_2 = Y_{24} - Y_{23}\tilde{a}_3,$$
$$\tilde{a}_1 = Y_{14} - Y_{12}\tilde{a}_2 - Y_{13}\tilde{a}_3,$$

применение которых в нашем случае дает

$$\tilde{a}_3 = -0{,}0671,$$
$$\tilde{a}_2 = -0{,}00937802 - (-0{,}01021141) \cdot (-0{,}0671) = -0{,}0101,$$
$$\tilde{a}_1 = -3{,}6246.$$

Однако при определении доверительных границ для a_i нам понадобятся диагональные элементы обратной матрицы $X^{-1} = C$.

Поясним способ отыскания элементов c_{ij} матрицы C. Сначала находим

$$c_{33} = \frac{1}{X_{33}} = \frac{1}{172{,}6696} = 0{,}0057914.$$

Далее,

$$c_{23} = -c_{33}Y_{23} = -0{,}0057914 \cdot (-0{,}01021141) = 0{,}00005914,$$
$$c_{13} = -c_{23}Y_{12} - c_{33}Y_{13} = -0{,}00005914 \cdot 11{,}75 - 0{,}0057914 \cdot 6{,}36 = \\ = -0{,}0375282.$$

Затем находим

$$c_{22} = \frac{1}{X_{22}} - c_{23}Y_{23} = 0{,}0019164,$$
$$c_{12} = -c_{22}Y_{12} - c_{23}Y_{13} = -0{,}0228938,$$

и, наконец,

$$c_{11} = \frac{1}{X_{11}} - c_{12}Y_{12} - c_{13}Y_{13} = 0{,}6076815.$$

Остальные элементы матрицы C выписываем, пользуясь условием симметрии $c_{ij} = c_{ji}$. Для контроля матрицы C вычисляем диагональные элементы матрицы $E = XC$ по формуле $d_{ii} = \sum_{k=1}^{3} x_{ik}c_{ki}$. Получим $d_{11} = 1{,}000000$, $d_{22} = 0{,}999982$, $d_{33} = 0{,}999998$. Решение системы находим по формуле $a_i = \sum_{j=1}^{3} c_{ij}b_j$; отсюда имеем

$$\tilde{a}_1 = -3{,}6249, \quad \tilde{a}_2 = -0{,}0101, \quad \tilde{a}_3 = -0{,}0671.$$

Мы видим, что ранее полученное решение отличается от решения, полученного с помощью обратной матрицы, лишь в последнем знаке $\tilde{a}_1$. Расхождение объясняется ошибками округления.

Замечание. В нормальных уравнениях рассматриваемого примера существенно колеблется порядок коэффициентов при неизвестных (от 10 до 1902). Такого рода неоднородность в коэффициентах встречается довольно часто и создает затруднения при вычислениях. Она вызывает необходимость увеличения количества значащих цифр для обеспечения требуемой точности окончательных результатов. Добиться большей однородности в порядке коэффициентов можно, уменьшая наблюденные значения некоторых из x_i в несколько раз, одновременно увеличивая во столько же раз значения соответствующих неизвестных. Например, в нашем случае можно уменьшить все значения x_2', x_3' в 10 раз и искать решение системы

$$10a_1' + 11{,}75a_2' + 6{,}36a_3' = -41{,}7,$$
$$11{,}75a_1' + 19{,}03a_2' + 7{,}42a_3' = -49{,}49,$$
$$6{,}36a_1' + 7{,}42a_2' + 5{,}772a_3' = -27{,}67.$$

Тогда решением исходной системы будет

$$a_1 = a_1', \quad a_2 = \frac{a_2'}{10}, \quad a_3 = \frac{a_3'}{10}.$$

Перейдем к отысканию доверительных интервалов для параметров a_1, a_2, a_3.

Доверительным интервалом для a_i является интервал

$$I_\gamma^{(i)} = \left[\tilde{a}_i - \gamma\sqrt{c_{ii}\frac{[\tilde{v}\tilde{v}]}{N-n}},\ \tilde{a}_i + \gamma\sqrt{c_{ii}\frac{[\tilde{v}\tilde{v}]}{N-n}}\right].$$

Здесь $\tilde{a}_i$ — точечная оценка параметра a_i, найденная из системы нормальных уравнений, N — число наблюдений, n — число параметров a_i, значение γ определяется по таблице распределения Стьюдента из условия $P\{|t| < \gamma\} = p$ (p — заданная надежность) при $k = N - n$ степенях свободы, c_{ii} — диагональный элемент матрицы $C = X^{-1}$ и, наконец,

$$[\tilde{v}\tilde{v}] = [b'b'] - \sum_{i=1}^{n}[x_i'b']\tilde{a}_i.$$

Вычислим $[\tilde{v}\tilde{v}]$ для нашего примера. Подставив в формулу значения $[b'b'] = 175{,}57$, $[x_1'b'] = b_1 = -41{,}7$, $[x_2'b'] = b_2 = -494{,}87$, $[x_3'b'] = b_3 = -276{,}75$, $\tilde{a}_1 = -3{,}6249$, $\tilde{a}_2 = -0{,}0101$, $\tilde{a}_3 = -0{,}0671$, получим

$$[\tilde{v}\tilde{v}] = 175{,}57 - (41{,}70 \cdot 3{,}6249 + 494{,}87 \cdot 0{,}0101 + $$
$$+ 276{,}75 \cdot 0{,}0671) = 0{,}8436.$$

Возьмем надежность $p=0{,}90$ и по таблице распределения Стьюдента найдем для $p=0{,}90$ и числа степеней свободы $k=N-n=$ $=10-3=7$ значение $\gamma=1{,}895$. Диагональные элементы матрицы C равны

$$c_{11}=0{,}6076815, \quad c_{22}=0{,}0019164, \quad c_{33}=0{,}0057914,$$

следовательно, доверительный интервал для a_1 определяется неравенством

$$-3{,}625-1{,}895\sqrt{0{,}6077\,\frac{0{,}8436}{7}}<a_1<-3{,}625+$$
$$+1{,}895\sqrt{0{,}6077\,\frac{0{,}8436}{7}},$$

или

$$-4{,}138<a_1<-3{,}112.$$

Произведя вычисления, получим доверительные интервалы для a_2, a_3 при надежности $p=0{,}90$:

$$I^{(2)}=[-0{,}0389;\ 0{,}0187],$$
$$I^{(3)}=[-0{,}1172;\ -0{,}0170].$$

Оценим точность наблюдений. Доверительный интервал для σ определяется из соотношения $P\{\gamma_1\tilde{s}<\sigma<\gamma_2\tilde{s}\}=P_2-P_1=p$. Здесь $\tilde{s}$ — точечная оценка для σ, $\tilde{s}=\sqrt{\frac{[\tilde{v}\tilde{v}]}{N-n}}$, p — заданная надежность, γ_1 и γ_2 определяются по табл. III γ-границ доверительного интервала для числа степеней свободы $k=N-n$.

Возьмем $p=0{,}90$. Вычисляем $\tilde{s}=\sqrt{\frac{0{,}8436}{7}}=0{,}347$. По табл. III для $k=N-n=7$ находим $\gamma_1=0{,}705$, $\gamma_2=1{,}797$. Значит, доверительный интервал для σ определяется неравенством

$$0{,}245<\sigma<0{,}624.$$

ГЛАВА VII

ОЦЕНИВАНИЕ ЛИНЕЙНЫХ ФОРМ ОТ ОСНОВНЫХ ПАРАМЕТРОВ ПРИ КОСВЕННЫХ НАБЛЮДЕНИЯХ. ТЕОРЕМЫ Ю. НЕЙМАНА — Ф. ДЭВИД*)

§ 1. Постановка задачи

Метод наименьших квадратов часто применяется к задачам оценивания параметров линейной регрессии. Впервые мы встретились с такой задачей во введении, (в § 1), где рассматривался вопрос о нахождении „выравнивающей" прямой линии для ряда точек по методу наименьших квадратов. Простой теоретико-вероятностный смысл подобная операция получает для задач о линейной регрессии, например, когда две случайные величины Y и X связаны соотношением

$$E(Y/X) = g_1 + g_2 X, \tag{7.1.1}$$

где слева стоит условное математическое ожидание Y при заданном X, g_1, g_2 — известны. Пусть $E(X) = a$, и нам нужно оценить линейную функцию $g_1 + g_2 a$ на основе наблюдений над значениями Y при заданных X. Это — задача оценивания линейной формы. Мы будем иметь несколько практических примеров подобной ситуации. Теперь дадим общую постановку задачи.

Пусть при тех же условиях, что и в гл. VI, производятся наблюдения линейных функций основных параметров $a_1, \ldots, a_n$;

$$y_r = l_r - \Delta_r = a_1 x_{r1} + a_2 x_{r2} + \ldots + a_n x_{rn} \quad (r = 1, 2, \ldots, N), \tag{7.1.2}$$

где наблюдаются величины l_r, а $\Delta_r \in N\left(0, \frac{\sigma}{\sqrt{p_r}}\right)$, p_r — известные нам веса.

Мы хотим произвести оценивание линейной функции параметров

$$a = g_1 a_1 + g_2 a_2 + \ldots + g_n a_n, \tag{7.1.3}$$

где величины g_i известны. Естественно возникает мысль произвести оценивание $a_1, a_2, \ldots, a_n$ по методу наименьших квадратов (согласно

*) Ю. Нейман — американский статистик. Ф. Дэвид — английский статистик, ученица Ю. Неймана.

изложенному в гл. VI) и полученные оценки $\tilde{a}_1, \ldots, \tilde{a}_n$ подставить вместо $a_1, \ldots, a_n$ в (7.1.3). Тогда для a получим следующую оценку:

$$\tilde{a} = g_1\tilde{a}_1 + \ldots + g_n\tilde{a}_n. \tag{7.1.4}$$

Очевидно, она будет несмещенной:

$$E(\tilde{a}) = E(g_1\tilde{a}_1 + \ldots + g_n\tilde{a}_n) = \sum_{j=1}^{n} g_j E(\tilde{a}_j) = \sum_{j=1}^{n} g_j a_j) = a.$$

Однако другие ее качества без дальнейшего исследования неясны. Мы покажем, далее, что оценка $\tilde{a}$ в известном смысле оптимальна.

§ 2. Теоремы Ю. Неймана — Ф. Дэвид

Пусть ранг матрицы X в уравнениях

$$Y = XA, \tag{7.2.1}$$

имеющих тот же смысл, что и в предыдущей главе, равен n. Введем еще матрицу

$$G = G_{1n} = \|g_1, g_2, \ldots, g_n\|,$$

так что

$$a = GA; \quad \tilde{a} = G\tilde{A}. \tag{7.2,2}$$

Таким образом,

$$E(\tilde{a}) = GE(\tilde{A}) = GA = a. \tag{7.2.3}$$

Далее, $B_{\tilde{A}} = \sigma^2 C^{-1}$, где $B_{\tilde{A}}$ — корреляционная матрица вектора $\tilde{A}$, откуда, согласно § 3 гл. II,

$$D(\tilde{a}) = \sigma^2 GC^{-1}G^T. \tag{7.2.4}$$

Естественно пытаться выяснить, насколько $D(\tilde{a})$ близка к нижней грани, доставляемой теорией оценивания (§ 3, гл. III).

Мы докажем далее, что в классе линейных несмещенных оценок вида $\lambda_1 l_1 + \lambda_2 l_2 + \ldots + \lambda_N l_N$ оценка $\tilde{a}$ имеет минимальную дисперсию, если даже погрешности Δ_i не подчиняются нормальному закону. (Это и есть основное утверждение Ю. Неймана и Ф. Дэвид.) Теперь же выведем некоторые формулы для $\tilde{a}$, аналогичные формулам гл. VI. Введем, как в § 6 гл. VI, величины и матрицы

$$H_h = \sum_{r=1}^{N} p_r l_r x_{rh} = \{X^T PL\}_{h1} \qquad (h = 1, 2, \ldots, n), \tag{7.2.5}$$

$$H = X^T PL,$$

$$C = X^T PX = \|c_{hk}\|; \qquad C^T = C, \tag{7.2.6}$$

и окаймления

$$C_{gg}=\left\|\begin{matrix} 0 & g_1 & g_2\dots g_n \\ g_1 & c_{11} & c_{12}\dots c_{1n} \\ \dots & \dots & \dots \\ g_n & c_{n1} & c_{n2}\dots c_{nn}\end{matrix}\right\| \quad (c_{ij}=c_{ji}), \tag{7.2.7}$$

$$C_{gH}=\left\|\begin{matrix} 0 & g_1 & g_2\dots g_n \\ H_1 & c_{11} & c_{12}\dots c_{1n} \\ \dots & \dots & \dots \\ H_n & c_{n1} & c_{n2}\dots c_{nn}\end{matrix}\right\|. \tag{7.2.8}$$

Докажем теорему (Ю. Нейман и Ф. Дэвид [15]).

Теорема 7.2.1. *Справедливы равенства*

$$\tilde{a}=-\frac{\det(C_{gH})}{\det(C)} \tag{7.2.9}$$

и

$$D(\tilde{a})=-\sigma^2\frac{\det(C_{gg})}{\det(C)}. \tag{7.2.10}$$

Для доказательства будем рассуждать, как в конце § 6, гл. VI об окаймлениях (7.2.7) и (7.2.8). Имеем

$$\det(C_{gH})=-\sum_{h,\,k=1}^{n} g_h H_k \{C^{-1}\}_{hk}\det(C)$$

и

$$-\frac{\det(C_{gH})}{\det(C)}=GC^{-1}X^TPL=G\tilde{A}=\tilde{a}, \tag{7.2.11}$$

что доказывает (7.2.9).

Далее, согласно (7.2.4), $D(\tilde{a})=\sigma^2GC^{-1}G^T$. Из окаймления (7.2.8) находим

$$\det(C_{gg})=-\sum_{h,\,k=1}^{n} g_h g_k \{C^{-1}\}_{hk}\det(C)$$

и

$$-\frac{(\det(C_{gg})}{\det(C)}=GC^{-1}G^T=\frac{D(\tilde{a})}{\sigma^2},$$

откуда следует (7.2.10).

Докажем теперь теорему, также принадлежащую Ю. Нейману и Ф. Дэвид.

Теорема 7.2.2. *Среди всех несмещенных линейных относительно наблюдений оценок*

$$\lambda_1 l_1+\lambda_2 l_2+\dots+\lambda_N l_N \tag{7.2.12}$$

для функции a оценка $\tilde{a}$ обладает минимальной дисперсией.

Доказательство проведем в матричной форме. Пусть

$$\Lambda = \| \lambda_1, \lambda_2, \ldots, \lambda_N \| . \tag{7.2.13}$$

Тогда изучаемые нами оценки имеют вид: $[\lambda l] = \Lambda L$, где L, как обычно, обозначает матрицу $\left\| \begin{matrix} l_1 \\ \vdots \\ l_N \end{matrix} \right\|$. Далее несмещенность должна означать, что

$$E([\lambda l]) = GA,$$

т. е.

$$\Lambda E(L) = \Lambda Y = \Lambda X A = GA. \tag{7.2.14}$$

Точный смысл требования несмещенности состоит в том, что равенство (7.2.14) должно выполняться при любом наборе параметров

$$A = \left\| \begin{matrix} a_1 \\ \vdots \\ a_n \end{matrix} \right\| .$$

Легко убедиться, что это возможно тогда и только тогда, когда

$$\Lambda X = G, \tag{7.2.15}$$

что мы и будем предполагать.

Далее, в силу (7.2.12), независимости наблюдений l_r и того, что

$$D(l_r) = \frac{\sigma^2}{p_r} \qquad (i = 1, 2, \ldots, N),$$

имеем

$$D([\lambda l]) = \sigma^2 \sum_{i=1}^{N} \frac{\lambda_i^2}{p_i} = \sigma^2 \Lambda P^{-1} \Lambda^T. \tag{7.2.16}$$

Из (7.2.4) видим, что $D(\tilde{a}) = \sigma^2 G C^{-1} G^T$ или, в силу (7.2.15),

$$D(\tilde{a}) = \sigma^2 \Lambda X C^{-1} X^T \Lambda^T. \tag{7.2,17}$$

Для доказательства того, что $D([\lambda l]) \geqslant D(\tilde{a})$, достаточно теперь доказать, что при любых $\Lambda = \| \lambda_1, \lambda_2, \ldots, \lambda_n \|$ будет

$$\Lambda P^{-1} \Lambda^T \geqslant \Lambda X C^{-1} X^T \Lambda^T = \Lambda X (X^T P X)^{-1} X^T \Lambda^T \tag{7.2.18}$$

в силу того, что $C = X^T P X$.

Для доказательства этого неравенства будем рассуждать, как в § 6, гл. VI. Там у нас рассматривалась матрица

$$U = X (X^T P X)^{-1} X^T P; \quad U^2 = U$$

[см. (6.6.3) и (6.6.5)].

Обозначая $UP^{-1}=X(X^TPX)^{-1}X^T=W$, видим, что $W^T=W$ и что правая часть (7.2.18) записывается в виде $\Delta W\Delta^T$.

Положим $\Delta P^{-\frac{1}{2}}=M$; тогда

$$M^T=\left(P^{-\frac{1}{2}}\right)^T\Delta^T=P^{-\frac{1}{2}}\Delta^T$$

и

$$\Delta W\Delta^T=MP^{\frac{1}{2}}WP^{\frac{1}{2}}M^T. \tag{7.2.19}$$

Матрица $P^{\frac{1}{2}}WP^{\frac{1}{2}}$, очевидно, симметрическая, и ввиду того, что

$$W=UP^{-1},\quad P^{\frac{1}{2}}WP^{\frac{1}{2}}=P^{\frac{1}{2}}UP^{\frac{1}{2}},\ \text{имеем}$$

$$\left(P^{\frac{1}{2}}WP^{\frac{1}{2}}\right)^2=P^{\frac{1}{2}}U^2P^{-\frac{1}{2}}=P^{\frac{1}{2}}UP^{-\frac{1}{2}}=P^{\frac{1}{2}}WP^{\frac{1}{2}}. \tag{7.2.20}$$

Подберем ортогональную матрицу $F=F_{NN}$, такую, что

$$F^TP^{\frac{1}{2}}WP^{\frac{1}{2}}F=D \tag{7.2.21}$$

— диагональная матрица. Тогда, ввиду $F^T=F^{-1}$ и (7.2.20), находим

$$D^2=D.$$

Полагая

$$D=\begin{Vmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & d_N \end{Vmatrix},$$

находим отсюда $d_i^2=d_i$, т. е. $d_i=0$ или $d_i=1$.

Таким образом, $MF\left(F^TP^{\frac{1}{2}}WP^{\frac{1}{2}}\right)FF^TM^T=MFDF^TM^T$; если

$$MF=\|\mu_1, \mu_2, \dots \mu_N\|, \tag{7.2.22}$$

то

$$MFDF^TM^T=\sum_{i=1}^N d_i\mu_i^2, \qquad d_i=0;\ 1. \tag{7.2.23}$$

Далее,

$$\Delta P^{-1}\Delta^T=\Delta P^{-\frac{1}{2}}P^{-\frac{1}{2}}\Delta^T=MM^T=MFF^TM^T=\sum_{i=1}^N \mu_i^2. \tag{7.2.24}$$

Отсюда, очевидно,

$$MFF^TM^T\geqslant MFDF^TM^T,\quad \text{или}\quad \Delta P^{-1}\Delta^T\geqslant\Delta W\Delta^T\ \text{при любых}\ \Delta,$$

что доказывает (7.2.18) и нашу теорему.

Остановимся еще на реальном смысле оценки $\tilde{a}$. Из того, что сказано в § 5 гл. VI, следует, что если $a_1', \ldots, a_n'$ принадлежат к весьма широкому классу асимптотически нормальных и асимптотически несмещенных оценок параметров $a_1, \ldots, a_n$, то линейная форма $[g\tilde{a}] = g_1\tilde{a}_1 + \ldots + g_n\tilde{a}_n$ будет в некотором смысле наивыгоднейшей оценкой $[ga]$ среди оценок вида $a' = g_1a_1' + \ldots + g_na_n'$ (при одних и тех же g), в том смысле, что при заданных $\varepsilon > 0$ и n и при больших N

$$P\{|\tilde{a} - [ga]| \leqslant \varepsilon\} \geqslant P\{|a' - [ga]| \leqslant \varepsilon\}. \qquad (7.2.25)$$

Если мы ограничиваемся лишь линейными несмещенными оценками a вида (7.2.12), то при нормальном векторе погрешностей Δ эти оценки нормальны, и, в силу теоремы 7.2.2, неравенства, аналогичные (7.2.25), будут выполняться для $N > n$, а не только для больших N, т. е. будет

$$P\{|\tilde{a} - a| \leqslant \varepsilon\} \geqslant P\{|[\lambda l] - a| \leqslant \varepsilon\}. \qquad (7.2.26)$$

§ 3. Оценивание линейной формы

Имеющихся у нас сведений о свойствах оценки $\tilde{a}$ достаточно для хорошего точечного оценивания $a = [ga]$. Однако для оценивания по методу доверительных интервалов и оценивания точности работы с $\tilde{a}$ нужно получить оценку для $D(\tilde{a})$.

Теорема 7.3.1. *Несмещенной оценкой для $D(\tilde{a})$ будет*

$$\tilde{S}^2 = -\frac{[p\tilde{v}\tilde{v}]}{N-n}\frac{\det(C_{gg})}{\det(C)} = -\frac{\det(C_{gg})\det(C_{HH})}{(N-n)(\det(C))^2}. \qquad (7.3.1)$$

Заметим, что во втором выражении (7.3.1) случайной является только величина $\det(C_{HH})$.

Доказательство непосредственно вытекает из предыдущих теорем. Из теоремы 7.2.1 [см. (7.2.10)] имеем

$$D(\tilde{a}) = -\sigma^2\frac{\det(C_{gg})}{\det(C)},$$

далее, из (6.9.3)

$$E\left(\frac{[p\tilde{v}\tilde{v}]}{N-n}\right) = \sigma^2, \text{ так что } E(\tilde{S}^2) = -\sigma^2\frac{\det(C_{gg})}{\det(C)} = D(\tilde{a}).$$

Далее, согласно теореме 6.6.7,

$$[p\tilde{v}\tilde{v}] = \frac{\det(C_{HH})}{\det(C)}.$$

Это полностью доказывает два равенства (7.3.1). Теперь переходим к конструкции доверительных интервалов для оценивания $a = [ga]$.

Мы видели в гл. VI, что случайные векторы $\tilde{A}$ и $\tilde{V}$ независимы; в частности, $\tilde{A}$ не зависит от $[p\tilde{v}\tilde{v}]$, и, стало быть, от $\tilde{S}$. Оценка $\tilde{a}=\sum_{i=1}^{n} g_i\tilde{a}_i$ есть линейная функция случайного вектора $\tilde{A}$ и потому статистически не зависит от $\tilde{S}$. По теореме 6.6.4 имеем

$$\frac{[p\tilde{v}\tilde{v}]}{\sigma^2}=\chi^2_{N-n}. \tag{7.3.2}$$

Далее, случайная величина

$$\frac{\tilde{a}-a}{\sigma}\sqrt{-\frac{\det(C)}{\det(C_{gg})}} \tag{7.3.3}$$

будет нормальной $N(0, 1)$. Поэтому имеет место следующая теорема:

Теорема 7.3.2. *Величина*

$$t_{N-n}=\frac{\tilde{a}-a}{\sqrt{-\frac{[p\tilde{v}\tilde{v}]\det(C_{gg})}{(N-n)\det(C)}}}=\frac{\tilde{a}-a}{\tilde{S}} \tag{7.3.4}$$

имеет распределение Стьюдента с $N-n$ степенями свободы.

Доверительный интервал для a строится теперь обычными приемами: по заданной надежности p_0 и числу степеней свободы $N-n$ находим в табл. I такое γ, что $P(|t_{N-n}|\leqslant\gamma)=p_0$. Доверительный интервал

$$I_\gamma=[\tilde{a}-\gamma\tilde{S},\quad \tilde{a}+\gamma\tilde{S}] \tag{7.3.5}$$

будет накрывать a с вероятностью p_0.

Дадим еще одно выражение для $\tilde{S}^2$ и $D(\tilde{a})$.

Теорема 7.3.3. *Если запишем $\tilde{a}$ в виде линейной функции наблюдений $l_1, l_2, \ldots, l_N$,* т. е.

$$G\tilde{A}=GC^{-1}X^TPL=\tilde{a}=[\lambda l]=\Lambda L, \tag{7.3.6}$$

то

$$D(\tilde{a})=\sigma^2\left[\frac{\lambda^2}{p}\right]=\sigma^2\sum_{i=1}^{N}\frac{\lambda_i^2}{p_i} \tag{7.3.7}$$

и

$$\tilde{S}^2=\frac{[p\tilde{v}\tilde{v}]}{N-n}=\left[\frac{\lambda^2}{p}\right]. \tag{7.3.8}$$

Для доказательства заметим, что из (7.3.1) следует

$$D(\tilde{a})=\sigma^2\Lambda P^{-1}\Lambda^T=\sigma^2\left[\frac{\lambda^2}{p}\right]. \tag{7.3.9}$$

Далее, из (7.3.1) в силу несмещенности $\tilde{S}^2$ и (6.9.3)

$$E(\tilde{S}^2)=D(\tilde{a})=\sigma^2\left(-\frac{\det(C_{gg})}{\det(C)}\right);$$

сравнивая это с (7.3.9), находим

$$-\frac{\det(C_{gg})}{\det(C)}=\left[\frac{\lambda^2}{p}\right], \tag{7.3.10}$$

что приводит к (7.3.8).

§ 4. Сводка формул и правила оценивания линейной функции параметров

Обозначения для основных уравнений и параметров здесь будут, как в гл. VI, § 6. Оценивается линейная форма

$$a=g_1a_1+\ldots+g_na_n, \tag{7.4.1}$$

где $g_1, \ldots, g_n$ заданы.

а) Точечное оценивание

$$\tilde{a}=g_1\tilde{a}_1+g_2\tilde{a}_2+\ldots+g_n\tilde{a}_n, \tag{7.4.2}$$

где $\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n$ оцениваются по методу наименьших квадратов. Оценка $\tilde{a}$ имеет также выражение с помощью симметричной матрицы системы нормальных уравнений

$$C=\left\|\begin{matrix} c_{11} & c_{12} & \ldots & c_{1n} \\ c_{21} & c_{22} & \ldots & c_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ c_{n1} & c_{n2} & \ldots & c_{nn} \end{matrix}\right\| \quad (c_{ij}=c_{ji}).$$

Составим величины

$$H_0=\sum_{r=1}^{N} p_r l_r^2,$$

$$H_h=\sum_{r=1}^{N} p_r l_r x_{rh} \qquad (h=1, 2, \ldots, n) \tag{7.4.3}$$

и окаймления матрицы C:

$$C_{HH}=\left\|\begin{matrix} H_0 & H_1 & \ldots & H_n \\ H_1 & c_{11} & \ldots & c_{1n} \\ \cdot & \cdot & \cdot & \cdot \\ H_n & c_{n1} & \ldots & c_{nn} \end{matrix}\right\|,$$

$$C_{gg}=\left\|\begin{matrix} 0 & g_1 & g_2 & \ldots & g_n \\ g_1 & c_{11} & c_{12} & \ldots & c_{1n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ g_n & c_{n1} & c_{n2} & \ldots & c_{nn} \end{matrix}\right\|,$$

$$C_{gH}=\left\|\begin{matrix} 0 & g_1 & g_2 & \ldots & g_n \\ H_1 & c_{11} & c_{12} & \ldots & c_{1n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ H_n & c_{n1} & c_{n2} & \ldots & c_{nn} \end{matrix}\right\|.$$

Тогда

$$\tilde{a} = -\frac{\det(C_{gH})}{\det(C)}. \tag{7.4.4}$$

Если оценки $\tilde{a}_1, \ldots, \tilde{a}_n$ по методу наименьших квадратов уже получены, то в вычислительном отношении выгодно использовать формулу (7.4.2).

Далее,

$$D(\tilde{a}) = -\sigma^2 \frac{\det(C_{gg})}{\det(C)} = \sum_{h,\, k=1}^{n} g_h g_k \{C^{-1}\}_{hk}. \tag{7.4.5}$$

Несмещенной оценкой для $D(\tilde{a})$ служит оценка

$$\tilde{S}^2 = -\frac{[p\tilde{v}\tilde{v}]}{N-n} \cdot \frac{\det(C_{gg})}{\det(C)} =$$

$$= -\frac{\det(C_{gg})\det(C_{HH})}{(N-n)(\det(C))^2} = -\frac{[p\tilde{v}\tilde{v}]}{N-n}\left[\frac{\lambda^2}{p}\right], \tag{7.4.6}$$

где λ_i таковы, что $\tilde{a} = [\lambda l]$.

Указанная формула годна и для случая, когда погрешности не нормальны, но обладают следующими свойствами:

$$E(\Delta_i) = 0; \quad D(\Delta_i) = \frac{\sigma^2}{p_i}.$$

Приближенная оценка штандарта $\sigma(\tilde{S})$ оценки $\tilde{S}$ при больших $(N-n)$ такова:

$$\sigma(\tilde{S}) \approx \sqrt{-\frac{\det(C_{gg})}{2\det(C)(N-n)}}. \tag{7.4.7}$$

б) Оценивание с помощью доверительных интервалов

Выбирается надежность p_0, имеющаяся в табл. I распределения Стьюдента. Вторым входом служит число степеней свободы $N-n$. По этим двум входам находим γ_0. Доверительный интервал

$$I_\gamma = [\tilde{a} - \gamma_0 \tilde{S},\ \tilde{a} + \gamma_0 \tilde{S}] \tag{7.4.8}$$

накрывает a с вероятностью p_0.

§ 5. Частные случаи, встречающиеся на практике. Задача о линейной регрессии

Пусть X — случайная величина, а Y — статистически связанная с ней случайная величина, так что (Y, X) есть случайный вектор. Часто встречаются случаи, когда Y имеет линейную регрессию на X, т. е. при заданном X

$$E(Y/X) = a_1 + a_2 X, \tag{7.5.1}$$

а условная дисперсия Y при заданном X постоянна. При этом a_1, a_2 неизвестны. На основании независимых наблюдений значений

$y_1, y_2, \ldots, y_N$, отвечающих $x_1, x_2, \ldots, x_N$, надо произвести оценивание

$$a = a_1 + a_2\alpha, \tag{7.5.2}$$

где α заданное число *). Мы будем иметь дело с практическими примерами, когда такая ситуация встречается. Если предполагать вектор (Y, X) нормальным, так что наблюдения $y_1, \ldots, y_N$ независимы и нормальны, то оценивание $a = a_1 + a_2\alpha$ по методу наименьших квадратов будет иметь реальный смысл, указанный в конце § 5 гл. VI. Если этого не предполагать, а допускать лишь наличие дисперсии y_i, то вся изложенная ранее рецептура будет годной, но полученные оценки будут хороши лишь в том смысле, что будут иметь минимальную дисперсию в том или ином классе несмещенных оценок, т. е., по существу, будут хороши в том же смысле наименьших квадратов, и, кроме того, будут получаться несложным в вычислительном отношении способом.

В настоящей задаче мы имеем ситуацию гл. VI, но прежние случайные величины l_i имеют здесь обозначения y_i. У нас $p_i = 1$, $g_1 = 1$, $g_2 = \alpha$, $x_{r1} = 1$, $x_{r2} = x_r$, так что

$$E(y_r) = a_1x_{r1} + a_2x_{r2} = a_1 + a_2x_r; \quad y_r = \Delta_r + E(y_r),$$

$$X = \begin{Vmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_N \end{Vmatrix}; \qquad X^T = \begin{Vmatrix} 1 & \ldots & 1 \\ x_1 & \ldots & x_N \end{Vmatrix};$$

$$C = X^TPX = X^TX = \begin{Vmatrix} N & [x] \\ [x] & [x^2] \end{Vmatrix},$$

$$\det(C) = N[x^2] - [x]^2 \tag{7.5.3}$$

(полезно сравнить с примером 1 во введении),

$$H_1 = \sum_{r=1}^{N} y_r = [y]; \qquad H_2 = \sum_{r=1}^{N} y_rx_r = [xy];$$

$$H_0 = \sum_{r=1}^{N} y_r^2 = [y^2];$$

$$C_{gg} = \begin{Vmatrix} 0 & 1 & \alpha \\ 1 & N & [x] \\ \alpha & [x] & [x^2] \end{Vmatrix}; \qquad C_{HH} = \begin{Vmatrix} [y^2] & [y] & [xy] \\ [y] & N & [x] \\ [xy] & [x] & [x^2] \end{Vmatrix};$$

$$C_{gH} = \begin{Vmatrix} 0 & 1 & \alpha \\ [y] & N & [x] \\ [xy] & [x] & [x^2] \end{Vmatrix}.$$

---

*) Иногда встречается случай, когда $\alpha = E(X)$. Из элементарного курса теории вероятностей известно, что

$$E(Y) = E(E(Y/X)) = a_1 + a_2E(X) = a_1 + a_2\alpha,$$

так что тогда ставится задача оценивания $E(Y)$.

Для вычисления

$$\tilde{a} = -\frac{\det(C_{gH})}{\det(C)}$$

находим $\det(C_{gH})$, разлагая этот определитель по элементам первой строки:

$$\det(C_{gH}) = [x][xy] - [x^2][y] + \alpha([x][y] - n[xy]).$$

Отсюда

$$\tilde{a} = \frac{[x^2][y] - [x][xy] + \alpha(N[xy] - [x][y])}{N[x^2] - [x]^2}. \tag{7.5.4}$$

Заметим, что при этих вычислениях несущественно, что $\alpha = E(X)$; α могло бы быть любым заданным числом.

Придадим выражению для $\tilde{a}$ вид, естественный в теории корреляции. Положим (см. § 1 введения) $\bar{x} = \frac{[x]}{N}$, $\bar{y} = \frac{[y]}{N}$ и

$$r_{xy} = \frac{[(x-\bar{x})(y-\bar{y})]}{\sqrt{[(x-\bar{x})^2]\,[(y-\bar{y})^2]}}; \tag{7.5.5}$$

величина r_{xy} называется выборочным коэффициентом корреляции между X и Y; имеем всегда $|r_{xy}| \leqslant 1$. Далее,

$$\begin{aligned} [(x-\bar{x})^2] &= [x^2] - N\bar{x}^2 = [x^2] - \frac{[x]^2}{N}, \\ [(y-\bar{y})^2] &= [y^2] - N\bar{y}^2 = [y^2] - \frac{[y]^2}{N}, \end{aligned} \tag{7.5.6}$$

$$[(x-\bar{x})(y-\bar{y})] = [xy] - \frac{1}{N}[x]\,[y]. \tag{7.5.7}$$

Положим теперь

$$s_x^2 = \frac{1}{N}[(x-\bar{x})^2]; \qquad s_y^2 = \frac{1}{N}[(y-\bar{y})^2]. \tag{7.5.8}$$

Отсюда находим

$$\tilde{a} = (\alpha - \bar{x})\,r_{xy}\frac{s_y}{s_x} + \frac{[x^2][y] - [x][xy] + \bar{x}(N[xy] - [x][y])}{N \cdot s_x^2}.$$

Второе слагаемое преобразуется в виде

$$\frac{[x^2]\bar{y} - \bar{x}[xy] + \bar{x}[xy] - \frac{1}{N}[x]^2\bar{y}}{s_x^2} = \bar{y},$$

так что получаем

$$\tilde{a} = \bar{y} + r_{xy}\frac{s_y}{s_x}(\alpha - \bar{x}), \tag{7.5.9}$$

[ср. (0.1.10) введения]. Далее, несмещенной оценкой для $D(\tilde{a})$, согласно (7.3.1), является оценка

$$\tilde{S}^2 = -\frac{\det(C_{gg})\det(C_{HH})}{(N-2)(\det(C))^2}. \tag{7.5.10}$$

Мы приведем здесь лишь окончательное выражение для $\tilde{S}^2$ (см. Ф. Дэвид [15], гл. XIII, стр. 160—178):

$$\tilde{S}^2 = \frac{s_y^2\left(1-r_{xy}^2\right)}{N-2}\left[1+\frac{(\alpha-\bar{x})^2}{s_x^2}\right]. \tag{7.5.11}$$

Если выбор абсцисс x_i в наших руках, т. е. если мы их можем назначить заранее, то из формулы (7.5.11) видно, что для уменьшения второго члена выгодно выбрать систему значений $x_1, \ldots, x_n$ так, чтобы $|\alpha-\bar{x}|$ было возможно меньше, а $s_x^2 \neq 0$ было возможно больше, т. е. расставлять x_i по возможности ближе к концам участка наблюдений x и так, чтобы их среднее арифметическое было возможно ближе к α (совпадало с α, если это возможно). При оценке коэффициентов линейной зависимости a_1 и a_2 порознь по методу наименьших квадратов также выгодно, чтобы s_x^2 было возможно больше, для чего можно делить наблюдения на две группы, расположенные вблизи краев участка значений x.

Общий вопрос о наивыгоднейшем распределении абсцисс x_i для применения метода наименьших квадратов, в случае если эти абсциссы в наших руках, пока мало изучен.

Пользуясь (7.5.11), можем записать в явной форме доверительный интервал для оценивания $a = a_1 + a_2\alpha$:

$$I_\gamma = [\tilde{a} - \gamma\tilde{S}, \quad \tilde{a} + \gamma\tilde{S}],$$

где

$$\tilde{a} = \bar{y} + r_{xy}\frac{s_y}{s_x}(\alpha - \bar{x});$$

$$\tilde{S} = s_y\sqrt{\frac{1-r_{xy}^2}{N-2}}\cdot\sqrt{1+\frac{(\alpha-\bar{x})^2}{s_x^2}}.$$

Обратимся еще к задачам множественной корреляции. Пусть имеется случайный вектор (X, Y, Z), причем при заданных X, Y условное математическое ожидание Z есть линейная функция X, Y:

$$E(Z/X, Y) = A + BX + CY, \tag{7.5.12}$$

где A, B, C неизвестны.

Условную дисперсию Z при фиксированных X и Y будем считать постоянной, но не известной нам. Нам будут заданы наблюдения x_r, y_r, z_r, причем

$$E(Z/x_r, y_r) = A + Bx_r + Cy_r,$$
$$z_r = \Delta_r + E(Z/x_r, y_r);$$

величины Δ_r независимы и имеют неизвестные дисперсии σ^2. Таким образом,

$$z_r + \Delta_r = A + Bx_r + Cy_r \qquad (r = 1, 2, \ldots, N), \tag{7.5.13}$$

и мы имеем обычную ситуацию метода наименьших квадратов. Задача состоит в точечном оценивании линейной формы $a = A + B\xi + C\eta$ с заданными ξ, η. Мы приведем без вывода явные формулы такого оценивания. В частности, полагая ξ, $\eta = 0$ и пользуясь симметрией формул, получим явные формулы для точечного оценивания A, B, C при равноточных наблюдениях.

Если погрешности Δ_r нормальны [например, если вектор (X, Y, Z) предполагается нормальным], то получаемые оценки будут иметь реальный смысл, указанный ранее; если же это не предполагается, можно лишь утверждать, что они будут хороши в отношении сравнительной малости их дисперсий.

Здесь

$$g_1 = 1, \quad g_2 = \xi, \quad g_3 = \eta; \quad x_{r1} = 1, \quad x_{r2} = x_r, \quad x_{r3} = y,$$
$$H_0 = [z^2]; \quad H_1 = [z]; \quad H_2 = [xz]; \quad H_3 = [yz];$$
$$X = \begin{Vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & y_N \end{Vmatrix}; \qquad X^T = \begin{Vmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_N \\ y_1 & y_2 & \ldots & y_N \end{Vmatrix};$$
$$C = X^T X = \begin{Vmatrix} N & [x] & [y] \\ [x] & [x^2] & [xy] \\ [y] & [xy] & [y^2] \end{Vmatrix}; \tag{7.5.14}$$
$$C_{gH} = \begin{Vmatrix} 0 & 1 & \xi & \eta \\ [z] & N & [x] & [y] \\ [xz] & [x] & [x^2] & [xy] \\ [yz] & [y] & [xy] & [y^2] \end{Vmatrix}. \tag{7.5.15}$$

Для величины a имеем оценку

$$\tilde{a} = -\frac{\det(C_{gH})}{\det(C)}. \tag{7.5.16}$$

Вводятся частные коэффициенты корреляции r_{xy}, r_{xz}, r_{yz} совершенно аналогично формулам (7.5.5), а также величины

$$s_x^2 = \frac{1}{N}[(x - \bar{x})^2]; \quad s_y^2 = \frac{1}{N}[(y - \bar{y})^2]; \quad s_z^2 = \frac{1}{N}[(z - \bar{z})^2].$$

После элементарных, но длинных выкладок из формулы (7.5.16) получается

$$\tilde{a} = \bar{z} + (\xi - \bar{x})\frac{s_z}{s_x}\left[\frac{r_{xz} - r_{xy}r_{yz}}{1 - r_{xy}^2}\right] + (\eta - \bar{y})\frac{s_z}{s_y}\left[\frac{r_{yz} - r_{xy}r_{xz}}{1 - r_{xy}^2}\right]. \tag{7.5.17}$$

В частности, при $\xi=\eta=0$ получаем оценку для A:

$$\tilde{A}=\bar{z}-\bar{x}\frac{s_z}{s_x}\left[\frac{r_{xz}-r_{xy}r_{yz}}{1-r_{xy}^2}\right]-\bar{y}\frac{s_z}{s_y}\left[\frac{r_{yz}-r_{xy}r_{xz}}{1-r_{xy}^2}\right], \tag{7.5.18}$$

Для получения $\tilde{B}$ достаточно взять $\eta=0$, получить оценку для $a=A+B\xi$, разделить ее на ξ и устремить ξ к ∞, т. е. найти $\lim\limits_{\xi\to\infty}\frac{\tilde{A}}{\xi}$. Тогда из (7.5.13) получим

$$\tilde{B}=\frac{s_z}{s_x}\left(\frac{r_{xz}-r_{xy}r_{yz}}{1-r_{xy}^2}\right); \tag{7.5.19}$$

аналогично при $\xi=0$, $\eta\to\infty$ получим

$$\tilde{C}=\frac{s_z}{s_y}\left(\frac{r_{yz}-r_{xy}r_{xz}}{1-r_{xy}^2}\right). \tag{7.5.20}$$

Далее,

$$\tilde{S}^2=-\frac{\det(C_{gg})\det(C_{HH})}{(N-3)(\det(C))^2};$$

после некоторых вычислений получаем

$$\tilde{S}^2=\frac{s_z^2}{(N-3)\left(1-r_{xy}^2\right)}\left(1-r_{xy}^2-r_{yz}^2-r_{xz}^2+2r_{xy}r_{yz}r_{xz}\right)\times$$
$$\times\left[1+\frac{1}{1-r_{xy}^2}\left(\left(\frac{\xi-\bar{x}}{s_x}\right)^2-2r_{xy}\left(\frac{\xi-\bar{x}}{s_x}\right)\left(\frac{\eta-\bar{y}}{s_y}\right)+\left(\frac{\eta-\bar{y}}{s_y}\right)^2\right)\right]. \tag{7.5.21}$$

Для несмещенных оценок дисперсий $\tilde{S}_A^2$, $\tilde{S}_B^2$, $\tilde{S}_C^2$ отсюда можно получить равенства

$$\left.\begin{aligned}\tilde{S}_A^2&=\frac{s_z^2}{(N-3)\left(1-r_{xy}^2\right)}R,\\ \tilde{S}_B^2&=\frac{s_z^2}{(N-3)\,s_x^2\left(1-r_{xy}^2\right)}R,\\ \tilde{S}_C^2&=\frac{s_z^2}{(N-3)\,s_y^2\left(1-r_{xy}^2\right)}R,\end{aligned}\right\} \tag{7.5.22}$$

где

$$R=1-r_{xy}^2-r_{yz}^2-r_{xz}^2+2r_{xy}r_{yz}r_{xz}. \tag{7.5.23}$$

§ 6. Примеры

Пример 1. В качестве первого примера возьмем данные Д. И. Менделеева из введения (стр. 9).

Прямая

$$y = 67{,}5 + 0{,}87x \tag{0.1.11}$$

была проведена по методу наименьших квадратов. Температура $x = 32°$ не входила в наблюдавшиеся значения; точечная оценка растворимости $NaNO_3$ при $x = 32°$ будет

$$\tilde{y} = 67{,}5 + 0{,}87 \cdot 32 = 95{,}3.$$

Построим доверительный интервал, накрывающий изучаемую нами растворимость при 32° с надежностью $p = 0{,}90$. Для этого найдем оценку $\tilde{S}$ по формуле

$$\tilde{S} = s_y \sqrt{\frac{1 - r_{xy}^2}{N-2}\left(1 + \frac{(\alpha - \bar{x})^2}{s_x^2}\right)}.$$

В нашем случае $N = 9$, $[x] = 234$, $[y] = 811{,}3$, $[x^2] = 10\,144$, $[y^2] = 76\,218{,}2$, $[xy] = 24\,628{,}6$.

Значит,

$$s_x^2 = \frac{1}{N}\left([x^2] - \frac{[x]^2}{N}\right) = \frac{1}{9}(10\,144 - 6084) = 451{,}1; \quad s_x = 21{,}24;$$

$$s_y^2 = \frac{1}{N}\left([y^2] - \frac{[y]^2}{N}\right) = 342{,}7; \quad s_y = 18{,}51;$$

$$r_{xy} = \frac{[xy] - \frac{[x][y]}{N}}{N s_x s_y} = 0{,}9990; \quad r_{xy}^2 = 0{,}9980.$$

Отсюда при $\alpha = 32\ (\bar{x} = 26)$ находим

$$\tilde{S} = 18{,}51 \sqrt{\frac{0{,}0020}{7}\left(1 + \frac{36}{451{,}1}\right)} = 0{,}33.$$

По табл. 1 t-распределения для $p = 0{,}90$ и $k = N - 2 = 7$ находим $\gamma = 1{,}895$.

Значит, доверительным интервалом для y при надежности 0,9 будет интервал $[\tilde{y} - 1{,}895\tilde{S};\ \tilde{y} + 1{,}895\tilde{S}]$, или $[94{,}6;\ 96{,}0]$.

Пример 2. В качестве второго примера возьмем пример 2 введения.

Допущения линейности корреляции скорости и пройденного расстояния, а также нормальности распределения пройденного расстояния при заданной скорости нуждаются в проверке. Такая проверка, вообще говоря, выходит из рамок метода наименьших квадратов и изучается в математической статистике. Мы не будем ею заниматься.

Допуская наличие указанных свойств для данных примера 2 введения, оценим значение y_i пройденного расстояния при скорости $x_i = 36$ *км/час* и построим доверительный интервал для y_i при надежности $p = 0{,}90$.

По формуле (0.1.13): $y = -5{,}362 + 0{,}7454\,x$, находим точечную оценку для y_i при $x_i = 36$. Имеем

$$\tilde{y}_i = -5{,}362 + 0{,}7454 \cdot 36 = 21{,}47.$$

Строим доверительный интервал для y_i. В нашем примере $N = 50$, $[x] = 1238{,}98$, $[y] = 655{,}43$, $[x^2] = 34\,248{,}4$, $[y^2] = 11\,618{,}5$, $[xy] = 18\,885{,}2$. Значит,

$$s_x^2 = \frac{[x^2] - \frac{[x]^2}{N}}{N} = \frac{34\,248{,}4 - 30\,701{,}4}{50} = 70{,}94; \quad s_x = 8{,}423,$$

аналогично $s_y^2 = 60{,}53$, $s_y = 7{,}780$, далее

$$r_{xy} = \frac{[xy] - \frac{[x]\,[y]}{50}}{50 \cdot s_x \cdot s_y} = 0{,}8069, \quad r_{xy}^2 = 0{,}6511.$$

Отсюда при $\alpha = 36\ (\bar{x} = 24{,}78)$

$$\tilde{S} = 7{,}78 \sqrt{\frac{0{,}3489}{48}\left(1 + \frac{(11{,}22)^2}{70{,}94}\right)} = 1{,}105.$$

По табл. 1 t-распределения находим для $p = 0{,}90$ и $k = N - 2 = 48$ значение $\gamma = 1{,}678$.

Значит, доверительным интервалом для y при $x = 36$ будет интервал $[\tilde{y} - 1{,}678\,\tilde{S};\ \tilde{y} + 1{,}678\,\tilde{S}]$, или $[19{,}61;\ 23{,}32]$.

Пример 3 *). При пробной стрельбе из орудия при возрастающей скорости ветра, перпендикулярного направлению стрельбы, получены значения бокового отклонения в 10 выстрелах, производившихся через равные промежутки времени τ (табл. 25) **):

Таблица 25

Момент времени t_i	Значения бокового отклонения y_i	Момент времени t_i	Значения бокового отклонения y_i
0	57,2	5	59,8
1	58,0	6	60,4
2	58,1	7	60,0
3	59,1	8	60,0
4	59,3	9	62,2

*) Данные примера заимствованы из книги [2], стр. 236.
**) Принимаем для простоты τ за единицу времени.

Предполагая боковое отклонение линейной функцией времени $y = a + bt$, оценим параметры a и b. Произведем также оценивание y_i в моменты $t = 2{,}5$; $6{,}5$; 10.

Находим оценки для a и b методом наименьших квадратов по формулам (0.1.9), переписанным в виде

$$b = \frac{[ty] - \frac{[t][y]}{N}}{[t^2] - \frac{[t]^2}{N}} \quad \text{и} \quad a = \bar{y} - b\bar{t}.$$

Вычисляем: $[t] = 45$, $[y] = 594{,}1$, $[t^2] = 285$, $[y^2] = 35\,313{,}79$, $[ty] = 2709{,}9$, $\bar{t} = 4{,}5$, $\bar{y} = 59{,}41$.

Значит, $b = 0{,}441$, $a = 57{,}4$, и зависимость бокового отклонения от времени выражается уравнением

$$y = 57{,}4 + 0{,}441\,t.$$

Находим точечные оценки для бокового отклонения (y_1, y_2, y_3) в моменты $t = 2{,}5$; $6{,}5$; 10. Они равны $\tilde{y}_1 = 58{,}5$, $\tilde{y}_2 = 60{,}3$, $\tilde{y}_3 = 61{,}8$.

Построим доверительные интервалы для y_1, y_2, y_3. Для этого вычисляем $s_t^2 = 8{,}250$, $s_t = 2{,}872$, $s_y^2 = 1{,}831$, $s_y = 1{,}353$, $r_{ty} = 0{,}9380$ и $r_{ty}^2 = 0{,}8798$. Оценки $\tilde{S}$, определяемые по формуле

$$\tilde{S} = s_y \sqrt{\frac{1 - r_{ty}^2}{N - 2}\left(1 + \frac{(\alpha - \bar{x})^2}{s_t^2}\right)}$$

для трех случаев ($\alpha_1 = 2{,}5$, $\alpha_2 = 6{,}5$, $\alpha_3 = 10$), соответственно равны $\tilde{S}_1 = 0{,}2021$, $\tilde{S}_2 = 0{,}2021$, $\tilde{S}_3 = 0{,}3583$.

Возьмем надежность $p = 0{,}90$. По табл. 1 t-распределения для $p = 0{,}90$ и $k = N - 2 = 8$ находим $\gamma = 1{,}860$.

Значит, доверительными интервалами для y_1, y_2, y_3 будут соответственно интервалы $[\tilde{y}_i - \gamma\tilde{S}_i;\ \tilde{y}_i + \gamma\tilde{S}_i]$, $i = 1, 2, 3$, или $[58{,}1;\ 58{,}9]$, $[59{,}9;\ 60{,}7]$, $[61{,}1;\ 62{,}5]$.

Пример 4 *). В табл. 26 помещены результаты измерения скоростей y (в *км/сек*) и расстояний x (в миллионах парсек) для десяти внегалактических туманностей.

Таблица 26

x	y	x	y
1,20	630	9,12	4 820
1,82	890	10,97	5 230
3,31	2 350	14,45	7 500
7,24	3 810	22,91	11 800
8,92	4 630	36,31	19 600

*) Данные примера заимствованы из книги [2], стр. 237.

Предполагая y линейной функцией от $x\,(y=a+bx)$, оценим параметры a и b, а также произведем оценивание y при $x=2;\ 8;\ 37$.

Оценки для a и b находим методом наименьших квадратов по формулам (0.1.9). Вычисляем $[x]=116{,}25$, $[y]=61\,260$, $[x^2]=2403{,}2945$, $[y^2]=672\,899\,800$, $[xy]=1\,270\,758{,}80$, $\bar{x}=11{,}625$, $\bar{y}=6126$. Значит, $b=531{,}06$, $a=-47{,}6$, и уравнение линейной регрессии y на x запишется в виде

$$y=-47{,}6+531{,}06x.$$

Точечными оценками для скоростей y_i $(i=1,\,2,\,3)$ при $x=2;\ 8$ и 37 будут соответственно

$$\tilde{y}_1=1014{,}5,\quad \tilde{y}_2=4\,200{,}9,\quad \tilde{y}_3=19\,601{,}6.$$

Найдем доверительные интервалы для y_i. Вычисляем $s_x^2=105{,}19$, $s_x=10{,}256$, $s_y^2=29\,762{,}104$, $s_y=5455{,}5$, $r_{xy}=0{,}99838$, $r_{xy}^2=0{,}99676$.

Значит, оценки $\tilde{S}$ для $\alpha_1=2$, $\alpha_2=8$, $\alpha_3=37$ соответственно равны $\tilde{S}_1=106$, $\tilde{S}_2=82{,}4$, $\tilde{S}_3=207$.

Возьмем надежность $p=0{,}90$. Для этой надежности и числа степеней свободы $k=N-2=8$ находим по таблице распределения Стьюдента $\gamma=1{,}860$.

Доверительными интервалами для y_1, y_2, y_3 будут интервалы $[\tilde{y}_i-\gamma\tilde{S}_i;\ \tilde{y}_i+\gamma\tilde{S}_i]$, $i=1,\,2,\,3$, или [817; 1212], [4047; 4354], [19 216; 19 987].

Пример 5. В табл. 27 приведены результаты экспериментального исследования количества тепла, выделяющегося при отвердевании портланд-цемента *). Полученные данные устанавливают связь между химическим составом клинкеров, из которых изготовлен кирпич, и выделяющимся теплом; x и y выражают соответственно содержание в клинкерах алюмината кальция $3CaO\cdot Al_2O_3$ и силиката кальция $3CaO\cdot SiO_2$ (в процентах от веса клинкеров), z — количество выделившегося тепла в калориях на грамм цемента.

Таблица 27

x	y	z
7	26	78,5
1	29	74,3
11	56	104,3
11	31	87,6
7	52	95,9
11	55	109,2
3	71	102,7
1	31	72,5
2	54	93,1
21	47	115,9
1	40	83,8
11	66	113,3
Сумма 87	558	1131,1

*) Данные заимствованы из книги [48], стр. 543.

Предполагая, что зависимость z от x и y линейна, произведем точечное оценивание z при $x = 10$, $y = 68$ и сравним полученный результат с наблюденным при этих значениях x и y (равным 109,4; см. последнее наблюдение в [48] стр. 543).

Произведем точечное оценивание z по формуле (7.5.17):

$$\tilde{z} = \bar{z} + (\xi - \bar{x}) \frac{s_z}{s_x} \left[\frac{r_{xz} - r_{xy} r_{yz}}{1 - r_{xy}^2} \right] + (\eta - \bar{y}) \frac{s_z}{s_y} \left[\frac{r_{yz} - r_{xy} r_{xz}}{1 - r_{xy}^2} \right].$$

Вычисляя входящие в формулу величины, получим

$$\begin{array}{lll} \bar{x} = 7{,}25; & \bar{y} = 46{,}50; & \bar{z} = 94{,}26; \\ s_x^2 = 34{,}02; & s_y^2 = 206{,}6; & s_z^2 = 208{,}7; \\ s_x = 5{,}833; & s_y = 14{,}37; & s_z = 14{,}45; \\ r_{xy} = 0{,}1954; & r_{xz} = 0{,}7292; & r_{yz} = 0{,}7996; \\ r_{xy}^2 = 0{,}0382. & & \end{array}$$

Полагая $\xi = 10$, $\eta = 68$ и подставляя найденные значения величин в формулу для $\tilde{z}$, находим точечную оценку величины z при $x = 10$, $y = 68$. Она равна $\tilde{z} = 113{,}09$.

ГЛАВА VIII

НЕПРЯМЫЕ (КОСВЕННЫЕ) УСЛОВНЫЕ ИЗМЕРЕНИЯ (УРАВНИВАНИЕ ПО ЭЛЕМЕНТАМ)

§ 1. Постановка задачи

Во введении мы встречались с такими задачами измерения величин, когда на оценки, даваемые нами величинам, естественно наложить некоторые условия. Типичным там являлся пример 4 измерения углов треугольника; более общая постановка давалась формулами (0.1.31) и (0.1.32). Здесь мы уточним эту постановку.

Пусть N измеряемых величин $\lambda_1, \lambda_2, \ldots, \lambda_N$ должны удовлетворять $q < N$ соотношениям, называемым *фундаментальными уравнениями*

$$F_j(\lambda_1, \lambda_2, \ldots, \lambda_N) = 0 \qquad (j = 1, 2, \ldots, q), \qquad (8.1.1)$$

где $F_j(\lambda_1, \ldots, \lambda_N)$ — функции, разлагающиеся в ряды Тейлора.

Пусть производятся измерения величин $\lambda_1, \ldots, \lambda_N$, имеющие случайные погрешности и дающие результаты $l_1, l_2, \ldots, l_N$. По внесении поправок $\delta_1, \delta_2, \ldots, \delta_N$ к этим результатам уравнения (8.1.1) должны точно удовлетворяться, так что $l_i + \delta_i = \lambda_i$ $(i = 1, 2, \ldots, N)$;

$$F_j(l_1 + \delta_1, l_2 + \delta_2, \ldots, l_N + \delta_N) = 0, \quad j = 1, 2, \ldots, q. \quad (8.1.2)$$

Числа $\delta_1, \delta_2, \ldots, \delta_N$ суть истинные поправки к наблюдениям, которых мы не знаем. Мы заменяем неизвестные нам величины λ_i на оценки $\tilde{l}_1, \tilde{l}_2, \ldots, \tilde{l}_N$ так, чтобы удовлетворялись фундаментальные уравнения

$$F_j(\tilde{l}_1, \tilde{l}_2, \ldots, \tilde{l}_N) = 0 \qquad (j = 1, 2, \ldots, q). \qquad (8.1.3)$$

Такая операция часто называется *уравниванием.* Для того чтобы отыскать нужные оценки, снова начнем с метода максимального правдоподобия.

Пусть величины $l_1, l_2, \ldots, l_N$ независимы, нормальны, и $l_i \in N\left(\lambda_i, \frac{\sigma}{\sqrt{p_i}}\right)$, так что $l_i = \lambda_i + \Delta_i$; $D(\Delta_i) = \frac{\sigma^2}{p_i}$; $p_i (i=1, 2, \ldots, N)$ — веса.

Составим функцию правдоподобия выборки $(l_1, l_2, \ldots, l_N)$:

$$L = (2\pi)^{-\frac{N}{2}} \sigma^{-N} p_1 p_2 \ldots p_N \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N} p_i (l_i - \lambda_i)^2\right\}. \quad (8.1.4)$$

Согласно принципу максимального правдоподобия, для оценивания величин $\lambda_1, \lambda_2, \ldots, \lambda_N$ надлежит взять вместо $\lambda_1, \lambda_2, \ldots, \lambda_N$ величины $\tilde{l}_1, \tilde{l}_2, \ldots, \tilde{l}_N$, доставляющие функции L условный максимум при условиях (8.1.3). При любом значении σ^2 такое предписание равносильно следующему:

$$\left.\begin{aligned} &\sum_{i=1}^{N} p_i\left(\tilde{l}_i - l_i\right)^2 = \min, \\ &F_j\left(\tilde{l}_1, \tilde{l}_2, \ldots, \tilde{l}_N\right) = 0, \quad j = 1, 2, \ldots, q \end{aligned}\right\} \quad (8.1.5)$$

т. е. предписанию наименьших квадратов, где берется условный минимум.

Заметим теперь, что фундаментальные уравнения (8.1.1) в дальнейшем будут предполагаться линейными, т. е. имеющими вид

$$f_{j0} + \lambda_1 f_{j1} + \lambda_2 f_{j2} + \ldots + \lambda_N f_{jN} = 0 \qquad (j = 1, 2, \ldots, q). \quad (8.1.6)$$

Основания к этому следующие. В тех случаях, когда будет возникать разбираемая задача (непрямые условные измерения) — задача уравнивания в геодезии, астрономии и других науках, — измеряемые величины, например углы в геодезии, будут в первом приближении известны. Мы будем знать приближение λ_i^0 к величинам λ_i и считать, что истинные погрешности $\Delta\lambda_i = \lambda_i - \lambda_i^0$ таковы, что нельзя пренебречь ими, но можно пренебречь их квадратами $(\Delta\lambda_i)^2$ сравнительно с выбранной единицей длины. Мы будем иметь в дальнейшем практические примеры такой ситуации. В таком случае, припоминая, что функции $F_j(\lambda_1, \ldots, \lambda_N)$ разложимы в ряд Тейлора и производя такое разложение в окрестности точки $\left(\lambda_1^0, \ldots, \lambda_N^0\right)$, найдем

$$F_j(\lambda_1, \ldots, \lambda_N) = F_j\left(\lambda_1^0, \lambda_2^0, \ldots, \lambda_N^0\right) + \Delta\lambda_1\left(\frac{\partial F_j}{\partial\lambda_1}\right)_0 + \Delta\lambda_2\left(\frac{\partial F_j}{\partial\lambda_2}\right)_0 +$$
$$+ \ldots + \Delta\lambda_N\left(\frac{\partial F_j}{\partial\lambda_N}\right)_0 + \text{величины порядка малости } (\Delta_1^2 + \Delta_2^2 + \ldots + \Delta_N^2).$$

Здесь частные производные $\left(\frac{\partial F_j}{\partial\lambda_i}\right)_0$ берутся в точке $(\lambda_1^0, \ldots, \lambda_N^0)$ и являются известными постоянными числами, как и $F_j\left(\lambda_1^0, \ldots, \lambda_N^0\right)$.

Основная идея дальнейшего заключается в том, что, поскольку числа $\lambda_1^0, \lambda_2^0, \ldots, \lambda_N^0$ нам известны, в качестве измеряемых величин можем взять не $\lambda_1, \ldots, \lambda_N$, а истинные поправки $\Delta\lambda_1, \Delta\lambda_2, \ldots, \Delta\lambda_N$.

Обозначая

$$F_j(\lambda_1^0, \ldots, \lambda_N^0) = f_{j0}, \left(\frac{\partial F_j}{\partial \lambda_i}\right)_0 = f_{ji} \qquad (i = 1, 2, \ldots, N),$$

придем вместо (8.1.1) к линейным фундаментальным уравнениям вида (8.1.6), где $\Delta\lambda_1, \ldots, \Delta\lambda_N$ для сокращения обозначены через $\lambda_1, \ldots, \lambda_N$. Эти уравнения верны с точностью до пренебрегаемого порядка малости $(\Delta_1^2 + \ldots + \Delta_N^2)$.

Из q уравнений (8.1.6) можно произвести исключение некоторых из чисел $\lambda_1, \lambda_2, \ldots, \lambda_N$, выразив их линейным образом через остальные. Число оставшихся „свободных" параметров будет определяться рангом матрицы $F = F_{qN}$ и будет равняться $m =$ ранг (F). Обозначим оставшиеся параметры $\xi_1, \xi_2, \ldots, \xi_n$, получим уравнения в параметрической форме

$$\lambda_r = \xi_1 x_{r1} + \xi_2 x_{r2} + \ldots + \xi_n x_{rn} \qquad (r = 1, 2, \ldots, N), \quad (8.1.7)$$

где $x_{rs}(r = 1, 2, \ldots, N, \ s = 1, 2, \ldots, n)$ — известные числа.

Если сделать над $\xi_1, \ldots, \xi_n$ неособенное линейное преобразование, переводящее эти параметры в $\xi_1', \ldots, \xi_n'$, то снова получим линейные выражения для λ_r типа (8.1.7) с известными коэффициентами. Поэтому в уравнениях (8.1.7) можно считать $\xi_1, \ldots, \xi_n$ параметрами, не обязательно совпадающими с какими-либо из λ_i. Уравнения (8.1.7) называются в этом случае *фундаментальными уравнениями в параметрической форме*, а параметры $\xi_1, \xi_2, \ldots, \xi_n$ — *элементами*. Оценивание $\lambda_1, \ldots, \lambda_N$ в этом случае называется *уравниванием с помощью элементов* (в противоположность уравниванию с помощью коррелят, которое будет рассмотрено позже).

Если мы переименуем $\xi_1, \xi_2, \ldots, \xi_n$ в $a_1, a_2, \ldots, a_n$ и λ_r — в y_r, то получим уравнения, близкие по форме и по существу к уравнениям (6.1.1) гл. VI:

$$y_r = x_{r0} + \sum_{j=1}^{n} a_j x_{rj} \qquad (r = 1, 2, \ldots, N; \ N > n), \quad (8.1.8)$$

где вместо y_r имеются данные $l_r = y_r + \Delta_r$, а погрешности Δ_r можно считать независимыми, случайными и нормальными $N\left(0, \frac{\sigma}{\sqrt{p_r}}\right)$. Отличие от уравнений (6.1.1) заключается в наличии „свободных членов" x_{r0}. Чтобы привести уравнения (8.1.8) к полному сходству с (6.1.1), запишем их в виде

$$y_r - x_{r0} = \sum_{j=1}^{n} a_j x_{rj}. \quad (8.1.9)$$

Здесь x_{r0} — известные числа. Наблюдениям l_r сопоставляем вспомогательные числа $l_r - x_{r0}$, которые можно считать значениями независимых случайных величин, нормальных $N\left(y_r - x_{r0}, \frac{\sigma}{\sqrt{p_r}}\right)$. Правда, в данном случае основной упор делается на оценивание левых частей $y_1, \ldots, y_N$, а не элементов $a_1, \ldots, a_n$, как в гл. VI, но ввиду линейных зависимостей (8.1.8) соответствующие оценки должны быть тесно связанными.

Точная постановка задачи уравнивания с помощью элементов будет — найти несмещенные оценки $\tilde{a}_1, \tilde{a}_2, \ldots, \tilde{a}_n$ и $\tilde{l}_1, \tilde{l}_2, \ldots, \tilde{l}_N$ так, чтобы удовлетворялись уравнения

$$\tilde{l}_r = x_{r0} + \sum_{j=1}^{n} \tilde{a}_j x_{rj} \qquad (r = 1, 2, \ldots, N), \tag{8.1.10}$$

которые называются *условными*. Эти оценки должны обладать возможно более хорошими качествами в отношении совместной эффективности.

§ 2. Уравнивание с помощью элементов по методу наименьших квадратов

Как было сказано ранее, постановка задачи уравнивания с помощью элементов, по существу, совпадает с постановкой задачи гл. VI. Если взять за оценки $\tilde{a}_1, \ldots, \tilde{a}_n$ оценки по методу наименьших квадратов, то они будут совместно эффективными. Реальная выгода оценивания по методу наименьших квадратов поясняется в § 5 гл. VI. Оценки же $\tilde{l}_1, \ldots, \tilde{l}_N$ для $l_1, \ldots, l_N$ получаются просто по формуле (8.1.9) при известных $\tilde{a}_1, \ldots, \tilde{a}_n$.

Для вычисления нужных оценок достаточно воспользоваться результатами гл. VI, заменив там числа y_r на числа $y_r - x_{r0}$. Выпишем получаемые таким образом результаты.

Введем матрицу

$$X_0 = \begin{Vmatrix} x_{10} \\ x_{20} \\ \vdots \\ x_{N0} \end{Vmatrix}.$$

Фундаментальные уравнения (8.1.8) запишутся в виде

$$Y - X_0 = X \cdot A, \tag{8.2.1}$$

где, как и в гл. VI,

$$X = X_{Nn} = \begin{Vmatrix} x_{11} & \ldots & x_{1n} \\ x_{21} & \ldots & x_{2n} \\ \cdot & \cdot\ \cdot\ \cdot & \cdot \\ x_{N1} & \ldots & x_{Nn} \end{Vmatrix}.$$

Мы считаем, как и в гл. VI, что ранг $(X)=n$. Далее полагаем (в обозначениях гл. VI)

$$\tilde{V}=XA+X_0-L=X\tilde{A}-(L-X_0), \tag{8.2.2}$$

где $\tilde{A}$ выбрана так, что

$$[p\tilde{v}\tilde{v}]=\min. \tag{8.2.3}$$

Все расчеты проходят так же, как в гл. VI, с заменой L на $L-X_0$.

Нормальные уравнения имеют вид

$$C\tilde{A}=X^TP(L-X_0); \quad C=X^TPX. \tag{8.2.4}$$

Далее, $\tilde{A}$ — несмещенные оценки для A, имеющие корреляционную матрицу

$$B_{\tilde{A}}=\sigma^2C^{-1}. \tag{8.2.5}$$

Оценками для $y_1, \ldots, y_N$ считаем величины $\tilde{l}_1, \ldots, \tilde{l}_N$, где

$$\tilde{l}_r=x_{r0}+\sum_{j=1}^{n}\tilde{a}_jx_{rj} \qquad (r=1, 2, \ldots, N), \tag{8.2.6}$$

т. е.

$$\tilde{L}=X_0+X\tilde{A}. \tag{8.2.7}$$

Оценки $\tilde{L}$ будут несмещенными оценками для Y, ибо

$$E(\tilde{L})=X_0+XE(\tilde{A})=X_0+XA=Y.$$

Далее, согласно § 7 гл. VI (здесь роль $\tilde{L}$ должна играть $\tilde{L}-X_0$), корреляционная матрица $B_{\tilde{L}}$ для $\tilde{L}$ такова:

$$B_{\tilde{L}}=\sigma^2XC^{-1}X^T. \tag{8.2.8}$$

Совершенно аналогично теореме 6.7.1, имеем и здесь соотношение

$$\frac{1}{\sigma^2}\sum_{r=1}^{N}p_rD(\tilde{l}_r)=\frac{1}{\sigma^2}[pD(\tilde{l})]=\sum\{XC^{-1}X^T\}_{rr}=n, \tag{8.2.9}$$

которое может быть употреблено для контроля правильности вычислений.

Кроме того, мы видим, что среднее взвешенное дисперсии оценок будет пропорционально $\frac{n}{N}$, так что отношение числа наблюдений к числу элементов выгодно иметь возможно бо́льшим.

Выведем еще формулы для вычисления $[p\tilde{v}\tilde{v}]$ или контроля таких вычислений. Обозначим $M=L-X_0$.

Теорема 8.2.1. *Имеют место формулы*

$$[p\tilde{v}\tilde{v}] = M^T PM - M^T PX\tilde{A} = [pmm] - [px_1 m]\tilde{a}_1 - \ldots - [px_n l]\tilde{a}_n, \tag{8.2.10}$$

$$[p\tilde{v}\tilde{v}] = M^T PM - \tilde{A}^T C\tilde{A}, \tag{8.2.11}$$

$$[p\tilde{v}\tilde{v}] = L^T PL - \tilde{L}^T P\tilde{L} + 2X_0^T P\tilde{V}. \tag{8.2.12}$$

Формула (8.2.10) получается непосредственно из (6.6.19) заменой L на $M = L - X_0$; (8.2.11) следует таким же образом из (6.6.18). Для вывода (8.2.12) запишем

$$\begin{aligned}[p\tilde{v}\tilde{v}] &= M^T PM - \tilde{A}^T C\tilde{A} = M^T PM - \tilde{A}^T X^T PX\tilde{A} = \\ &= M^T PM - (\tilde{L} - X_0)^T P(\tilde{L} - X_0) = \\ &= L^T PL - L^T PX_0 - X_0^T PL - \tilde{L}^T P\tilde{L} + X_0^T P\tilde{L} + \tilde{L}^T PX_0.\end{aligned}$$

Далее, $L^T PX_0$ есть число, т. е. матрица первого порядка; стало быть,

$$(L^T PX_0)^T = X_0^T PL = L^T PX_0;$$

$$(\tilde{L}^T PX_0)^T = X_0^T P\tilde{L} = \tilde{L}^T PX_0,$$

так что

$$\begin{aligned}[p\tilde{v}\tilde{v}] &= L^T PL - \tilde{L}^T P\tilde{L} - 2X_0^T PL + 2X_0^T P\tilde{L} = \\ &= L^T PL - \tilde{L}^T P\tilde{L} + 2X_0^T P(\tilde{L} - L).\end{aligned}$$

Далее,

$$\tilde{L} = X_0 + X\tilde{A}; \quad \tilde{V} = X_0 + X\tilde{A} - L = \tilde{L} - L,$$

так что

$$[p\tilde{v}\tilde{v}] = L^T PL - \tilde{L}^T P\tilde{L} + 2X_0^T P\tilde{V},$$

что и требовалось доказать.

Доверительные интервалы для оценивания величин $a_1, a_2, \ldots, a_n$ строятся так же, как в гл. VI, с заменой L на $L - X_0$ в оценках $\tilde{A}$. При этом

$$\tilde{A} = C^{-1} X^T P(L - X_0), \tag{8.2.13}$$

$$I_\gamma^{(i)} = \left[\tilde{a}_i - \gamma\sqrt{\{C^{-1}\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}},\quad \tilde{a}_i + \gamma\sqrt{\{C^{-1}\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}}\right]. \tag{8.2.14}$$

Для оценивания y_i $(i = 1, 2, \ldots, N)$ получаются доверительные интервалы

$$J_\gamma^{(i)} = \left[\tilde{l}_i - \gamma\sqrt{\{XC^{-1}X^T\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}},\quad \tilde{l}_i + \gamma\sqrt{\{XC^{-1}X^T\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}}\right]. \tag{8.2.15}$$

Длина доверительного интервала $J_\gamma^{(i)}$ является случайной величиной

$$d_\gamma^{(i)} = 2\gamma s \sqrt{\{XC^{-1}X^T\}_{ii}}.$$

Как и в гл. VI, можем рассмотреть величину

$$\frac{1}{N} E\left(p_i \sum_{i=1}^{N} (d_\gamma^{(i)})^2\right) = \frac{4\gamma^2\sigma^2 n}{N}. \tag{8.2.16}$$

При $N - n \geqslant 20$ можно считать

$$\frac{4\gamma^2\sigma^2 n}{N} \approx \frac{3{,}24\sigma^2 n}{N}.$$

Для оценивания точности измерений в качестве точечной оценки σ употребляем

$$\tilde{s} = \sqrt{\frac{[p\tilde{v}\tilde{v}]}{N-n}}. \tag{8.2.17}$$

Доверительный интервал для σ берем в виде $[\gamma_1\tilde{s}, \gamma_2\tilde{s}]$, как описано в § 9 гл. VI.

§ 3. Правила уравнивания по элементам

Фундаментальные уравнения:

$$y_r = x_{r0} + \sum_{j=1}^{n} a_j x_{rj} \quad (r = 1, 2, \ldots, N, \ N > n). \tag{8.3.1}$$

Для точечного оценивания a_j и y_r составляются нормальные уравнения: при $x_j = (x_{1j}, x_{2j}, \ldots, x_{Nj})$

$$\sum_{\nu=1}^{n} [px_i x_\nu]\, \tilde{a}_\nu = [px_i(l - x_0)]. \tag{8.3.2}$$

Решая нормальные уравнения (8.3.2), получаем оценки $\tilde{a}_1, \ldots, \tilde{a}_n$. Оценки для y_r получаются по формулам

$$\tilde{y}_r = x_{r0} + \sum_{j=1}^{n} \tilde{a}_j x_{rj} \qquad (r = 1, 2, \ldots, N, \ N > n). \tag{8.3.3}$$

Далее, надлежит вычислить $[p\tilde{v}\tilde{v}]$, где

$$\tilde{v}_r = \sum_{i=1}^{n} \tilde{a}_i x_{ri} + x_{r0} - l_r. \tag{8.3.4}$$

Для вычисления $[p\tilde{v}\tilde{v}]$ или контроля таких вычислений можно использовать формулы

$$[p\tilde{v}\tilde{v}] = [pmm] - [px_1m]\,\tilde{a}_1 - \ldots - [px_nm]\,\tilde{a}_n, \tag{8.3.5}$$

где $m = (m_1, \ldots, m_N)$; $m_r = l_r - x_{r0}$;

$$[p\tilde{v}\tilde{v}] = [pmm] - \sum_{h,\,k=1}^{n} [px_hx_k]\,\tilde{a}_h\tilde{a}_k. \tag{8.3.6}$$

Если положим

$$\tilde{l}_r = x_{r0} + \sum_{j=1}^{n} \tilde{a}_j x_{rj}, \tag{8.3.7}$$

то

$$[p\tilde{v}\tilde{v}] = [pll] - [p\tilde{l}\,\tilde{l}\,] - 2\,[p\tilde{v}x_0]. \tag{8.3.8}$$

Для оценивания точности наблюдений составляется величина

$$\tilde{s} = \sqrt{\frac{[p\tilde{v}\,\tilde{v}]}{N-n}}, \tag{8.3.9}$$

которая служит оценкой для σ. Эта оценка — асимптотически несмещенная относительно σ при возрастании $N-n$. Ее штандарт $\sigma(\tilde{s}) \approx \frac{\sigma}{2(N-n)}$ при больших $N-n$. Он оценивается приближенно через $\frac{\tilde{s}}{N-n}$.

Для оценивания штандартов $\sigma(\tilde{a}_i)$ оценок $\tilde{a}_i$ надо найти диагональные элементы матрицы, обратной к матрице C нормальных уравнений (8.3.2), т. е. числа $\{C^{-1}\}_{ii}$. После этого $\sigma(\tilde{a}_i)$ оценивается величиной

$$\tilde{s}\sqrt{\{C^{-1}\}_{ii}}. \tag{8.3.10}$$

Далее, для оценок l_r параметров y_r надо составить матрицу

$$X^TC^{-1}X, \quad \text{где} \quad X = \begin{Vmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{21} & \ldots & x_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ x_{N1} & x_{N2} & \ldots & x_{Nn} \end{Vmatrix},$$

в которой нам понадобятся лишь диагональные элементы $\{X^TC^{-1}X\}_{rr}$. Тогда

$$\sigma(\tilde{l}_r) = \sigma\sqrt{\{X^TC^{-1}X\}_{rr}}. \tag{8.3.11}$$

Приближенной оценкой $\sigma(\tilde{l}_r)$ при больших $N-n$ будет величина

$$\tilde{s}\sqrt{\{X^TC^{-1}X\}_{rr}}. \tag{8.3.12}$$

Для контроля вычислений может служить формула

$$\frac{1}{\sigma^2}[pD(\tilde{l})] = \sum_{r=1}^{N} \{XC^{-1}X^T\}_{rr} = n. \qquad (8.3.13)$$

Отсюда видно также, что точность наблюдений при заданных весах и σ^2 будет зависеть от отношения $\frac{N}{n}$, которое выгодно иметь возможно бо́льшим.

Оценивание с помощью доверительных интервалов

Используем табл. I распределения Стьюдента. Выбираем надежность p_0 из имеющихся в таблице, вторым входом будет служить число степеней свободы $N-n$. По этим двум входам находим число γ. Доверительный интервал

$$I_\gamma^{(i)} = \left[\tilde{a}_i - \gamma\sqrt{\{C^{-1}\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}}, \quad \tilde{a}_i + \gamma\sqrt{\{C^{-1}\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}}\right] \qquad (8.3.14)$$

будет накрывать a_i с вероятностью p_0.

Для оценивания y_r составляем доверительный интервал

$$J_\gamma^{(r)} = \left[\tilde{l}_r - \gamma_0\sqrt{(X^TC^{-1}X)_{rr}\frac{[p\tilde{v}\tilde{v}]}{N-n}}, \ \tilde{l}_r + \gamma_0\sqrt{(X^TC^{-1}X)_{rr}\frac{[p\tilde{v}\tilde{v}]}{N-n}}\right], \qquad (8.3.15)$$

накрывающий y_r с вероятностью p_0.

Для оценивания σ с помощью доверительных интервалов употребляем табл. III приложений. По заданной надежности p_0 находим числа γ_1 и γ_2. Доверительный интервал

$$I = \left[\gamma_1\sqrt{\frac{[p\tilde{v}\tilde{v}]}{N-n}}, \quad \gamma_2\sqrt{\frac{[p\tilde{v}\tilde{v}]}{N-n}}\right] \qquad (8.3.16)$$

накрывает σ с вероятностью p_0.

Примеры применения этого метода будут даны совместно с примерами по применению метода коррелят в гл. IX.

ГЛАВА IX

УРАВНИВАНИЕ С ПОМОЩЬЮ КОРРЕЛЯТ

§ 1. Постановка задачи

Метод уравнивания с помощью коррелят отличается от метода уравнивания с помощью элементов лишь в аналитическом отношении; математико-статистическая постановка и смысл метода остаются прежними.

Мы должны произвести оценивание параметров $y_1, \ldots, y_N$, на которые накладывается $q\,(q < N)$ связей

$$x_{j0} + y_1 x_{j1} + y_2 x_{j2} + \ldots + y_N x_{jN} = 0 \quad (j = 1, 2, \ldots, q). \qquad (9.1.1)$$

Мы считаем эти связи линейными по соображениям, изложенным в гл. VIII; переход от нелинейных связей к линейным с небольшой погрешностью будет показан на конкретных примерах.

Нам надлежит найти оценки $\tilde{l}_1, \ldots, \tilde{l}_N$ для $y_1, \ldots, y_N$ на основании наблюдений $l_1, l_2, \ldots, l_N$, которые мы будем считать имеющими случайные погрешности $\Delta_1, \ldots, \Delta_N$, так что l_r будут независимыми случайными величинами, нормальными $N\left(y_r, \frac{\sigma}{\sqrt{p_r}}\right)$. Согласно гл. VIII мы будем искать оценки $\tilde{l}_1, \ldots, \tilde{l}_N$ по предписанию

$$\left.\begin{aligned} &\sum_{i=1}^{N} p_i \left(\tilde{l}_i - l_i\right)^2 = \left[p\left(\tilde{l} - l\right)^2\right] = \min \\ &x_{j0} + \tilde{l}_1 x_{j1} + \tilde{l}_2 x_{j2} + \ldots + \tilde{l}_N x_{jN} = 0 \quad (j = 1, 2, \ldots, q). \end{aligned}\right\} \qquad (9.1.2)$$

Полученные оценки $\tilde{l}_1, \ldots, \tilde{l}_N$ должны совпадать с оценками, найденными с помощью элементов и обозначенными так же в гл. VIII. Именно, там тоже отыскивались оценки $\tilde{l}_1, \ldots, \tilde{l}_N$ под условием $\left[p\left(\tilde{l} - l\right)^2\right] = \min$, причем $\tilde{l}_i$ удовлетворяют соотношениям (8.1.10), вполне равносильным линейным связям в (9.1.2). Соотношения (8.1.10) так и выбирались, чтобы они были равносильны этим связям, т. е. чтобы всякий набор чисел $\tilde{l}_1, \ldots, \tilde{l}_N$, изображаемый в виде (8.1.10) с какими-либо параметрами $\tilde{a}_1, \ldots, \tilde{a}_n$, удовлетворял бы q линейным связям в (9.1.2), и обратно, всякий набор чисел $\tilde{l}_1, \ldots, \tilde{l}_N$, удовлетворяющий линейным связям в (9.1.2) [и не обязательно удовлетворяющий первому из соотношений (9.1.2)], мог быть изображен в виде (8.1.10) для каких-либо значений параметров $\tilde{a}_1, \ldots, \tilde{a}_n$ (не обязательно вычисляемых нами оценок). Таким образом, при нахождении условного минимума в (9.1.2) мы получим старые значения оценок. Однако мы не будем это делать старым способом — путем параметрической записи линейных соотношений (9.1.2) с помощью элементов, так как это потребовало бы операции исключения переменных $\tilde{l}_r$ из уравнений (9.1.2), а рассмотрим процесс вычисления оценок $\tilde{l}_r$, т. е. нахождения условного минимума $\left[p\left(\tilde{l}-l\right)^2\right]$, непосредственно используя линейные связи (условные уравнения) (9.1.1). Этот процесс и составляет суть метода коррелят. Он выполняется по методу Лагранжа для нахождения условного минимума.

Напомним формулировку метода Лагранжа. Пусть $\Phi(z_1, \ldots, z_N)$ — всюду дважды дифференцируемая функция; $F_1(z_1, \ldots, z_N), \ldots, F_q(z_1, \ldots, z_N)$, $(q < N)$ — функции с тем же свойством. Надлежит найти значения $z_1, \ldots, z_N$, для которых

$$\Phi(z_1, \ldots, z_N) = \min \quad \text{и} \quad F_j(z_1, \ldots, z_N) = 0 \qquad (j = 1, 2, \ldots, q).$$

Для нахождения таких значений составляем вспомогательную функцию

$$\Psi(z_1, \ldots, z_N) = \Phi - k_1F_1 - k_2F_2 - \ldots - k_qF_q,$$

где $k_1, \ldots, k_q$ — неопределенные постоянные множители Лагранжа. Далее, приравниваем к нулю полный дифференциал Ψ: $d\Psi = 0$, т. е.

$$\frac{\partial\Psi}{\partial z_1} = \frac{\partial\Psi}{\partial z_2} = \ldots = \frac{\partial\Psi}{\partial z_N} = 0.$$

К полученной системе уравнений присоединяем q условных уравнений $F_j(z_1, \ldots, z_N) = 0$ $(j = 1, 2, \ldots, q)$, что дает $N + q$ уравнений для определения $z_1, z_2, \ldots, z_N$ и $k_1, k_2, \ldots, k_q$. Среди решений $(z_1, \ldots z_N)$ этих уравнений и надо отыскивать нужные нам решения.

§ 2. Вычисление оценок с помощью коррелят

Будем вести расчеты в матричной записи. Для записи условных уравнений (9.1.1) введем матрицы:

$$X_0 = \begin{Vmatrix} x_{10} \\ x_{20} \\ \vdots \\ x_{q0} \end{Vmatrix}; \quad X = X_{qN} = \begin{Vmatrix} x_{11} & x_{12} & \ldots & x_{1N} \\ \cdot & \cdot & \cdot & \cdot \\ x_{q1} & x_{q2} & \ldots & x_{qN} \end{Vmatrix} \tag{9.2.1}$$

Матрицы L и Y будут иметь тот же смысл, что и в гл. VIII. Условные уравнения (9.1.1) запишутся в виде

$$X_0 + XY = 0. \tag{9.2.2}$$

Мы будем считать, что ранг $(X) = q$, в противном случае в условных уравнениях будут зависимости, которые можно исключить, введя новую сокращенную систему условных уравнений.

Неизвестные нам поправки к L для получения Y обозначим $V = V_{N1}$;

$$L + V = Y. \tag{9.2.3}$$

Положим еще

$$X_0 + XL = M. \tag{9.2.4}$$

Тогда, подставляя (9.2.3) в (9.2.2), найдем

$$M + XV = 0. \tag{9.2.5}$$

Это будут q условных линейных уравнений для истинных поправок $v_1, v_2, \ldots, v_N$. При этих условиях надлежит выбрать такое $V = \tilde{V}$, чтобы было $[p\tilde{v}\tilde{v}] = \min$; мы докажем, что это V существует и единственно.

Действуя по методу Лагранжа, мы должны ввести q множителей $k_1, k_2, \ldots, k_q$, называемых коррелятами. Одностолбцовую матрицу коррелят обозначим $K = K_{q1}$;

$$K = K_{q1} = \begin{Vmatrix} k_1 \\ k_2 \\ \vdots \\ k_q \end{Vmatrix}. \tag{9.2.6}$$

Тогда функция Лагранжа будет иметь вид

$$\Psi = [pvv] - K^T(M + XV) = V^T PV - K^T(M + XV). \tag{9.2.7}$$

Составим полный дифференциал Ψ:

$$d\Psi = d\,[pvv] - d\,[K^T(M + XV)]. \tag{9.2.8}$$

Имеем, очевидно,

$$d\,[pvv] = 2\,[pv\,dv].$$

Обозначим

$$dV = \begin{Vmatrix} dv_1 \\ dv_2 \\ \vdots \\ dv_N \end{Vmatrix}. \tag{9.2.9}$$

Легко проверить, что

$$2\,[pv\,dv] = 2V^T P\,dV. \tag{9.2.10}$$

Далее, легко находим

$$d\,[K^T(M + XV)] = dK^T M + dK^T XV = K^T X\,dV. \tag{9.2.11}$$

При составлении функции Лагранжа Ψ удвоим неопределенные множители k_i для удобства записи.

Получим

$$d\Psi = 2V^T P\,dV - 2K^T X\,dV = 2(V^T P - K^T X)\,dV = 0 \quad (9.2.12)$$

и

$$V^T P - K^T X = 0. \quad (9.2.13)$$

Или, более подробно,

$$(V_{N1})^T P_{NN} - (K_{q1})^T X_{qN} = 0_{1N}. \quad (9.2.14)$$

Это дает N уравнений с неизвестными V и K ($N+q$ неизвестных). К ним присоединяются условные уравнения

$$M + XV = 0, \quad \text{или} \quad M_{q1} + X_{qN} V_{N1} = 0_{q1}, \quad (9.2.15)$$

что дает еще q уравнений относительно V. Покажем, что эти уравнения однозначно разрешимы. Имеем из (19.2.13)

$$V^T P = K^T X.$$

Транспонируя, находим

$$PV = X^T K, \quad (9.2.16)$$

$$V = P^{-1} X^T K. \quad (9.2.17)$$

Подставляя в (9.2.15), получаем

$$M + XP^{-1}X^T K = 0. \quad (9.2.18)$$

Матрица $XP^{-1}X^T$ будет симметрической $q \times q$ матрицей. Ввиду того, что ранг $(X) = q$ и P^{-1} — диагональная матрица ранга N (обратная матрица весов), она будет неособенной. Это следует из теоремы 1.3.1.

Обозначим

$$XP^{-1}X^T = G = G_{qq}; \quad (9.2.19)$$

тогда G — неособенная, G^{-1} — существует, и из (9.2.18) следует

$$K = -G^{-1}M. \quad (9.2.20)$$

Таким образом, K определяется однозначно, и из (9.2.17) получаем явное выражение для V:

$$V = \tilde{V} = -P^{-1}X^T G^{-1} M; \quad G = XP^{-1}X^T. \quad (9.2.21)$$

V определяется однозначно, так определенное V мы и обозначим $\tilde{V}$ и возьмем в качестве поправок к L.

Равенство (9.2.20), записанное в виде

$$GK + M = 0,$$

дает нормальную систему уравнений для q коррелят с симметрической матрицей $G = XP^{-1}X^T$. Запишем ее подробнее.

Имеем

$$M = X_0 + XL = \begin{Vmatrix} m_1 \\ \vdots \\ m_q \end{Vmatrix}.$$

Так как

$$X = \begin{Vmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ \dots & \dots & \dots & \dots \\ x_{q1} & x_{q2} & \dots & x_{qN} \end{Vmatrix}; \qquad X_0 = \begin{Vmatrix} x_{10} \\ x_{20} \\ \vdots \\ x_{q0} \end{Vmatrix},$$

то

$$m_i = x_{i0} + x_{i1}l_1 + x_{i2}l_2 + \dots + x_{iN}l_N \qquad (i = 1, 2, \dots, q). \quad (9.2.22)$$

Введя q полученных из строк X векторов,

$$x_i = (x_{i1}, x_{i2}, \dots, x_{iN}) \qquad (i = 1, 2, \dots, q), \quad (9.2.23)$$

можем написать

$$m_i = x_{i0} + [x_i l] \qquad (i = 1, 2, \dots, q). \quad (9.2.24)$$

С помощью векторов x_i запишем $XP^{-1}X^T$. Имеем

$$X^T = \begin{Vmatrix} x_{11} & x_{21} & \dots & x_{q1} \\ x_{12} & x_{22} & \dots & x_{q2} \\ \dots & \dots & \dots & \dots \\ x_{1N} & x_{2N} & \dots & x_{qN} \end{Vmatrix};$$

$$P^{-1}X^T = \begin{Vmatrix} \frac{x_{11}}{p_1} & \frac{x_{21}}{p_1} & \dots & \frac{x_{q1}}{p_1} \\ \frac{x_{12}}{p_2} & \frac{x_{22}}{p_2} & \dots & \frac{x_{q2}}{p_2} \\ \dots & \dots & \dots & \dots \\ \frac{x_{1N}}{p_N} & \frac{x_{2N}}{p_N} & \dots & \frac{x_{qN}}{p_N} \end{Vmatrix};$$

$$G = XP^{-1}X^T = \begin{Vmatrix} \left[\frac{x_1x_1}{p}\right] & \left[\frac{x_1x_2}{p}\right] & \dots & \left[\frac{x_1x_q}{p}\right] \\ \left[\frac{x_2x_1}{p}\right] & \left[\frac{x_2x_2}{p}\right] & \dots & \left[\frac{x_2x_q}{p}\right] \\ \dots & \dots & \dots & \dots \\ \left[\frac{x_qx_1}{p}\right] & \left[\frac{x_qx_2}{p}\right] & \dots & \left[\frac{x_qx_q}{p}\right] \end{Vmatrix}.$$

Поэтому нормальные уравнения для коррелят (9.2.18) запишутся в виде

$$\left[\frac{x_\nu x_1}{p}\right] k_1 + \left[\frac{x_\nu x_2}{p}\right] k_2 + \ldots + \left[\frac{x_\nu x_q}{p}\right] k_q + [x_\nu l] + x_{\nu 0} = 0 \qquad (9.2.25)$$
$$(\nu = 1, 2, \ldots, q).$$

В таком виде эти уравнения получены еще Гауссом. Найдя корреляты $k_1, \ldots, k_q$ из нормальных уравнений (9.2.25), мы можем найти кажущиеся поправки из (9.2.17): $\tilde{V} = P^{-1}X^T K$, или, подробнее,

$$\tilde{v}_r = \frac{1}{p_r}(x_{1r}k_1 + x_{2r}k_2 + \ldots + x_{qr}k_q). \qquad (9.2.26)$$

Оценки для Y имеют вид

$$\tilde{L} = L + \tilde{V}. \qquad (9.2.27)$$

Как было пояснено выше, эти оценки $\tilde{L}$ совпадают с полученными с помощью элементов, и потому $\tilde{V}$ действительно доставляет условный минимум $[p\tilde{v}\tilde{v}]$, а оценки $\tilde{L}$ будут обладать несмещенностью и другими выгодными свойствами. Мы выведем, однако, некоторые из этих свойств новым способом, чтобы иметь обоснование уравнивания по коррелятам, независимое от теории уравнивания с помощью элементов.

§ 3. Доказательство минимальности

Теорема 9.3.1. *Найденные по формуле* (9.2.21) *кажущиеся поправки*

$$\tilde{V} = -P^{-1}X^T G^{-1} M \qquad (9.3.1)$$

дают единственный минимум $[pvv]$ *при условии*

$$M + X\tilde{V} = 0. \qquad (9.3.2)$$

Доказательство. Имеем

$$\begin{aligned}\tilde{V}^T P \tilde{V} &= M^T (G^{-1})^T X P^{-1} P P^{-1} X^T G^{-1} M = \\ &= M^T (G^{-1})^T X P^{-1} X^T G^{-1} M = M^T (G^{-1})^T M = \\ &= M^T (G^T)^{-1} M = M^T (X P^{-1} X^T)^{-1} M \end{aligned} \qquad (9.3.3)$$

в силу (9.2.21). Далее, в силу (9.3.2)

$$\tilde{V}^T P \tilde{V} = \tilde{V}^T X^T (X P^{-1} X^T)^{-1} X \tilde{V}. \qquad (9.3.4)$$

Какова бы ни была матрица $V = V_{N1}$, удовлетворяющая условию

$$M + XV = 0, \qquad (9.3.5)$$

будем иметь

$$V^T X^T (X P^{-1} X^T)^{-1} X V = M^T (X P^{-1} X^T)^{-1} M. \qquad (9.3.6)$$

Докажем теперь, что для любой матрицы $V = V_{N1}$

$$V^T P V \geqslant V^T X^T (XP^{-1} X^T) XV. \tag{9.3.7}$$

Если это неравенство будет доказано, то для матрицы V, удовлетворяющей (9.3.5), в силу (9.3.6) будем иметь:

$$V^T P V \geqslant M^T (XP^{-1} X^T)^{-1} M = \tilde{V}^T P \tilde{V}, \tag{9.3.8}$$

чем и будет доказано желаемое.

Неравенство (9.3.7) доказывается совершенно аналогично неравенству (7.2.18).

Вводим матрицу

$$U = X^T (XP^{-1} X^T)^{-1} XP^{-1}, \tag{9.3.9}$$

так что

$$U^2 = U. \tag{9.3.10}$$

Полагая $UP = X^T (XP^{-1} X^T)^{-1} X = W$, видим, что $W^T = W$ и что правая часть (9.3.7) записывается в виде

$$V^T W V.$$

Положим $P^{\frac{1}{2}} V = Z$, $V^T P^{\frac{1}{2}} = Z^T$, так что

$$V^T P V = Z^T Z \tag{9.3.11}$$

и

$$V^T W V = Z^T P^{-\frac{1}{2}} W P^{-\frac{1}{2}} Z. \tag{9.3.12}$$

Поскольку W—симметрическая матрица, то и $P^{-\frac{1}{2}} W P^{-\frac{1}{2}}$ —симметрическая. Так как $W = UP$, то $P^{-\frac{1}{2}} W P^{-\frac{1}{2}} = P^{-\frac{1}{2}} U P^{\frac{1}{2}}$, откуда в силу (9.3.10) имеем

$$\left(P^{-\frac{1}{2}} W P^{-\frac{1}{2}}\right)^2 = P^{-\frac{1}{2}} W P^{-\frac{1}{2}}. \tag{9.3.13}$$

Подберем ортогональную матрицу $F = F_{qq}$ такую, что

$$F\left(P^{-\frac{1}{2}} W P^{-\frac{1}{2}}\right) F^T = D = D_{qq} \tag{9.3.14}$$

диагональная матрица. Так как $F^T = F^{-1}$, из (9.3.10) находим: $D^2 = D$. Полагая

$$D = \begin{Vmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ \cdot & \cdot & \cdot \\ 0 & 0 & d_q \end{Vmatrix}, \tag{9.3.15}$$

находим $d_i^2 = d_i$, т. е. $d_i = 0;\ 1$ $(i = 1,\ 2,\ \ldots,\ q)$. Отсюда

$$Z^T F^T F \left(P^{-\frac{1}{2}} W P^{-\frac{1}{2}}\right) F^T F Z = Z^T F^T D F Z. \qquad (9.3.16)$$

Если положим $FZ = \begin{Vmatrix} \mu_1 \\ \mu_2 \\ \cdot \\ \cdot \\ \cdot \\ \mu_q \end{Vmatrix}$,

то

$$Z^T F^T D F Z = \sum_{i=1}^{q} d_i \mu_i^2; \quad d_i = 0;\ 1. \qquad (9.3.17)$$

Отсюда, очевидно,

$$Z^T F^T F Z = \sum_{i=1}^{q} \mu_i^2 \geqslant \sum_{i=1}^{q} d_i \mu_i^2 = Z^T F^T D F Z.$$

Или, в силу (9.3.9) и определения W и Z,

$$V^T P V \geqslant V^T X^T \left(X P^{-1} X^T\right) X V,$$

что совпадает с (9.3.7) и доказывает теорему. Докажем теперь несмещенность оценок $\tilde{L}$. Имеем

$$\tilde{V} = -P^{-1} X^T G^{-1} M; \quad M = X_0 + XL.$$

Отсюда

$$E(\tilde{V}) = -P^{-1} X^T G^{-1} \left(X_0 + XE(L)\right) = \\ = -P^{-1} X^T G^{-1} (X_0 + XY) = 0$$

силу (9.2.2).

Значит,

$$E\tilde{V} = 0, \text{ и } E\tilde{L} = E\left(L + \tilde{V}\right) = EL = Y, \qquad (9.3.18)$$

что и требовалось доказать.

Укажем еще явное выражение $\tilde{L}$ через L.

Теорема 9.3.2. *Имеет место равенство*

$$\tilde{L} = -P^{-1} X^T G^{-1} X_0 + \left(E - P^{-1} X^T G^{-1} X\right) L. \qquad (9.3.19)$$

Для доказательства замечаем, что

$$\tilde{L} = L + \tilde{V} = L - P^{-1} X^T G^{-1} M.$$

Подставляя выражение $M = X_0 + XL$, получаем (9.3.19).

§ 4. Статистическое поведение коррелят и оценок

Выражение (9.2.20) для матрицы коррелят

$$K = -G^{-1}M$$

после подстановки выражения для M принимает вид

$$K = -G^{-1}X_0 - G^{-1}XL = -G^{-1}(X_0 + XL), \tag{9.4.1}$$

откуда видно, что $K = \begin{Vmatrix} k_1 \\ k_2 \\ \cdot \\ \cdot \\ k_q \end{Vmatrix}$ представляет собой случайный нормальный вектор. Его свойства указываются следующей теоремой.

Теорема 9.4.1. *Нормальный случайный вектор коррелят*

$$K = \begin{Vmatrix} k_1 \\ \cdot \\ \cdot \\ k_q \end{Vmatrix}$$

n-мерен, имеет нулевой вектор средних и корреляционную матрицу

$$B_K = \sigma^2 G^{-1}; \quad G = XP^{-1}X^T. \tag{9.4.2}$$

Для доказательства заметим, что

$$K = -G^{-1}M; \quad M = X_0 + XL; \quad EM = X_0 + XY = 0,$$

так что

$$EK = -G^{-1}EM = 0.$$

Далее,

$$K = -G^{-1}(X_0 + XL) = -G^{-1}X_0 - G^{-1}XL. \tag{9.4.3}$$

Корреляционная матрица K по теореме 2.3.1 имеет вид

$$B_K = -G^{-1}XB_LX^TG^{-1} = G^{-1}X\sigma^2P^{-1}X^TG^{-1} = \\ = \sigma^2G^{-1}GG^{-1} = \sigma^2G^{-1}, \tag{9.4.4}$$

что и требовалось доказать.

Найдем теперь корреляционную матрицу $B_{\tilde{L}}$ вектора оценок $\tilde{L}$.

Теорема 9.4.2. *Имеем*

$$B_{\tilde{L}} = \sigma^2\left(P^{-1} - P^{-1}X^TG^{-1}XP^{-1}\right). \tag{9.4.5}$$

Доказательство. Так как $E(\tilde{L}) = Y$, то $B_{\tilde{L}} = E[(\tilde{L} - Y)(\tilde{L} - Y)^T]$.

Из (9.3.19) (теорема 9.3.2) находим

$$\tilde{L} - Y = -P^{-1}X^TG^{-1}X_0 + (E - P^{-1}X^TG^{-1}X)(Y + \Delta) - Y =$$
$$= -P^{-1}X^TG^{-1}(X_0 + XY) + (E - P^{-1}X^TG^{-1}X)\Delta.$$

В силу (9.2.2) имеем $X_0 + XY = 0$, так что

$$\tilde{L} - Y = (E - P^{-1}X^TG^{-1}X)\Delta. \tag{9.4.6}$$

Далее, $B_\Delta = \sigma^2 P^{-1}$.

Обозначая

$$U = P^{-1}X^TG^{-1}X, \tag{9.4.7}$$

из равенства $G = XP^{-1}X^T$ находим

$$U^2 = U; \quad U^T = PUP^{-1}. \tag{9.4.8}$$

Теперь (9.4.6) записываем в виде

$$\tilde{L} - Y = (E - U)\Delta. \tag{9.4.9}$$

По теореме 2.3.1 имеем

$$B_{\tilde{L}} = \sigma^2(E - U)P^{-1}(E - U)^T = \sigma^2(E - U)P^{-1}(E - PUP^{-1}) =$$
$$= \sigma^2(P^{-1} - UP^{-1} - UP^{-1} + U^2P^{-1}) =$$
$$= \sigma^2(P^{-1} - UP^{-1} - UP^{-1} + UP^{-1}) =$$
$$= \sigma^2(P^{-1} - UP^{-1}) = \sigma^2(P^{-1} - P^{-1}X^TG^{-1}XP^{-1}),$$

что и доказывает (9.4.5).

Остановимся теперь на реальном смысле оценок $\tilde{l}_1, \ldots, \tilde{l}_N$, полученных для измеряемых величин $y_1, \ldots, y_N$ при уравнивании с помощью коррелят. Само уравнивание, как показывает его название, приводит к оценкам $\tilde{l}_1, \ldots, \tilde{l}_N$, удовлетворяющим q условным уравнениям, за счет введения в наблюдения $l_1, \ldots, l_N$, вообще говоря, не удовлетворяющие этим уравнениям, кажущихся поправок $\tilde{v}_i = \tilde{l}_i - l_i$. Далее, оценки $\tilde{l}_1, \ldots, \tilde{l}_N$ могут быть линейно выражены через некоторые $N - q$ из них; это будет отвечать введению $N - q$ элементов из числа самих параметров $y_1, \ldots, y_N$. Оценки, соответствующие этим параметрам, будут вести себя соответственно сказанному в § 5 гл. VI. Именно, в весьма широком классе асимптотически нормальных и асимптотически несмещенных оценок этих параметров каждая оценка $\tilde{l}_i$ порознь (и любая линейная комбинация оценок $\tilde{l}_i$) будет уклоняться от соответствующего параметра (или линейной комбинации этих параметров) менее чем на заданное число $\varepsilon > 0$ с бо́льшей вероятностью, чем другая оценка (соответственно, линейная комбинация оценок).

Оценки же, линейно выражающиеся через $N-q$ указанных оценок, будут в определенном смысле экстремальными, как указывалось в § 2 гл. VII.

Случай, когда число условных уравнений q невелико сравнительно с N и, стало быть, с числом $N-q$ основных параметров (иногда называемых свободными), близок к тривиальному случаю $q=0$, когда условных уравнений нет и все параметры являются свободными. В этом случае оценками y_r по методу наименьших квадратов будут сами l_r; при этом $D(l_r)=\frac{\sigma^2}{p_r}$, так что качество оценки будет зависеть лишь от веса наблюдения. Правда, и здесь оценка l_r будет экстремальной в широком классе асимптотически нормальных и асимптотически несмещенных оценок y'_r параметра y_r, так что

$$P\{|l_r-y_r|\leqslant\varepsilon\}>P\{|y'_r-y_r|\leqslant\varepsilon\} \qquad (9.4.10)$$

для каждого r порознь, при достаточно большом N. В промежуточном случае, когда числа q и N оба велики, асимптотические свойства оценок по методу наименьших квадратов среди других оценок не изучены. То же касается уравнивания по элементам, когда числа элементов и наблюдений оба велики.

О самой же точности таких оценок см. ниже.

§ 5. Различные выражения $[p\tilde{v}\tilde{v}]$ и его статистическое поведение

Статистическое поведение $[p\tilde{v}\tilde{v}]$ может быть изучено путем сведения уравнивания с помощью коррелят к уравниванию с помощью элементов. Так как число наблюдений равно N, а число (независимых) условных уравнений есть q, то число элементов (получаемых, например, исключением) будет $N-q$. Это число играет роль числа n предыдущей главы; отсюда $N-n=q$, и $\frac{1}{\sigma^2}[p\tilde{v}\tilde{v}]$ должно быть распределено, как χ_q^2. Мы, однако, приведем вывод этого свойства, независимый от теории уравнивания с помощью элементов, и дадим некоторые выражения, полезные для вычисления $[p\tilde{v}\tilde{v}]$ и контроля такого вычисления.

Теорема 9.5.1. *Имеют место формулы:*

$$[p\tilde{v}\tilde{v}]=M^TG^{-1}M, \qquad (9.5.1)$$

$$[p\tilde{v}\tilde{v}]=-M^TK, \qquad (9.5.2)$$

$$[p\tilde{v}\tilde{v}]=K^TGK. \qquad (9.5.3)$$

Первая из этих формул совпадает с (9.3.3). Вторая следует из нее с помощью (9.2.20): $K = -G^{-1}M$; наконец, (9.5.3) получается из формулы (9.5.1), записанной в виде

$$[p\tilde{v}\tilde{v}] = M^T G^{-1} G G^{-1} M; \quad G^T = G; \quad K^T = -M^T G^{-1}.$$

Докажем теперь, что величина $\frac{1}{\sigma^2}[p\tilde{v}\tilde{v}]$ распределена по χ_q^2.

Теорема 9.5.2.

$$\frac{1}{\sigma^2}[p\tilde{v}\tilde{v}] = \chi_q^2.$$

Доказательство. Из (9.3.19) находим

$$\tilde{V} = \tilde{L} - L = -P^{-1}X^T G^{-1}(X_0 + XL), \tag{9.5.4}$$

$$L = Y + \Delta, \tag{9.5.5}$$

$\Delta = \begin{Vmatrix} \Delta_1 \\ \cdot \\ \vdots \\ \Delta_N \end{Vmatrix}$ — нормальный вектор с $E(\Delta) = 0$, $B_\Delta = \sigma^2 P^{-1}$. Далее, $X_0 + XY = 0$, так что

$$\tilde{V} = P^{-1}X^T G^{-1} X\Delta. \tag{9.5.6}$$

Обозначая

$$U = P^{-1}X^T G^{-1} X, \tag{9.5.7}$$

имеем

$$U^2 = U; \quad U^T = PUP^{-1}. \tag{9.5.8}$$

Положим еще

$$\Delta^{(0)} = P^{-\frac{1}{2}}\Delta, \tag{9.5.9}$$

так что

$$E\Delta^{(0)} = 0; \quad B_{\Delta^{(0)}} = \sigma^2 E_{NN}. \tag{9.5.10}$$

Тогда получим из (9.5.6) и (9.5.7)

$$-\tilde{V} = UP^{\frac{1}{2}}\Delta^{(0)} \tag{9.5.11}$$

и

$$\begin{aligned} [p\tilde{v}\tilde{v}] &= \tilde{V}^T P \tilde{V} = (\Delta^{(0)})^T P^{-\frac{1}{2}} U^T P U P^{-\frac{1}{2}} \Delta^{(0)} = \\ &= (\Delta^{(0)})^T P^{-\frac{1}{2}} P U^2 P^{-\frac{1}{2}} \Delta^{(0)} = (\Delta^{(0)})^T P^{\frac{1}{2}} U P^{-\frac{1}{2}} \Delta^{(0)}. \end{aligned} \tag{9.5.12}$$

Положим

$$W = P^{\frac{1}{2}} U P^{-\frac{1}{2}}. \tag{9.5.13}$$

Матрица $W = P^{-\frac{1}{2}} X^T G^{-1} X P^{-\frac{1}{2}}$ — симметрическая, причем

$$W^2 = W. \tag{9.5.14}$$

Согласно § 4 гл. I, можно подобрать ортогональную матрицу F такую, что

$$FWF^T = D = \left\| \begin{matrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & d_q \end{matrix} \right\|. \tag{9.5.15}$$

Так как $F^T = F^{-1}$, то $D^2 = D$, так что $d_i = 0$ или 1. Далее, ранг (D) равен количеству единиц среди чисел d_i. Но ранг $(D) =$ $=$ ранг (U).

Докажем, что ранг $(U) = q$. В самом деле, матрица G — неособенная, так что ранг $(G^{-1}) = q$. Далее, ранг $(X) =$ ранг $(X^T) = q$. Согласно неравенству (1.2.18),

$$\text{ранг } (U) \leqslant q. \tag{9.5.16}$$

Но из выражения (9.5.7) видим, что $XU = X$, так что ранг $(X) =$ $= q \leqslant$ ранг (U). Отсюда ранг $(U) = q$ и $d_i = 1$ $(i = 1, 2, \ldots, q)$.

Таким образом,

$$FWF^T = E. \tag{9.5.17}$$

Для (9.5.12) находим отсюда

$$[p\tilde{v}\tilde{v}] = \left(\Delta^{(0)}\right)^T F^T FWF^T F\Delta^{(0)} = Z^T Z,$$

где $Z = F\Delta^{(0)}$ — ортогональное преобразование $\Delta^{(0)}$, и, стало быть, по теореме Фишера 2.3.2, Z есть нормальный вектор с независимыми компонентами $z_i \in N(0, \sigma)$ $(i = 1, 2, \ldots, q)$.

Таким образом,

$$[p\tilde{v}\tilde{v}] = \sum_{i=1}^{q} z_i^2 = \chi_q^2,$$

что и требовалось доказать.

Теорема 9.5.3. *Случайные нормальные векторы $\tilde{V} = \tilde{L} - L$ и $\tilde{L}$ стохастически независимы. При этом $\tilde{V}$ есть $(N - q)$-мерный, а $\tilde{L}$ есть q-мерный нормальный вектор.*

Доказательство. На основании теоремы 2.3.3 и равенства $E(\tilde{V}) = 0$ для доказательства независимости $\tilde{V}$ и $\tilde{L}$ достаточно показать, что

$$E\tilde{V}(\tilde{L} - Y)^T = 0. \tag{9.5.18}$$

Имеем из (9.5.6)

$$\tilde{V} = -U\Delta; \quad U = P^{-1}X^T G^{-1} X. \tag{9.5.19}$$

Из (9.4.9) получаем

$$\tilde{L} - Y = (E - U)\Delta, \tag{9.5.20}$$

$$E\left(\Delta\Delta^T\right) = \sigma^2 P^{-1}. \tag{9.5.21}$$

Отсюда

$$E\tilde{V}(\tilde{L}-Y) = -U\left(E\left(\Delta\Delta^T\right)(E-U)^T\right) = -\sigma^2 UP^{-1}\left(E-PUP^{-1}\right) \tag{9.5.22}$$

в силу соотношений $U^2 = U$; $U^T = PUP^{-1}$.

Далее, правая часть (9.5.22) равна

$$-\sigma^2\left(UP^{-1}-U^2P^{-1}\right) = -\sigma^2\left(UP^{-1}-UP^{-1}\right) = 0,$$

что и доказывает (9.5.18).

Далее, $\tilde{V}$ есть q-мерный вектор, ибо матрица U имеет ранг q (см. теорему 2.3.1). Для доказательства того, что $\tilde{L}-Y$ есть $(N-q)$-мерный вектор, достаточно показать, что ранг $E-U$ равен $N-q$.

Положим $W = P^{\frac{1}{2}}UP^{-\frac{1}{2}}$; $W^2 = W$, W — симметрическая матрица. Подберем ортогональную матрицу F такую, что матрица $FWF^T = D$ — диагональная; тогда $D^2 = D$, так что

$$D = \begin{Vmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & d_N \end{Vmatrix},$$

где $d_i^2 = d_i$; $d_i = 0$; 1.

Так как ранг (U) = ранг $(D) = q$, то число d_i, равных единице, есть q, а число d_i, равных нулю, есть $N-q$. Далее, матрица

$$FP^{\frac{1}{2}}(E-U)P^{-\frac{1}{2}}F^T = E-D$$

имеет тот же ранг, что $E-U$; ранг $(E-D)$, очевидно, равен $N-q$, что и доказывает требуемое.

Найдем еще корреляционную матрицу $B_{\tilde{V}}$ вектора $\tilde{V}$.

Теорема 9.5.4.

$$B_{\tilde{V}} = \sigma^2 P^{-1}X^TG^{-1}XP^{-1}. \tag{9.5.23}$$

Доказательство. Согласно (9.5.19), имеем

$$\tilde{V} = -U\Delta; \quad U = P^{-1}X^TG^{-1}X; \quad E\tilde{V} = 0.$$

Отсюда

$$B_{\tilde{V}} = UB_\Delta U^T = \sigma^2 UP^{-1}PUP^{-1} = \sigma^2 U^2P^{-1} =$$
$$= \sigma^2 UP^{-1} = \sigma^2 P^{-1}X^TG^{-1}XP^{-1}.$$

Приведем еще одну формулу, полезную при контроле вычислений.

Теорема 9.5.5.

$$\sum_{i=1}^{N} p_i D(\tilde{l}_i) = [pD(\tilde{l})] = (N-q)\sigma^2. \tag{9.5.24}$$

Доказательство. Из (9.4.5) имеем (корреляционная матрица $\tilde{L}$ обозначена через $B_{\tilde{L}}$)

$$B_{\tilde{L}} = \sigma^2 \left(P^{-1} - P^{-1}X^T G^{-1} X P^{-1}\right),$$

$$D(\tilde{l}_i) = \sigma^2 \left\{P^{-1} - P^{-1}X^T G^{-1} X P^{-1}\right\}_{ii}; \qquad i = 1, 2, \ldots, N. \tag{9.5.25}$$

$$\sum_{i=1}^{N} p_i D(\tilde{l}_i) = [pD(\tilde{l})] = \sigma^2 \sum_{i=1}^{N} \left\{E - P^{-1}X^T G^{-1} X\right\}_{ii} = \sigma^2 \operatorname{Sp}(E-U), \tag{9.5.26}$$

где $U = P^{-1}X^T G^{-1} X$.

Как и в доказательстве теоремы 9.5.2, замечаем, что при помощи неособенной матрицы $Q = FP^{\frac{1}{2}}$ можно привести матрицу U к диагональному виду:

$$D = \begin{Vmatrix} d_1 & 0 & \ldots & 0 \\ . & . & . & . \\ 0 & 0 & \ldots & d_N \end{Vmatrix},$$

где d_i равны нулю или единице и число единиц среди d_i равно q.

Поэтому

$$Q(E-U)Q^{-1} = E - D = \begin{Vmatrix} 1-d_1 & 0 & \ldots & 0 \\ . & . & . & . \\ 0 & 0 & \ldots & 1-d_N \end{Vmatrix}$$

есть диагональная матрица, имеющая ровно $N-q$ единиц. Отсюда, согласно § 3 гл. 1,

$$\operatorname{Sp}(E-U) = \operatorname{Sp}(E-D) = N-q.$$

Это и доказывает (9.5.24).

Мы имеем

$$\sigma(\tilde{l}_i) = \sqrt{\left(P^{-1} - P^{-1}X^T G^{-1} X P^{-1}\right)_{ii}}\ \sigma. \tag{9.5.27}$$

Так как $\frac{1}{\sigma^2}[p\tilde{v}\tilde{v}] = \chi_q^2$, то, очевидно,

$$E[p\tilde{v}\tilde{v}] = q\sigma^2; \quad E\frac{[p\tilde{v}\tilde{v}]}{q} = \sigma^2,$$

т. е. $\frac{[p\tilde{v}\tilde{v}]}{q}$ является несмещенной оценкой для σ^2. Для σ асимптотически несмещенной при возрастании q будет оценка

$$s=\sqrt{\frac{[p\tilde{v}\tilde{v}]}{q}}. \tag{9.5.28}$$

При этом, при больших q, $\sigma(s)\approx\frac{\sigma}{\sqrt{2q}}$.

Таким образом, для $\sigma(\tilde{l}_i)$ получаем оценку, асимптотически несмещенную при больших q:

$$\sigma(\tilde{l}_i)\approx\sqrt{(P^{-1}-P^{-1}X^TG^{-1}XP^{-1})_{ii}}\cdot\sqrt{\frac{[p\tilde{v}\tilde{v}]}{q}}. \tag{9.5.29}$$

Как и в гл. VIII, теорема 9.5.5 дает возможность составить понятие об общей точности оценок $\tilde{l}_i$. Из (9.5.24) видим, что

$$\frac{1}{N}\sum p_i D(\tilde{l}_i)=\left(1-\frac{q}{N}\right)\sigma^2, \tag{9.5.30}$$

так что выгодно иметь число независимых условных уравнений возможно бо́льшим, а число свободных параметров $N-q$ возможно меньшим сравнительно с N. Разумеется, это согласуется с соответствующим результатом гл. VIII, так как $N-q$ отвечает числу элементов.

Для оценивания σ с помощью доверительных интервалов пользуемся тем, что

$$\frac{[p\tilde{v}\tilde{v}]}{\sigma^2}=\chi_q^2.$$

По табл. III при q степенях свободы находим γ_1 и γ_2.

Доверительный интервал

$$I=\left[\gamma_1\sqrt{\frac{[p\tilde{v}\tilde{v}]}{N-n}},\quad \gamma_2\sqrt{\frac{[p\tilde{v}\tilde{v}]}{N-n}}\right]$$

накрывает σ с вероятностью p_0.

§ 6. Оценивание y_i и σ с помощью доверительных интервалов

Мы видели (теорема 9.5.3), что нормальный вектор $\tilde{V}$ не зависит от $\tilde{L}$, так что и $[p\tilde{v}\tilde{v}]$ не зависит от $\tilde{L}$. Далее $\frac{1}{\sigma^2}[p\tilde{v}\tilde{v}]=\chi_q^2$, поэтому

$$\frac{1}{\sigma}\sqrt{\frac{[p\tilde{v}\tilde{v}]}{q}}=\frac{1}{\sigma}s=\sqrt{\frac{\chi_q^2}{q}}. \tag{9.6.1}$$

Составим величину $\tilde{l}_i - y_i$. Согласно (9.5.27), будем иметь

$$\frac{\tilde{l}_i - y_i}{\sigma\sqrt{(P^{-1} - P^{-1}X^TG^{-1}XP^{-1})_{ii}}} \in N(0,\ 1). \tag{9.6.2}$$

Эта случайная величина не зависит от $\frac{1}{\sigma}s$. По определению распределения Стьюдента, дробь

$$t_q = \frac{\tilde{l}_i - y_i}{s\sqrt{(P^{-1} - P^{-1}X^TG^{-1}XP^{-1})_{ii}}} \tag{9.6.3}$$

будет иметь распределение Стьюдента $S_q(t)$. Для построения доверительного интервала для оценивания y_i по заданной надежности p, имеющейся в таблицах, и числу степеней свободы q находим число γ_0. Доверительный интервал

$$J_\gamma^{(i)} = \Big[\tilde{l}_i - \gamma_0 s\sqrt{(P^{-1} - P^{-1}X^TG^{-1}XP^{-1})_{ii}},$$
$$\tilde{l}_i + \gamma_0 s\sqrt{(P^{-1} - P^{-1}X^TG^{-1}XP^{-1})_{ii}}\Big] \tag{9.6.4}$$

будет накрывать y_i с вероятностью p.

Длина доверительного интервала $J_\gamma^{(i)}$ будет случайной и равной

$$d_\gamma^{(i)} = 2\gamma_0 s\sqrt{(P^{-1} - P^{-1}X^TG^{-1}XP^{-1})_{ii}}. \tag{9.6.5}$$

Как и ранее, мы можем оценить взвешенную среднюю квадратичную длину этих интервалов

$$\frac{1}{N}E\left[\sum p_i\left(d_\gamma^{(i)}\right)^2\right] = \left(1 - \frac{q}{N}\right)\sigma^2.$$

При заданной надежности доверительные интервалы будут в среднем тем у́же, чем больше q сравнительно с N.

§ 7. Оценивание линейной функции от измеряемых параметров при косвенных наблюдениях

Как мы видим на примерах, часто приходится оценивать функции от измеряемых величин при непрямых измерениях. При этом нелинейные функции во многих случаях можно линеаризовать, как пояснялось в § 1 гл. VII.

Для примера пусть дан треугольник ABC, где мы измеряем углы A, B, C и сторону BC, а после уравнивания измерений углов оцениваем стороны CA и BA. Задача сводится к оцениванию линейной функции измеряемых величин.

Пусть имеется линейная функция наблюдений

$$y = h_1y_1 + h_2y_2 + \ldots + h_Ny_N = [hy]. \tag{9.7.1}$$

Положим

$$H = \|h_1, \ldots, h_N\|, \tag{9.7.2}$$

$$Y = \left\| \begin{matrix} y_1 \\ \vdots \\ y_N \end{matrix} \right\|. \tag{9.7.3}$$

Тогда

$$y = HY. \tag{9.7.4}$$

Надо произвести оценивание y. Как и в случае безусловных непрямых наблюдений (гл. VI), в качестве оценки для y возьмем

$$\tilde{y} = h_1\tilde{l}_1 + h_2\tilde{l}_2 + \ldots + h_N\tilde{l}_N = [h\tilde{l}] = H\tilde{L}. \tag{9.7.5}$$

Мы не будем здесь говорить о качествах такой оценки, а найдем лишь ее дисперсию и построим доверительные интервалы. Имеем

$$E(\tilde{y}) = HY = y, \tag{9.7.6}$$

т. е. $\tilde{y}$ есть несмещенная оценка. Далее, по теореме 2.3.1,

$$D(\tilde{y}) = HB_{\tilde{L}}H^T = \sigma^2 H\left(P^{-1} - P^{-1}X^TG^{-1}XP^{-1}\right)H^T =$$

$$= \sigma^2\left[\sum_{i=1}^{N}\frac{h_i^2}{p_i} - HP^{-1}X^TG^{-1}XP^{-1}H^T\right]. \tag{9.7.7}$$

Полагая $H_1 = HP^{-1} = \left\|\frac{h_1}{p_1}, \frac{h_2}{p_2}, \ldots, \frac{h_N}{p_N}\right\|$, имеем

$$HP^{-1}X^TG^{-1}XP^{-1}H^T = H_1X^TG^{-1}XH_1^T. \tag{9.7.8}$$

Мы можем составить и доверительный интервал для $\tilde{y}$. Именно, случайный вектор $\tilde{L}$ не зависит от $\frac{1}{\sigma^2}[p\tilde{v}\tilde{v}] = \chi_q^2$. Поэтому и $\tilde{y}$ не зависит от этой величины. Дробь

$$t_q = \frac{\tilde{y} - y}{s\sqrt{\left(\sum_{i=1}^{N}\frac{h_i^2}{p_i} - H_1X^TG^{-1}XH_1^T\right)}} \tag{9.7.9}$$

имеет распределение Стьюдента с q степенями свободы. Выбирая надежность p, по числу степеней свободы q находим число γ и строим доверительный интервал

$$J_\gamma^{(i)} = [\tilde{y} - sh, \tilde{y} + sh], \quad \text{где} \quad h = \sqrt{\sum_{i=1}^{N}\frac{h_i^2}{p_i} - H_1X^TG^{-1}XH^T}.$$

Этот интервал накрывает y с вероятностью p.

§ 8. Сравнение уравниваний с помощью элементов и коррелят

При уравнивании по элементам основная вычислительная трудность состоит в решении нормальных уравнений, т. е. в обращении матрицы $C = C_{nn} = X^T P X$ n-го порядка. При уравнивании с помощью коррелят основная вычислительная трудность падает на решение нормальных уравнений для коррелят, т. е. на обращение матрицы $G = G_{qq} = XP^{-1}X^T$ q-го порядка; вычисление кажущихся поправок $\tilde{V}$ при известных коррелятах требует уже не столь трудоемкой операции умножения на заданную матрицу $P^{-1}X^T$. При этом число элементов равно $n = N - q$. Если составление условных уравнений и фундаментальных уравнений по элементам требует примерно одного и того же труда (или вообще незатруднительно), то мы видим, что при $N - q > q$, т. е. при $q < \frac{N}{2}$, вообще говоря, предпочтительнее производить уравнивание по коррелятам, а при $q > \frac{N}{2}$, т. е. при $q > N - q$, предпочтительнее производить уравнивание по элементам.

§ 9. Сводка формул. Правила уравнивания с помощью коррелят

а) Точечное оценивание

Условные уравнения:

$$x_{j0} + y_1 x_{j1} + y_2 x_{j2} + \ldots + y_N x_{jN} = 0 \qquad (j = 1, 2, \ldots, q). \quad (9.9.1)$$

При наблюдениях l_i над y_i с весами p_i составим

$$m_i = x_{i0} + x_{i1} l_1 + x_{i2} l_2 + \ldots + x_{iN} l_N = x_{i0} + [x_i l], \quad (9.9.2)$$

где

$$x_i = (x_{i1}, x_{i2}, \ldots, x_{iN}) \qquad (i = 1, 2, \ldots, q). \quad (9.9.3)$$

Далее, составляем нормальные уравнения для коррелят $k_1, k_2, \ldots, k_q$:

$$\left[\frac{x_\nu x_1}{p}\right] k_1 + \left[\frac{x_\nu x_2}{p}\right] k_2 + \ldots + \left[\frac{x_\nu x_q}{p}\right] k_q + [x_\nu l] + x_{\nu 0} = 0, \quad (9.9.4)$$

$$\nu = 1, 2, \ldots, q.$$

После решения этих нормальных уравнений для коррелят находятся кажущиеся поправки по формуле

$$\tilde{v}_r = \frac{1}{p_r}(x_{1r} k_1 + x_{2r} k_2 + \ldots + x_{qr} k_q). \quad (9.9.5)$$

Оценки для величин y_i даются формулами

$$\tilde{l}_i = l_i + \tilde{v}_i. \quad (9.9.6)$$

Далее вычисляется величина $[p\tilde{v}\tilde{v}]$. Для ее вычисления или контроля вычислений можно пользоваться формулой [см. (9.5.2)].

$$-[p\tilde{v}\tilde{v}] = m_1k_1 + m_2k_2 + \ldots + m_qk_q = [mk]. \tag{9.9.7}$$

Оценивание σ происходит по формуле

$$\sigma \approx s = \sqrt{\frac{[p\tilde{v}\tilde{v}]}{q}}, \tag{9.9.8}$$

s — асимптотически несмещенная оценка σ при возрастающих q. При этом штандарт ее равен

$$\sigma(s) \approx \frac{\sigma}{\sqrt{2q}}. \tag{9.9.9}$$

Для вычисления штандартов $\sigma(\tilde{l}_i)$ оценок $\tilde{l}_i$ надлежит найти i-й диагональный элемент матрицы

$$P^{-1}X^TGXP^{-1}, \tag{9.9.10}$$

где

$$G = XP^{-1}X^T = \begin{Vmatrix} \left[\frac{x_1x_1}{p}\right] & \left[\frac{x_1x_2}{p}\right] & \ldots & \left[\frac{x_1x_q}{p}\right] \\ \cdot & \cdot & \cdot & \cdot \\ \left[\frac{x_qx_1}{p}\right] & \left[\frac{x_qx_2}{p}\right] & \ldots & \left[\frac{x_qx_q}{p}\right] \end{Vmatrix}$$

есть матрица системы нормальных уравнений для коррелят,

$$X = \begin{Vmatrix} x_{11} & x_{12} & \ldots & x_{1N} \\ \cdot & \cdot & \cdot & \cdot \\ x_{q1} & x_{q2} & \ldots & x_{qN} \end{Vmatrix}; \quad X^T = \begin{Vmatrix} x_{11} & \ldots & x_{q1} \\ x_{12} & \ldots & x_{q2} \\ \cdot & \cdot & \cdot \\ x_{1N} & \ldots & x_{qN} \end{Vmatrix}; \quad P^{-1} = \begin{Vmatrix} \frac{1}{p_1} & 0 & \ldots & 0 \\ 0 & \frac{1}{p_2} & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & \frac{1}{p_N} \end{Vmatrix}.$$

Тогда

$$\sigma(\tilde{l}_1) \approx s\sqrt{\left(\frac{1}{p_i} - \{P^{-1}X^TGXP^{-1}\}_{ii}\right)}. \tag{9.9.11}$$

Для контроля полезна формула

$$\sum_{i=1}^{N} p_i\left(\frac{1}{p_i} - \{P^{-1}X^TGXP^{-1}\}_{ii}\right) = N - q. \tag{9.9.12}$$

Выгодно иметь число независимых условных уравнений возможно бо́льшим,

б) Оценивание с помощью доверительных интервалов

Пользуемся табл. I распределения Стьюдента.

Выбираем надежность p_0, и при q степенях свободы находим γ_0. Составляем доверительный интервал

$$J_{\gamma_0}^{(i)} = [\tilde{l}_i - \gamma_0 s \sqrt{\{P^{-1} - P^{-1}X^T G^{-1} X P^{-1}\}_{ii}},$$
$$\tilde{l}_i + \gamma_0 s \sqrt{\{P^{-1} - P^{-1}X^T G^{-1} X P^{-1}\}_{ii}}],$$

где $s = \frac{[p\tilde{v}\tilde{v}]}{q}$. Он накрывает $\tilde{l}_i$ с вероятностью p_0. Для оценивания σ с помощью доверительных интервалов в табл. I при надежности p_0 и числе степеней свободы q выбираем γ_1 и γ_2 такие, что доверительный интервал $[\gamma_1 s, \gamma_2 s]$ будет накрывать σ с надежностью p_0.

§ 10. Примеры

Пример 1. На земельном участке (рис. 5) измерены разности высот пяти точек (стрелки на чертеже показывают направления убывания высот). * Измеренные значения (в метрах) равны

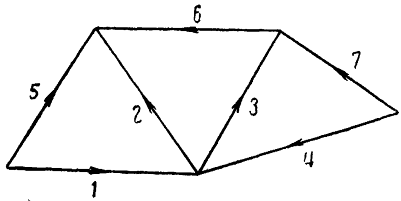

Рис. 5.

$$L = \begin{Vmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \\ l_5 \\ l_6 \\ l_7 \end{Vmatrix} = \begin{Vmatrix} 20,21 \\ 40,07 \\ 34,17 \\ 35,84 \\ 60,40 \\ 5,87 \\ 69,99 \end{Vmatrix}.$$

Веса считаются обратно пропорциональными соответствующим расстояниям. Матрица весов имеет вид

$$P = \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1,1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{Vmatrix}.$$

Принимая за элементы первые четыре разности высот, найдем наилучшие оценки для разностей высот методом уравнивания по элементам.

* Данные заимствованы из книги [2], стр. 204.

Фундаментальные уравнения запишутся, очевидно, в виде

$$\lambda_1=\xi_1, \qquad \lambda_5=\xi_1+\xi_2$$
$$\lambda_2=\xi_2, \qquad \lambda_6=\xi_2-\xi_3,$$
$$\lambda_3=\xi_3, \qquad \lambda_7=\xi_3+\xi_4.$$
$$\lambda_4=\xi_4,$$

Или, если переименовать λ_r в y_r, ξ_i в a_i, то в матричной записи фундаментальные уравнения запишутся в виде

$$Y=XA$$

[см. (8. 2. 1); $X_0=0$].

У нас

$$Y=\begin{Vmatrix} y_1\\ y_2\\ y_3\\ y_4\\ y_5\\ y_6\\ y_7 \end{Vmatrix}, \quad X=\begin{Vmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 1 & 1 & 0 & 0\\ 0 & 1 & -1 & 0\\ 0 & 0 & 1 & 1 \end{Vmatrix}, \quad A=\begin{Vmatrix} a_1\\ a_2\\ a_3\\ a_4 \end{Vmatrix}.$$

Согласно формуле (8.2.4), нормальные уравнения несмещенных оценок $\tilde{A}$ для A имеют вид

$$C\tilde{A}=X^TPL, \text{ где } C=X^TPX, \text{ или } \tilde{A}=C^{-1}X^TPL.$$

Мы имеем

$$C=X^TPX=\begin{Vmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 0 & 1 & 1 & 0\\ 0 & 0 & 1 & 0 & 0 & -1 & 1\\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{Vmatrix}\times$$

$$\times\begin{Vmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0{,}9 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 1{,}1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0{,}8 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0{,}9 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{Vmatrix}\cdot X=\begin{Vmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0{,}9 & 0 & 0 & 1 & 0.9 & 0\\ 0 & 0 & 1{,}1 & 0 & 0 & -0{,}9 & 1\\ 0 & 0 & 0 & 0{,}8 & 0 & 0 & 1 \end{Vmatrix}\times$$

$$\times\begin{Vmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 1 & 1 & 0 & 0\\ 0 & 1 & -1 & 0\\ 0 & 0 & 1 & 1 \end{Vmatrix}=\begin{Vmatrix} 2 & 1 & 0 & 0\\ 1 & 2{,}8 & -0{,}9 & 0\\ 0 & -0{,}9 & 3 & 1\\ 0 & 0 & 1 & 1{,}8 \end{Vmatrix}$$

Найдем теперь матрицу C^{-1}. Элементы обратной матрицы можно вычислять по формуле

$$(C^{-1})_{sr} = \frac{D_{rs}}{\det(C)},$$

где D_{rs} — алгебраическое дополнение элемента c_{rs} в $\det(C)$. Имеем

$$\det(C) = \begin{vmatrix} 2 & 1 & 0 & 0 \\ 1 & 2{,}8 & -0{,}9 & 0 \\ 0 & -0{,}9 & 3 & 1 \\ 0 & 0 & 1 & 1{,}8 \end{vmatrix} = 17{,}324.$$

Далее, матрица алгебраических дополнений элементов определителя C имеет вид

$$D = \begin{Vmatrix} 10{,}862 & -4{,}4 & -1{,}62 & 0{,}9 \\ -4{,}4 & 8{,}8 & 3{,}24 & -1{,}8 \\ -1{,}62 & 3{,}24 & 8{,}28 & -4{,}6 \\ 0{,}9 & -1{,}8 & -4{,}6 & 12{,}18 \end{Vmatrix},$$

значит,

$$C^{-1} = \begin{Vmatrix} 0{,}626991 & -0{,}253983 & -0{,}093512 & 0{,}051951 \\ -0{,}253983 & 0{,}507966 & 0{,}187024 & -0{,}103902 \\ -0{,}093512 & 0{,}187024 & 0{,}477950 & -0{,}265528 \\ 0{,}051951 & -0{,}103902 & -0{,}265528 & 0{,}703071 \end{Vmatrix}.$$

Имеем

$$X^T PL = \begin{Vmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0{,}9 & 0 & 0 & 1 & 0{,}9 & 0 \\ 0 & 0 & 1{,}1 & 0 & 0 & -0{,}9 & 1 \\ 0 & 0 & 0 & 0{,}8 & 0 & 0 & 1 \end{Vmatrix} \cdot \begin{Vmatrix} 20{,}21 \\ 40{,}07 \\ 34{,}17 \\ 35{,}84 \\ 60{,}40 \\ 5{,}87 \\ 69{,}99 \end{Vmatrix} = \begin{Vmatrix} 80{,}610 \\ 101{,}746 \\ 102{,}294 \\ 98{,}662 \end{Vmatrix}.$$

Следовательно, матрица $\tilde{A}$ имеет вид

$$\tilde{A} = C^{-1} \cdot \begin{Vmatrix} 80{,}610 \\ 101{,}746 \\ 102{,}294 \\ 98{,}662 \end{Vmatrix} = \begin{Vmatrix} 20{,}260 \\ 40{,}090 \\ 34{,}185 \\ 35{,}821 \end{Vmatrix}.$$

Тогда оценки $\tilde{l}_r$ для разностей высот запишутся, согласно формуле (8.2.7), в виде матрицы:

$$\tilde{L} = X\tilde{A} = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1 \end{Vmatrix} \cdot \begin{Vmatrix} 20{,}260 \\ 40{,}090 \\ 34{,}185 \\ 35{,}821 \end{Vmatrix} = \begin{Vmatrix} 20{,}260 \\ 40{,}090 \\ 34{,}185 \\ 35{,}821 \\ 60{,}350 \\ 5{,}905 \\ 70{,}006 \end{Vmatrix}.$$

Пример 2. Измерение четырех углов четырехугольника дало следующие результаты *):

$$L = \begin{Vmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \end{Vmatrix} = \begin{Vmatrix} 50^\circ\,12'\,37'' \\ 112^\circ\,17'\,19'' \\ 120^\circ\,47'\,26'' \\ 76^\circ\,46'\,18'' \end{Vmatrix},$$

при этом матрица весов измерений имеет вид

$$P = \begin{Vmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{Vmatrix}.$$

Принимая за элементы первые три угла, найдем наилучшие оценки углов методом уравнивания по элементам.

Фундаментальные уравнения запишутся тогда в виде

$$\begin{aligned} y_1 &= a_1, \\ y_2 &= a_2, \\ y_3 &= a_3, \\ y_4 &= 360^\circ - a_1 - a_2 - a_3. \end{aligned}$$

Или, в матричной записи,

$$Y = X_0 + XA,$$

где

$$Y = \begin{Vmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{Vmatrix}, \quad X_0 = \begin{Vmatrix} 0 \\ 0 \\ 0 \\ 360^\circ \end{Vmatrix}, \quad X = \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{Vmatrix}, \quad A = \begin{Vmatrix} a_1 \\ a_2 \\ a_3 \end{Vmatrix}.$$

*) Данные заимствованы из книги [2], стр. 237, упр. 78.

Согласно формуле (8.2.4), нормальные уравнения несмещенных оценок $\tilde{A}$ для A имеют вид

$$C\tilde{A} = X^T P(L - X_0), \quad C = X^T PX,$$

или

$$\tilde{A} = C^{-1} X^T P(L - X_0).$$

Имеем

$$C = X^T PX = \begin{Vmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{Vmatrix} \cdot \begin{Vmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{Vmatrix} \cdot X =$$

$$= \begin{Vmatrix} 3 & 0 & 0 & -2 \\ 0 & 4 & 0 & -2 \\ 0 & 0 & 2 & -2 \end{Vmatrix} \cdot \begin{Vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -1 & -1 \end{Vmatrix} = \begin{Vmatrix} 5 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 4 \end{Vmatrix}.$$

Найдем матрицу C^{-1} (см. пример 1).

$$\det(C) = \begin{vmatrix} 5 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 4 \end{vmatrix} = 76.$$

Матрица алгебраических дополнений элементов C имеет вид

$$K = \begin{Vmatrix} 20 & -4 & -8 \\ -4 & 16 & -6 \\ -8 & -6 & 26 \end{Vmatrix}.$$

Значит,

$$C^{-1} = \begin{Vmatrix} 0{,}26315790 & -0{,}05263158 & -0{,}10526316 \\ -0{,}05263158 & 0{,}21052632 & -0{,}07894737 \\ -0{,}10526316 & -0{,}07894737 & 0{,}34210527 \end{Vmatrix}.$$

Далее вычисляем

$$\tilde{A} = C^{-1} X^T P(L - X_0) = C^{-1} X^T P \begin{Vmatrix} 50^\circ\, 12'\, 37'' \\ 112^\circ\, 17'\, 19'' \\ 120^\circ\, 47'\, 26'' \\ -283^\circ\, 13'\, 42'' \end{Vmatrix} = \begin{Vmatrix} 50^\circ\, 11'\, 50'',7 \\ 112^\circ\, 16'\, 44'',3 \\ 120^\circ\, 46'\, 16'',5 \end{Vmatrix}.$$

А оценки для углов запишутся в виде матрицы (для четвертого угла оценка, очевидно, получается добавлением 360° к вычисленному значению $-283^\circ\, 14'\, 51'',5$):

$$\tilde{L} = \begin{Vmatrix} 50^\circ\, 11'\, 50'',7 \\ 112^\circ\, 16'\, 44'',3 \\ 120^\circ\, 46'\, 16'',5 \\ 76^\circ\, 45'\, 8'',5 \end{Vmatrix}.$$

Пример 3. Решим теперь пример 1, применяя уравнивание с помощью коррелят. Истинные значения разностей высот точек связаны тремя соотношениями:

$$\lambda_1+\lambda_2-\lambda_5=0,$$
$$\lambda_2-\lambda_3-\lambda_6=0,$$
$$\lambda_3+\lambda_4-\lambda_7=0.$$

В целях удобства, записи будем по-прежнему вести в матричной форме. Тогда оценки для y, согласно (9.2.27), равны

$$\tilde{L}=L+\tilde{V},$$

где L — одностолбцовая матрица наблюденных значений разностей высот, $\tilde{V}$ — матрица кажущихся поправок, определяемая по формуле (9.2.21);

$$\tilde{V}=-P^{-1}X^TG^{-1}M=P^{-1}X^TK,$$

где $G=XP^{-1}X^T$, $K=-G^{-1}M$ — матрица коррелят,

$$M=X_0+XL.$$

У нас

$$X_0=0,$$

$$X=\left\|\begin{matrix} 1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 \end{matrix}\right\|.$$

Значит,

$$M=XL=\left\|\begin{matrix} 1 & 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1. & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & -1 \end{matrix}\right\|\cdot\left\|\begin{matrix} 20.21 \\ 40,07 \\ 34,17 \\ 35,84 \\ 60,40 \\ 5,87 \\ 69,99 \end{matrix}\right\|=\left\|\begin{matrix} -0,12 \\ 0,03 \\ 0,02 \end{matrix}\right\|.$$

Очевидно, $\det(P)=0,7128$ и

$$P^{-1}=\left\|\begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1,11111 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,90909 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1,25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1,11111 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{matrix}\right\|.$$

Имеем

$$G = XP^{-1}X^T = \begin{Vmatrix} 1 & 1,11111 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1,11111 & -0,90909 & 0 & 0 & -1,11111 & 0 \\ 0 & 0 & 0,90909 & 1,25 & 0 & 0 & -1 \end{Vmatrix} \times$$

$$\times \begin{Vmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & -1 & 1 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{Vmatrix} = \begin{Vmatrix} 3,11111 & 1,11111 & 0 \\ 1,11111 & 3,13131 & -0,90909 \\ 0 & -0,90909 & 3,15909 \end{Vmatrix}$$

и $\det(G) = 24,3041$, матрица D алгебраических дополнений элементов G равна

$$D = \begin{Vmatrix} 9,06565 & -3,51010 & -1,01010 \\ -3,51010 & 9,82828 & 2,82828 \\ -1,01010 & 2,82828 & 8,50728 \end{Vmatrix},$$

поэтому

$$G^{-1} = \begin{Vmatrix} 0,37301 & -0,14442 & -0,04156 \\ -0,14442 & 0,40439 & 0,11637 \\ -0,04156 & 0,11637 & 0,35003 \end{Vmatrix}.$$

Выпишем теперь матрицу коррелят

$$K = -G^{-1}M = -G^{-1} \cdot \begin{Vmatrix} -0,12 \\ 0,03 \\ 0,02 \end{Vmatrix} = \begin{Vmatrix} 0,04993 \\ -0,03179 \\ -0,01548 \end{Vmatrix}.$$

Далее находим матрицу кажущихся поправок (выписываем три десятичных знака).

$$\tilde{V} = P^{-1}X^TK = \begin{Vmatrix} 1 & 0 & 0 \\ 1,11111 & 1,11111 & 0 \\ 0 & -0,90909 & 0,90909 \\ 0 & 0 & 1,25 \\ -1 & 0 & 0 \\ 0 & -1,11111 & 0 \\ 0 & 0 & -1 \end{Vmatrix} \cdot \begin{Vmatrix} 0,04993 \\ -0,03179 \\ -0,01548 \end{Vmatrix} = \begin{Vmatrix} 0,050 \\ 0,020 \\ 0,015 \\ -0,019 \\ -0,050 \\ 0,035 \\ 0,015 \end{Vmatrix},$$

откуда

$$\tilde{L} = L + \tilde{V} = \begin{Vmatrix} 20,260 \\ 40,090 \\ 34,185 \\ 35,821 \\ 60,350 \\ 5,905 \\ 70,004 \end{Vmatrix}.$$

ГЛАВА X

НЕКОТОРЫЕ СЛУЧАИ ОБРАБОТКИ НАБЛЮДЕНИЙ В ГЕОДЕЗИИ

§ 1. Уравнивание одиночного нивелирного хода*)

Пусть имеется одиночный нивелирный ход в n станций между точками A и B с известными отметками H_A, H_B (рис. 6). Ищется отметка точки E (k-й станции, считая от A). Для этого производится отметка точки E от точки A и от точки B:

$$H'_E = H_A + \sum_{i=1}^{k} h_i; \quad H''_E = H_B - \sum_{i=k+1}^{n} h_i. \tag{10.1.1}$$

Требуется произвести оценивание H_E на основании полученных данных.

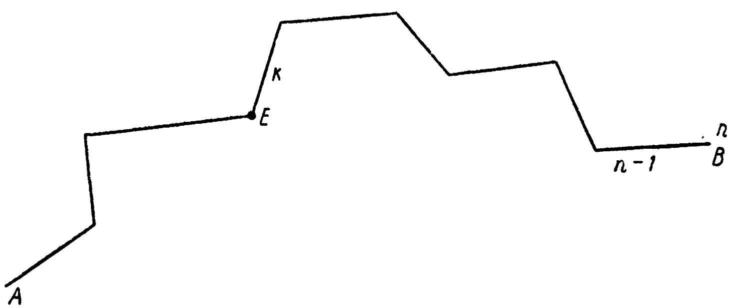

Рис. 6.

Формально — это задача на уравнивание условных измерений, так как сумма истинных значений приближений должна быть равна $H_B - H_A$. Но в данном случае проще свести задачу к прямым измерениям. Будем считать, что $h_i = y_i + \Delta_i$, где y_i — истинные превышения, а Δ_i — случайные погрешности, независимые и нормальные $N(0,1)$. Таким образом, мы предполагаем, что системати-

*) Постановки задач в § 1 и 2 взяты из книги А. С. Чеботарёва [51]. В составлении § 1 и 2 участвовал О. В. Шалаевский.

ческие погрешности отсутствуют, что само по себе является гипотезой, нуждающейся в проверке. Вопросы проверки гипотез принадлежат к математической статистике и трактуются в соответствующих курсах. Некоторые случаи такой проверки изложены в § 6 гл. IV и § 3 гл. XII.

Имеем, обозначая через y_E истинную отметку E,

$$H'_E = y_E + \Delta_1 + \Delta_2 + \ldots + \Delta_k, \qquad (10.1.2)$$

$$H''_E = y_E - \Delta_{k+1} - \ldots - \Delta_n. \qquad (10.1.3)$$

Далее, в силу независимости Δ_i [см. (2.1.18)]

$$\left.\begin{aligned} &D(\Delta_1 + \ldots + \Delta_k) = k\sigma^2, \\ &D(-\Delta_{k+1} - \ldots - \Delta_n) = (n-k)\sigma^2. \end{aligned}\right\} \qquad (10.1.4)$$

Теперь равенства (10.1.2) и (10.1.3) возвращают нас к ситуации гл. V. Полагая $x_1 = H'_E$, $x_2 = H''_E$, видим, что для оценивания y_E произведены прямые неравноточные измерения с весами $p_1 = \frac{1}{k}$ и $p_2 = \frac{1}{n-k}$.

По формуле (5.2.3) непосредственно находим оценку для y_E:

$$\tilde{y} = \frac{p_1 x_1 + p_2 x_2}{p_1 + p_2} = \frac{(n-k) H'_E + k H''_E}{n}. \qquad (10.1.5)$$

Далее, имеем по формуле (5.2.7)

$$D(\tilde{y}) = \frac{\sigma^2}{p_1 + p_2} = \frac{k(n-k)\sigma^2}{n}. \qquad (10.1.6)$$

Эти формулы годны для любых станций ($k = 1, 2, \ldots, n$). Найдем k, при котором $D(\tilde{y})$ максимально. В случае четного n это будет станция $k = \frac{n}{2}$, а в случае нечетного — две станции:

$$k = \frac{n+1}{2} \quad \text{и} \quad k = \frac{n-1}{2}.$$

В самом деле,

$$k(n-k) = \frac{n^2}{4} - \left(\frac{n}{2} - k\right)^2,$$

откуда следует наше утверждение. Итак, наименее точные результаты оценивания получаются в середине нивелирного хода. В этом случае для четного n имеем

$$D(\tilde{y}) = \frac{n}{4}\sigma^2; \quad \sigma(\tilde{y}) = \frac{\sigma\sqrt{n}}{2}. \qquad (10.1.7)$$

Величина $f_h = H'_E - H''_E$ есть невязка данного нивелирного хода. Из (10.1.5) видим, что

$$\tilde{y} = H'_E + \frac{k(H'_E - H''_E)}{n} = H'_E + \frac{k}{n} f_h. \qquad (10.1.8)$$

Эта формула годна для любых значений k и представляет собой правило распределения невязок по станциям нивелирного хода. Для оценивания σ, если для этого нет каких-либо предыдущих данных, можно произвести обратный нивелирный ход. Пусть наряду с превышениями $h_1, h_2, \ldots, h_n$ прямого хода, обратный ход дает соответствующие превышения $h'_n, h'_{n-1}, \ldots, h'_1$ [h'_k отвечает $(k-1)$ станции и т. д]. Предполагая отсутствие систематической ошибки, положим $h'_i = -y_i + \delta_i$, где величины δ_i $(i = 1, 2, \ldots, n)$ можем считать статистически независимыми в совокупности и независимыми от Δ_j. Они распределены так же, как Δ_i, т. е. нормальны $N(0, \sigma)$. Имеем тогда

$$h_i + h'_i = \Delta_i + \delta_i.$$

Величины $\Delta_i + \delta_i$ в силу независимости Δ_i и δ_i будут нормальными $N(0, \sigma\sqrt{2})$. Имеем

$$E\frac{1}{n}\sum_{i=1}^{n}(h_i + h'_i)^2 = E\frac{1}{n}[(h + h')^2] = 2\sigma^2. \qquad (10.1.9)$$

Поэтому величина

$$\frac{1}{2n}\sum_{i=1}^{n}(h_i + h'_i)^2 = \frac{1}{2n}[(h + h')^2] = q^2 \qquad (10.1.10)$$

будет несмещенной оценкой для σ^2, а

$$\frac{n}{4}q^2 = \frac{1}{8}[(h + h')^2] \qquad (10.1.11)$$

— несмещенной оценкой для $D(\tilde{y})$. Соответствующие штандарты будут иметь асимптотически несмещенные оценки

$$q = \sqrt{\frac{1}{2n}[(h + h')^2]}, \qquad (10.1.12)$$

$$\sqrt{\frac{1}{8}[(h + h')^2]} \quad [\text{оценка для } \sigma(\tilde{y})]. \qquad (10.1.13)$$

Далее, ввиду того, что величины $\Delta_i + \delta_i$ нормальны $N(0, \sigma\sqrt{2})$, величина $\frac{nq^2}{\sigma^2}$ будет распределена по закону χ^2_n. Отсюда следует, что при больших n

$$\sigma(q) \approx \frac{\sigma}{\sqrt{2n}}. \qquad (10.1.14)$$

Для оценивания σ можно построить доверительные интервалы, как в гл. IV. Именно,

$$\frac{nq^2}{\sigma^2}=\chi_n^2.$$

Ввиду этого для построения доверительного интервала можно пользоваться табл. III приложений. По заданной надежности p_0 при числе степеней свободы $k=n$ находим γ_1 и γ_2 такие, что $P(\gamma_1 q<\sigma<\gamma_2 q)=p_0$. Доверительный интервал $[\gamma_1 q, \gamma_2 q]$ будет накрывать σ с вероятностью p_0.

При наличии обратного нивелирного хода информацию, даваемую этим ходом, совместно с информацией, даваемой прямым ходом, можно использовать для более точного оценивания отметки y на k-й станции и для построения доверительных интервалов. Для этого к равенствам (10.1.1) присоединяют равенства

$$x_3=H_B+h'_n+h'_{n-1}+\ldots+h'_{k+1}, \tag{10.1.15}$$

$$x_4=H_A-h'_1-h'_2-\ldots-h'_k. \tag{10.1.16}$$

Их можно рассматривать как значения y с погрешностями весов соответственно $\frac{1}{n-k}$ и $\frac{1}{k}$, а потому, на основании теории оценивания прямых неравноточных измерений (гл. V), можем ввести новую оценку $\approx y$ для y:

$$\tilde{\tilde{y}}=\frac{\frac{1}{k}(x_1+x_4)+\frac{1}{n-k}(x_2+x_3)}{\frac{2}{k}+\frac{2}{n-k}}=\frac{(n-k)(x_1+x_4)+k(x_2+x_3)}{2n}. \tag{10.1.17}$$

Имеем [см. (5.2.7)]

$$D(\tilde{\tilde{y}})=\frac{\sigma^2}{[p]}=\frac{k(n-k)\sigma^2}{2n}, \tag{10.1.18}$$

т. е. дисперсия новой оценки уменьшается вдвое по сравнению со старой, что естественно. Конструкция доверительных интервалов в данном случае несколько кропотлива, и мы ее не будем приводить. Сомкнутый нивелирный ход ($H_A=H_B$) можно получить, уравнивая по тем же правилам.

§ 2. Уравнивание нивелирных ходов, опирающихся на марки

Пусть из нескольких точек $A_1, \ldots, A_k$ с известными отметками $H_{A_1}, \ldots, H_{A_k}$ проложены нивелирные ходы в точку E (рис. 7), причем ход из точки A_i имеет n_i станций. Таким образом, для хода из точки A_i получаем отметку в точке E:

$$x_i=H_{A_i}+h_1+h_2+\ldots+h_{n_i}. \tag{10.2.1}$$

При отсутствии систематической ошибки имеем, как и в § 1,

$$x_i = y + \delta_{1i} + \ldots + \delta_{n_i i}, \tag{10.2.2}$$

где δ_{ji} независимы в совокупности ($i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, n_i$), δ_{ij} нормальны $N(0, \sigma)$, а y—истинное значение отметки в точке E.

Согласно сказанному в § 2 гл. V, в качестве оценки y выбираем

$$\tilde{y} = \frac{\sum\limits_{i=1}^{k} \frac{1}{n_i} x_i}{\sum\limits_{i=1}^{k} \frac{1}{n_i}}. \tag{10.2.3}$$

При этом

$$D(\tilde{y}) = \frac{\sigma^2}{\sum\limits_{i=1}^{k} \frac{1}{n_i}}; \quad \sigma(\tilde{y}) = \frac{\sigma}{\sqrt{\sum\limits_{i=1}^{k} \frac{1}{n_i}}}. \tag{10.2.4}$$

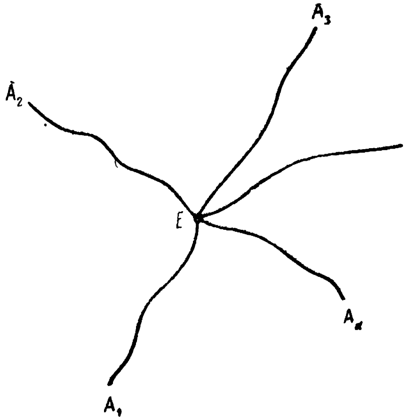

Рис. 7.

Для оценивания σ надлежит либо иметь предварительную информацию о точности работы инструментов в данных условиях, либо проводить обратные нивелирные ходы; далее следует действовать как в § 1.

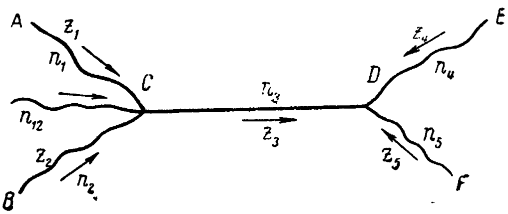

Рис. 8.

Рассмотрим еще случай двух точек, изображенный на рис. 8. Числа станций показаны на соответствующих ходах. Мы могли бы построить оценку отметки в C по сходящимся в ней ходам и далее, исходя из нее и полученного для нее веса, оценивать отметку точки D из получившихся трех ходов.

Таким образом часто и действуют; оценивание получается несложное и, как мы увидим, результаты получаются оптимальные.

Подойдем к этой задаче, как к задаче уравнивания по элементам, в результате чего оптимальный способ действия (в смысле гл. VII) будет установлен.

Пусть H_C и H_D — истинные превышения C и D над отметкой A; H_B, H_E, H_F — истинные превышения точек B, E, F над точкой A. Обозначим истинные превышения следующим образом: $(C, A) = \xi_1$,

$$(C, B) = \xi_2; \quad (D, C) = \xi_3; \quad (E, D) = \xi_4; \quad (F, D) = \xi_5.$$

Имеем уравнения в элементах

$$\begin{aligned} H_C &= \xi_1 = H_B + \xi_2, \\ H_D &= \xi_1 + \xi_3 = H_E + \xi_4 = H_F + \xi_5. \end{aligned}$$

Введем новые обозначения $\xi_1 = \eta_1$, $\xi_3 = \eta_2$. Имеем два независимых элемента η_1 и η_2, через которые выражаются остальные, и пять измерений

$$\begin{aligned} l_1 + \Delta_1 &= \eta_1, \\ l_2 + \Delta_2 &= \eta_1, \\ l_3 + \Delta_3 &= \eta_2, \\ l_4 + \Delta_4 &= \eta_1 + \eta_2, \\ l_5 + \Delta_5 &= \eta_1 + \eta_2. \end{aligned} \tag{10.2.5}$$

Матрица X будет имет вид (см. гл. VI)

$$X = X_{52} = \begin{Vmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ 1 & 1 \end{Vmatrix}; \quad X^T = \begin{Vmatrix} 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{Vmatrix}. \tag{10.2.6}$$

Матрица весов такова:

$$P = \begin{Vmatrix} \frac{1}{n_1} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{n_2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{n_3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{n_4} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{n_5} \end{Vmatrix}; \tag{10.2.7}$$

$$X^T P = \begin{Vmatrix} \frac{1}{n_1} & \frac{1}{n_2} & 0 & \frac{1}{n_4} & \frac{1}{n_5} \\ 0 & 0 & \frac{1}{n_3} & \frac{1}{n_4} & \frac{1}{n_5} \end{Vmatrix},$$

$$C = X^T P X = \left\| \begin{matrix} \frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_4}+\frac{1}{n_5}, & \frac{1}{n_4}+\frac{1}{n_5} \\ \frac{1}{n_4}+\frac{1}{n_5}, & \frac{1}{n_3}+\frac{1}{n_4}+\frac{1}{n_5} \end{matrix} \right\|,$$

$$X^T P L = \left\| \begin{matrix} \frac{l_1}{n_1}+\frac{l_2}{n_2}+\frac{l_4}{n_4}+\frac{l_5}{n_5} \\ \frac{l_3}{n_3}+\frac{l_4}{n_4}+\frac{l_5}{n_5} \end{matrix} \right\|.$$

Нормальные уравнения для η_1 и η_2 примут вид

$$\eta_1\left(\frac{1}{n_1}+\frac{1}{n_2}+\frac{1}{n_4}+\frac{1}{n_5}\right)+\eta_2\left(\frac{1}{n_4}+\frac{1}{n_5}\right)=\frac{l_1}{n_1}+\frac{l_2}{n_2}+\frac{l_4}{n_4}+\frac{l_5}{n_5},$$

$$\eta_1\left(\frac{1}{n_4}+\frac{1}{n_5}\right)+\eta_2\left(\frac{1}{n_3}+\frac{1}{n_4}+\frac{1}{n_5}\right)=\frac{l_3}{n_3}+\frac{l_4}{n_4}+\frac{l_5}{n_5}.$$

Положим $p_i=\frac{1}{n_i}$ $(i=1,\ 2,\ 3,\ 4,\ 5)$. Имеем

$$d=\det(C)=(p_1+p_2+p_4+p_5)(p_3+p_4+p_5)-(p_4+p_5)^2,$$

$$C^{-1}=\frac{1}{d}\left\| \begin{matrix} (p_3+p_4+p_5) & -(p_4+p_5) \\ -(p_4+p_5) & (p_1+p_2+p_4+p_5) \end{matrix} \right\|.$$

Отсюда получаем оценки $\tilde{\eta}_1$ и $\tilde{\eta}_2$ для элементов η_1 и η_2:

$$\tilde{\eta}_1=\frac{1}{d}[(p_3+p_4+p_5)(p_1l_1+p_2l_2+p_4l_4+p_5l_5)-$$
$$-(p_4+p_5)(p_3l_3+p_4l_4+p_5l_5)]=\frac{1}{d}[(p_3+p_4+p_5)(p_1l_1+p_2l_2)-$$
$$-(p_4+p_5)p_3l_3+p_3(p_4l_4+p_5l_5)], \quad (10.2.8)$$

$$\tilde{\eta}_2=\frac{1}{d}[-(p_4+p_5)(p_1l_1+p_2l_2+p_4l_4+p_5l_5)+$$
$$+(p_1+p_2+p_4+p_5)(p_3l_3+p_4l_4+p_5l_5)]=$$
$$=\frac{1}{d}[-(p_4+p_5)(p_1l_1+p_2l_2)+$$
$$+(p_1+p_2+p_4+p_5)p_3l_3+(p_1+p_2)(p_4l_4+p_5l_5)]. \quad (10.2.9)$$

Величина $\tilde{\eta}_1$ и является оценкой отметки точки C, для отметки же точки D имеем оценку $\tilde{\eta}_1+\tilde{\eta}_2$:

$$\tilde{\eta}_1+\tilde{\eta}_2=\frac{1}{d}[p_3(p_1l_1+p_2l_2)+$$
$$+(p_1+p_2)p_3l_3+(p_1+p_2+p_3)(p_4l_4+p_5l_5)]. \quad (10.2.10)$$

Как и следует ожидать, в оценках превышений C и D над A имеется симметрия: при замене $(p_4,\ p_5)\longleftrightarrow(p_1,\ p_2)$, $(l_4,\ l_5)\longleftrightarrow(l_1,\ l_2)$, $p_1\longleftrightarrow p_3$, $l_1\longleftrightarrow l_3$ одна оценка переходит в другую.

Подсчитаем еще дисперсию оценки $\tilde{\eta}_1$; по формуле (8.2.5) имеем

$$D(\tilde{\eta}_1)=\sigma^2\{C^{-1}\}_{11}=\sigma^2\frac{p_3+p_4+p_5}{d}=$$
$$=\sigma^2\frac{p_3+p_4+p_5}{p_3(p_1+p_2+p_4+p_5)+(p_1+p_2)(p_4+p_5)}. \qquad (10.2.11)$$

Из указанных выше соображений симметрии получаем

$$D(\tilde{\eta}_1+\tilde{\eta}_2)=\sigma^2\frac{p_1+p_2+p_3}{p_3(p_1+p_2+p_4+p_5)+(p_1+p_2)(p_4+p_5)}. \qquad (10.2.12)$$

Можно заметить, что присоединение к данным от двух ходов из точек A и B в точку C информации от других ходов повышает точность оценивания. При оценивании отметки C только по двум ходам мы получили бы оценку η' с дисперсией (см. предыдущий параграф)

$$D(\eta_1')=\frac{\sigma^2}{p_1+p_2}. \qquad (10.2.13)$$

Покажем, что

$$D(\tilde{\eta}_1)<D(\eta_1'). \qquad (10.2.14)$$

Имеем

$$\frac{p_3+p_4+p_5}{p_3(p_1+p_2+p_4+p_5)+(p_1+p_2)(p_4+p_5)}-\frac{1}{p_1+p_2}=$$
$$=\frac{(p_3+p_4+p_5)(p_1+p_2)-p_3(p_1+p_2+p_4+p_5)-(p_1+p_2)(p_4+p_5)}{(p_1+p_2)\{p_3(p_1+p_2+p_4+p_5)+(p_1+p_2)(p_4+p_5)\}}.$$

Числитель равен $-p_3(p_4+p_5)<0$, что и доказывает (10.2.14). Как пояснено в § 2 гл. VIII, полученные результаты являются оптимальными в определенном смысле. Покажем, что иной способ, часто употребляемый геодезистами и описанный в книге А. С. Чеботарёва [51], приводит в данном случае к тем же оценкам и, стало быть, является также оптимальным *).

Будем пользоваться следующими обозначениями (см. рис. 8):

l_1 — наблюденная отметка C по ходу z_1,
l_2 — наблюденная отметка C по ходу z_2,
l_3 — наблюденное превышение D над C по ходу z_3,
l_4 — наблюденная отметка D по ходу z_4,
l_5 — наблюденная отметка D по ходу z_5.

Определим предварительную оценку для отметки C по ходам z_1 и z_2

$$l_{12}=\frac{l_1p_1+l_2p_2}{p_1+p_2}. \qquad (10.2.15)$$

Вес l_{12} есть $p_{12}=p_1+p_2$.

*) Это замечание сообщено автору О. В. Шалаевским.

Вместо двух ходов z_1 и z_2 берем один ход z_{12} в C, дающий тот же вес p_{12}. Этот ход имеет

$$n_{12}=\frac{1}{p_{12}}=\frac{1}{p_1+p_2}$$

станций.

Считаем, что в точку D проложен один сложный ход z_{123}, состоящий из ходов z_{12} и z_3, с числом станций $n_{12}+n_3$.

Ход z_{123} дает наблюденную отметку D, равную

$$l_{123}=l_{12}+l_3.$$

Вес хода z_{123} есть

$$p_{123}=\frac{1}{n_{12}+n_3}=\frac{1}{\dfrac{1}{p_1+p_2}+\dfrac{1}{p_3}}=\frac{(p_1+p_2)\,p_3}{p_1+p_2+p_3}. \qquad (10.2.16)$$

Теперь определяем (окончательную) оценку отметки D по трем ходам z_{123}, z_4, z_5 по обычному правилу

$$\bar{l}_D=\frac{l_{123}p_{123}+l_4p_4+l_5p_5}{p_{123}+p_4+p_5}. \qquad (10.2.17)$$

Запишем это подробнее. Поскольку

$$l_{123}=l_{12}+l_3=\frac{l_1p_1+l_2p_2}{p_1+p_2}+l_3=\frac{l_1p_1+l_2p_2+l_3(p_1+p_2)}{p_1+p_2},$$

то

$$\bar{l}_D=\frac{\dfrac{[l_1p_1+l_2p_2+l_3(p_1+p_2)]}{(p_1+p_2)}\,\dfrac{(p_1+p_2)\,p_3}{(p_1+p_2+p_3)}+l_4p_4+l_5p_5}{\dfrac{(p_1+p_2)\,p_3}{p_1+p_2+p_3}+p_4+p_5}=$$

$$=\frac{(l_1p_1+l_2p_2)\,p_3+l_3(p_1+p_2)\,p_3+(l_4p_4+l_5p_5)(p_1+p_2+p_3)}{(p_1+p_2)\,p_3+(p_1+p_2+p_3)(p_4+p_5)}=$$

$$=\frac{(l_1p_1+l_2p_2)\,p_3+(p_1+p_2)\,p_3l_3+(p_1+p_2+p_3)(l_4p_4+l_5p_5)}{(p_1+p_2)(p_4+p_5)+p_3(p_1+p_2+p_4+p_5)}.$$

Это совпадает с формулой (10.2.10). Окончательную оценку для отметки C находим следующим образом.

Рассмотрим ход z_{123} и невязку этого хода $l_{123}-\bar{l}_D$.

Величина

$$-\frac{l_{123}-\bar{l}_D}{n_{12}+n_3} \qquad (10.2.18)$$

представляет собой поправку на одну станцию хода z_{123}, следовательно,

$$\bar{l}_C=l_{12}-\frac{l_{123}-\bar{l}_D}{n_{12}+n_3}\,n_{12}.$$

Напишем это подробнее.

$$\bar{l}_C = \frac{l_1p_1+l_2p_2}{p_1+p_2} -$$

$$- \frac{\dfrac{l_1p_1+l_2p_2+(p_1+p_2)\,l_3}{p_1+p_2} - \dfrac{(l_1p_1+l_2p_2)\,p_3+(p_1+p_2)\,p_3l_3+(p_1+p_2+p_3)\,(l_4p_4+l_5p_5)}{(p_1+p_2)\,p_3+(p_1+p_2+p_3)\,(p_4+p_5)}}{\dfrac{(p_1+p_2+p_3)}{p_3\,(p_1+p_2)}} \cdot \frac{1}{(p_1+p_2)} =$$

$$= \frac{l_1p_1+l_2p_2+l_3p_3}{p_1+p_2+p_3} + \frac{(l_1p_1+l_2p_2)\,p_3^2+(p_1+p_2)\,p_3^2l_3+p_3\,(p_1+p_2+p_3)\,(l_4p_4+l_5p_5)}{(p_1+p_2+p_3)\,[(p_1+p_2)\,p_3+(p_1+p_2+p_3)\,(p_4+p_5)]} =$$

$$= \frac{(p_1+p_2+p_3)\,(p_3+p_4+p_5)\,(l_1p_1+l_2p_2)-(p_1+p_2+p_3)\,(p_4+p_5)\,p_3l_3+p_3\,(p_1+p_2+p_3)\,(l_4p_4+l_5p_5)}{(p_1+p_2+p_3)\,[(p_1+p_2)\,p_3+(p_1+p_2+p_3)\,(p_4+p_5)]} =$$

$$= \frac{(p_3+p_4+p_5)\,(l_1p_1+l_2p_2)-p_3\,(p_4+p_5)\,l_3+p_3\,(l_4p_4+l_5p_5)}{(p_1+p_2)\,(p_4+p_5)+p_3\,(p_1+p_2+p_4+p_5)}.$$

Это также совпадает с формулой (10.2.8). Ввиду совпадения оценок точности этих оценок будут одинаковыми.

Совершенно аналогично ставятся и решаются задачи по выравниванию угломерных ходов.

§ 3. Измерение горизонтальных углов по способу Гаусса — Шрейбера

Пусть из точки O измеряются k углов на ориентиры $A_1, A_2, \ldots, A_k$ (рис. 9). При этом часто действуют по способу, предложенному К. Ф. Гауссом и разработанному германским офицером Шрейбером.

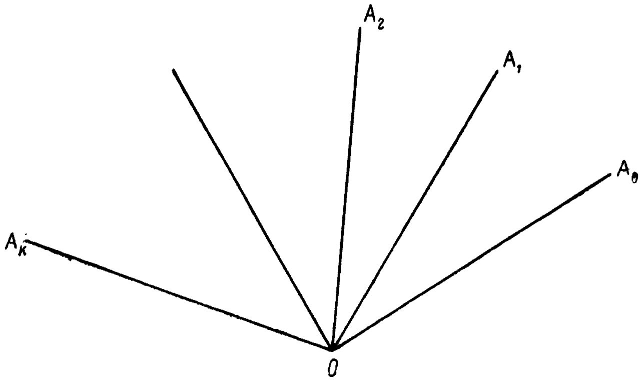

Рис. 9.

Этот способ состоит в том, что измеряются сперва углы между OA_0 и остальными направлениями $OA_1, OA_2, \ldots, OA_k$, затем углы между OA_1 и направлениями $OA_2, \ldots, OA_k$, и т. д., всего $\frac{k(k+1)}{2}$ углов. Среди них отметим k независимых:

$$\alpha_1 = \angle A_0OA_1, \quad \alpha_2 = \angle A_0OA_2, \quad \ldots, \quad \alpha_k = \angle A_0OA_k;$$

эти углы могут служить элементами при уравнивании по элементам. Измерения углов предположим для простоты равноточными.

Мы получим $\frac{k(k+1)}{2}$ фундаментальных уравнений в элементах. Эти уравнения естественным образом разобьются на k систем из $k, k-1, \ldots, 1$ уравнений. Присвоим этим системам номера $0, 1, 2, \ldots, (k-1)$.

Нулевая система будет иметь вид

$$l_1^{(0)} + \Delta_1^{(0)} = \alpha_1,$$
$$l_2^{(0)} + \Delta_2^{(0)} = \alpha_2,$$
$$\cdots\cdots\cdots$$
$$l_k^{(0)} + \Delta_k^{(0)} = \alpha_k,$$

первая система:

$$l_1^{(1)} + \Delta_1^{(1)} = \alpha_2 - \alpha_1,$$
$$\cdots\cdots\cdots\cdots$$
$$l_{k-1}^{(1)} + \Delta_{k-1}^{(1)} = \alpha_k - \alpha_1,$$

j-я система:

$$l_1^{(j)} + \Delta_1^{(j)} = \alpha_{j+1} - \alpha_j,$$
$$\cdots\cdots\cdots\cdots$$
$$l_{k-j}^{(j)} + \Delta_{k-j}^{(j)} = \alpha_k - \alpha_j,$$
$$j = 1, 2, \ldots, (k-1).$$

Матрицу, отвечающую нулевой системе относительно элементов $\alpha_1, \ldots, \alpha_k$, обозначим через $R_{kk}^{(0)}$, Имеем

$$R_{kk}^{(0)} = \begin{Vmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 1 \end{Vmatrix} = E_{kk}. \tag{10.3.1}$$

Для j-й системы ($j = 1, 2, \ldots, k-1$) имеем

$$R_{k-j,\,k}^{(j)} = \left.\begin{Vmatrix} \overbrace{0 \ldots 0}^{(j-1)} & \overbrace{-1}^{j} & \overbrace{1\ 0 \ldots 0}^{(k-j)} \\ 0 \ldots 0 & -1 & 0\ 1 \ldots 0 \\ \cdots\cdots & \cdot & \cdots\cdots \\ 0 \ldots 0 & -1 & 0 \ldots\ldots 1 \end{Vmatrix}\right\}(k-j) \tag{10.3.2}$$

Матрица всей системы наблюдений получит вид

$$X = X_{\frac{k(k+1)}{2},\, k} = \left\| \begin{matrix} R_{kk}^{(0)} \\ \cdots \\ R_{k-1,k}^{(1)} \\ \cdots \\ R_{k-2,k}^{(2)} \\ \cdots \\ \cdots \\ R_{1,k}^{(k-1)} \end{matrix} \right\| \tag{10.3.3}$$

(см. § 2 гл. I).

Матрица системы k нормальных уравнений для оценок элементов $\tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_k$ такова: $C = C_{kk} = X^T X$ (см. гл. VIII). Из записи фундаментальных уравнений в форме (8.3.1) легко замечаем, что эта матрица имеет вид

$$\begin{aligned} C = C_{kk} = X^T X = \left(R_{kk}^{(0)}\right)^T R_{kk}^{(0)} + \left(R_{k-1,k}^{(1)}\right)^T R_{k-1,k}^{(1)} + \ldots + \\ + \left(R_{k-j,k}^{(j)}\right)^T R_{k-j,k}^{(j)} + \ldots + \left(R_{1,k}^{(k-1)}\right)^T R_{1,k}^{(k-1)}. \end{aligned} \tag{10.3.4}$$

Написанные произведения матриц вычисляются непосредственно. Имеем

$$\left(R_{kk}^{(0)}\right)^T R_{kk}^{(0)} = E_{kk} = \left.\left\| \begin{matrix} \overbrace{1 \quad 0 \ \ldots \ 0}^{(k)} \\ 0 \quad 1 \ \ldots \ 0 \\ \cdots\cdots \\ 0 \quad 0 \ \ldots \ 1 \end{matrix} \right\|\right\}(k), \tag{10.3.5}$$

$$\left(R_{k-j,k}^{(j)}\right)^T R_{k-j,k}^{(j)} = \begin{array}{r} \\ (j-1)\left\{\begin{array}{c} \\ \\ \\ \\ \end{array}\right. \\ (k-j+1)\left\{\begin{array}{c} \\ \\ \\ \\ \end{array}\right. \end{array} \begin{array}{c} \begin{matrix} \overbrace{\quad\quad\quad}^{(j-1)} & \overbrace{\quad}^{j} & \overbrace{\quad\quad\quad}^{(k-j)} \end{matrix} \\ \left\| \begin{matrix} 0 \ldots 0 & 0 & 0 \ldots & 0 \\ 0 \ldots 0 & 0 & 0 \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 \ldots 0 & 0 & 0 \ldots & 0 \\ 0 \ldots 0 & k-j & -1 \ldots & -1 \\ 0 \ldots 0 & -1 & 1 \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 \ldots 0 & -1 & 0 \ldots & 1 \end{matrix} \right\| \end{array}, \quad j = 1, 2, \ldots, k-1 \tag{10.3.6}$$

(разумеется, это симметрическая матрица). Отсюда получаем с помощью (10.3.4):

$$C = C_{kk} = \left\| \begin{matrix} k & -1 & \ldots & -1 \\ -1 & k & \ldots & -1 \\ \cdots & \cdots & \cdots & \cdots \\ -1 & -1 & \ldots & k \end{matrix} \right\|.$$

Такой простой вид матрицы C позволит нам дать явное выражение оценок $\tilde{\alpha}_1, \ldots, \tilde{\alpha}_k$ и построить для них доверительные интервалы. Нужно только подсчитать правые части нормальных уравнений [см. (8.3.2)], которые в записи (8.3.2) имеют вид $[x_i l]$ ($i = 1, 2, \ldots, k$; $p_i = 1$). С помощью (10.3.2) и (10.3.3) находим

$$\left.\begin{aligned} &[x_i l] = d_i = \\ &= l_i^{(0)} + l_{i-1}^{(1)} + l_{i-2}^{(2)} + \ldots + l_1^{(i-1)} - \left(l_1^{(i)} + l_2^{(i)} + \ldots + l_{k-i}^{(i)}\right), \\ &\qquad i = 1, 2, \ldots, k-1, \\ &[x_k l] = d_k = l_k^{(0)} + l_{k-1}^{(1)} + \ldots + l_1^{(k-1)}. \end{aligned}\right\} \qquad (10.3.7)$$

Остановимся на геометрическом смысле полученных выражений.

Из определения $l_\gamma^{(\delta)}$, имеем $l_\gamma^{(\delta)} = \alpha_{\delta+\gamma} - \alpha_\gamma$. Будем временно обозначать разность углов α_μ и α_ν через (μ, ν): $\alpha_\mu - \alpha_\nu = (\mu, \nu)$ $(\mu > \nu > 1)$.

Положим еще $\alpha_0 = 0$, так что $(\mu, 0) = \alpha_\mu$. Тогда

$$\left.\begin{aligned} &d_1 = (1, 0) - (2, 1) - (3, 1) - \ldots - (k, 1); \\ &d_i = (i, 0) + (i, 1) + (i, 2) + \ldots \\ &\qquad \ldots + (i, i-1) - (i+1, i) - \ldots - (k, i), \\ &\qquad i = 2, 3, \ldots, k-1, \\ &d_k = (k, 0) + (k, 1) + \ldots + (k, k-1) \end{aligned}\right\} \qquad (10.3.8)$$

Имеем

$$\sum_{i=1}^{k} d_i = (1, 0) + (2, 0) + \ldots + (k, 0). \qquad (10.3.9)$$

В самом деле, в сумме $\sum\limits_{i=1}^{k} d_i$ каждое выражение $(\mu, 0)$ встретится один раз со знаком $+$ и ни разу со знаком $-$. Далее, при $\nu > 0$ выражение (μ, ν) $(\mu > \nu)$ встретится один раз со знаком $+$ и один раз со знаком $-$. Для того чтобы оно встретилось со знаком $+$, необходимо и достаточно, чтобы $i = \mu$, $d_i = d_\mu$; это будет выражение (i, ν). Для вхождения его со знаком $-$ мы должны иметь $i = \nu$, $d_i = d_\nu$; это будет выражение (μ, i), чем и доказано (10.3.9).

Обозначая $S = \sum\limits_{i=1}^{k} d_i$, получаем далее

$$S + d_j = 2(j, 0) + \sum_{\substack{\mu=1 \\ j>1}}^{j-1} \big((j, \mu) + (\mu, 0)\big) - \sum_{\nu=j+1}^{k} \big((\nu, j) + (\nu, 0)\big). \qquad (10.3.10)$$

Составим нормальные уравнения

$$\begin{aligned} k\tilde{\alpha}_1 - \tilde{\alpha}_2 - \tilde{\alpha}_3 - \ldots - \tilde{\alpha}_k &= d_1 \\ -\tilde{\alpha}_1 + k\tilde{\alpha}_2 - \tilde{\alpha}_3 - \ldots - \tilde{\alpha}_k &= d_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ -\tilde{\alpha}_1 - \tilde{\alpha}_2 - \ldots\ldots + k\tilde{\alpha}_k &= d_k. \end{aligned} \qquad (10.3.11)$$

Для решения их можно не находить обратной матрицы C^{-1} (хотя диагональные элементы $\{C^{-1}\}_{ii}$ нам понадобятся для вычисления дисперсий оценок), а поступить следующим образом. Прибавим 2-е, 3-е, ..., k-е уравнение к первому; получим новую систему

$$
\begin{aligned}
\tilde{\alpha}_1 + \tilde{\alpha}_2 + \ldots + \tilde{\alpha}_k &= d_1 + d_2 + \ldots + d_k \\
-\tilde{\alpha}_1 + k\tilde{\alpha}_2 - \ldots - \tilde{\alpha}_k &= d_2 \\
\ldots\ldots\ldots\ldots\ldots \\
-\tilde{\alpha}_1 - \tilde{\alpha}_2 - \ldots + k\tilde{\alpha}_k &= d_k.
\end{aligned}
\tag{10.3.12}
$$

Теперь прибавляем первое уравнение ко всем остальным; получаем

$$
\begin{aligned}
\tilde{\alpha}_1 + \tilde{\alpha}_2 + \ldots + \tilde{\alpha}_k &= d_1 + d_2 + \ldots + d_k \\
(k+1)\,\tilde{\alpha}_2 &= d_1 + 2d_2 + d_3 + \ldots + d_k \\
(k+1)\,\tilde{\alpha}_3 &= d_1 + d_2 + 2d_3 + d_4 + \ldots + d_k \\
\ldots\ldots\ldots\ldots\ldots \\
(k+1)\,\tilde{\alpha}_k &= d_1 + d_2 + \ldots + d_{k-1} + 2d_k.
\end{aligned}
\tag{10.3.13}
$$

Положим, далее, $S = d_1 + d_2 + \ldots + d_k$, тогда имеем

$$\tilde{\alpha}_j = \frac{S + d_j}{k+1}, \quad j = 2,\ 3,\ \ldots,\ k. \tag{10.3.14}$$

Отсюда

$$\tilde{\alpha}_1 = S - \sum_{i=2}^{k} \frac{S + d_i}{k+1} = \frac{S + d_1}{k+1}, \tag{10.3.15}$$

так что формула (10.3.14) годна и для $j = 1$.

На основании (10.3.10) получаем явное выражение оценок

$$\tilde{\alpha}_j = \frac{1}{k+1}\left(2\,(j,\,0) + \sum_{\substack{\mu=1 \\ j>1}}^{j-1} \big((j,\mu) + (\mu,0)\big) - \sum_{\nu=j+1}^{k} \big((\nu,\,j) + (\nu,\,0)\big)\right). \tag{10.3.16}$$

Найдем дисперсии оценок $\tilde{\alpha}_j$. Имеем [см. (8.2.5)]

$$D(\tilde{\alpha}_j) = \sigma^2 \{C^{-1}\}_{jj}. \tag{10.3.17}$$

Далее,

$$C = \begin{Vmatrix} k & -1 & \ldots & -1 \\ -1 & k & \ldots & -1 \\ \cdot & \cdot & \cdot & \cdot \\ -1 & -1 & \ldots & k \end{Vmatrix}.$$

Величина $\{C^{-1}\}_{jj}$ равна j-му диагональному минору матрицы C, деленному на $\det(C)$.

Для вычисления $\det(C)$ прибавляем 2-ю, ..., k-ю строку к первой; получим

$$\begin{Vmatrix} 1 & 1 & \ldots\ldots & 1 \\ -1 & k & \ldots\ldots & -1 \\ \cdot & \cdot & \cdot\cdot\cdot\cdot\cdot\cdot & \cdot \\ -1 & -1 & \ldots\ldots & k \end{Vmatrix},$$

и затем прибавляем первую строку ко всем остальным. Тогда

$$\det(C) = \begin{Vmatrix} 1 & 1 & \ldots\ldots & & 1 \\ 0 & k+1 & 0 & \ldots & 0 \\ 0 & 0 & k+1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \ldots & \ldots & \ldots & k+1 \end{Vmatrix} = (k+1)^{k-1}. \qquad (10.3.18)$$

Далее, все диагональные миноры равны между собой. Первый из них равен детерминанту

$$\left.\begin{Vmatrix} \overbrace{k \quad -1 \ \ldots \ -1}^{(k-1)} \\ -1 \quad k \ \ldots \ -1 \\ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \ \cdot \\ -1 \quad \ldots\ldots \quad k \end{Vmatrix}\right\}(k-1).$$

Прибавляя 2-ю, ..., $(k-1)$-ю строку к первой, получаем

$$\begin{Vmatrix} 2 & 2 & \ldots & 2 \\ -1 & k & \ldots & -1 \\ \cdot & \cdot & \cdot & \cdot \\ -1 & \ldots & \ldots & k \end{Vmatrix} = 2 \begin{Vmatrix} 1 & \ldots & \ldots & 1 \\ -1 & k & \ldots & -1 \\ \cdot & \cdot & \cdot & \cdot \\ -1 & -1 & \ldots & k \end{Vmatrix} = 2(k+1)^{k-2}. \qquad (10.3.19)$$

Последнее равенство можно получить, прибавляя первую строку ко всем остальным. Отсюда, деля (10.3.19) на (10.3.18), находим

$$\{C^{-1}\}_{jj} = \frac{2}{k+1} \qquad (10.3.20)$$

и

$$D(\tilde{\alpha}_j) = \frac{2\sigma^2}{k+1}; \quad \sigma(\tilde{\alpha}_j) = \sigma\sqrt{\frac{2}{k+1}}, \qquad (10.3.21)$$
$$j = 1, 2, \ldots, k.$$

Мы видим, что штандарт получаемых оценок пропорционален σ — штандарту единичного измерения и обратно пропорционален $\sqrt{k+1}$, где k — число направлений на ориентиры.

Перейдем к построению доверительных интервалов для оценивания $\tilde{\alpha}_j$. Для этого надлежит построить выражение $[p\tilde{v}\tilde{v}]$. Из (8.3.5) видим, что

$$[p\tilde{v}\tilde{v}] = [\tilde{v}\tilde{v}] = [ll] - [x_1 l]\,\tilde{a}_1 - \ldots - [x_k l]\,\tilde{a}_k.$$

Далее, согласно (9.3.7),

$$[x_i l] = d_i,$$
$$[ll] = \sum_{\substack{i, j=0 \\ i>j}}^{k} \left((i, j)\right)^2. \tag{10.3.22}$$

Таким образом, найдя оценки $\tilde{a}_1, \ldots, \tilde{a}_k$, вычисляем $[p\tilde{v}\tilde{v}]$ по формуле

$$[p\tilde{v}\tilde{v}] = \sum_{\substack{i, j=0 \\ i>j}}^{k} \left((i, j)\right)^2 - \sum_{i=1}^{k} d_i \tilde{a}_i. \tag{10.3.23}$$

После этого можем построить доверительные интервалы $I_\gamma^{(i)}$ для оценивания $\tilde{a}_i$ с помощью формулы (8.3.14). Мы должны взять в ней

$$n = k; \quad N = \frac{k(k+1)}{2}; \quad N - n = \frac{k(k-1)}{2}, \tag{10.3.24}$$
$$\{C^{-1}\}_{ii} = \frac{2}{k+1}.$$

Пользуясь табл. I распределения Стьюдента, по заданной надежности p_0 и при числе степеней свободы $N - n = \frac{k(k-1)}{2}$ находим γ_0. Доверительный интервал

$$I_{\gamma_0}^{(i)} = \left[\tilde{a}_i - 2\gamma_0 \sqrt{\frac{[p\tilde{v}\tilde{v}]}{k(k^2-1)}}, \quad \tilde{a}_i + 2\gamma_0 \sqrt{\frac{[p\tilde{v}\tilde{v}]}{k(k^2-1)}}\right] \tag{10.3.25}$$

будет накрывать измеряемый угол α_i с вероятностью p_0. Для оценивания σ с помощью доверительных интервалов употребляем табл. III. По заданной надежности p_0 находим числа γ_1 и γ_2. Доверительный интервал

$$I = \left[\gamma_1 \sqrt{\frac{2[p\tilde{v}\tilde{v}]}{k(k-1)}}, \quad \gamma_2 \sqrt{\frac{2[p\tilde{v}\tilde{v}]}{k(k-1)}}\right] \tag{10.3.26}$$

накрывает σ с вероятностью p_0.

Помимо углов $\alpha_1, \ldots, \alpha_k$, нам еще могут быть нужны углы $\alpha_\mu - \alpha_\nu$ $(\mu > \nu)$ между направлениями на ориентиры. Оценка $\alpha_\mu - \alpha_\nu$ получается в виде $\tilde{\alpha}_\mu - \tilde{\alpha}_\nu$ (см. гл. VII). Найдем ее дисперсию и построим для нее доверительные интервалы, пользуясь результатами гл. VII.

Пусть $\tilde{\alpha} = \tilde{\alpha}_\mu - \tilde{\alpha}_\nu$. Мы должны прежде всего подсчитать $D(\tilde{\alpha})$. Имеем

$$D(\tilde{\alpha}) = E\left(\tilde{\alpha}_\mu - \alpha_\mu - (\tilde{\alpha}_\nu - \alpha_\nu)\right)^2 =$$
$$= D(\tilde{\alpha}_\mu) + D(\tilde{\alpha}_\nu) - 2E(\tilde{\alpha}_\mu - \alpha_\mu)(\tilde{\alpha}_\nu - \alpha_\nu) =$$
$$= \sigma^2\left(\{C^{-1}\}_{\mu\mu} + \{C^{-1}\}_{\nu\nu} - 2\{C^{-1}\}_{\mu\nu}\right). \tag{10.3.27}$$

Таким образом, нам нужно найти все элементы обратной матрицы C^{-1}, помимо ранее найденных диагональных элементов.

Матрица C имела вид

$$C = \begin{Vmatrix} k & -1 & \dots & -1 \\ -1 & k & \dots & -1 \\ \cdot & \cdot & \cdot & \cdot \\ -1 & -1 & \dots & k \end{Vmatrix}.$$

Вычислим элемент $\{C^{-1}\}_{ji}$. Пусть $i < j$. Положим в C все элементы j-го столбца равными нулю, кроме элемента i-й строки этого столбца, который равен единице, и рассмотрим матрицу

$$\begin{array}{cc} & \begin{array}{cccccccccc} & & & i\downarrow & & & j\downarrow & & & \end{array} \\ \begin{array}{c} \\ \\ \\ i\rightarrow \\ \\ \\ \\ \\ \\ \\ \end{array} & \begin{Vmatrix} k & -1 & \dots & -1 & \dots & -1 & 0 & -1 & \dots & -1 \\ -1 & k & \dots & -1 & \dots & -1 & 0 & -1 & \dots & -1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & -1 & \dots & k & \dots & -1 & 1 & -1 & \dots & -1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & -1 & \dots & -1 & \dots & k & 0 & -1 & \dots & -1 \\ -1 & -1 & \dots & -1 & \dots & -1 & 0 & -1 & \dots & -1 \\ -1 & -1 & \dots & -1 & \dots & -1 & 0 & k & \dots & -1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & -1 & \dots & -1 & \dots & -1 & 0 & -1 & \dots & k \end{Vmatrix} \end{array}.$$

Прибавим 2-ю, ..., k-ю строку к первой строке, а затем прибавим получившуюся первую ко всем остальным. Получим

$$\begin{array}{cc} & \begin{array}{cccccccccc} & & & i\downarrow & & & j\downarrow & & & \end{array} \\ \begin{array}{c} \\ \\ \\ i\rightarrow \\ \\ \\ \\ \\ \\ \\ \end{array} & \begin{Vmatrix} 1 & 1 & \dots & 1 & \dots & 1 & 1 & 1 & \dots & 1 \\ 0 & k+1 & \dots & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & k+1 & \dots & 0 & 2 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 0 & \dots & k+1 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 0 & 1 & k+1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 0 & \dots & 0 & 1 & 0 & \dots & k+1 \end{Vmatrix} \end{array}.$$

Все элементы j-й строки, кроме элемента (jj), равны нулю, следовательно, j-й столбец и j-ю строку можно вычеркнуть. Останется определитель $(k-1)$-порядка, у которого ниже главной диагонали — нули, а на главной диагонали $(k-2)$ элементов равны $(k+1)$. Этот определитель равен $(k+1)^{k-2}$, но

$$\det(C) = (k+1)^{k-1}.$$

Таким образом,

$$\{C^{-1}\}_{ij} = \{C^{-1}\}_{ji} = \frac{1}{k+1}.$$

Пользуясь формулами (10.3.27) и (10.3.20), находим

$$D(\tilde{\alpha}) = \frac{\sigma^2}{k+1}(2+2-2) = \frac{2\sigma^2}{k+1}, \tag{10.3.28}$$

$$\sigma(\tilde{\alpha}) = \sigma\sqrt{\frac{2}{k+1}}. \tag{10.3.29}$$

Сравнивая с формулой (10.3.21), видим, что

$$D(\tilde{\alpha}_\mu - \tilde{\alpha}_\nu) = D(\tilde{\alpha}_\mu) = D(\tilde{\alpha}_\nu) = \frac{2\sigma^2}{k+1} (\mu > \nu),$$

т. е. оценки углов статистически связаны так, что оценка для разности углов $\tilde{\alpha}$ имеет ту же дисперсию, что и оценка для каждого угла.

Перейдем теперь к построению доверительных интервалов для $\tilde{\alpha}$. Пользуясь формулой (7.4.5), составляем величину

$$\tilde{S}^2 = \frac{[p\tilde{v}\tilde{v}]}{N-n} \frac{D(\tilde{\alpha})}{\sigma^2} = \frac{4[p\tilde{v}\tilde{v}]}{k(k^2-1)}. \tag{10.3.30}$$

Формула (7.4.8) показывает, что при выборе по надежности p_0 и $\frac{k(k-1)}{2}$ степеням свободы в табл. I распределения Стьюдента числа γ_0, получаем доверительный интервал

$$I_\gamma = \left[\tilde{\alpha} - 2\gamma_0\sqrt{\frac{[p\tilde{v}\tilde{v}]}{k(k^2-1)}},\ \tilde{\alpha} + 2\gamma_0\sqrt{\frac{[p\tilde{v}\tilde{v}]}{k(k^2-1)}}\right], \tag{10.3.31}$$

накрывающий $\alpha_\mu - \alpha_\nu$ с вероятностью p_0.

ГЛАВА XI

ОЦЕНИВАНИЕ РЕЗУЛЬТАТОВ ПРЯМЫХ И ОБРАТНЫХ ЗАСЕЧЕК

§ 1. Прямая засечка более чем с двух пунктов. Доверительные области*)

Как было указано во введении, задачи оценивания результатов прямой засечки часто возникают на практике. В задачах геодезии, в артиллерийской инструментальной разведке, в задачах звукометрии и радиолокации часто возникает необходимость определения пункта при помощи засечек на него с нескольких твердых пунктов (пунктов с известными координатами).

Пусть из n твердых точек $A_1, A_2, \ldots, A_n$ с заданными координатами (рис. 10) измерены углы между линиями $A_1A_1', A_2A_2', \ldots$ $\ldots, A_nA_n'$ с известными дирекционными углами и направлением на искомую точку O. В силу случайных погрешностей в измерении углов эти линии не пересекутся в одной точке и не дадут точки O, так что возникнет задача оценивания положения точки O по данным наблюдений. Считая погрешности малыми, мы можем вычислить сначала по элементарным формулам тригонометрии приближенные координаты (x_0', y_0') при помощи какого-либо из получившихся $n-1$ треугольников, все углы которого не менее 30°. Для линеаризации проблемы будем считать поправки $x_0'-x_0$, $y_0'-y_0$ малыми, так что их квадратами можно пренебречь сравнительно с выбранной единицей длины.

Точку O' с координатами (x_0', y_0') можно рассматривать как вспомогательную твердую точку; длины линий A_iO' мы обозначим через d_i' и дирекционные углы их $\alpha_i^{(0)}$ можно считать точно известными. Если α_i есть истинное значение дирекционного угла A_iO, то

$$\alpha_i = \alpha_i^{(0)} + \Delta\alpha_i. \tag{11.1.1}$$

---

*) Геометрическая часть задачи о прямой и обратной засечке здесь трактуется по книге А. С. Чеботарёва „Геодезия", ч. II, М. 1949. Добавлена лишь конструкция доверительных областей.

Имеем для соответствующим образом выбранной системы координат

$$\operatorname{tg}\alpha_i^{(0)} = \frac{y_0' - y_i}{x_0' - x_i}; \quad y_0 - y_0' = \Delta y_0, \quad x_0 - x_0' = \Delta x_0. \tag{11.1.2}$$

Будем измерять углы в радиальной мере; тогда получим

$$\frac{\Delta\alpha_i}{\cos^2\alpha_i^{(0)}} \approx \frac{(x_0' - x_i)\,\Delta y_0 - (y_0' - y_i)\,\Delta x_0}{(x_0' - x_i)^2} \tag{11.1.3}$$

(равенство верно с точностью до величин второго порядка малости).

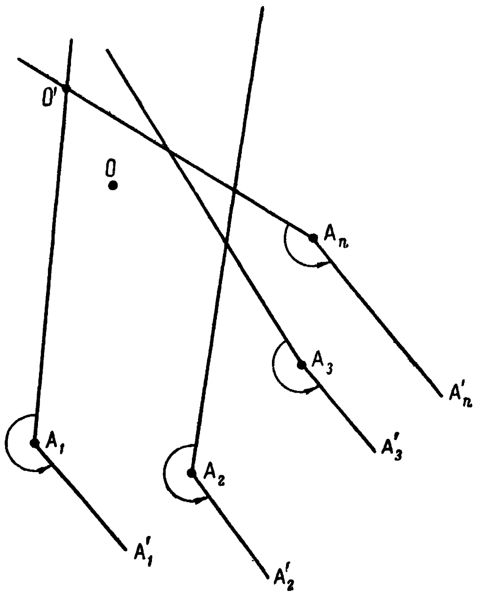

Рис. 10.

Далее, в указанной системе координат

$$x_0' - x_i = d_i' \cos\alpha_i^{(0)};$$
$$y_0' - y_i = d_i' \sin\alpha_i^{(0)}. \tag{11.1.4}$$

Ввиду (11.1.3) получим

$$\Delta\alpha_i \approx \frac{\cos\alpha_i^{(0)}}{d_i'}\Delta y_0 - \frac{\sin\alpha_i^{(0)}}{d_i'}\Delta x_0. \tag{11.1.5}$$

Нам известны, однако, не разности $\Delta\alpha_i = \alpha_i - \alpha_i^{(0)}$, а значения $l_i = \alpha_i - \alpha_i^{(0)} + \Delta_i = \Delta\alpha_i + \Delta_i$ $(i = 1, 2, \ldots, n)$, поскольку наши наблюдения несут погрешности Δ_i. Эти погрешности будем считать равноточными, несмещенными, независимыми и нормальными $N(0, \sigma)$.

Обозначим еще

$$-\frac{\sin\alpha_i^{(0)}}{d_i'} = a_i; \quad \frac{\cos\alpha_i^{(0)}}{d_i'} = b_i \quad (i = 1, 2, \ldots, n). \tag{11.1.6}$$

Тогда получим (с точностью до величин второго порядка малости)

$$l_i - \Delta_i = a_i\,\Delta x_0 + b_i\,\Delta y_0 \quad (i = 1, 2, \ldots, n). \tag{11.1.7}$$

Этим задача сводится к задаче, рассматривавшейся в гл. VI (непрямые безусловные измерения). Наличие лишь двух величин Δx_0 и Δy_0 позволяет дать явные выражения для точечных оценок; мы дадим их, а затем построим доверительные области для пары $(\Delta x_0, \Delta y_0)$.

Матрица коэффициентов имеет вид

$$X=\begin{Vmatrix} a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_n & b_n \end{Vmatrix}. \qquad (11.1.8)$$

Будем считать, как обычно, $n>2$, так что имеем по крайней мере три твердые точки.

Имеем: ранг $(X)=2$. В самом деле,

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}=\begin{vmatrix} -\frac{\sin\alpha_1^{(0)}}{d_1'} & \frac{\cos\alpha_1^{(0)}}{d_1'} \\ -\frac{\sin\alpha_2^{(0)}}{d_2'} & \frac{\cos\alpha_2^{(0)}}{d_2'} \end{vmatrix}=\frac{1}{d_1'd_2'}\left(\sin\alpha_2^{(0)}\cos\alpha_1^{(0)}-\sin\alpha_1^{(0)}\cos\alpha_2^{(0)}\right)=$$

$$=\frac{1}{d_1'd_2'}\sin\left(\alpha_2^{(0)}-\alpha_1^{(0)}\right)\neq 0 .$$

Для составления нормальных уравнений воспользуемся сводкой формул гл. VI (§ 11). Из (6.11.2), принимая во внимание, что $p_i=1$, получаем нормальные уравнения

$$\left.\begin{aligned} [aa]\widetilde{\Delta x}+[ab]\widetilde{\Delta y}&=[al], \\ [ab]\widetilde{\Delta x}+[bb]\widetilde{\Delta y}&=[bl], \end{aligned}\right\} \qquad (11.1.9)$$

где через $\widetilde{\Delta x}$, $\widetilde{\Delta y}$ обозначены оценки для Δx_0 и Δy_0. Отсюда находим непосредственно

$$\left.\begin{aligned} \widetilde{\Delta x}&=\frac{[bb][al]-[ab][bl]}{[aa][bb]-[ab]^2}, \\ \widetilde{\Delta y}&=\frac{[aa][bl]-[ab][al]}{[aa][bb]-[ab]^2}. \end{aligned}\right\} \qquad (11.1.10)$$

Мы получили точечную оценку $(\widetilde{\Delta x}, \widetilde{\Delta y})$ для истинных поправок Δx_0, Δy_0. Найдем корреляционную матрицу вектора оценок $(\widetilde{\Delta x}, \widetilde{\Delta y})$, которая равна $\sigma^2 C^{-1}$. Имеем

$$C=\begin{Vmatrix} [aa] & [ab] \\ [ab] & [bb] \end{Vmatrix}. \qquad (11.1.11)$$

Положим $D=[aa][bb]-]ab]^2$.

Искомая корреляционная матрица будет

$$B=\sigma^2 C^{-1}=\frac{\sigma^2}{D}\begin{Vmatrix} [bb] & -[ab] \\ -[ab] & [aa] \end{Vmatrix}. \qquad (11.1.12)$$

Отсюда

$$D(\widetilde{\Delta x}) = \frac{\sigma^2 [bb]}{D}; \quad \sigma(\widetilde{\Delta x}) = \sigma \sqrt{\frac{[bb]}{D}}; \tag{11.1.13}$$

$$D(\widetilde{\Delta y}) = \frac{\sigma^2 [aa]}{D}; \quad \sigma(\widetilde{\Delta y}) = \sigma \sqrt{\frac{[aa]}{D}}. \tag{11.1.14}$$

Из этих формул, характеризующих точность оценивания, видим, что при небольшом числе n твердых точек $A_1, \ldots, A_n$ точность достигается лишь за счет уменьшения σ — штандарта погрешности в измерении углов.

Для дальнейшего мы должны получить выражение $[p\tilde{v}\tilde{v}]$. Пользуясь формулой (6.11.3), находим ($\tilde{v}$ отвечает обозначениям гл. VI):

$$[p\tilde{v}\tilde{v}] = [\tilde{v}\tilde{v}] = [ll] - [al]\,\widetilde{\Delta x} - [bl]\,\widetilde{\Delta y}, \tag{11.1.15}$$

где $\widetilde{\Delta x}$ и $\widetilde{\Delta y}$ надлежит взять из формул (11.1.10).

Мы имеем далее (см. гл. VI, § 9)

$$E\left(\frac{[p\tilde{v}\tilde{v}]}{n-2}\right) = \sigma^2, \tag{11.1.16}$$

так что $\frac{[p\tilde{v}\tilde{v}]}{n-2}$ будет несмещенной оценкой для σ^2. Как указано в гл. VI, § 9, мы можем применить $[p\tilde{v}\tilde{v}]$ для оценивания σ с помощью доверительных интервалов; вряд ли, однако, это является целесообразным в случае засечек, когда точность работы измеряющих инструментов естественнее оценивать по предыдущей информации о работе их в данных условиях. Более целесообразной является задача построения доверительных областей, накрывающих искомую истинную поправку $(\Delta x_0, \Delta y_0)$. Здесь будут предложены две конструкции таких доверительных областей. Начнем сперва с построения доверительных интервалов порознь для Δx_0 и для Δy_0 по формуле (6.11.10).

Здесь пользуемся табл. I распределения Стьюдента с $k = n - 2$ степенями свободы. Из (11.1.15) при помощи (11.1.10) вычисляем $[p\tilde{v}\tilde{v}] = [\tilde{v}\tilde{v}]$. При надежности p находим число γ, беря вторым входом $(n-2)$ степеней свободы. Далее,

$$C^{-1} = \frac{1}{D} \begin{Vmatrix} [bb] & -[ab] \\ -[ab] & [aa] \end{Vmatrix}. \tag{11.1.17}$$

Таким образом, получаем для Δx_0 доверительный интервал

$$I_\gamma^{(\Delta x_0)} = \left[\widetilde{\Delta x} - \gamma \sqrt{\frac{[bb]}{D} \frac{[\tilde{v}\tilde{v}]}{n-2}}, \quad \widetilde{\Delta x} + \gamma \sqrt{\frac{[bb]}{D} \frac{[\tilde{v}\tilde{v}]}{n-2}}\right]. \tag{11.1.18}$$

Для Δy_0 доверительный интервал имеет вид

$$I_\gamma^{(\Delta y_0)} = \left[\widetilde{\Delta y} - \gamma \sqrt{\frac{[aa]}{D} \frac{[\tilde{v}\tilde{v}]}{n-2}}, \quad \widetilde{\Delta y} + \gamma \sqrt{\frac{[aa]}{D} \frac{[\tilde{v}\tilde{v}]}{n-2}}\right]. \tag{11.1.19}$$

Пусть $\mathfrak{A}$ — событие, состоящее в том, что интервал $I_\gamma^{(\Delta x_0)}$ накрывает Δx_0 (что будем записывать так: $\Delta x_0 \in I_\gamma^{(\Delta x_0)}$) и $\mathfrak{B}$ — событие $\Delta y_0 \in I_\gamma^{(\Delta y_0)}$.

Вероятности каждого из событий порознь равны p:

$$P(\mathfrak{A}) = p; \quad P(\mathfrak{B}) = p.$$

Совмещение событий $\mathfrak{A}$ и $\mathfrak{B}$ будет означать, что точка $(\Delta x_0, \Delta y_0)$ накрывается прямоугольником Q_γ, проекции которого на оси координат будут $I_\gamma^{(\Delta x_0)}$ и $I_\gamma^{(\Delta y_0)}$ (чем он и определяется). Однако это совмещение событий, т. е. событие ($\mathfrak{A}$ и $\mathfrak{B}$), не будет иметь вероятность, равную произведению $P(\mathfrak{A})$ и $P(\mathfrak{B})$, так как события $\mathfrak{A}$ и $\mathfrak{B}$ зависимы. Мы можем, однако, написать, пользуясь элементарными теоремами теории вероятностей:

$$P(\mathfrak{A} \text{ и } \mathfrak{B}) = 1 - P(\bar{\mathfrak{A}} \text{ или } \bar{\mathfrak{B}}) =$$
$$= 1 - P(\bar{\mathfrak{A}}) - P(\bar{\mathfrak{B}}) + P(\bar{\mathfrak{A}} \text{ и } \bar{\mathfrak{B}}) \geqslant 1 - P(\mathfrak{A}) - P(\mathfrak{B}), \tag{11.1.20}$$

где через $\bar{\mathfrak{A}}$ и $\bar{\mathfrak{B}}$ обозначены события, противоположные событиям $\mathfrak{A}$ и $\mathfrak{B}$.

Имеем, далее,

$$P(\bar{\mathfrak{A}}) = 1 - P(\mathfrak{A}) = 1 - p; \quad P(\bar{\mathfrak{B}}) = 1 - p.$$

Таким образом, из (11.1.20) получаем

$$P(\mathfrak{A} \text{ и } \mathfrak{B}) \geqslant 1 - (1-p) - (1-p) = 2p - 1. \tag{11.1.21}$$

Например, если $p = 0{,}9$, то $P(\mathfrak{A}$ и $\mathfrak{B}) \geqslant 0{,}8$, а если $p = 0{,}95$, то $P(\mathfrak{A}$ и $\mathfrak{B}) \geqslant 0{,}9$.

Таким образом, мы можем сформулировать следующую теорему.

Теорема 11.1.1. *Прямоугольная область Q_γ, определяемая своими проекциями на оси координат — доверительными интервалами* (11.1.18) *и* (11.1.19) — *накрывает точку $(\Delta x_0, \Delta y_0)$ с вероятностью не меньшей, чем $2p - 1$, где p — надежность, взятая для доверительных интервалов.*

Область Q_γ будем называть *доверительным прямоугольником.*

Как видно из формул (11.1.18) и (11.1.19), ширина и высота доверительного прямоугольника равны соответственно

$$2\gamma \sqrt{\frac{[bb]}{D} \frac{[\tilde{v}\tilde{v}]}{n-2}} \quad \text{и} \quad 2\gamma \sqrt{\frac{[aa]}{D} \frac{[\tilde{v}\tilde{v}]}{n-2}}. \tag{11.1.22}$$

Заметим, что при надежности $p = 0{,}95$ и небольшом числе степеней свободы $(n-2)$ число γ может оказаться сравнительно большим. Так как $E\left(\frac{[\tilde{v}\tilde{v}]}{n-2}\right) = \sigma^2$, то из (11.1.22) видим, что при таком значении p применение доверительного прямоугольника Q_γ будет целесообразным лишь при достаточно большой точности основных измерений (достаточно малом σ^2).

Укажем теперь еще один вид доверительной области, накрывающей точку $(\Delta x_0, \Delta y_0)$ с заданной вероятностью p. Как явствует из теоремы 6.6.3, случайный вектор $(\widetilde{\Delta x}, \widetilde{\Delta y})$ статистически независим от вектора $[\tilde{v}\tilde{v}]$, определяемого формулой (11.1.15). Корреляционная матрица этого вектора равна $\sigma^2 C^{-1}$. Найдем плотность распределения нормального вектора $(\widetilde{\Delta x}-\Delta x_0, \widetilde{\Delta y}-\Delta y_0)$. Согласно формуле (2.3.10), эта плотность распределения $f(\xi, \eta)$ имеет вид

$$f(\xi, \eta)=\frac{\sqrt{D}}{2\pi\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}\left[[aa]\xi^2+2[ab]\xi\eta+[bb]\eta^2\right]\right\}. \qquad (11.1.23)$$

Положим $Q(\xi, \eta)=([aa]\xi^2+2[ab]\xi\eta+[bb]\eta^2)$ и докажем, что величина $\frac{1}{\sigma^2}Q(\xi, \eta)$ распределена по закону χ_2^2. Имеем

$$\varphi(x)=P\left(\frac{1}{\sigma^2}Q(\xi, \eta)<x\right)=\iint\limits_{\frac{1}{\sigma^2}Q(\xi, \eta)<x} f(\xi, \eta)\,d\xi\,d\eta=$$

$$=\frac{\sqrt{D}}{2\pi\sigma^2}\iint\limits_{\frac{1}{\sigma^2}Q(\xi, \eta)<x}\exp\left\{-\frac{1}{2\sigma^2}Q(\xi, \eta)\right\}d\xi\,d\eta. \qquad (11.1.24)$$

Полагая $\frac{\xi}{\sigma}=u_1$, $\frac{\eta}{\sigma}=u_2$, имеем

$$\varphi(x)=\frac{\sqrt{D}}{2\pi}\iint\limits_{Q(u_1, u_2)<x}\exp\left\{-\frac{1}{2}Q(u_1, u_2)\right\}du_1\,du_2. \qquad (11.1.25)$$

Введем матрицу $U=U_{21}=\begin{Vmatrix} u_1 \\ u_2 \end{Vmatrix}$, тогда (см. гл. II, § 3) получим

$$\varphi(x)=\frac{\sqrt{D}}{2\pi}\iint\limits_{U^TCU<x}\exp\left\{-\frac{1}{2}U^TCU\right\}du_1\,du_2. \qquad (11.1.26)$$

Введем ортогональное преобразование F, полагая

$$U'=\begin{Vmatrix} u_1' \\ u_2' \end{Vmatrix}=FU \quad \text{так, что} \quad F^TCF=\begin{Vmatrix} d_1 & 0 \\ 0 & d_2 \end{Vmatrix}$$

есть диагональная матрица. При этом $d_1 d_2=D$. Получим

$$\varphi(x)=\frac{\sqrt{D}}{2\pi}\iint\limits_{d_1 u_1'^2+d_2 u_2'^2<x}\exp\left\{-\frac{1}{2}\left(d_1 u_1'^2+d_2 u_2'^2\right)\right\}du_1'\,du_2'.$$

Положим еще $w_1 = u_1' \sqrt{d_1}$; $w_2 = u_2' \sqrt{d_2}$, тогда

$$\varphi(x) = \frac{1}{2\pi} \iint\limits_{w_1^2 + w_2^2 < x} \exp\left\{-\frac{1}{2}\left(w_1^2 + w_2^2\right)\right\} dw_1\, dw_2. \qquad (11.1.27)$$

Но справа стоит величина, по определению равная $P(\chi_2^2 < x)$ (см. гл. II, § 6). Таким образом,

$$\varphi(x) = P(\chi_2^2 < x)$$

и

$$\frac{1}{\sigma^2} Q(\xi, \eta) = \chi_2^2, \qquad (11.1.28)$$

что и требовалось доказать.

Теперь мы можем построить доверительную область для $(\Delta x_0, \Delta y_0)$. Имеем

$$\frac{1}{\sigma^2} Q(\xi, \eta) = \chi_2^2 \quad \text{и} \quad \frac{[\tilde{v}\tilde{v}]}{\sigma^2} = \chi_{n-2}^2,$$

причем эти величины независимы одна от другой. Таким образом, величина

$$\frac{\frac{1}{\sigma^2} Q(\xi, \eta)}{\frac{[\tilde{v}\tilde{v}]}{\sigma^2}} = \frac{Q(\xi, \eta)}{[\tilde{v}\tilde{v}]} \qquad (11.1.29)$$

имеет распределение Фишера (см. гл. II, § 6) $F_{2, n-2}(x)$. Записывая подробнее выражение $Q(\xi, \eta)$ и $[\tilde{v}\tilde{v}]$, получаем следующую теорему.

Теорема 11.1.2*). *Величина*

$$\frac{Q(\xi, \eta)}{[\tilde{v}\tilde{v}]} = \frac{[aa](\widetilde{\Delta x} - \Delta x_0)^2 + 2[ab](\widetilde{\Delta x} - \Delta x_0)(\widetilde{\Delta y} - \Delta y_0) + [bb](\widetilde{\Delta y} - \Delta y_0)^2}{[ll] - [al]\widetilde{\Delta x} - [bl]\widetilde{\Delta y}} \qquad (11.1.30)$$

имеет распределение Фишера $F_{2, n-2}(x)$.

Последнее распределение имеет весьма простой вид. Из формулы (2.6.13) видим, что

$$F'_{2, n-2}(x) = f_{2, n-2}(x) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma(1)\Gamma\left(\frac{n}{2} - 1\right)} \cdot \frac{1}{(1+x)^{\frac{n}{2}}} = \frac{\frac{n}{2} - 1}{(1+x)^{\frac{n}{2}}} \quad (x > 0). \qquad (11.1.31)$$

*) Эта теорема, возможно, является новой.

Таким образом, при $\gamma > 0$ имеем

$$F_{2,\,n-2}(\gamma)=\left(\frac{n-2}{2}\right)\int\limits_0^{\gamma}\frac{dx}{(1+x)^{\frac{n}{2}}}=1-\frac{1}{(1+\gamma)^{\frac{n-2}{2}}}. \qquad (11.1.32)$$

Поэтому

$$P\left(\frac{Q(\xi,\,\eta)}{[\tilde{v}\tilde{v}]}<\gamma\right)=P(Q(\xi,\,\eta)<\gamma\,[\tilde{v}\tilde{v}])=1-\frac{1}{(1+\gamma)^{\frac{n-2}{2}}}. \qquad (11.1.33)$$

Если задаться надежностью p, то при $F_{2,\,n-2}(\gamma)=p$ имеем

$$p=1-\frac{1}{(1+\gamma)^{\frac{n-2}{2}}};\qquad \gamma=\frac{1}{(1-p)^{\frac{2}{n-2}}}-1. \qquad (11.1.34)$$

Событие, состоящее в том, что $Q(\xi,\,\eta)<\gamma\,[\tilde{v}\tilde{v}]$, можно записать в виде

$$Q(\Delta x_0-\widetilde{\Delta x};\quad \Delta y_0-\widetilde{\Delta y})<\gamma\,([ll]-[al]\,\widetilde{\Delta x}-[bl]\,\widetilde{\Delta y}) \qquad (11.1.35)$$

и истолковать следующим образом: при заданных $\widetilde{\Delta x}$, $\widetilde{\Delta y}$ точка $(\Delta x_0,\ \Delta y_0)$ лежит внутри некоторого эллипса с центром $(\widetilde{\Delta x},\ \widetilde{\Delta y})$ с известным отношением полуосей и положением на плоскости, определяемым квадратичной формой $Q(\xi,\,\eta)=[aa]\xi^2+2\,[ab]\,\xi\eta+[bb]\,\eta^2$ с известными коэффициентами, и с большой полуосью, определенной правой частью (11.1.35). Указанный эллипс, который мы обозначим $Э_\gamma$ и который имеет уравнение

$$Q(\Delta x-\widetilde{\Delta x},\ \Delta y-\widetilde{\Delta y})=\gamma\,([ll]-[al]\,\widetilde{\Delta x}-[bl]\,\widetilde{\Delta y}) \qquad (11.1.36)$$

(Δx, Δy — текущие координаты), накрывает точку $(\Delta x_0,\ \Delta y_0)$. В изучаемой нами ситуации, когда $(\widetilde{\Delta x},\ \widetilde{\Delta y})$ — случайный вектор, этот эллипс $Э_\gamma$ является случайным и накрывает точку $(\Delta x_0,\ \Delta y_0)$ с вероятностью p_0, где p_0 и γ связаны соотношением (11.1.34). Это приводит нас к теореме 11.1.2.

Рассмотрим еще, что произойдет при увеличении n. При больших n будем иметь

$$\left(\frac{1}{1-p_0}\right)^{\frac{2}{n-2}}=\exp\frac{2}{n-2}\ln\frac{1}{1-p_0}=1+\frac{2}{n-2}\ln\frac{1}{1-p_0}+\delta_n,$$

где $|\delta_n|\leqslant\frac{K(p_0)}{n^2}$. Здесь $K(p_0)$ — постоянная, зависящая только от p_0. Таким образом,

$$\gamma=\frac{2}{n-2}\ln(1-p_0)+\delta_n\approx\frac{2}{n-2}\ln\frac{1}{1-p_0}$$

при больших n.

Ввиду этого при больших n доверительный эллипс будет иметь приближенно вид

$$Q(\Delta x - \widetilde{\Delta x},\ \Delta y - \widetilde{\Delta y}) \approx \frac{2}{n-2} \ln \frac{1}{1-p_0} [\tilde{v}\tilde{v}]. \qquad (11.1.37)$$

Теорема 11.1.3. *При заданной надежности p_0 доверительный эллипс $Э_\gamma$ с уравнением*

$$Q(\Delta x - \widetilde{\Delta x},\ \Delta y - \widetilde{\Delta y}) = \gamma([ll] - [al]\widetilde{\Delta x} - [bl]\widetilde{\Delta y}), \qquad (11.1.38)$$

где Δx, Δy — текущие координаты, $Q(\xi, \eta) = [aa]\xi^2 + 2[ab]\xi\eta + [bb]\eta^2$ и $\gamma = \dfrac{1}{(1-p_0)^{\frac{2}{n-2}}} - 1$, накрывает точку $(\Delta x_0, \Delta y_0)$ с вероятностью p_0.

Заметим, что эллипс $Q(\xi, \eta) = 1$ является корреляционным эллипсом случайного вектора $(\widetilde{\Delta x}, \widetilde{\Delta y})$ (см. гл. II, § 4). Сделаем еще следующее замечание о точности работы с доверительным эллипсом $Э_\gamma$. Его размеры определяются правой частью (11.1.38), которая равна $\gamma[\tilde{v}\tilde{v}]$; при этом

$$E\gamma[\tilde{v}\tilde{v}] = (n-2)\gamma\sigma^2. \qquad (11.1.39)$$

Таким образом, при заданном n хорошая точность получается (как и во всех других способах оценивания) лишь при малом σ^2, т. е. при больших весах исходных наблюдений, например при достаточном числе повторных равноточных измерений углов.

Остановимся еще на вопросах наивыгоднейшего выбора точек $A_1, \ldots, A_n$, если их возможно выбирать. Из формул (11.1.13) и (11.1.14) видим, что $D(\widetilde{\Delta x})$ и $D(\widetilde{\Delta y})$ пропорциональны отношениям $[aa]$ и $[bb]$ к $D = [aa][bb] - ([ab])^2$; величины a_i и b_i имеют геометрический смысл, выявляемый формулами (11.1.6). Если за меру качества работы взять сумму $D(\widetilde{\Delta x}) + D(\widetilde{\Delta y})$, то видим, что желателен такой выбор точек $A_i (i = 1, 2, \ldots, n)$, при котором величина

$$\frac{[aa] + [bb]}{[aa][bb] - [ab]^2} \qquad (11.1.40)$$

возможно меньшая.

§ 2. Прямая засечка с двух пунктов с повторными наблюдениями

Рассмотрим случай простой прямой засечки с двух твердых пунктов A_1 и A_2, причем предположим для простоты, что в каждом из них производится по $n/2$ равноточных наблюдений (n — четное). Мы можем сделать те же выводы, что и в § 1; формулы (11.1.1) — (11.1.6) будут иметь место при $i = 1, 2$; в частности, a_i

и $b_i (i=1, 2)$ будут иметь значения, даваемые равенствами (11.1.6). Фундаментальные уравнения в элементах Δx_0, Δy_0 примут вид

$$l_i - \Delta_i = a_1 \Delta x_0 + b_1 \Delta y_0 \qquad \left(i = 1, 2, \ldots, \frac{n}{2}\right),$$
$$l_i - \Delta_i = a_2 \Delta x_0 + b_2 \Delta y_0 \qquad \left(i = \frac{n}{2} + 1, \ldots, n\right).$$

Таким образом, основная матрица коэффициентов X будет

$$X = \left\| \begin{matrix} a_1 & b_1 \\ a_1 & b_1 \\ \vdots & \vdots \\ a_1 & b_1 \\ a_2 & b_2 \\ \vdots & \vdots \\ a_2 & b_2 \end{matrix} \right\| \begin{matrix} \left.\vphantom{\begin{matrix} a \\ a \\ \vdots \\ a \end{matrix}}\right\} \frac{n}{2} \\ \left.\vphantom{\begin{matrix} a \\ \vdots \\ a \end{matrix}}\right\} \frac{n}{2} \end{matrix} . \tag{11.2.1}$$

Нормальные уравнения для оценок $\widetilde{\Delta x}$ и $\widetilde{\Delta y}$ величин Δx_0, Δy_0 будут иметь вид (11.1.9):

$$\begin{aligned} [aa]\widetilde{\Delta x} + [ab]\widetilde{\Delta y} &= [al], \\ [ab]\widetilde{\Delta x} + [bb]\widetilde{\Delta y} &= [bl], \end{aligned} \tag{11.2.2}$$

где

$$\left.\begin{aligned} &[aa] = \frac{n(a_1^2 + a_2^2)}{2}; \ [bb] = \frac{n(b_1^2 + b_2^2)}{2}; \\ &[ab] = \frac{n(a_1 b_1 + a_2 b_2)}{2}; \ [al] = a_1\left(l_1 + \ldots + l_{\frac{n}{2}}\right) + \\ &\qquad + a_2\left(l_{\frac{n}{2}+1} + \ldots + l_n\right); \\ &[bl] = b_1\left(l_1 + \ldots + l_{\frac{n}{2}}\right) + b_2\left(l_{\frac{n}{2}+1} + \ldots + l_n\right). \end{aligned}\right\} \tag{11.2.3}$$

Ввиду этого

$$D = [aa][bb] - [ab]^2 = \frac{n^2}{4}\left((a_1^2 + a_2^2)(b_1^2 + b_2^2) - (a_1 b_1 + a_2 b_2)^2\right) =$$
$$= \frac{n^2}{4}(a_1 b_2 - a_2 b_1)^2 = \left(\frac{n}{2} D_0\right)^2,$$

где

$$D_0 = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}. \tag{11.2.4}$$

Обозначим еще

$$l_1 + \ldots + l_{\frac{n}{2}} = l^{(1)}; \quad l_{\frac{n}{2}+1} + \ldots + l_n = l^{(2)}. \tag{11.2.5}$$

Решения нормальных уравнений (11.1.10) примут вид

$$\left.\begin{aligned}\widetilde{\Delta x} &= \frac{(b_1^2+b_2^2)(a_1l^{(1)}+a_2l^{(2)})-(a_1b_1+a_2b_2)(b_1l^{(1)}+b_2l^{(2)})}{D_0^2},\\ \widetilde{\Delta y} &= \frac{(a_1^2+a_2^2)(b_1l^{(1)}+b_2l^{(2)})-(a_1b_1+a_2b_2)(a_1l^{(1)}+a_2l^{(2)})}{D_0^2}.\end{aligned}\right\} \qquad (11.2.6)$$

Элементарный подсчет позволяет упростить эти выражения:

$$\left.\begin{aligned}\widetilde{\Delta x} &= \frac{b_2l^{(1)}-b_1l^{(2)}}{D_0},\\ \widetilde{\Delta y} &= \frac{-a_2l^{(1)}+a_1l^{(2)}}{D_0}.\end{aligned}\right\} \qquad (11.2.7)$$

Далее, (11.1.13) и (11.1.14) дают

$$D(\widetilde{\Delta x}) = \frac{\sigma^2 n}{2}\cdot\frac{b_1^2+b_2^2}{D_0^2}, \qquad (11.2.8)$$

$$D(\widetilde{\Delta y}) = \frac{\sigma^2 n}{2}\cdot\frac{a_1^2+a_2^2}{D_0^2}, \qquad (11.2.9)$$

$$\sigma(\widetilde{\Delta x}) = \frac{\sigma}{|D_0|}\sqrt{\frac{n(b_1^2+b_2^2)}{2}}; \quad \sigma(\widetilde{\Delta y}) = \frac{\sigma}{|D_0|}\sqrt{\frac{n(a_1^2+a_2^2)}{2}}. \qquad (11.2.10)$$

В дальнейшем мы припомним геометрический смысл некоторых из употребляемых величин.

Согласно формулам (11.1.15) и (11.2.3), имеем

$$[\tilde v\,\tilde v] = [ll] - (a_1l^{(1)}+a_2l^{(2)})\widetilde{\Delta x} - (b_1l^{(1)}+b_2l^{(2)})\widetilde{\Delta y}, \qquad (11.2.11)$$

где $\widetilde{\Delta x}$ и $\widetilde{\Delta y}$ определяются из формул (11.2.7).

Имеем

$$\frac{1}{\sigma^2}[\tilde v\,\tilde v] = \chi^2_{N-2}; \quad E[\tilde v\,\tilde v] = (n-2)\sigma^2. \qquad (11.2.12)$$

Оценивание σ можно провести, как указано в гл. V. При возрастающем n асимптотически несмещенной оценкой σ будет

$$\sqrt{\frac{[\tilde v\,\tilde v]}{n-2}}.$$

Оценивание σ с помощью доверительных интервалов можно произвести, используя таблицу III. Выбирая надежность p и беря число степеней свободы $n-2$, находим γ_1 и γ_2 такие, что доверительный интервал

$$\left[\gamma_1\sqrt{\frac{[\tilde v\,\tilde v]}{n-2}},\quad \gamma_2\sqrt{\frac{[\tilde v\,\tilde v]}{n-2}}\right] \qquad (11.2.13)$$

будет накрывать σ с вероятностью p.

Доверительные области для точки $(\Delta x_0, \Delta y_0)$ строятся, как в § 1. Доверительные интервалы для Δx_0 и Δy_0 порознь, согласно (11.1.17), (11.1.18), (11.2.3), получат вид

$$I_\gamma^{(\Delta x_0)} = \left[\widetilde{\Delta x} - \frac{\gamma}{|D_0|}\sqrt{\frac{2}{n(n-2)}(b_1^2+b_2^2)[\tilde v \tilde v]},\right.$$
$$\left.\widetilde{\Delta x} + \frac{\gamma}{|D_0|}\sqrt{\frac{2}{n(n-2)}(b_1^2+b_2^2)[\tilde v \tilde v]}\right], \qquad (11.2.14)$$

$$I_\gamma^{(\Delta y_0)} = \left[\widetilde{\Delta y} - \frac{\gamma}{|D_0|}\sqrt{\frac{2}{n(n-2)}(a_1^2+a_2^2)[\tilde v \tilde v]},\right.$$
$$\left.\widetilde{\Delta y} + \frac{\gamma}{|D_0|}\sqrt{\frac{2}{n(n-2)}(a_1^2+a_2^2)[\tilde v \tilde v]}\right], \qquad (11.2.15)$$

где γ находят по надежности p и числу степеней свободы, как указано в § 1.

Доверительный прямоугольник Q_γ, определяемый своими проекциями на оси координат $I_\gamma^{(\Delta x_0)}$ и $I_\gamma^{(\Delta y_0)}$, накрывает точку $(\Delta x_0, \Delta y_0)$ с надежностью не меньшей, чем $2p-1$.

Длина и ширина прямоугольника Q_γ суть случайные величины

$$\frac{2\gamma}{|D_0|}\sqrt{\frac{2}{n(n-2)}(b_1^2+b_2^2)[\tilde v \tilde v]} \quad \text{и} \quad \frac{2\gamma}{|D_0|}\sqrt{\frac{2}{n(n-2)}(a_1^2+a_2^2)[\tilde v \tilde v]}. \qquad (11.2.16)$$

При больших n математические ожидания этих величин имеют асимптотические выражения

$$\frac{\gamma\sigma}{|D_0|}\frac{2}{\sqrt{n}}\sqrt{2(b_1^2+b_2^2)} \quad \text{и} \quad \frac{\gamma\sigma}{|D_0|}\frac{2}{\sqrt{n}}\sqrt{2(a_1^2+a_2^2)}, \qquad (11.2.17)$$

откуда видно, что средние размеры доверительного прямоугольника Q_γ при большом числе наблюдений n прямо пропорциональны σ и обратно пропорциональны $\sqrt{n}$.

Перейдем к построению доверительного эллипса $Э_\gamma$. Все расчеты проводятся, как в § 1. Матрица C системы нормальных уравнений здесь имеет вид

$$C = \frac{n}{2}\begin{Vmatrix} (a_1^2+a_2^2) & (a_1b_1+a_2b_2) \\ (a_1b_1+a_2b_2) & (b_1^2+b_2^2) \end{Vmatrix}, \qquad (11.2.18)$$

а соответствующая ей квадратичная форма имеет вид

$$Q(\xi, \eta) = \frac{n}{2}\left((a_1^2+a_2^2)\xi^2 + 2(a_1b_1+a_2b_2)\xi\eta + (b_1^2+b_2^2)\eta^2\right) =$$
$$= \frac{n}{2}\left((a_1\xi+b_1\eta)^2 + (a_2\xi+b_2\eta)^2\right). \qquad (11.2.19)$$

Доверительный эллипс $Э_\gamma$ имеет уравнение [см. (11.1.36)]

$$Q(\Delta x - \widetilde{\Delta x}, \Delta y - \widetilde{\Delta y}) = \gamma [\tilde v \tilde v]\frac{2}{n-2}, \qquad (11.2.20)$$

где Δx, Δy — текущие координаты, $[\tilde{v}\tilde{v}]$ находится по формуле (11.2.11), а γ связана с надежностью p простой формулой

$$\gamma = \frac{1}{(1-p)^{\frac{2}{n-2}}} - 1. \tag{11.2.21}$$

Эллипс $Э_\gamma$ будет накрывать искомую точку $(\Delta x_0, \Delta y_0)$ с вероятностью p. При помощи формул (11.2.19), (11.2.11) и (11.2.5) можем выписать подробное уравнение доверительного эллипса $Э_\gamma$:

$$(a_1^2 + a_2^2)(\Delta x - \widetilde{\Delta x})^2 + 2(a_1 b_1 + a_2 b_2)(\Delta x - \widetilde{\Delta x})(\Delta y - \widetilde{\Delta y}) + \\ + (b_1^2 + b_2^2)(\Delta y - \widetilde{\Delta y})^2 = \frac{4\gamma\,[\tilde{v}\tilde{v}]}{n^2}, \tag{11.2.22}$$

$$[\tilde{v}\tilde{v}] = [ll] - (a_1 l^{(1)} + a_2 l^{(2)})\widetilde{\Delta x} - (b_1 l^{(1)} + b_2 l^{(2)})\widetilde{\Delta y};$$
$$l^{(1)} = l_1 + \ldots + l_{\frac{n}{2}}; \quad l^{(2)} = l_{\frac{n}{2}+1} + \ldots + l_n.$$

Ввиду того, что $E[\tilde{v}\tilde{v}] = (n-2)\cdot\sigma^2$, средний размер полуосей доверительного эллипса $Э_\gamma$ при возрастании n будет пропорционален σ и асимптотически пропорционален $\frac{1}{\sqrt{n}}$. Остановимся еще на

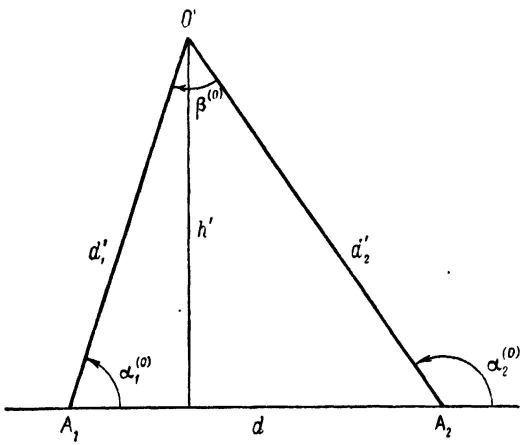

Рис. 11.

геометрическом смысле отдельных коэффициентов, входящих в наши выражения.

Допустим для простоты, что твердые пункты A_1 и A_2 лежат на одной прямой, параллельной линии, от которой отсчитываются дирекционные углы (например, осевому меридиану; рис. 11). Из формул (11.1.6) при $i=1, 2$ находим

$$D_0=\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}=\frac{1}{d_1'd_2'}\begin{vmatrix} -\sin\alpha_1^{(0)} & \cos\alpha_1^{(0)} \\ -\sin\alpha_2^{(0)} & \cos\alpha_2^{(0)} \end{vmatrix}=$$

$$=\frac{1}{d_1'd_2'}(\sin\alpha_2^{(0)}\cos\alpha_1^{(0)}-\sin\alpha_1^{(0)}\cos\alpha_2^{(0)})=\frac{1}{d_1'd_2'}\sin(\alpha_2^{(0)}-\alpha_1^{(0)}). \quad (11.2.23)$$

Если угол при вспомогательной точке обозначим $\beta^{(0)}$, то

$$D_0=\frac{\sin\beta^{(0)}}{d_1'd_2'}. \quad (11.2.24)$$

Из рис. 11 видим, что $\alpha_2^{(0)}=\alpha_1^{(0)}+\beta^{(0)}$. Отсюда имеем

$$a_1^2+a_2^2=\frac{\sin^2\alpha^{(0)}}{d_1'^2}+\frac{\sin^2(\alpha^{(0)}+\beta^{(0)})}{d_2'^2}, \quad (11.2.25)$$

$$b_1^2+b_2^2=\frac{\cos^2\alpha^{(0)}}{d_1'^2}+\frac{\cos^2(\alpha^{(0)}+\beta^{(0)})}{d_2'^2}. \quad (11.2.26)$$

Выражения для дисперсий оценок $\widetilde{\Delta x}$ и $\widetilde{\Delta y}$ (11.2.8) и (11.2.9) можно записать следующим образом:

$$D(\widetilde{\Delta x})=\frac{\sigma^2 n}{2}\cdot\frac{d_1'^2\cos^2\alpha^{(0)}+d_2'^2\cos^2(\alpha^{(0)}+\beta^{(0)})}{\sin^2\beta^{(0)}},$$

$$D(\widetilde{\Delta y})=\frac{\sigma^2 n}{2}\cdot\frac{d_1'^2\sin^2\alpha^{(0)}+d_2'^2\sin^2\left(\alpha^{(0)}+\beta^{(0)}\right)}{\sin^2\beta^{(0)}}.$$

Отсюда

$$D(\widetilde{\Delta x})+D(\widetilde{\Delta y})=\frac{\sigma^2 n}{2}\,\frac{d_1'^2+d_2'^2}{\sin^2\beta^{(0)}}. \quad (11.2.27)$$

Если d есть расстояние между точками A_1 и A_2, то (рис. 11)

$$\frac{d}{\sin\beta^{(0)}}=\frac{d_1'}{|\sin\alpha_2^{(0)}|}=\frac{d_2'}{|\sin\alpha_1^{(0)}|}, \quad (11.2.28)$$

поэтому

$$D(\widetilde{\Delta x})+D(\widetilde{\Delta y})=\frac{\sigma^2 n}{2}\cdot d^2\cdot\frac{(\sin^2\alpha_1^{(0)}+\sin^2(\alpha_1^{(0)}+\beta^{(0)})}{\sin^4\beta^{(0)}}.$$

Если мы можем выбрать точки A_1 и A_2 на одной и той же прямой и с заданным значением угла $\beta^{(0)}$ при вспомогательной точке O', то минимум $D(\widetilde{\Delta x})+D(\widetilde{\Delta y})$ получится, когда $f(\alpha_0)=\sin^2\alpha_1^{(0)}+$ $+\sin^2(\alpha_1^{(0)}+\beta^{(0)})$ минимально. Для вычисления минимума составляем уравнение

$$\frac{df(\alpha_0^{(0)})}{d\alpha^{(0)}}=\frac{1}{2}\left(\sin 2\alpha_0^{(0)}+\sin 2(\alpha^{(0)}+\beta^{(0)})\right)=0.$$

Далее, $\alpha^{(0)}+\beta^{(0)}=\alpha_2^{(0)}$, откуда видим, что должно быть $\pi-\alpha_2^{(0)}=\alpha_1^{(0)}$, т. е. минимум $D(\widetilde{\Delta x})+D(\widetilde{\Delta y})$ получается, если треугольник $A_1O'A_2$ равнобедренный. То, что здесь получился минимум, видно из равенства

$$f''(\alpha^{(0)})\equiv\cos 2\alpha^{(0)}-\cos 2(\alpha^{(0)}+\beta^{(0)})=-\cos 2\alpha^{(0)}-\cos(2\pi-2\alpha^{(0)})=$$
$$=-2\cos 2\alpha_1^{(0)}>0,$$

если

$$\alpha_1^{(0)}<\frac{\pi}{4}.$$

Вернемся еще к уравнению доверительного эллипса $Э_\gamma$ (11.2.22). Выражение (11.2.19) для $Q(\xi,\ \eta)$ с помощью (11.1.6) может быть записано так:

$$Q(\xi,\ \eta)=\frac{n}{2}\left((a_1\xi+b_1\eta)^2+(a_2\xi+b_2\eta)^2\right)=$$
$$=\frac{n}{2}\left(\frac{(-\xi\sin\alpha_1^{(0)}+\eta\cos\alpha_1^{(0)})^2}{d_1'^2}+\frac{(-\xi\sin\alpha_2^{(0)}+\eta\cos\alpha_2^{(0)})^2}{d_2'^2}\right). \quad (11.2.29)$$

Считая $\alpha_1^{(0)}$ острым, а $\alpha_2^{(0)}$—тупым углом, вводя угол $\pi-\alpha_2^{(0)}$ и пользуясь формулами (11.2.28), получим

$$Q(\xi,\ \eta)=\frac{n\sin^2\beta^{(0)}}{2d^2}\left((-\xi+\eta\operatorname{ctg}\alpha_1^{(0)})^2+\left[\xi-\eta\operatorname{ctg}(\pi-\alpha_2^{(0)})\right]^2\right). \quad (11.2.30)$$

Это дает несколько более удобное выражение для доверительного эллипса $Э_\gamma$.

§ 3. Обратная засечка на многие пункты. Доверительные области

Обратная засечка более чем на два пункта, как известно, является весьма полезным средством определения места; обратная засечка по трем пунктам — известная задача Потенота *) — применяется уже более 300 лет и не теряет своего практического значения (например, в задачах геодезии и навигации). Решение задачи Потенота („определение четвертой точки по трем данным") в предположении отсутствия погрешностей измерений осуществляется

*) Л. Потенот (1660—1732) — французский математик.

средствами элементарной тригонометрии и излагается почти во всех учебниках геодезии (см., например, Ф. Н. Красовский [26], стр. 188—190).

Пусть из искомой точки O измеряются углы между направлениями на твердые точки $A_1, A_2, \ldots, A_n$ $(n \geqslant 3)$, причем имеются небольшие случайные погрешности. Выбирая три подходящие точки и решая задачу Потенота, получим приближенное (в силу наличия погрешностей измерения) положение O' для точки O; разности координат точек O и O' считаем такими, квадратами которых можно пренебречь сравнительно со взятой единицей длин (рис. 12).

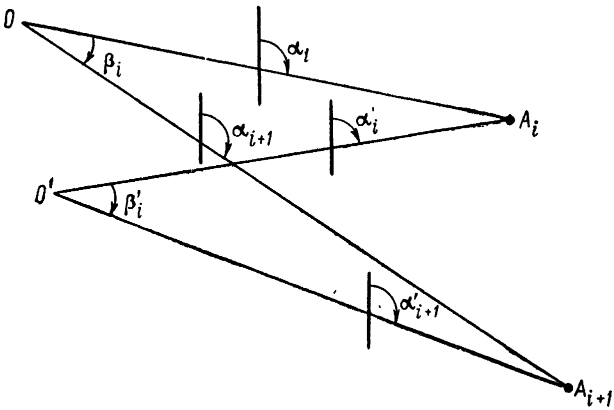

Рис. 12.

Введя обозначения углов соответственно рисунку (α_i — дирекционные углы; $i = 1, 2, \ldots, n$), находим

$$\beta_i = \alpha_{i+1} - \alpha_i, \quad \beta_i^{(0)} = \alpha_{i+1}^{(0)} - \alpha_i^{(0)}.$$

Положим $\Delta\alpha_{i+1} = \alpha_{i+1} - \alpha_{i+1}^{(0)}$; $\Delta\alpha_i = \alpha_i - \alpha_i^{(0)}$. Отсюда

$$\Delta\beta_i = \beta_i - \beta_i^{(0)} = \Delta\alpha_{i+1} - \Delta\alpha_i. \tag{11.3.1}$$

Нам известны, однако, не разности $\Delta\beta_i = \beta_i - \beta_i^{(0)}$, а

$$l_i = \beta_i + \Delta_i - \beta_i^{(0)} = \Delta\beta_i + \Delta_i \quad (i = 1, 2, \ldots, n), \tag{11.3.2}$$

где Δ_i — погрешности измерений углов. Как и ранее, для случая прямой засечки (§ 1), будем их считать равноточными, несмещенными, независимыми и нормальными $N(0, \sigma)$. По формуле (11.1.5), вводя разности координат Δx_0 и Δy_0, точек O и O', находим:

$$\left.\begin{aligned} \Delta\alpha_i &\approx -\frac{\cos\alpha_i^{(0)}}{d'_i}\Delta y_0 + \frac{\sin\alpha_i^{(0)}}{d'_i}\Delta x_0, \\ \Delta\alpha_{i+1} &\approx \frac{\cos\alpha_{i+1}^{(0)}}{d'_{i+1}}\Delta y_0 + \frac{\sin\alpha_{i+1}^{(0)}}{d'_{i+1}}\Delta x_0 \end{aligned}\right\} \tag{11.3.3}$$

(углы измеряются в радиальной мере), где равенства соблюдаются с точностью до второго порядка малости. Знаки изменены по сравнению с (11.1.5), ибо здесь варьирует вершина угла. Подставляя в (11.3.1) и пользуясь (11.3.2), получаем

$$l_i - \Delta_i = \left(\frac{\sin\alpha_{i+1}^{(0)}}{d'_{i+1}} - \frac{\sin\alpha_i^{(0)}}{d'_i}\right)\Delta x_0 + \left(-\frac{\cos\alpha_{i+1}^{(0)}}{d'_{i+1}} + \frac{\cos\alpha_i^{(0)}}{d'_i}\right)\Delta y_0, \quad (11.3.4)$$
$$i = 1, 2, \ldots, n.$$

Обозначим, далее,

$$a_i = \frac{\sin\alpha_{i+1}^{(0)}}{d'_{i+1}} - \frac{\sin\alpha_i^{(0)}}{d'_i}; \quad b_i = -\frac{\cos\alpha_{i+1}^{(0)}}{d'_{i+1}} + \frac{\cos\alpha_i^{(0)}}{d'_i}. \quad (11.3.5)$$

Тогда получаем следующую систему фундаментальных уравнений в элементах Δx_0, Δy_0:

$$l_i - \Delta_i = a_i \Delta x_0 + b_i \Delta y_0 \qquad (i = 1, 2, \ldots, n-1). \quad (11.3.6)$$

Эта система имеет совершенно такой же вид, как (11.1.7); только величины a_i и b_i имеют другой смысл.

Разумеется, все выводы, сделанные в § 1 для оценок $\widetilde{\Delta x}$ и $\widetilde{\Delta y}$, сохраняют силу. Все формулы для точечных оценок $\widetilde{\Delta x}$ и $\widetilde{\Delta y}$, данные в § 1, имеют место при замене n на $n-1$ и (11.1.6) на (11.3.5). Доверительный прямоугольник Q_γ и доверительный эллипс $Э_\gamma$ строятся так же, как в § 1.

§ 4. Доверительные области в задаче Потенота при многократных измерениях

Пусть число твердых точек для обратной засечки равно трем, и задача Потенота решается однозначно (три твердые точки и искомая не лежат на одной окружности и не близки к ней одновременно). Из искомой точки O два угла между твердыми точками A_1, A_2, A_3 пусть замеряются по $\frac{n}{2}$ раз (n — четное). Тогда приходим к фундаментальным уравнениям вида (11.2.2), где a_i и b_i нужно подставить из формулы (11.3.5). Доверительный эллипс $Э_{\gamma_0}$ имеет вид (11.2.22), где a_i и b_i берутся из формул (11.3.5). При заданном p_0 и $\gamma_0 = \frac{1}{(1-p_0)^{\frac{2}{n-2}}} - 1$ он будет накрывать $(\Delta x_0, \Delta y_0)$ с вероятностью p_0.

ГЛАВА XII

ПАРАБОЛИЧЕСКОЕ ИНТЕРПОЛИРОВАНИЕ ПО МЕТОДУ НАИМЕНЬШИХ КВАДРАТОВ

§ 1. Постановка задачи

Параболой n-го порядка называется кривая, имеющая в декартовых координатах уравнение вида

$$y = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n. \qquad (12.1.1)$$

Согласно теореме Вейерштрасса (см., например, Г. М. Фихтенгольц [44]), любую непрерывную функцию можно приблизить на конечном интервале сколь угодно точно такими параболами n-го порядка; таким образом можно изобразить достаточно точно практически любую непрерывную экспериментальную кривую. Пусть x принимает значения x_i $(i = 1, 2, \ldots, N)$, а при каждом $x = x_i$ значение $y = y_i$ наблюдается с погрешностями Δ_i, которые мы предположим независимыми в совокупности, для простоты равноточными и нормальными $N(0, \sigma)$. При этом задача оценивания коэффициентов по N наблюдениям при $N > n + 1$ будет частным случаем известной нам задачи оценивания по методу наименьших квадратов с помощью элементов; здесь $x_{ij} = x_i^j$ $\quad (j = 1, 2, \ldots, n)$,

$$X = \left\| \begin{matrix} 1 & x_1 & \ldots & x_1^n \\ 1 & x_2 & \ldots & x_2^n \\ \cdot & \cdot & \cdot & \cdot \\ 1 & x_N & \ldots & x_N^n \end{matrix} \right\| \qquad (12.1.2)$$

и ранг $(X) = n + 1$ при $N > n + 1$. В самом деле, если точки x_i различны (что предполагается), то, как известно из курсов алгебры, определитель Вандермонда не равен нулю:

$$\begin{vmatrix} 1 & x_1 & \ldots & x_1^n \\ \cdot & \cdot & \cdot & \cdot \\ 1 & x_{n+1} & \ldots & x_{n+1}^n \end{vmatrix} = \prod_{j<i} (x_i - x_j) \neq 0.$$

При указанных условиях оценивание коэффициентов $a_0, a_1, \ldots a_n$, по методу наименьших квадратов будет иметь определенный реальный вероятностный смысл, описанный выше. Само оценивание называется тогда *нахождением криволинейной линии регрессии* (12.1.1) или *параболическим интерполированием по методу наименьших квадратов*. Если и не предполагать нормальности вектора погрешностей $\Delta = \left\| \begin{matrix} \Delta_1 \\ \cdot \\ \cdot \\ \cdot \\ \Delta_N \end{matrix} \right\|$, то все же расчетные правила оценивания по методу наименьших квадратов остаются в силе.

Часто встречается случай, когда мы не знаем вначале, какого порядка нужно взять параболу, чтобы, найдя ее коэффициенты по методу наименьших квадратов, хорошо описать эмпирический ряд наблюдений. Так, в примерах 2 и 3 введения (расстояние, пройденное автомобилем после сигнала об остановке; содержание темных стеклянистых ядрышек в зернах пшеницы) мы могли бы начать с выравнивания ряда наблюдений по прямой $a_0 + a_1 x$, затем пробовать выравнивание по параболе второго порядка $a_0 + a_1 x + a_2 x^2$, затем, в случае надобности, перейти к параболе третьего порядка $a_0 + a_1 x + a_2 x^2 + a_3 x^3$ и т. д.

При этом естественно возникает вопрос о том, как проверить, что парабола n-го порядка хорошо описывает наш ряд наблюдений, в то время как парабола $(n-1)$ порядка неудовлетворительна в этом отношении. Этот вопрос, относящийся к теории испытания статистических гипотез и изучаемый математической статистикой, будет рассмотрен далее.

§ 2. Нормальные уравнения. Ортогональные полиномы Чебышева

Для составления системы нормальных уравнений с матрицей X воспользуемся формулой (6.11.2). Полагая

$$X_0 = (1, \ldots, 1); \quad X_j = \left(x_1^j, x_2^j, \ldots, x_N^j\right) \qquad (j = 1, \ldots, n)$$

и

$$\alpha_\nu = \sum_{i=1}^{N} x_i^\nu = [x^\nu], \quad \nu = 0, 1, \ldots, n \qquad (\alpha_0 = N + 1),$$

$$\alpha_{\nu 1} = \sum_{i=1}^{N} x_i^\nu l_i = [x^\nu l] \qquad (\nu = 0, 1, \ldots, n),$$

мы перепишем систему нормальных уравнений

$$\sum_{\nu=0}^{n} [X_i x_\nu] \tilde{a}_\nu = [X_i l] \qquad (i = 0, 1, \ldots, n) \tag{12.2.1}$$

в виде

$$\left.\begin{aligned} a_0\alpha_0 + a_1\alpha_1 + \ldots + a_n\alpha_n &= \alpha_{01} \\ a_0\alpha_1 + a_1\alpha_2 + \ldots + a_n\alpha_{n+1} &= \alpha_{11} \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ a_0\alpha_n + a_1\alpha_{n+1} + \ldots + a_n\alpha_{2n} &= \alpha_{n1}. \end{aligned}\right\} \tag{12.2.2}$$

Ввиду того, что ранг $(X) = n+1$, имеем

$$\Delta_{n+1} = \begin{vmatrix} \alpha_0 & \alpha_1 & \dots & \alpha_n \\ \alpha_1 & \alpha_2 & \dots & \alpha_{n+1} \\ \dots & \dots & \dots & \dots \\ \alpha_n & \alpha_{n+1} & \dots & \alpha_{2n} \end{vmatrix} \neq 0$$

(см. § 3 гл. I). Решение системы (12.2.1) может быть найдено по известным правилам, излагаемым в любом курсе алгебры. Имеем

$$a_j = \frac{\Delta_{j,\, n+1}}{\Delta_{n+1}} \qquad (j = 0,\ 1,\ 2,\ \dots,\ n), \tag{12.2.3}$$

где $\Delta_{j,\, n+1}$ — минор, получающийся заменой в Δ_{n+1} соответствующего $(j+1)$-го столбца на $\begin{matrix} \alpha_{01} \\ \vdots \\ \alpha_{n1} \end{matrix}$. Если запишем уравнение кривой

$$\tilde{y} = \tilde{a}_0 + \tilde{a}_1 x + \dots + \tilde{a}_n x_n^n, \tag{12.2.4}$$

отличающейся от кривой (12.1.1) тем, что коэффициенты a_i заменены их оценками по методу наименьших квадратов, то в силу (12.2.3) это уравнение примет вид

$$\begin{vmatrix} \tilde{y} & 1 & x & \dots & x^n \\ \alpha_{01} & \alpha_0 & \alpha_1 & \dots & \alpha_n \\ \alpha_{11} & \alpha_1 & \alpha_2 & \dots & \alpha_{n+1} \\ \dots & \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_n & \alpha_{n+1} & \dots & \alpha_{2n} \end{vmatrix} = 0. \tag{12.2.5}$$

При $n = 1,\ 2,\ 3,\ 4$ это уравнение может быть использовано непосредственно; далее расчеты становятся более громоздкими. Однако при такой схеме вычислений для каждого значения n приходится выписать все оценки $\tilde{a}_0,\ \dots,\ \tilde{a}_n$. Естественно возникает желание иметь такого рода схему оценивания коэффициентов парабол последовательных порядков, описывающих данный ряд наблюдений, при которой найденные оценки для коэффициентов не будут изменяться при переходе от параболы $(n-1)$-го порядка к параболе n-го порядка, так что нужно будет лишь добавить вычисления по оцениванию коэффициента a_n. Для того чтобы это сделать, вместо основной (базисной) системы одночленов $1,\ x,\ x^2,\ \dots,\ x^n,\ \dots$ следует взять подходящую систему полиномов $\varphi_0(x) = \text{const},\ \varphi_1(x),\ \dots$ $\dots,\ \varphi_n(x)$, причем $\varphi_n(x)$ — полином порядка n, и вместо (12.1.1) употреблять разложение вида

$$y = c_0\varphi_0(x) + c_1\varphi_1(x) + \dots + c_n\varphi_n(x). \tag{12.2.6}$$

Пусть соответственно значениям $x = x_1,\ x_2,\ \dots,\ x_N$ наблюдения дают значения $l = l_1,\ l_2,\ \dots,\ l_N$. (В случае точного выполнения статистической схемы, рассматривавшейся ранее, $l_i = y_i + \Delta_i$, где Δ_i — случайные погрешности, независимые в совокупности и нормальные $N(0, \sigma)$. Если такая схема и не выполняется в точности, то все же правила оценивания остаются теми же и приводят к однозначным результатам, реальный смысл которых будет зависеть от характера погрешностей $\Delta_1, \ldots, \Delta_N$). Пусть последовательность полиномов

$$\varphi_0(x),\ \varphi_1(x),\ \ldots,\ \varphi_n(x),\ \ldots \tag{12.2.7}$$

образует ортогональную систему функций для системы точек x_1, $x_2, \ldots, x_N$. Это означает, что

$$\sum_{i=1}^{N} \varphi_\mu(x_i)\,\varphi_\nu(x_i) = [\varphi_\mu(x)\cdot\varphi_\nu(x)] = 0 \text{ при } \mu \neq \nu. \tag{12.2.8}$$

Такие соотношения приводят к большому упрощению системы нормальных уравнений для оценивания коэффициентов $c_0, c_1, \ldots, c_n$ в (12.2.6). Система типа (12.2.1) принимает вид

$$[\varphi_i(x)\,\varphi_i(x)]\,\tilde{c}_i = [\varphi_i(x)\,l] \qquad (i = 0,\ 1,\ \ldots,\ n), \tag{12.2.9}$$

откуда

$$\tilde{c}_i = \frac{[\varphi_i(x)\,l]}{[\varphi_i(x)\,\varphi_i(x)]}. \tag{12.2.10}$$

Покажем, как строится ортогональная система полиномов для системы точек $x_1. \ldots, x_N$, называемая *системой полиномов Чебышева*. Положим $\varphi_0(x) = 1$. В качестве $\varphi_1(x)$ возьмем линейную комбинацию x и $\varphi_0(x)$: $\varphi_1(x) = x + b\varphi_0(x)$, где b подберем так, чтобы $[\varphi_0(x)\,\varphi_1(x)] = [x] + b\,[1] = 0$. Отсюда и из равенства $[1] = N$ имеем

$$b = -\frac{[x]}{N} = -\bar{x},$$

$$\varphi_1(x) = x - \bar{x} = x - \frac{[x]}{N}. \tag{12.2.11}$$

Далее, положим

$$\varphi_2(x) = x^2 + b_1\varphi_1(x) + b_0\varphi_0(x) = x^2 + b_1(x - \bar{x}) + b_0.$$

Для определения коэффициентов b_1 и b_0 используем условия ортогональности $[\varphi_2(x)\,\varphi_0(x)] = 0$, $[\varphi_2(x)\,\varphi_1(x)] = 0$. Получаем

$$[x^2] + b_0\,[1] = 0, \qquad [x^2\varphi_1(x)] + b_1\,[(\varphi_1(x))^2] = 0;$$

отсюда

$$b_0 = -\frac{[x^2]}{N}; \qquad b_1 = -\frac{[x^2\varphi_1(x)]}{[(x-\bar{x})^2]}$$

и

$$\varphi_2(x) = x^2 - \frac{[x^2(x-\bar{x})]}{[(x-\bar{x})^2]}(x - \bar{x}) - \frac{[x^2]}{N}.$$

Мы можем действовать таким образом и далее, определяя $\varphi_3(x)$, $\varphi_4(x)$, ... Можно найти формулу, дающую общий вид $\varphi_m(x)$.

Теорема 12.2.1.

$$\varphi_m(x) = x^m - \frac{[x^m \varphi_{m-1}(x)]}{[\varphi_{m-1}(x)\,\varphi_{m-1}(x)]}\,\varphi_{m-1}(x) - \ldots - \frac{[x^m]}{N}. \qquad (12.2.12)$$

Доказательство. Положим

$$\varphi_m(x) = x^m + b_{m-1}\varphi_{m-1}(x) + b_{m-2}\varphi_{m-2}(x) + \ldots + b_0\varphi_0(x).$$

Умножая на $\varphi_k(x)$ ($k = m-1, \ldots, 0$) и суммируя по $x = x_1, x_2, \ldots$ $\ldots, x_N$, воспользуемся условием $[\varphi_m(x)\,\varphi_k(x)] = 0$. Получим

$$[x^m \varphi_k(x)] = b_k\,[\varphi_k(x)\,\varphi_k(x)],$$

$$b_k = -\frac{[x^m \varphi_k(x)]}{[\varphi_k(x)\,\varphi_k(x)]},$$

что и доказывает теорему. Мы видим, что нет нужды, чтобы все точки были различными. Для однозначного определения коэффициентов нужно лишь, чтобы $[\varphi_k(x)\,\varphi_k(x)] \neq 0$ для всех употребляемых значений k.

Легко заметить, что каждую степень x^k, а значит, и каждый полином $P(x)$, можно записать в виде линейной комбинации полиномов Чебышева. Из формул (12.2.10), в которых под $\varphi_i(x)$ можно понимать полиномы Чебышева, видно, что использование этих полиномов весьма удобно: решения $\tilde{c}_i$ нормальных уравнений выражаются непосредственно через $[\varphi_i(x)\,l]$ и $[\varphi_i(x)\,\varphi_i(x)]$. Правда, надо заметить, что $\varphi_i(x)$ сами зависят от абсцисс $x_1, \ldots, x_N$ и должны быть составлены по ним заранее. Далее, весьма важно, что при увеличении числа n, т. е. введении коэффициентов $c_{n+1}, c_{n+2}, \ldots$ при тех же наблюдениях $l_1, \ldots, l_N$, надо вычислить лишь оценки для новых коэффициентов $c_{n+1}, c_{n+2}, \ldots$, оценки же $\tilde{c}_0, \tilde{c}_1, \ldots, \tilde{c}_n$ для коэффициентов $c_0, \ldots, c_n$ остаются в силе, их не надо вычислять вновь.

Далее, сравнительно удобные выражения получаются для корреляционных матриц $B_{\tilde{C}}$ оценок $\tilde{c}_0, \ldots, \tilde{c}_n$, доверительных интервалов и доверительных областей. Матрица нормальных уравнений

$$C = \|[\varphi_\mu(x)\,\varphi_\nu(x)]\|$$

в силу ортогональности полиномов Чебышева превращается в диагональную матрицу:

$$C = \begin{Vmatrix} [\varphi_0(x)\,\varphi_0(x)] & 0 & \ldots & 0 \\ 0 & [\varphi_1(x)\,\varphi_1(x)] & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & [\varphi_n(x)\,\varphi_n(x)] \end{Vmatrix}. \qquad (12.2.13)$$

Таким образом [см. (8.2.5)],

$$B_C = \sigma^2 C^{-1} = \sigma^2 \begin{Vmatrix} \frac{1}{[\varphi_0(x)\varphi_0(x)]} & 0 & \ldots & 0 \\ 0 & \frac{1}{[\varphi_1(x)\varphi_1(x)]} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \frac{1}{[\varphi_n(x)\varphi_n(x)]} \end{Vmatrix}, \tag{12.2.14}$$

откуда следует

Теорема 12.2.2. *Оценки $\tilde{c}_0, \tilde{c}_1, \ldots, \tilde{c}_n$ нормальны, независимы в совокупности (в случае нормального вектора погрешностей) и имеют дисперсии*

$$D(\tilde{c}_i) = \frac{\sigma^2}{[\varphi_i(x)\varphi_i(x)]}. \tag{12.2.15}$$

В случае невыполнения для вектора погрешностей Δ нормального закона, можно утверждать лишь некоррелированность оценок $\tilde{c}_i$. Для оценивания σ^2 и построения доверительных интервалов и областей (вектор погрешностей Δ предполагается нормальным) надлежит вычислить $[p\tilde{v}\tilde{v}] = [\tilde{v}\tilde{v}]$. В данном случае это можно сделать непосредственно. Имеем (см. гл. VIII)

$$\tilde{V} = X \begin{Vmatrix} \tilde{c}_0 \\ \tilde{c}_1 \\ \vdots \\ \tilde{c}_n \end{Vmatrix} - L, \tag{12.2.16}$$

где

$$X = \begin{Vmatrix} \varphi_0(x_1) & \varphi_1(x_1) & \ldots & \varphi_n(x_1) \\ \varphi_0(x_2) & \varphi_1(x_2) & \ldots & \varphi_n(x_2) \\ \ldots & \ldots & \ldots & \ldots \\ \varphi_0(x_N) & \varphi_1(x_N) & \ldots & \varphi_n(x_N) \end{Vmatrix}. \tag{12.2.17}$$

Полагая

$$\tilde{L} = X \begin{Vmatrix} \tilde{c}_0 \\ \vdots \\ \tilde{c}_n \end{Vmatrix},$$

так что

$$\tilde{l}_r = \varphi_0(x_r)\tilde{c}_0 + \varphi_1(x_r)\tilde{c}_1 + \ldots + \varphi_n(x_r)\tilde{c}_n \qquad (r = 1, 2, \ldots, N), \tag{12.2.18}$$

имеем $\tilde{V} = \tilde{L} - L$; $\tilde{v}_r = \tilde{l}_r - l_r$, поэтому

$$[\tilde{v}\tilde{v}] = [(\tilde{l} - l)^2]. \tag{12.2.19}$$

Согласно теореме 6.6.4,

$$E\frac{[\tilde{v}\tilde{v}]}{N-n-1}=\sigma^2, \tag{12.2.20}$$

так что $\frac{[\tilde{v}\tilde{v}]}{N-n-1}$ является несмещенной оценкой для σ^2. Далее, величина $\frac{1}{\sigma^2}[\tilde{v}\tilde{v}]$ имеет χ^2-распределение с $N-n$ степенями свободы; способ построения доверительных интервалов для $c_0, \ldots, c_n$ порознь указан в § 3 гл. VIII.

Аналогично конструкции доверительных эллипсов $Э_{\gamma_0}$ в главе XI, мы можем строить доверительный эллипсоид $Э_{\gamma_0}$ для $c_0, \ldots, c_n$ в совокупности; в данном случае такая конструкция сравнительно удобна. Как известно из общей теории уравнивания по элементам (см. гл. VIII), нормальный случайный вектор $(\tilde{c}_0, \ldots, \tilde{c}_n)$ статистически независим от $[\tilde{v}\tilde{v}]$. Составим новый нормальный случайный вектор

$$\left(\frac{\tilde{c}_0-c_0}{\sigma\sqrt{[\varphi_0(x)\varphi_0(x)]}}, \ldots, \frac{\tilde{c}_n-c_n}{\sigma\sqrt{[\varphi_n(x)\varphi_n(x)]}}\right).$$

Он также независим от $[\tilde{v}\tilde{v}]$ и имеет нулевой вектор средних и единичную корреляционную матрицу;

$$B=E=\left\|\begin{matrix}1 & 0 & \ldots & 0\\ \cdot & \cdot & \cdot & \cdot\\ 0 & 0 & \ldots & 1\end{matrix}\right\|.$$

Ввиду этого квадратичная форма

$$\frac{1}{\sigma^2}Q(\tilde{c}_0-c_0, \ldots, \tilde{c}_n-c_n)=\frac{1}{\sigma^2}\sum_{i=0}^{n}\frac{(\tilde{c}_i-c_i)^2}{[\varphi_i(x)\varphi_i(x)]} \tag{12.2.21}$$

будет распределена по χ^2_{n+1}, а величина

$$q=\frac{\frac{1}{\sigma^2}Q(\tilde{c}_0-c_0, \ldots, \tilde{c}_n-c_n)}{\frac{1}{\sigma^2}[\tilde{v}\tilde{v}]}=\frac{Q(\tilde{c}_0-c_0, \ldots, \tilde{c}_n-c_n)}{[\tilde{v}\tilde{v}]} \tag{12.2.22}$$

будет иметь распределение Фишера $F_{n+1, N-n-1}(x)$. Если зададим надежность p_0 и выберем по таблице распределения Фишера γ_0 так, что $F_{n+1, N-n-1}(\gamma_0)=p_0$, то можем заметить (как и в § 1, гл. XI), что эллипсоид $Э_\gamma$ с уравнением

$$Q(z_0-\tilde{c}_0, \ldots, z_n-\tilde{c}_n)=\gamma_0[\tilde{v}\tilde{v}] \tag{12.2.23}$$

будет накрывать точку $(\tilde{c}_0, \tilde{c}_1, \ldots, \tilde{c}_n)$ с вероятностью p_0.

Случай, когда абсциссы наблюдений $x_1, x_2, \ldots, x_N$ принимают равноотстоящие значения, т. е. $x_{i+1} - x_i = d$ $(i = 1, 2, \ldots, N-1)$, был исследован в свое время П. Л. Чебышевым. Для этого случая существуют таблицы, облегчающие технику вычислений (см. В. С. Немчинов [39]). Приведем без доказательства соотношение П. Л. Чебышева

$$\varphi_{k+1}(x) = (x - \bar{x})\,\varphi_k(x) - \frac{k^2(N^2 - k^2)}{4(4k^2 - 1)}\,\varphi_{k-1}(x). \qquad (12.2.24)$$

Первые четыре полинома имеют вид

$$\varphi_1(x) = x - \bar{x}, \quad \varphi_2(x) = (x - \bar{x})^2 - \frac{N^2 - 1}{12},$$

$$\varphi_3(x) = (x - \bar{x})^3 - \frac{3N^2 - 7}{20}(x - \bar{x}),$$

$$\varphi_4(x) = (x - \bar{x})^4 - \frac{3N^2 - 13}{14}(x - \bar{x})^2 + \frac{3(N^2 - 1)(N^2 - 9)}{560};$$

$$[\varphi_k(x)\,\varphi_k(x)] = \frac{(k!)^2 N (N^2 - 1)(N^2 - 4)\ldots(N^2 - k^2)}{(1\cdot 3\cdot 5\ldots(2k-1))^2\cdot 2^{2k}(2k+1)}.$$

Для абсцисс $x = 0, 1, 2, \ldots, n-1$ (к которым сводятся равноотстоящие абсциссы в общем случае) в таблицах Фишера и Иэйтса даны значения первых пяти полиномов при $n \leqslant 52$. Первые шесть полиномов Чебышева для таких абсцисс имеют вид (Кендалл [20], стр. 160)

$$\varphi_1 = x - \frac{n-1}{2}, \quad \varphi_2 = \varphi_1^2 - \frac{n^2 - 1}{12},$$

$$\varphi_3 = \varphi_1^3 - \frac{3n^2 - 7}{20}\varphi_1,$$

$$\varphi_4 = \varphi_1^4 - \frac{3n^2 - 13}{14}\varphi_1^2 + \frac{3(n^2 - 1)(n^2 - 9)}{560},$$

$$\varphi_5 = \varphi_1^5 - \frac{5(n^2 - 7)}{18}\varphi_1^3 + \frac{15n^4 - 230n^2 + 407}{1008}\varphi_1,$$

$$\varphi_6 = \varphi_1^6 - \frac{5(3n^2 - 31)}{44}\varphi_1^4 + \frac{5n^4 - 110n^2 + 329}{176}\varphi_1^2 - \frac{5(n^2-1)(n^2-9)(n^2-25)}{14784}.$$

Аналогично рассматривается вопрос о множественной корреляции, когда основные уравнения имеют вид

$$l_r - \Delta_r = a_1 z_1^{(r)} + a_2 z_2^{(r)} + \ldots + a_n z_n^{(r)} \qquad (r = 1, 2, \ldots, N)$$

(линейный случай). Подобные случаи нам уже встречались в гл. VII.

§ 3. Проверка гипотезы о наличии параболической регрессии данного порядка. Примеры

Рассмотрим задачу проверки гипотезы о том, что имеют место равенства

$$l_r + \Delta_r = c_0\varphi_0(x_r) + \ldots + c_n\varphi_n(x_r) \qquad (r = 1, 2, \ldots, N), \quad (12.3.1)$$

где $\varphi_i(x)$ — ортогональные полиномы Чебышева, а вектор погрешностей $\Delta = \begin{Vmatrix} \Delta_1 \\ \vdots \\ \Delta_N \end{Vmatrix}$ — нормальный с независимыми одинаково распределенными компонентами. Такая гипотеза называется гипотезой о наличии параболической регрессии n-го порядка.

В правой части (12.3.1) может стоять линейная комбинация любых полиномов степени, меньшей или равной n; как было пояснено выше, ее можно привести к линейной комбинации полиномов Чебышева, т. е. к виду правой части (12.3.1). Для проверки указанной гипотезы полезно иметь при каждой абсциссе x_r ($r = 1, 2, \ldots, N$) не одно, а несколько наблюдений $l_{r1}, l_{r2}, \ldots, l_{rm}$; для простоты допустим, что число $m > 1$ одно и то же для всех абсцисс x_r. Это равносильно тому, что каждая абсцисса x_r имеет по m повторений, так что полное число их, считая повторения, будет $N_1 = mN$. Все правила расчета, изложенные в § 2, остаются при этом в силе. Выражения $[\varphi_i(x)\varphi_i(x)]$ и $[\varphi_i(x)l]$, входящие в формулу (12.2.10) для вычисления $\tilde{c}_i$, будут иметь вид

$$[\varphi_i(x)\varphi_i(x)] = m\sum_{r=1}^{N}\left(\varphi_i(x_r)\right)^2, \tag{12.3.2}$$

$$[\varphi_i(x)l] = \sum_{r=1}^{N}\varphi_i(x_r)\,l^{(r)}, \tag{12.3.3}$$

где

$$l^{(r)} = l_{r1} + l_{r2} + \ldots + l_{rm}. \tag{12.3.4}$$

Полагая

$$\bar{l}_r = \frac{1}{m}\,l^{(r)} = \frac{l_{r1} + \ldots + l_{rm}}{m}, \tag{12.3.5}$$

получим

$$\tilde{c}_i = \frac{\sum\limits_{r=1}^{N}\varphi_i(x_r)\,\bar{l}_r}{\sum\limits_{r=1}^{N}\left(\varphi_i(x_r)\right)^2}. \tag{12.3.6}$$

Рассмотрим теперь выражение

$$\tilde{l}_r = \tilde{c}_0\varphi_0(x_r) + \tilde{c}_1\varphi_1(x_r) + \ldots + \tilde{c}_n\varphi_n(x_r). \tag{12.3.7}$$

Теорема 12.3.1. *Если испытуемая гипотеза верна, то нормальные случайные векторы*

$$(l_{11} - \bar{l}_1, \ldots, l_{1m} - \bar{l}_1, l_{21} - \bar{l}_2, \ldots, l_{2m} - \bar{l}_2, \ldots$$
$$\ldots, l_{N1} - \bar{l}_N, \ldots, l_{Nm} - \bar{l}_N) \tag{12.3.8}$$

и

$$(\bar{l}_1 - \tilde{l}_1, \bar{l}_2 - \tilde{l}_2, \ldots, \bar{l}_N - \tilde{l}_N) \tag{12.3.9}$$

независимы.

Доказательство. На основании теоремы 2.3.3 достаточно доказать, что

$$E\left((l_{\nu\mu}-\bar{l}_\nu)(\bar{l}_\rho-\tilde{l}_\rho)\right)=0 \qquad (12.3.10)$$

при любых ν, $\rho=1, 2, \ldots, N$ и $\mu=1, 2, \ldots, m$.

Заметим, что при любых μ, ν и ρ имеет место равенство

$$E(l_{\nu\mu}-\bar{l}_\nu)\bar{l}_\rho=0. \qquad (12.3.11)$$

В самом деле, $l_{\nu\mu}$ и $\bar{l}_\nu$ при $\rho\neq\nu$ не зависят от $l_{\rho1}, l_{\rho2}, \ldots, l_{\rho m}$ в совокупности, ибо погрешности наблюдений $\Delta_1, \ldots, \Delta_N$ предполагаются независимыми в совокупности. Если же $\rho=\nu$, то

$$E(l_{\nu\mu}-\bar{l}_\nu)\bar{l}_\rho=E(l_{\nu\mu}-\bar{l}_\nu)\bar{l}_\nu=E(\Delta_{\nu\mu}-\bar{\Delta}_\nu)\bar{\Delta}_\nu,$$

где Δ_i — погрешности, отвечающие l_i. Но

$$E(\Delta_{\nu\mu}\bar{\Delta}_\nu)=\frac{\sigma^2}{m}; \quad E(\bar{\Delta}_\nu^2)=\frac{\sigma^2}{m},$$

что и доказывает (12.3.11).

Остается доказать, что

$$E\tilde{l}_\rho(l_{\nu\mu}-\bar{l}_\nu)=0. \qquad (12.3.12)$$

Ввиду (12.3.6) и (12.3.7) достаточно обнаружить, что для всех значений τ

$$E\left(\bar{l}_\tau(l_{\nu\mu}-\bar{l}_\nu)\right)=0. \qquad (12.3.13)$$

Это равносильно (12.3.11) и потому верно. Тем самым верно (12.3.12) и (12.3.10), что и доказывает теорему.

Обратим внимание на то, что векторы (12.3.8) и (12.3.9) дают нам различные несмещенные оценки для дисперсии σ^2, если гипотеза верна. Для составления этих оценок рассмотрим сперва случайный вектор (12.3.9). Имеем из (12.3.6)

$$\tilde{c}_i=\frac{\sum\varphi_i(x_r)\bar{l}_r}{\sum(\varphi_i(x_r))^2}, \qquad (12.3.14)$$

что можно истолковать как решение нормальных уравнений для системы фундаментальных уравнений вида

$$\bar{l}_r-\bar{\Delta}_r=c_0\varphi_0(x_r)+\ldots+c_n\varphi_n(x_r) \quad (r=1, 2, \ldots, N), \qquad (12.3.15)$$

где $\bar{\Delta}_r$ — независимые случайные погрешности, нормальные $N\left(0, \frac{\sigma}{\sqrt{m}}\right)$.

В таком случае $\bar{l}_r-\tilde{l}_r=\tilde{v}_r$ $(r=1, 2, \ldots, N)$ и $\frac{m}{\sigma^2}[\tilde{v}\tilde{v}]$ должно быть распределено, как χ^2_{N-n-1} (см. гл. VIII). Таким образом,

$$q_1=\frac{m}{N-n-1}[\tilde{v}\tilde{v}] \qquad (12.3.16)$$

будет несмещенной оценкой для σ^2.

Вектор (12.3.8) доставляет нам другую несмещенную оценку. Группы $l_{r1}, l_{r2}, \ldots, l_{rm}$ с групповыми средними $\bar{l}_r$ дают непосредственно несмещенную оценку для дисперсии σ^2 (см. § 2 гл. III):

$$\frac{1}{m-1}\sum_{i=1}^{m}(l_{ri}-\bar{l}_r)^2.$$

Складывая такие оценки для $r=1, 2, \ldots, N$ и деля на N, получим несмещенную оценку для σ^2:

$$q_2=\frac{1}{N(m-1)}\sum_{r=1}^{N}\sum_{i=1}^{m}(l_{ri}-\bar{l}_r)^2. \qquad (12.3.17)$$

При этом величина $\frac{1}{\sigma^2}\sum_{i=1}^{m}(l_{ri}-\bar{l}_r)^2$ распределена по χ^2_{m-1} (см. гл. III) и при разных r величины $l_{ri}-\bar{l}_r$ независимы. Таким образом, величина $\frac{1}{\sigma^2}N(m-1)q_2$ распределена как сумма N независимых χ^2_{m-1} и по формуле (2.6.4) имеем $\frac{1}{\sigma^2}N(m-1)q_2=\chi^2_{N(m-1)}$. При этом q_1 не зависит от q_2.

Заметим, что оценка (12.3.16) получена лишь с помощью наблюдений l_r без учета фундаментальных уравнений (12.3.1), в то время как оценка (12.3.17) использует испытуемую гипотезу и фундаментальные уравнения. Для построения критерия принятия или отбрасывания испытуемой гипотезы составим частное

$$w=\frac{m[\tilde{v}\tilde{v}]}{N(m-1)q_2}=\frac{\frac{m}{m-1}[\tilde{v}\tilde{v}]}{Nq_2}=\frac{\chi^2_{N-n-1}}{\chi^2_{N(m-1)}}. \qquad (12.3.18)$$

Числитель и знаменатель этой дроби независимы. Таким образом, получаем распределение Фишера $F_{N-n-1,\,N(m-1)}(x)$.

Составим вспомогательную величину $w_1=\frac{N(m-1)}{N-n-1}w$ и по таблице II приложений найдем γ_0 такое, что

$$p_0=0{,}95=P\left(\frac{1}{\gamma_0}<w_1<\gamma_0\right)=P(|\log w_1|<\log\gamma_0)=$$
$$=P\left(\frac{N(m-1)}{N-n-1}\gamma_1\leqslant w_1\leqslant\frac{N(m-1)}{N-n-1}\gamma_2\right).$$

Если теперь найденное из (12.3.18) значение w не будет удовлетворять неравенству

$$\gamma_1<w<\gamma_2,$$

то мы отвергаем гипотезу о наличии параболической регрессии n-го порядка. Обоснование целесообразности такого, а не иного критерия имеется в подробных курсах математической статистики (см., например, Кендалл [20]). Надо заметить, что уточнение критерия Аббе, изложенное в § 6 гл. IV, по существу является частным случаем изложенных здесь соображений.

Обратимся теперь к примерам.

Пример 1. В табл. 28 (столбцы 1—3) приведены данные о распределении валиков по овальности до и после шлифования *). Здесь x обозначает овальность валика до шлифования, $y(x)$ — вычисленную по наблюдениям среднюю овальность валика после шлифования при фиксированном x (x и y измерены в сотых долях миллиметра), n_i — наблюденное число x_i.

Таблица 28

x_i	n_i	y_i	$n_i y_i$	x_i^2	x_i^3	$\varphi_1(x_i)$	$\varphi_1^2(x_i)$	$\varphi_2(x_i)$	$\varphi_2^2(x_i)$
1	21	0,10	2,10	1	1	—2,18	4,75	2,24	5,02
2	8	0,19	1,52	4	8	—1,18	1,39	—1,41	1,99
3	13	0,24	3,12	9	27	—0,18	0,03	—3,06	9,36
4	11	0,32	3,52	16	64	0,82	0,67	—2,71	7,34
5	11	0,39	4,29	25	125	1,82	3,31	—0,36	0,13
6	10	0,48	4,80	36	216	2,82	7,95	3,99	15,92

Найдем зависимость между y и x в форме параболы второй степени

$$y = a_0 + a_1 x + a_2 x^2,$$

пользуясь методом наименьших квадратов.

Запишем уравнение искомой параболы в виде

$$y = c_0\varphi_0(x) + c_1\varphi_1(x) + c_2\varphi_2(x),$$

где $\varphi_0(x)$, $\varphi_1(x)$, $\varphi_2(x)$ — ортогональная система полиномов для системы точек x_i, и оценим коэффициенты c_0, c_1, c_2. Значения ортогональных функций вычисляем по формулам:

$$\varphi_0(x) = 1,$$

$$\varphi_1(x) = x - \frac{[x]}{N},$$

$$\varphi_2(x) = x^2 - \frac{[x^3] - \frac{[x][x^2]}{N}}{[x^2] - \frac{[x]^2}{N}}\left(x - \frac{[x]}{N}\right) - \frac{[x^2]}{N}.$$

*) Данные заимствованы из книги [14], стр. 429.

В нашем случае

$$N=\sum_{i=1}^{6} n_i = 74, \quad [x]=\sum_{i=1}^{6} n_i x_i = 235,$$

$$[x^2]=\sum_{i=1}^{6} n_i x_i^2 = 981, \quad [x^3]=4675,$$

значит, полиномы Чебышева имеют вид

$$\varphi_0(x)=1,$$
$$\varphi_1(x)=x-3,18,$$
$$\varphi_2(x)=x^2-6,65x+7,89.$$

Значения полиномов $\varphi_i(x)$ и их квадратов при различных x_i приведены в столбцах 7—10 табл. 28.

Оценки $\tilde{c}_i$ для коэффициентов c_i вычисляем по формуле

$$\tilde{c}_i=\frac{[\varphi_i(x)\,y]}{[\varphi_i(x)\,\varphi_i(x)]}.$$

Получим

$$\tilde{c}_0=\frac{[\varphi_0(x)\,y]}{[\varphi_0(x)\,\varphi_0(x)]}=\frac{\sum\limits_{i=1}^{6} n_i y_i}{[\varphi_0^2(x)]}=\frac{19,35}{74}=0,261,$$

$$\tilde{c}_1=\frac{[\varphi_1(x)\,y]}{[\varphi_1(x)\,\varphi_1(x)]}=\frac{\sum\limits_{i=1}^{6} \varphi_1(x_i)\,n_i y_i}{\sum\limits_{i=1}^{6} n_i\varphi_1^2(x_i)}=\frac{17,30}{234,5}=0,074,$$

$$\tilde{c}_2=\frac{[\varphi_2(x)\,y]}{[\varphi_2(x)\,\varphi_2(x)]}=\frac{\sum\limits_{i=1}^{6} \varphi_2(x_i)\,n_i y_i}{\sum\limits_{i=1}^{6} n_i\varphi_2^2(x_i)}=\frac{1,082}{484,4}=0,0022.$$

Следовательно, уравнение искомой параболы имеет вид

$$y=0,26+0,074(x-3,18)+0,0022(x^2-6,65x+7,89)=$$
$$=0,042+0,059x+0,0022x^2.$$

Пример 2. Данные табл. 29 показывают зависимость между потерей в весе (в процентах) и температурой (в градусах Цельсия) для 16 проб почвы *).

*) Данные заимствованы из книги [20], стр. 150, при этом значения температуры переведены на шкалу Цельсия.

Таблица 29

t_i	y_i	x_i	x_i^2	x_i^3	$\varphi_1(x)$	$\varphi_2(x)$	$\varphi_3(x)$
37,8	3,71	0,378	0,14288	0,05401	—0,36163	0,08383	—0,01997
40,6	3,81	0,406	0,16484	0,06692	—0,33363	0,06236	—0,01005
43,3	3,86	0,433	0,18749	0,08118	—0,30663	0,04313	—0,00217
46,1	3,93	0,461	0,21252	0,09797	—0,27863	0,02473	0,00435
49,4	3,96	0,494	0,24404	0,12055	—0,24563	0,00506	0,01004
55,6	4,20	0,556	0,30914	0,17188	—0,18363	—0,02600	0,01564
62,2	4,34	0,622	0,38688	0,24064	—0,11763	—0,05064	0,01574
67,2	4,51	0,672	0,45158	0,30346	—0,06763	—0,06349	0,01274
72,8	4,73	0,728	0,52998	0,38583	—0,01163	—0,07195	0,00731
81,7	5,35	0,817	0,66749	0,54534	0,07737	—0,07248	—0,00336
88,3	5,74	0,883	0,77969	0,68847	0,14337	—0,06265	—0,01070
95,0	6,14	0,950	0,90250	0,85738	0,21037	—0,04377	—0,01565
100,0	6,51	1,000	1,00000	1,00000	0,26037	—0,02382	—0,01669
107,8	6,98	1,078	1,16208	1,25273	0,33837	0,01728	—0,01174
113,9	7,44	1,139	1,29732	1,47765	0,39937	0,05790	—0,00071
121,7	7,76	1,217	1,48109	1,80249	0,47737	0,12069	0,02487
Сумма .	82,97	11,834	9,91952	9,14650			

Требуется найти линию регрессии y на t, пользуясь методом наименьших квадратов.

Чтобы избежать больших чисел при вычислениях, заменим t величиной $x = \frac{t}{100}$.

Начав с выравнивания ряда наблюдений по прямой, перейдем к выравниванию по параболам второго и третьего порядка. Ищем прямую $y = c_0\varphi_0(x) + c_1\varphi_1(x)$, где $\varphi_0(x)$, $\varphi_1(x)$ — ортогональные полиномы Чебышева для системы точек x_i, $\varphi_0(x) = 1$, $\varphi_1(x) = x - \frac{[x]}{N}$, а оценки для c_i вычисляются по формуле $\tilde{c}_i = \frac{[\varphi_i(x)\,y]}{[\varphi_i^2(x)]}$. В нашем случае $N = 16$, $[x] = 11{,}834$, значит $\varphi_1(x) = x - 0{,}73963$ [значения $\varphi_1(x)$ приведены в шестом столбце табл. 29]. Далее вычисляем накоплением суммы $[\varphi_1(x)y] = 5{,}75522$, $[\varphi_1^2(x)] = 1{,}16680$. Следовательно,

$$\tilde{c}_0 = \frac{82{,}97}{16} = 5{,}1856, \qquad \tilde{c}_1 = \frac{5{,}75522}{1{,}1668} = 4{,}932,$$

и уравнение искомой прямой имеет вид

$$y = 5{,}186 + 4{,}932\,(x - 0{,}73963) = 1{,}538 + 4{,}932x.$$

Для отыскания парабол второго и третьего порядка

$$y = c_0\varphi_0(x) + c_1\varphi_1(x) + c_2\varphi_2(x),$$
$$y = c_0\varphi_0(x) + c_1\varphi_1(x) + c_2\varphi_2(x) + c_3\varphi_3(x)$$

находим полиномы $\varphi_2(x)$ и $\varphi_3(x)$ по формулам

$$\varphi_2(x) = x^2 - \frac{[x^3] - \frac{[x][x^2]}{N}}{[x^2] - \frac{[x]^2}{N}}\left(x - \frac{[x]}{N}\right) - \frac{[x^2]}{N},$$

$$\varphi_3(x) = x^3 - \frac{[x^3\varphi_2(x)]}{[\varphi_2^2(x)]}\varphi_2(x) - \frac{[x^3\varphi_1(x)]}{[\varphi_1^2(x)]}\varphi_1(x) - \frac{[x^3]}{N}.$$

Подставляем значения соответствующих величин в формулу для $\varphi_2(x)$ и находим

$$\varphi_2(x) = x^2 - 1{,}55107x + 0{,}52725.$$

Значения $\varphi_2(x_i)$ приведены в седьмом столбце табл. 29. Вычисляем накоплением суммы $[x^3\varphi_2(x)] = 0{,}13229$, $[\varphi_2^2(x)] = 0{,}05574$, $[x^3\varphi_1(x)] = 2{,}24757$, $[\varphi_1^2(x)] = 1{,}16680$. Получим выражение для $\varphi_3(x)$

$$\varphi_3(x) = x^3 - 2{,}373\ \varphi_2(x) - 1{,}9263\ \varphi_1(x) - 0{,}57166.$$

Значения $\varphi_3(x_i)$ приведены в последнем столбце табл. 29.

Находим оценки для коэффициентов c_2 и c_3:

$$\tilde{c}_2 = \frac{[\varphi_2(x)\,y]}{[\varphi_2^2(x)]} = \frac{0{,}1934}{0{,}05574} = 3{,}470,$$

$$\tilde{c}_3 = \frac{[\varphi_3(x)\,y]}{[\varphi_3^2(x)]} = \frac{-0{,}01624}{0{,}002738} = -5{,}931.$$

Следовательно, уравнения искомых парабол второго и третьего порядка имеют вид

$$y = 3{,}368 - 0{,}450x + 3{,}470x^2,$$
$$y = 5{,}73 - 10{,}86x + 17{,}54x^2 - 5{,}93x^3.$$

Абсолютные величины наибольшего отклонения наблюденного значения от выравненного по прямой и по параболам второго и третьего порядка соответственно равны 0,40, 0,20, 0,11.

Заметим, что при последовательном вычислении значений $\varphi_i(x)$ по рекуррентным формулам происходит исчезновение значащих цифр, а потому целесообразно вести вычисления с некоторым запасом значащих цифр.

ГЛАВА XIII

НЕКОТОРЫЕ ИССЛЕДОВАНИЯ А. ВАЛЬДА. ПРЯМАЯ ОРТОГОНАЛЬНОЙ РЕГРЕССИИ И ЕЕ ПРИМЕНЕНИЯ

§ 1. Постановка задачи. Состоятельные оценки

Во введении и в гл. VII подробно рассматривалась задача оценивания коэффициентов α и β линии регрессии $y=\alpha x+\beta$ на основании ряда наблюдений l_r, соответственно фундаментальным уравнениям

$$l_r-\Delta_r=\alpha x_r+\beta \qquad (r=1,\ 2,\ \ldots,\ N).$$

Эта задача является важным частным случаем уравнивания с помощью элементов при матрице коэффициентов

$$X=\left\|\begin{matrix} x_1 & 1 \\ x_2 & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ x_N & 1 \end{matrix}\right\|.$$

Здесь, как и в общей схеме уравнивания, абсциссы x_i предполагаются точно известными, а случайные погрешности Δ_r предполагаются лишь при измерении ординат y_r.

Часто, однако, встречаются случаи, когда целесообразно считать, что и при измерении абсцисс имеются случайные погрешности δ_r. В примере 1 введения (данные Д. И. Менделеева) так получится, например, если признать, что не только при измерении растворимости $NaNO_3$, но и при измерении температуры имеются ощутимые случайные погрешности. Схема такого рода исследована А. Вальдом *) в 1940 г.

А. Вальд указывает, что подобная ситуация встречается, например, в задачах экономики капитализма. Пусть t_1, t_2, ..., t_n — моменты времени, когда отмечается цена x_i некоторого товара, и y_i — спрос на этот товар. Можно считать, что для каждого t_i существуют

*) А. Вальд (1902—1950) — американский математик-статистик.

математические ожидания $E(x_i)=X_i$ и $E(y_i)=Y_i$. Если X_i и Y_i связаны линейной зависимостью

$$Y_i=\alpha X_i+\beta,$$

то мы приходим к описанной ситуации.

Мы изложим здесь работу А. Вальда [5] как исследование задачи, тесно связанной с методом наименьших квадратов. Дадим точную постановку задачи.

Дано N пар наблюдений $(x_1, y_1), \ldots, (x_N, y_N)$, которые рассматриваются как отдельные значения случайных векторов, причем $E(x_i)=X_i$ и $E(y_i)=Y_i$. Величины X_i и Y_i будем называть *истинными значениями* x_i и y_i, а величины $\delta_i=x_i-X_i$ и $\Delta_i=y_i-Y_i$ — *погрешностями*. Будем считать выполненными следующие условия.

1. Случайные величины $\delta_1, \delta_2, \ldots, \delta_N$ одинаково распределены и некоррелированы: $E(\delta_i\delta_j)=0$ $(i\neq j)$; $E(\delta_i^2)$ существует $(i=1, 2, \ldots, N)$.

2. $\Delta_1, \ldots, \Delta_N$ одинаково распределены и некоррелированы: $E(\Delta_i\Delta_j)=0$ $(i\neq j)$; $E(\Delta_i^2)$ существует $(i=1, 2, \ldots, N)$.

3. Случайные погрешности δ_i и Δ_j некоррелированы: $E(\delta_i\Delta_j)=0$ $i, j=1, 2, \ldots, N)$.

4. Между истинными значениями X и Y существует линейная зависимость

$$Y_i=\alpha X_i+\beta. \tag{13.1.1}$$

Кроме того, из $E(x_i)=X_i$; $E(y_i)=Y_i$ следует, что $E(\delta_i)=0$, $E(\Delta_i)=0$.

Пусть δ — случайная величина, распределенная как δ_i, Δ — случайная величина, распределенная как Δ_i.

Требуется на основании наблюдений

$$(x_1, y_1), \ldots, (x_N, y_N) \tag{13.1.2}$$

произвести оценивание коэффициентов α, β и штандартов $\sigma(\delta)=\sigma_\delta$, $\sigma(\Delta)=\sigma_\Delta$ случайных величин δ и Δ.

Мы начнем с установления оценок для α и β таких, что при неограниченном увеличении числа наблюдений оценки уклоняются от соответствующих параметров менее чем на заданную величину ε с вероятностью, сколь угодно близкой к единице. (Такие оценки называются *состоятельными*, или *согласными*.) Для удобства изложения будем считать N четным, $N=2m$.

Введем статистики

$$\left.\begin{aligned} a_1&=\frac{(x_1+\ldots+x_m)-(x_{m+1}+\ldots+x_N)}{N},\\ a_2&=\frac{(y_1+\ldots+y_m)-(y_{m+1}+\ldots+y_N)}{N}. \end{aligned}\right\} \tag{13.1.3}$$

Для оценивания α будем употреблять статистику

$$a = \frac{a_2}{a_1} = \frac{(y_1 + \ldots + y_m) - (y_{m+1} + \ldots + y_N)}{(x_1 + \ldots + x_m) - (x_{m+1} + \ldots + x_N)}. \tag{13.1.4}$$

Наложим еще одно условие (пятое).*)

Условие 5:

$$\left| \frac{(X_1 + \ldots + X_m) - (X_{m+1} + \ldots + X_N)}{N} \right| > c_0 > 0, \tag{13.1.5}$$

где c_0 — положительная константа. Заметим, что так, например, будет, если $X_i - X_{i+1} = c_0 > 0$. Докажем, что при указанных условиях a является состоятельной оценкой для α, т. е.

$$P\{|a - \alpha| < \varepsilon\} \to 1 \tag{13.1.6}$$

при заданном $\varepsilon > 0$ и $N \to \infty$.

Обозначим $E(a_1) = \alpha_1$; $E(a_2) = \alpha_2$. Очевидно,

$$\left. \begin{aligned} \alpha_1 &= \frac{(X_1 + \ldots + X_m) - (X_{m+1} + \ldots + X_N)}{N}, \\ \alpha_2 &= \frac{(Y_1 + \ldots + Y_m) - (Y_{m+1} + \ldots + Y_N)}{N}. \end{aligned} \right\} \tag{13.1.7}$$

В силу (13.1.1) имеем

$$\alpha_2 = \alpha\alpha_1; \quad \frac{\alpha_2}{\alpha_1} = \alpha. \tag{13.1.8}$$

Далее,

$$D(a_1) = \frac{N}{N^2} D(x_1) = \frac{\sigma_\delta^2}{N},$$

и аналогично

$$D(a_2) = \frac{\sigma_\Delta^2}{N}.$$

Докажем теперь (13.1.6). Пусть задано $\varepsilon > 0$. Согласно неравенству Чебышева, имеем

$$\left. \begin{aligned} P\{|a_1 - \alpha_1| \leqslant \varepsilon\} &\geqslant 1 - \frac{\sigma_\delta^2}{N\varepsilon^2}, \\ P\{|a_2 - \alpha_2| \leqslant \varepsilon\} &\geqslant 1 - \frac{\sigma_\Delta^2}{N\varepsilon^2}. \end{aligned} \right\} \tag{13.1.9}$$

*) Для упрощения изложения здесь несколько усилено соответствующее условие А. Вальда (см. [5]).

В силу этого имеем для вероятности совмещения событий [см. (11.1.20)]

$$P\{|a_1-\alpha_1|\leqslant\varepsilon,\ |a_2-\alpha_2|\leqslant\varepsilon\}\geqslant 1-\frac{\sigma_\delta^2+\sigma_\Delta^2}{N\varepsilon^2}. \qquad (13.1.10)$$

Ввиду этого, с вероятностью, большей или равной $1-\frac{\sigma_\delta^2+\sigma_\Delta^2}{N\varepsilon^2}$, будем иметь

$$\frac{\alpha_2-\varepsilon}{\alpha_1+\varepsilon}\leqslant\frac{a_2}{a_1}\leqslant\frac{\alpha_2+\varepsilon}{\alpha_1-\varepsilon}. \qquad (13.1.11)$$

Ввиду пятого условия, $\alpha_1>c_0$. В силу $\alpha=\frac{\alpha_2}{\alpha_1}$ имеем

$$\left|\frac{a_2}{a_1}-\alpha\right|\leqslant\varepsilon' \qquad (13.1.12)$$

с вероятностью, большей или равной $1-\frac{\sigma_\delta^2+\sigma_\Delta^2}{N\varepsilon^2}$. Здесь $\varepsilon'=\varepsilon'(\varepsilon)$ сколь угодно мало вместе с ε. При заданном ε и $N\to\infty$ вероятность события (13.1.12) будет сколь угодно близка к единице.

Далее, начальную ординату β прямой $Y=\alpha X+\beta$ оцениваем следующим образом:

$$b=\bar{y}-a\bar{x}, \qquad (13.1.13)$$

где, как обычно, $\bar{y}=\frac{[y]}{N}$; $\bar{x}=\frac{[x]}{N}$. В силу закона больших чисел, $\bar{y}$ сходится по вероятности к $\bar{Y}=\frac{[Y]}{N}$, $\bar{x}$ сходится по вероятности к $\bar{X}=\frac{[X]}{N}$ и потому, в силу доказанного относительно a, оценка b будет сходиться по вероятности к $\bar{Y}-\alpha\bar{X}=\beta$ при $N\to\infty$, так что b будет состоятельной оценкой для β.

Теперь займемся отысканием состоятельных оценок для σ_δ^2 и σ_Δ^2. Введем выборочные штандарты

$$s_x=\sqrt{\frac{1}{N}\sum(x_i-\bar{x})^2}\ ; \quad s_y=\sqrt{\frac{1}{N}\sum(y_i-\bar{y})^2}\ ,$$

и „выборочную ковариацию“

$$s_{xy}=\sum\frac{(x_i-\bar{x})(y_i-\bar{y})}{N}.$$

Через s_X, s_Y, s_{XY} будем означать те же величины для истинных значений $X_1,\ldots,X_N$; $Y_1,\ldots,Y_N$.

Имеем

$$E(s_x^2)=s_X^2+\sigma_\delta^2\frac{N-1}{N}, \qquad (13.1.14)$$

$$E(s_y^2)=s_Y^2+\sigma_\Delta^2\frac{N-1}{N}, \qquad (13.1.15)$$

$$E(s_{xy})=s_{XY}. \qquad (13.1.16)$$

Докажем эти формулы.

Имеем

$$(x_i-\bar{x})=(x_i-X_i)+(X_i-\bar{X})-(\bar{x}-\bar{X}),$$

откуда

$$E\sum_{i=1}^{N}(x_i-\bar{x})^2=E\sum_{i=1}^{N}(x_i-X_i)^2+E\sum_{i=1}^{N}(X_i-\bar{X})^2+$$
$$+NE(\bar{x}-\bar{X})^2+2E\sum_{i=1}^{N}(X_i-\bar{X})(x_i-X_i)-2E\sum_{i=1}^{N}(X_i-\bar{X})(\bar{x}-\bar{X})-$$
$$-2E\sum_{i=1}^{N}(x_i-X_i)(\bar{x}-\bar{X}). \qquad (13.1.17)$$

Далее, $X_i-x_i=\delta_i$, $\bar{x}-\bar{X}=\bar{\delta}$, так что

$$E\sum_{i=1}^{N}\delta_i\bar{\delta}=N\frac{\sigma_\delta^2}{N}=\sigma_\delta^2;\quad E(\bar{\delta}^2)=\frac{\sigma_\delta^2}{N}.$$

Учитывая, что $X_i-\bar{X}$ суть величины постоянные, и $E(x_i-X_i)=$ $=E(\delta_i)=0$; $E(\bar{x}-\bar{X})=0$, переписываем (13.1.17) в виде

$$NE(s_x^2)=N\sigma_\delta^2+Ns_X^2+\sigma_\delta^2-2\sigma_\delta^2=(N-1)\sigma_\delta^2+Ns_X^2.$$

Деление на N дает (13.1.14). Равенство (13.1.15) доказывается так же.

Для доказательства (13.1.16) пишем

$$(x_i-\bar{x})(y_i-\bar{y})=\big((x_i-X_i)+(X_i-\bar{X})-(\bar{x}-\bar{X})\big)\times$$
$$\times[(y_i-Y_i)+(Y_i-\bar{Y})-(\bar{y}+\bar{Y})].$$

Беря математические ожидания и учитывая некоррелированность x_i и y_i, находим

$$E(x_i-\bar{x})(y_i-\bar{y})=(X_i-\bar{X})(Y_i-\bar{Y}),$$

что и приводит к (13.1.16).

Так как $Y_i=\alpha X_i+\beta$, то

$$s_Y=\alpha s_X, \qquad (13.1.18)$$

$$s_{XY}=\alpha s_X^2. \qquad (13.1.19)$$

Из этих формул и (13.1.16) находим

$$s_X^2=\frac{E(s_{xy})}{\alpha}, \qquad (13.1.20)$$

$$s_Y^2=\alpha E(s_{xy}). \qquad (13.1.21)$$

Подставляя эти выражения в (13.1.14) и (13.1.15), получаем

$$\sigma_\delta^2 = \frac{N}{N-1}\left(E\left(s_x^2\right) - \frac{E\left(s_{xy}\right)}{a}\right), \tag{13.1.22}$$

$$\sigma_\Delta^2 = \frac{N}{N-1}\left(E\left(s_y^2\right) - aE\left(s_{xy}\right)\right). \tag{13.1.23}$$

По закону больших чисел s_x^2, s_y^2 и s_{xy} сходятся по вероятности к своим математическим ожиданиям, и мы получаем состоятельные оценки для σ_δ^2 и σ_Δ^2 соответственно:

$$\frac{N}{N-1}\left(s_x^2 - \frac{s_{xy}}{a}\right), \tag{13.1.24}$$

$$\frac{N}{N-1}\left(s_y^2 - as_{xy}\right). \tag{13.1.25}$$

§ 2. Доверительные интервалы

Допустим теперь, что погрешности δ_r и Δ_r распределены по нормальным законам $N(0, \sigma_\delta)$ и $N(0, \sigma_\Delta)$. Будем считать выполненными все предыдущие условия, за исключением условия 5. Введем статистики

$$\bar{x}_1 = \frac{x_1 + \ldots + x_m}{m}; \qquad \bar{y}_1 = \frac{y_1 + \ldots + y_m}{m};$$

$$\bar{x}_2 = \frac{x_{m+1} + \ldots + x_N}{m}; \qquad \bar{y}_2 = \frac{y_{m+1} + \ldots + y_N}{m};$$

$$\left(s_x'\right)^2 = \frac{\sum\limits_{i=1}^{m}(x_i - \bar{x}_1)^2 + \sum\limits_{j=m+1}^{N}(x_j - \bar{x}_2)^2}{N};$$

$$\left(s_y'\right)^2 = \frac{\sum\limits_{i=1}^{m}(y_i - \bar{y}_1)^2 + \sum\limits_{j=m+1}^{N}(y_j - \bar{y}_2)^2}{N};$$

$$s_{xy}' = \frac{\sum\limits_{i=1}^{m}(x_i - \bar{x}_1)(y_i - \bar{y}_1) + \sum\limits_{j=m+1}^{N}(x_j - \bar{x}_2)(y_j - \bar{y}_2)}{N};$$

величины $\bar{X}_1$, $\bar{X}_2$, $\left(s_X'\right)^2$, $\left(s_Y'\right)^2$, s_{XY}' означают такие же функции от истинных значений $X_1, \ldots, X_N$; $Y_1, \ldots, Y_N$.

Выражения s_x', s_y', s_{xy}' несколько отличаются от s_x, s_y, s_{xy}. Они вводятся потому, что являются независимыми от углового коэффициента $a = \frac{a_2}{a_1} = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1}$, в то время как прежние величины s_x, s_y, s_{xy}, вообще говоря, зависят от a.

Докажем независимость s'_x, s'_y, s'_{xy} от a. Достаточно доказать, что $x_i-\bar{x}_1$, $y_i-\bar{y}_1$ $(i=1, 2, \ldots, m)$; $x_j-\bar{x}_2$, $y_j-\bar{y}_2$, $(j=m+1, \ldots, N)$ не зависят от $\bar{x}_2-\bar{x}_1$ и $\bar{y}_2-\bar{y}_1$.

Заметим, что все указанные случайные величины в совокупности нормальны. Достаточно доказать, что математические ожидания соответствующих попарных произведений равны нулю. Имеем, например,

$$E\left[\left(x_i-\bar{x}_1-E(x_i-\bar{x}_1)\right)\left(\bar{x}_2-\bar{x}_1-E(\bar{x}_2-\bar{x}_1)\right)\right]=$$
$$=E(x_i-\bar{x}_1)E(\bar{x}_2)-E(x_i\bar{x}_1)+E(\bar{x}_1^2)-(X_i-\bar{X}_1)(\bar{X}_2-\bar{X}_1)-$$
$$-(X_i-\bar{X}_1)(\bar{X}_2-\bar{X}_1)+(X_i-\bar{X}_1)(\bar{X}_2-\bar{X}_1)=(X_i-\bar{X}_1)\bar{X}_2-$$
$$-X_i\bar{X}_1-\frac{\sigma_\delta^2}{m}+\bar{X}_1^2+\frac{\sigma_\delta^2}{m}-(X_i-\bar{X}_1)(\bar{X}_2-\bar{X}_1)=0.$$

Подсчет, совершенно аналогичный тому, который приводит к (13.1.22) и (13.1.23), дает

$$\sigma_\delta^2=\left[E\left(s'_x\right)^2-\frac{E\left(s'_{xy}\right)}{\alpha}\right]\frac{N}{N-2}, \tag{13.2.1}$$

$$\sigma_\Delta^2=\left[E\left(s'_y\right)^2-\alpha E\left(s'_{xy}\right)\right]\frac{N}{N-2}. \tag{13.2.2}$$

Заметим, что эти формулы отличаются от (13.1.22) и (13.1.23) лишь заменой $N-1$ на $N-2$.

Если α известно, то для σ_δ^2 и σ_Δ^2 получаются несмещенные оценки

$$\left[\left(s'_x\right)^2-\frac{s'_{xy}}{\alpha}\right]\frac{N}{N-2}, \tag{13.2.3}$$

$$\left[\left(s'_y\right)^2-\alpha s'_{xy}\right]\frac{N}{N-2}. \tag{13.2.4}$$

Отсюда при известном α получаем для $\sigma_\Delta^2+\alpha^2\sigma_\delta^2$ несмещенную оценку

$$s^2=\left(\left(s'_y\right)^2+\alpha^2\left(s'_x\right)^2-2\alpha s'_{xy}\right)\frac{N}{N-2}=$$
$$=\frac{1}{N-2}\left[\sum_{i=1}^{m}\left((y_i-\alpha x_i)-(\bar{y}_1-\alpha\bar{x}_1)\right)^2+\right.$$
$$\left.+\sum_{j=m+1}^{N}\left((y_j-\alpha x_j)-(\bar{y}_2-\alpha\bar{x}_2)\right)^2\right]. \tag{13.2.5}$$

Докажем теперь, что величина

$$\frac{(N-2)s^2}{\sigma_\Delta^2+\alpha^2\sigma_\delta^2} \tag{13.2.6}$$

распределена по χ^2_{N-2}.

В самом деле,

$$(y_i-\alpha x_i)-(\overline{y}_1-\alpha\overline{x}_1)=\Delta_i-\alpha\delta_i-(\overline{\Delta}_1-\alpha\overline{\delta}_1) \qquad (i=1,\ \ldots,\ m),$$

ибо

$$\overline{y}_1-\alpha\overline{x}_1=\overline{\Delta}_1-\alpha\overline{\delta}_1+\overline{Y}_1-\alpha\overline{X}_1=\overline{\Delta}_1-\alpha\overline{\delta}_1+\beta,$$

и аналогично

$$(y_j-\alpha x_j)-(\overline{y}_2-\alpha\overline{x}_2)=\Delta_j-\alpha\delta_j-(\overline{\Delta}_2-\alpha\overline{\delta}_2),\quad (j=m+1,\ \ldots,\ N),$$

где

$$\overline{\Delta}_1=\frac{\Delta_1+\ldots+\Delta_m}{m};\quad \overline{\Delta}_2=\frac{\Delta_{m+1}+\ldots+\Delta_N}{m};$$

$$\overline{\delta}_1=\frac{\delta_1+\ldots+\delta_m}{m};\quad \overline{\delta}_2=\frac{\delta_{m+1}+\ldots+\delta_N}{m}.$$

Далее, $D(\Delta_k-\alpha\delta_k)=\sigma_\Delta^2+\alpha^2\sigma_\delta^2$, и $\Delta_k-\alpha\delta_k$ не зависит от $\Delta_l-\alpha\delta_l$ $(k\neq l)$. Поэтому величина

$$\frac{(N-2)\,s^2}{\sigma_\Delta^2+\alpha^2\sigma_\delta^2}=\frac{1}{\sigma_\Delta^2+\alpha^2\sigma_\delta^2}\left(\sum_{i=1}^{m}\left(\Delta_i-\alpha\delta_i-(\overline{\Delta}_i-\alpha\overline{\delta}_i)\right)^2+\right.$$

$$\left.+\sum_{j=m+1}^{N}\left((\Delta_j-\alpha\delta_j)-(\overline{\Delta}_j-\alpha\overline{\delta}_j)\right)^2\right)$$

есть сумма двух независимых величин, распределенных по χ^2_{m-1}, т. е. она распределена по $\chi^2_{2m-2}=\chi^2_{N-2}$, что и требовалось доказать.

Покажем теперь, что величина

$$\frac{\sqrt{N}\,a_1(a-\alpha)}{\sqrt{\sigma_\Delta^2+\alpha^2\sigma_\delta^2}}=\frac{\sqrt{N}\,(a_2-\alpha a_1)}{\sqrt{\sigma_\Delta^2+\alpha^2\sigma_\delta^2}} \tag{13.2.7}$$

нормальна $N(0,\ 1)$.

В самом деле, из (13.1.3)—(13.1.7) видно, что

$$a_1(a-\alpha)=a_2-\alpha a_1=\alpha_2+\frac{\overline{\Delta}_1-\overline{\Delta}_2}{2}-a_1\frac{\alpha_2}{\alpha_1}=$$

$$=\alpha_2+\frac{\overline{\Delta}_1-\overline{\Delta}_2}{2}-\left(\alpha_1+\frac{\overline{\delta}_1-\overline{\delta}_2}{2}\right)\frac{\alpha_2}{\alpha_1}=\frac{\overline{\Delta}_1-\overline{\Delta}_2}{2}-\alpha\,\frac{\overline{\delta}_1-\overline{\delta}_2}{2}.$$

Последнее выражение нормально, и его дисперсия равна

$$\frac{1}{4}D(\overline{\Delta}_1-\overline{\Delta}_2)+\alpha^2D(\overline{\delta}_1-\overline{\delta}_2)=\frac{1}{4}\left(\frac{2\sigma_\Delta^2}{m}+\alpha^2\frac{2\sigma_\delta^2}{m}\right)=\frac{\sigma_\Delta^2+\alpha^2\sigma_\delta^2}{N}.$$

Таким образом, величина (13.2.7) нормальна $N(0, 1)$. Далее, величина (13.2.7) не зависит от (13.2.6), ибо величины

$$(\Delta_i - \overline{\Delta}_1) - \alpha(\delta_i - \overline{\delta}_1) \qquad (i = 1, 2, \ldots, m),$$
$$(\Delta_j - \overline{\Delta}_2) - \alpha(\delta_j - \overline{\delta}_2) \qquad (j = m+1, \ldots, N)$$

не зависят от $\frac{\overline{\Delta}_1 - \overline{\Delta}_2}{2} - \alpha(\overline{\delta}_1 - \overline{\delta}_2)$, что можно доказать, рассматривая произведения соответствующих математических ожиданий.

Ввиду этого, величина

$$t = \frac{\sqrt{N}\, a_1 (a - \alpha)}{s} \tag{13.2.8}$$

имеет распределение Стьюдента с $N-2$ степенями свободы. Подставляя значение s из (13.2.5), находим, что

$$t = \frac{a_1 (a - \alpha)\sqrt{N-2}}{\sqrt{(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}}}. \tag{13.2.9}$$

По заданной надежности p_0 и числу степеней свободы $N-2$ в таблице распределения Стьюдента выбирается число γ такое, что $P(|t_0| \leqslant \gamma) = p_0$. Тогда с вероятностью p_0 имеем

$$a_1^2 (a - \alpha)^2 \geqslant \left[(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}\right] \frac{t_0^2}{N-2}, \tag{13.2.10}$$

и доверительная область для α получится при решении такого неравенства относительно α.

Покажем теперь, что при условии

$$a_1^2 > \frac{(s_x)^2 t_0^2}{N-2} \tag{13.2.11}$$

уравнение

$$a_1^2 (a - \alpha)^2 = \left((s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}\right) \frac{t_0^2}{N-2} \tag{13.2.12}$$

будет иметь лишь реальные корни k_1 и k_2 ($k_2 \geqslant k_1$) и доверительным интервалом для α с надежностью p_0 будет интервал

$$I = [k_1, k_2]. \tag{13.2.13}$$

Из (13.2.5) мы имеем

$$(s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy} > 0 \text{ для всех } \alpha.$$

Значит, при $\alpha = a$ левая часть (13.2.12), равная нулю, меньше, чем правая часть. Таким образом, выражение

$$a_1^2 (a - \alpha)^2 - \left((s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy}\right) \frac{t_0^2}{N-2} \tag{13.2.14}$$

будет отрицательно при $\alpha = a$. Если выполнено (13.2.11), то после умножения на $\frac{t_0^2}{N-2}$ выражение (13.2.14) не превосходит

$$\begin{aligned}(a-\alpha)^2 (s'_x)^2 - \left((s'_y)^2 + \alpha^2 (s'_x)^2 - 2\alpha s'_{xy} \right) &= \\ = \left((a-\alpha)^2 - \alpha^2 \right) (s'_x)^2 - (s'_y)^2 - 2\alpha s'_{xy} &= \\ = (a^2 - 2\alpha a) \left(s'^2_x + \frac{s'_{xy}}{a} \right) - a\,(s'_{xy}) - (s'_y)^2 &. \end{aligned} \tag{13.2.15}$$

Предполагая, что $s'^2_x + \frac{s'_{xy}}{a} \neq 0$, видим, что найдется $\alpha = a'$, при котором выражения (13.2.15) и, стало быть, (13.2.14), положительны. Приравняв (13.2.14) к нулю, мы получим уравнение второй степени относительно α, которому соответствует парабола, при $\alpha = a$ расположенная ниже оси $O\alpha$ и имеющая одну абсциссу a', при которой она выше оси $O\alpha$. Следовательно, найдется и вторая точка с абсциссой $\alpha = a''$ такая, что, не нарушая общности, можем считать $a' < a < a''$. Тогда и точки пересечения оси $O\alpha$, т. е. корни выражения (13.2.14) k_1 и k_2, будут реальными, причем $k_1 < a < k_2$. Далее, неравенство (13.2.10) будет соблюдаться при $k_1 \leqslant \alpha \leqslant k_2$, так что (13.2.13) действительно дает нужный доверительный интервал.

Заметим теперь, что неравенство (13.2.11) будет иметь место с вероятностью, сколь угодно близкой к единице, если N велико и условие 5 выполнено, ибо тогда a_1 будет сходиться по вероятности к некоторой положительной постоянной, а $\frac{(s_x)^2 t_0}{N-2}$ будет сходиться по вероятности к нулю.

Теперь будем строить доверительный интервал для β, если α считать точно известным.

Рассмотрим величины

$$b_\alpha = \bar{y} - \alpha\bar{x}; \quad \bar{x} = \frac{[x]}{N}; \quad \bar{y} = \frac{[y]}{N}.$$

Имеем

$$b_\alpha - \beta = (\bar{y} - \bar{Y}) - \alpha(\bar{x} - \bar{X}) = \bar{\Delta} - \alpha\bar{\delta},$$

где

$$\bar{\Delta} = \frac{[\Delta]}{N}; \quad \bar{\delta} = \frac{[\delta]}{N}.$$

Отсюда заключаем, что величина

$$\frac{(b_\alpha - \beta)\sqrt{N}}{\sqrt{\sigma_\Delta^2 + \alpha^2\sigma_\delta^2}} \tag{13.2.16}$$

нормальна $N(0, 1)$. Она не зависит от (13.2.6), так что величина

$$t=\frac{\sqrt{N-2}\,(b_\alpha-\beta)}{\sqrt{\left(s_y'\right)^2+\alpha'\left(s_x'\right)^2-2\alpha s_{xy}'}}$$

имеет распределение Стьюдента с $N-2$ степенями свободы. По заданной надежности p_0 и числу степеней свободы $N-2$ выбираем в таблице распределения Стьюдента число γ и получаем доверительный интервал для оценивания β:

$$I=\left[b_\alpha-\gamma\sqrt{\frac{\left(s_y'\right)^2+\alpha^2\left(s_x'\right)^2-2\alpha s_{xy}'}{N-2}},\right.$$
$$\left.b_\alpha+\gamma\sqrt{\frac{\left(s_y'\right)^2+\alpha^2\left(s_x'\right)^2-2\alpha s_{xy}'}{N-2}}\right].$$

§ 3. Группировка наблюдений

В § 1 и 2 мы производили оценивание α и β на основе разбиения наблюдений на две группы — G_1 и G_2. Группа G_1 состояла из пар наблюдений $(x_1, y_1), \ldots, (x_m, y_m)$, а G_2 — из $(x_{m+1}, y_{m+1}), \ldots$ $\ldots, (x_N, y_N)$. Однако совершенно такие же выводы можно было делать на основе любого разбиения на равные по объему группы G_1 и G_2 при одном лишь условии, что правило разбиения не зависит от погрешностей $\delta_1, \ldots, \delta_N$; $\Delta_1, \ldots, \Delta_N$, т. е. от самого набора наблюдений (x_k, y_k) $(k=1, 2, \ldots, N)$. Мы можем сохранить те же группы G_1 и G_2, но по-разному нумеровать наблюдения. Это равносильно тому же разбиению на две группы. Вопрос о наилучшем разбиении сам по себе не имеет однозначного ответа, так как надо знать, в каком смысле мы понимаем слово „наилучшее“. Если мы поставим вопрос о качестве оценивания α, то ввиду того, что для α давалась точечная оценка $a=\frac{a_2}{a_1}$, естественно стараться сделать a_1 (и тем самым a_2) возможно большим с вероятностью, близкой к единице, ибо тогда флуктуации знаменателя дадут наименее чувствительные флуктуации оценки a.

Мы имеем

$$a_1=\frac{(x_1+\ldots+x_m)-(x_{m+1}+\ldots+x_N)}{N}. \tag{13.3.1}$$

Если бы разрешалось делать разбиение на две группы на основе самих наблюдений, то можно было бы заметить, что максимум $|a_1|$ мы получим, если перенумеруем наблюдения в порядке возрастания: $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_N$. В самом деле, если половина чисел x_n должна иметь знак $+$, а вторая половина знак $-$, то наибольшее значение $|a_1|$ получится, если m наименьших чисел имеют один знак, а m наибольших — другой. Хотя это правило вносит зависимость от наблюдений, практически при больших N и небольших дисперсиях погрешностей им можно пользоваться, используя предыдущие формулы.

§ 4. Линия ортогональной регрессии (градиент) и ее применение

В § 1 введения мы знакомились с задачей проведения прямой линии, у которой сумма квадратов расстояний от точек заданной системы минимальна. Уточним математико-статистическую постановку задачи. Пусть система точек $P_1, \ldots, P_N$ плоскости отличается следующим свойством: существует прямая линия l такая, что если отсчитывать расстояния Δ_i от P_i до l, придавая Δ_i знак плюс по одну сторону прямой l и знак минус по другую, то величины $\Delta_1, \ldots, \Delta_N$ в совокупности нормальны $N(0, \sigma)$ и независимы.

Возьмем прямоугольную систему координат XOY; пусть точка P_i имеет в ней координаты (x_i, y_i), а искомая линия l имеет уравнение

$$y = \alpha x + \beta, \tag{13.4.1}$$

где α и β неизвестны. Требуется оценить α и β по наблюдениям (x_i, y_i) $(i = 1, 2, \ldots, N)$.

Мы укажем в дальнейшем пример, когда такая постановка задачи естественна. По известной формуле аналитической геометрии имеем

$$\Delta_i = \frac{y_i - \alpha x_i - \beta}{\sqrt{\alpha^2 + 1}}. \tag{13.4.2}$$

Поэтому функция правдоподобия для $\Delta_1, \ldots, \Delta_N$ будет иметь вид

$$L = (2\pi)^{-\frac{N}{2}} \sigma^{-N} \exp\left\{-\frac{1}{2(\alpha^2+1)\sigma^2} \sum_{i=1}^{N} (y_i - \alpha x_i - \beta)^2\right\}. \tag{13.4.3}$$

Условие максимального правдоподобия сводится к тому, чтобы при всяком σ было

$$\sum_{i=1}^{N} \left(\frac{y_i - \alpha x_i - \beta}{\sqrt{\alpha^2+1}}\right)^2 = \min, \tag{13.4.4}$$

т. е. α и β надо определить так, чтобы сумма квадратов расстояний от точек (x_i, y_i) до прямой $y = \alpha x + \beta$ была минимальной. Такая задача рассмотрена в примере 6 введения; решение (вообще говоря, неоднозначное) дается формулами [см. (0.1.29)]

$$\operatorname{tg} 2\alpha' = \frac{2\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2 - \sum_{i=1}^{N}(y_i - \bar{y})^2}, \tag{13.4.5}$$

$$\bar{y} - \alpha'\bar{x} - \beta' = 0. \tag{13.4.6}$$

В наиболее актуальных случаях точки P_i располагаются в довольно длинной и узкой полосе, так что $\operatorname{tg} 2\alpha'$ определяется однозначно, и искомое значение из двух возможных α' и $\alpha' + \frac{\pi}{2}$ (отвечающее, как пояснялось во введении, оси с минимальным моментом инерции) находится однозначно. Экстремальные в определенном смысле качества и асимптотическая нормальность оценок (α', β') следуют из общих свойств оценок по способу максимального правдоподобия (см. гл. III). Получаемая линия называется *линией ортогональной регрессии*, или *линией градиента*.

В качестве примера, когда такая линия играет существенную роль, можно указать задачу построения средней линии профиля обработанной поверхности (см. Ю. В. Линник и А. П. Хусу [33]). При исследовании профиля поверхности, имеющей лишь шероховатость и не имеющей волнистости, обычно имеется сильно увеличенная фотография микронеровностей профиля. Соображения теории вероятностных процессов (случайных кривых) заставляют предположить, что должна существовать ось l такая, что если отнести к ней профиль, как к оси абсцисс, то по оси ординат получится реализация нормальной стационарной случайной кривой. Это предположение хорошо подтверждается наблюдениями.

Если брать ряд абсцисс $\xi_1, \ldots, \xi_N$ на оси l на таком расстоянии, чтобы разрушались корреляционные связи соответствующих ординат $\eta_1, \ldots, \eta_N$, то получим ряд точек $P_1, \ldots, P_N$, которые будут подчиняться указанной в начале параграфа схеме. Таким образом, нахождение оси l, которую можно назвать *средней линией профиля*, приведется к оцениванию коэффициентов для линии ортогональной регрессии по формулам (13.4.5) и (13.4.6). Линия ортогональной регрессии применяется также в некоторых вопросах теории корреляции (см., например, применение к исследованию цен на картофель в Швеции в книге Г. Крамера [25], стр. 505, и теоретическое исследование Моники Кризи [27]).

ГЛАВА XIV

ДОПОЛНИТЕЛЬНЫЕ СВЕДЕНИЯ О МЕТОДЕ НАИМЕНЬШИХ КВАДРАТОВ

§ 1. Доверительные эллипсоиды

Вернемся к схеме уравнивания по элементам (гл. VIII), сохраняя обозначения этой главы. Имеем

$$Y = X^{(0)} + XA, \tag{14.1.1}$$

где $X^{(0)} = X^{(0)}_{N1}$ и $X = X_{Nn}$ — известные матрицы; ранг $(X) = n$ $(N > n)$; $Y = Y_{N1}$ — неизвестно, $A = A_{n1} = \begin{Vmatrix} a_1 \\ \vdots \\ a_n \end{Vmatrix}$ — матрица неизвестных параметров (элементов).

Матрица наблюдений имеет вид $L \doteq L_{N1} = Y + \Delta$, где $\Delta = \Delta_{N1} = \begin{Vmatrix} \Delta_1 \\ \vdots \\ \Delta_N \end{Vmatrix}$ — нормальный вектор случайных погрешностей, причем

$$E(\Delta) = 0; \quad B_\Delta = E(\Delta\Delta^T) = \sigma^2 P^{-1}, \tag{14.1.2}$$

$P = P_{NN} = \begin{Vmatrix} p_1 & 0 \ldots & 0 \\ \cdot & \cdot \; \cdot \; \cdot & \cdot \\ 0 & 0 \ldots & p_N \end{Vmatrix}$; $p_i > 0$ — матрица известных весов наблюдений; σ^2 — неизвестный параметр.

Ранее (гл. VIII) мы строили доверительные интервалы для оценивания $a_1, a_2, \ldots, a_n$ порознь; в гл. VII рассматривались также интервалы для оценивания линейной формы $l = g_1 a_1 + \ldots + g_n a_n$ от наших параметров с известными коэффициентами.

Пусть теперь нам желательно оценить m линейных функций

$$h_r = g_{r1} a_1 + \ldots + g_{rn} a_n; \quad r = 1, 2, \ldots, m; \quad 1 \leqslant m \leqslant n, \tag{14.1.3}$$

от элементов a_i с известными коэффициентами g_{rj}.

Введем матрицу

$$G = G_{mn} = \|g_{rj}\| \tag{14.1.4}$$

и переписываем (14.1.3) в виде

$$H = H_{m1} = GA; \qquad H = \left\| \begin{matrix} h_1 \\ \cdot \\ \cdot \\ \cdot \\ h_m \end{matrix} \right\|. \tag{14.1.5}$$

Не нарушая общности, будем считать, что ранг $(G) = m$. Будем строить доверительную область для $H = H_{m1}$. Через $\tilde{A}$, как и ранее (гл. VIII), обозначим матрицу $\left\| \begin{matrix} \tilde{a}_1 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{a}_n \end{matrix} \right\|$ оценок A по методу наименьших квадратов.

***Теорема 14.1.1* (*Ю. В. Линник* [32]).** *Пусть* $\tilde{H} = G\tilde{A}$, $\tilde{K} = K_{mm} = GC^{-1}G^T$ *и* $Z = \left\| \begin{matrix} z_1 \\ \cdot \\ \cdot \\ \cdot \\ z_m \end{matrix} \right\|$ *— вектор текущих координат* $(z_1, \ldots, z_m)$. *Тогда матрица K неособенная, и доверительный эллипсоид* $Э_{\gamma_0}$:

$$(Z - \tilde{H})^T K^{-1} (Z - \tilde{H}) = \gamma_0 [p\tilde{v}\tilde{v}] \tag{14.1.6}$$

накрывает точку $Z = H$ с вероятностью p_0, причем

$$F_{m,\, N-n}(\gamma_0) = p_0, \tag{14.1.7}$$

где

$$F_{m,\, N-n}(x) = \frac{\Gamma\left(\frac{N-n+m}{2}\right)}{\Gamma\left(\frac{N-n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \int_0^x \frac{u^{\frac{m}{2}-1}}{(1+u)^{\frac{N-n+m}{2}}}\, du \tag{14.1.8}$$

— функция распределения Фишера (см. § 6 гл. II).

Доказательство. Как и в гл. VIII, обозначаем

$$C = X^T P X, \tag{14.1.9}$$

$C = C_{nn}$ — неособенная матрица, и

$$\tilde{A} - A = C^{-1} X^T P \Delta. \tag{14.1.10}$$

Далее, $\tilde{A}$ есть n-мерный нормальный вектор, и ранг $(G) = m \leqslant n$. Ввиду этого, $\tilde{H} - H = G(\tilde{A} - A)$ по теореме 2.3.1 будет m-мерным нормальным вектором. При этом

$$E(\tilde{H} - H) = 0, \tag{14.1.11}$$

а корреляционная матрица имеет вид

$$B_{\tilde{H}}=B_{(\tilde{H}-H)}=E\left[G(\tilde{A}-A)\left(G(\tilde{A}-A)\right)^T\right]=$$
$$=GB_{\tilde{A}}G^T=\sigma^2GC^{-1}G^T=\sigma^2K. \qquad (14.1.12)$$

Так как $\tilde{H}$ есть m-мерный нормальный вектор и $K=K_{mm}$, то K должна быть неособенной матрицей. Составим теперь положительную квадратичную форму

$$Q=\frac{1}{\sigma^2}(\tilde{H}-H)^TK^{-1}(\tilde{H}-H). \qquad (14.1.13)$$

и докажем, что она распределена, как χ^2_m.

Пусть $f(x_1, \ldots, x_m)$ — плотность вероятности для вектора $\tilde{H}-H$. Как известно, при $X=\left\|\begin{matrix} x_1 \\ \vdots \\ x_m \end{matrix}\right\|$

$$f(x_1, \ldots, x_m)=(2\pi)^{-\frac{m}{2}}\sigma^{-m}\left(\det(K)\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}X^TK^{-1}X\right\} \qquad (14.1.14)$$

(см. § 2 гл. II). Далее, при всяком $x>0$ имеем

$$P\{Q\leqslant x\}=\int\limits_{\sigma^2X^TK^{-1}X\leqslant x}\ldots\int f(x_1, \ldots, x_m)\,dx_1\ldots dx_m.$$

Пусть $F=F_{mm}$ — ортогональная матрица такая, что квадратичная форма $\sigma^2X^TK^{-1}X$ подстановкой $X=FZ$ приводится к диагональному виду Z^TDZ, где

$$D=\sigma^2F^TK^{-1}F=\left\|\begin{matrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & d_n \end{matrix}\right\|.$$

Тогда $d_i>0$ и, ввиду того что $\det(F)=\pm 1$ и соответствующий якобиан равен единице, получаем

$$P\{Q\leqslant x\}=$$
$$=(2\pi)^{-\frac{m}{2}}\left(\det(\sigma^2K)\right)^{-\frac{1}{2}}\int\limits_{Z^TDZ\leqslant x}\ldots\int\exp\left\{-\frac{1}{2}Z^TDZ\right\}dz_1.\ \ldots,\ dz_m.$$

Далее, $\det(D)=d_1d_2\ldots d_m=\left(\det(\sigma^2K)\right)^{-1}$, так что

$$P\{Q\leqslant x\}=$$
$$=(2\pi)^{-\frac{m}{2}}(d_1d_2\ldots d_m)^{\frac{1}{2}}\int\limits_{d_1z_1^2+\ldots+d_mz_m^2\leqslant x}\ldots\int\exp\left\{-\frac{1}{2}\left(\sum_{j=1}^m d_jz_j^2\right)\right\}dz_1\ldots dz_m.$$

Полагая $z_j\sqrt{d_j}=w_j$ $(j=1, 2, \ldots, m)$, находим

$$P\{Q\leqslant x\}=$$
$$=(2\pi)^{-\frac{m}{2}}\underset{w_1^2+\ldots+w_m^2\leqslant x}{\int\ldots\int}\exp\left\{-\frac{1}{2}\left(w_1^2+\ldots+w_m^2\right)\right\}dw_1\ldots dw_m. \tag{14.1.15}$$

По определению χ_m^2 имеем

$$P\{Q\leqslant x\}=P\left\{\chi_m^2\leqslant x\right\},$$

что и требовалось доказать.

Как известно из гл. VIII, вектор $\tilde{A}$ статистически не зависит от $[p\tilde{v}\tilde{v}]$, это же касается и вектора $\tilde{H}-H$ и его функции—квадратичной формы Q. Ввиду этого имеем

$$\frac{Q}{[p\tilde{v}\tilde{v}]\,\sigma^{-2}}=\frac{(H-\tilde{H})^TK^{-1}(H-\tilde{H})}{[p\tilde{v}\tilde{v}]}=\frac{(\chi_m')^2}{\chi_{N-n}^2}, \tag{14.1.16}$$

где χ^2-распределения в числителе и знаменателе независимы.

Таким образом, при заданном $\gamma_0>0$ вероятность события

$$(H-\tilde{H})^TK^{-1}(H-\tilde{H})\leqslant\gamma_0[p\tilde{v}\tilde{v}] \tag{14.1.17}$$

(обозначим это событие через $\mathfrak{A}$) есть

$$P(\mathfrak{A})=F_{m,\,N-n}(\gamma_0), \tag{14.1.18}$$

где $F_{m,\,N-n}(x)$ — соответствующее распределение Фишера. Накрытие эллипсоидом $Э_{\gamma_0}$ [см. (14.1.6)] точки $Z=H$ совпадает с событием $\mathfrak{A}$, и теорема доказана.

В гл. XI мы уже применяли частные случаи эллипсоида $Э_{\gamma_0}$ к задачам прямых и обратных засечек. Рассмотрим еще некоторые частные случаи. Если $G=E=E_{nn}$, $H=A$, то мы ищем доверительную область для всех параметров $(a_1, a_2, \ldots, a_n)$. В этом случае $K=C^{-1}$; $K^{-1}=C$, и эллипсоид $Э_{\gamma_0}$ имеет вид

$$(Z-\tilde{A})^TC(Z-\tilde{A})=\gamma_0[p\tilde{v}\tilde{v}]. \tag{14.1.19}$$

Другой крайний случай — выделение одного параметра, скажем, a_i. Полагаем $G=\|0, 0, \ldots, 1, 0, \ldots, 0\|$ (единица на i-м месте). Тогда

$$K=GC^{-1}G^T=\left\{C^{-1}\right\}_{ii}. \tag{14.1.20}$$

Доверительный эллипсоид $Э_{\gamma_0}$ принимает вид

$$\frac{(z-\tilde{a}_i)^2}{\left\{C^{-1}\right\}_{ii}}=\gamma_0\,[p\tilde{v}\tilde{v}], \tag{14.1.21}$$

а накрытие им параметра a_i есть событие

$$\frac{|a_i - \tilde{a}_i|}{\sqrt{\{C^{-1}\}_{ii}\,[p\tilde{v}\tilde{v}]}} \leqslant \sqrt{\gamma_0}. \tag{14.1.22}$$

Если ввести в рассмотрение равносильное событие $\mathfrak{A}$:

$$\frac{|a_i - \tilde{a}_i|}{\sqrt{\{C^{-1}\}_{ii}\frac{[p\tilde{v}\tilde{v}]}{N-n}}} \leqslant \sqrt{\gamma_0(N-n)}, \tag{14.1.23}$$

то $P(\mathfrak{A})$ будет подсчитываться по закону Стьюдента с $(N-n)$ степенями свободы, и мы получим те же доверительные интервалы для оценивания a_i порознь, которые описывались в гл. VIII.

Сделаем еще некоторые замечания по поводу решения уравнения (14.1.7):

$$F_{m,\,N-n}(\gamma_0) = p_0.$$

Нетрудно обнаружить, что распределение Фишера $F_{m,\,N-n}(x)$ всегда выражается в элементарных функциях (см. Ю. В. Линник [32]).

В случае $m=n=2$, который нам встречался в гл. XI, в задачах засечек на плоскостях, непосредственно получалось [см. (11.1.33) и (11.1.34)]:

$$\gamma_0 = \left(\frac{1}{1-p_0}\right)^{\frac{2}{N-2}} - 1;$$

$\gamma_0 = \frac{2}{N-2}\ln\frac{1}{1-p_0} + O\left(\frac{1}{N^2}\right)$ при возрастании N. В других случаях выражения получаются громоздкими, и мы получим лишь асимптотическое выражение для γ_0 при возрастании N. Проще всего это сделать, исходя из вероятностного смысла распределения Фишера. Уравнение (14.1.7) запишем в виде

$$F_{m,\,N-n}(\gamma_0) = P\left\{(\chi'_m)^2 \leqslant \gamma_0\chi^2_{N-n}\right\} = p_0. \tag{14.1.24}$$

Здесь будем считать $m \leqslant n$, n ограничено; $N\to\infty$. Величины $(\chi'_m)^2$ и χ^2_{N-n} статистически независимы. Полагаем при заданном n и $N\to\infty$

$$\chi^2_{N-n} = N - n + \xi_N, \quad \text{где } E\xi_n = 0 \quad \text{и} \quad \sigma(\xi_N) = O(\sqrt{N});$$

тогда (14.1.24) примет вид

$$F_{m,\,N-n}(\gamma_0) = P\left\{(\chi'_m)^2 \leqslant \gamma_0\left(1+\frac{\xi_N}{N-n}\right)(N-n)\right\} = p_0. \tag{14.1.25}$$

Пренебрегая флуктуациями $\frac{\xi_N}{N-n}$ видим, что

$$\gamma_0 \sim \frac{\xi_0}{N-n},$$

где ξ_0 находится по таблице распределения χ_m^2: если $P\{\chi_m^2 \leqslant x\} = K_m(x)$, то

$$K_m(\xi_0) = p_0. \tag{14.1.26}$$

Отсюда нетрудно составить представление о размерах полуосей доверительного эллипсоида $Э_{\gamma_0}$.

Заметим еще, что те же результаты относительно доверительного эллипсоида $Э_\gamma$ будут верны, если $X = X_{Nn}$ и абсциссы x_{rj} ($r = 1, 2, \ldots, N$; $j = 1, 2, \ldots, n$) выбираются вообще случайно, под единственным условием, что с вероятностью, равной единице, ранг $(X) = n$. При этом $C = X^TPX$ становится тоже случайной матрицей и $\tilde{A} = C^{-1}X^TPL$ не будет, вообще говоря, уже нормальным вектором при нормальном L. Однако эллипсоид $Э_{\gamma_0}$ остается доверительной областью с той же надежностью. В самом деле, при всякой фиксированной матрице X ранга n она имеет одну и ту же (условную) надежность p_0; поэтому и априорная надежность накрытия ею A будет p_0. Эти результаты не следует смешивать с результатами А. Вальда (гл. XII), где X, по существу, случайна, но уравнения $Y = XA$ уже не имеют места в силу погрешностей при измерении абсцисс.

§ 2. Зависимые наблюдения

Рассмотрим ту же схему уравнивания с помощью элементов

$$Y = X_0 + XA; \quad L - X_0 - \Delta = XA, \tag{14.2.1}$$

где вектор погрешностей $\Delta = \begin{Vmatrix} \Delta_1 \\ \vdots \\ \Delta_N \end{Vmatrix}$ будем считать нормальным, но уже не обязательно с независимыми компонентами Δ_i. Это будет отвечать схеме статистически зависимых наблюдений; погрешности у них статистически связаны. Будем считать вектор средних нулевым (погрешности Δ_i несмещенными), а корреляционную матрицу B_Δ вектора Δ

$$B_\Delta = \sigma^2 B_0 \tag{14.2.2}$$

известной с точностью до σ^2, т. е. с известной матрицей B_0. Тогда мы можем свести дело к уже изученному случаю независимых погрешностей Δ_i.

Пусть $G = G_{NN}$ — неособенная матрица. Умножая на нее слева (14.2.1), получим

$$GY = GX_0 + GXA; \quad GL - GX_0 - G\Delta = GXA. \tag{14.2.3}$$

Здесь роль Y, X_0, X, L, Δ будут играть те же матрицы, умноженные слева на G. Положим

$$\Delta^{(1)} = \begin{Vmatrix} \Delta_1^{(1)} \\ \vdots \\ \Delta_N^{(1)} \end{Vmatrix} = G\Delta. \tag{14.2.4}$$

Корреляционная матрица нового нормального вектора $\Delta^{(1)}$ по теореме 2.3.1 равна

$$B_{\Delta^{(1)}} = GB_\Delta G^T = \sigma^2 GB_0G^T. \tag{14.2.5}$$

Мы можем выбрать G неособенной (даже ортогональной) матрицей так, чтобы

$$GB_0G^T = \begin{Vmatrix} d_1 & 0 \dots & 0 \\ \cdot & \cdot \; \cdot \; \cdot & \cdot \\ 0 & 0 \dots & d_N \end{Vmatrix} \tag{14.2.6}$$

стала диагональной матрицей. Тогда

$$B_{\Delta^{(1)}} = \sigma^2 GB_0G^T = \begin{Vmatrix} d_1\sigma^2 & 0 & \dots & 0 \\ 0 & d_2\sigma^2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & d_N\sigma^2 \end{Vmatrix}. \tag{14.2.7}$$

Полагая $d_r = \dfrac{1}{p_r}$ $(r = 1, 2, \dots, N)$, видим, что вектор $\Delta^{(1)}$ имеет независимые компоненты с весами $p_1, p_2, \dots, p_N$, и соотношения (14.2.3) дают обычную схему уравнивания с помощью элементов, где вместо Y, X_0, X, L, Δ надлежит брать матрицы GY, GX_0, GX, GL, $G\Delta$. Все формулы и выводы гл. VIII остаются в силе. В частности, система нормальных уравнений для $\tilde{A}$ примет вид

$$C\tilde{A} = X^TG^TPGL; \quad C = X^TG^TPGX, \tag{14.2.8}$$

и т. д.

При построении матрицы G у нас будет довольно большая свобода действий, поскольку мы не требуем ее ортогональности. Этот прием можно использовать, когда при уравнивании почему-либо удобно вводить наблюдения с зависимыми ошибками (например, при уравнивании нивелирных ходов, когда ходы опираются на какую-либо марку, высота которой оценена с помощью наблюдений, включающих некоторые из этих ходов, а не дана точно).

Мы могли бы пытаться совершить операцию умножения на неособенную матрицу G слева, переводящую уравнения (14.2.1) в (14.2.3) и в прежних схемах, например в случае независимых равноточных наблюдений. Однако это не привело бы к полезным результатам. Если матрица G неортогональна, то мы нарушили бы таким образом статистическую независимость наблюдений, если же она ортогональна, то новый вектор наблюдений имел бы независимые компоненты, но система нормальных уравнений просто не изменилась бы. В самом деле, если G — ортогональная матрица, то $G^T = G^{-1}$, $P = E$, (14.2.8) совпадает с $C\widetilde{A} = X^T L$; $C = X^T X$, т. е. остается прежней.

§ 3. Роль нормального закона в теории метода наименьших квадратов

В гл. VI (§ 4—5) и гл. VIII мы ознакомились с некоторыми экстремальными свойствами метода наименьших квадратов при нормальном векторе Δ погрешностей с независимыми компонентами. Они вытекали из совместной эффективности оценок $(\widetilde{a}_1, \ldots, \widetilde{a}_n)$ для элементов $a_1, \ldots, a_n$ по методу наименьших квадратов. Возникает естественный вопрос о том, могут ли быть оценки $(\widetilde{a}_1, \ldots, \widetilde{a}_n)$ совместно эффективными в случае, когда погрешности наблюдений независимы, но подчиняются какому-либо другому закону, отличному от нормального.

Ответ на этот вопрос отрицателен. Как оказывается, в весьма широком классе несмещенных оценок оценки, найденные по методу наименьших квадратов, могут быть совместно эффективными лишь если вектор погрешностей Δ нормален (В. В. Петров [40], стр. 56—57). Таким образом, наличие оптимальных свойств у метода наименьших квадратов тесно связано с нормальностью вектора погрешностей.

Чтобы разобрать этот вопрос, будем рассматривать схему уравнивания с помощью элементов (гл. VIII) в обозначениях этой главы. Будем считать сперва для простоты наблюдения равноточными, и $X^{(0)} = 0$. Нам будет нужна теорема 5, стр. 93. Допустим, что для функции правдоподобия L наблюдений $l_1, l_2, \ldots, l_N$ соблюдены условия, указанные в этой теореме. Мы должны иметь

$$\ln L = \sum_{r=1}^{N} \ln f\left(l_r - \sum_{j=1}^{n} a_r x_{rj},\ \sigma\right), \tag{14.3.1}$$

где плотность вероятности $f(x)$ предполагается непрерывной, не обращающейся в нуль и дважды дифференцируемой. Уравнения максимального правдоподобия для оценок $a'_1, \ldots, a'_n$ параметров $a_1, \ldots, a_n$ получат вид

$$\frac{\partial \ln L}{\partial a'_j} = \sum_{1}^{N} \frac{\partial}{\partial a'_j} \ln f_r(\delta_r;\ \sigma) = 0, \tag{14.3.2}$$

где

$$\delta_r = l_r - \sum_{j=1}^{n} a' x_{rj}. \tag{14.3.3}$$

Введем матрицы

$$\delta = \begin{Vmatrix} \delta_1 \\ \vdots \\ \delta_N \end{Vmatrix}; \quad A' = \begin{Vmatrix} a'_1 \\ \vdots \\ a'_n \end{Vmatrix};$$

X и L будут иметь смысл гл. VI. Равенства (14.3.3) можно записать в виде $XA' = L - \delta$. Умножая на X^T слева, находим

$$X^T X A' = X^T L - X^T \delta. \tag{14.3.4}$$

Так как $X^T X = C$ (матрица нормальной системы), то отсюда

$$CA' = X^T L - X^T \delta$$

и

$$A' = C^{-1} X^T L - C^{-1} X^T \delta. \tag{14.3.5}$$

Далее,

$$C^{-1} X^T L = \tilde{A} = \begin{Vmatrix} \tilde{a}_1 \\ \vdots \\ \tilde{a}_n \end{Vmatrix}$$

есть одностолбцовая матрица оценок по методу наименьших квадратов. Таким образом,

$$A' = \tilde{A} - C^{-1} X^T \delta. \tag{14.3.6}$$

Совпадение оценок A' с оценками по методу наименьших квадратов $\tilde{A}$ равносильно, таким образом, соотношению

$$C^{-1} X^T \delta = 0,$$

или, ввиду неособенности C^{-1},

$$X^T \delta = 0. \tag{14.3.7}$$

При таком совпадении, уравнения максимального правдоподобия (14.3.2) должны иметь единственным решением $(\tilde{a}_1, \ldots, \tilde{a}_n)$. Имеем далее

$$\frac{\partial \ln f_r(\delta_r; \sigma)}{\partial a'_j} = \frac{f'_r(\delta_r; \sigma)}{f_r(\delta_r, \sigma)} \frac{\partial \delta_r}{\partial a'_j} = -\frac{f'_r(\delta_r; \sigma)}{f_r(\delta_r; \sigma)} x_{rj}. \tag{14.3.8}$$

Таким образом, эти уравнения переписываются в виде

$$\sum_{r=1}^{N} \frac{f'_r(\delta_r; \sigma)}{f_r(\delta_r; \sigma)} x_{rj} = 0. \tag{14.3.9}$$

Введя матрицу

$$\varphi(\delta)=\varphi_{n1}=\left\|\begin{matrix}\dfrac{f_1'}{f_1}(\delta_1;\ \sigma)\\ \vdots\\ \dfrac{f_N'}{f_N}(\delta_N;\ \sigma)\end{matrix}\right\|=\left\|\begin{matrix}\varphi_1(\delta_r)\\ \vdots\\ \varphi_N(\delta_N)\end{matrix}\right\|, \tag{14.3.10}$$

записываем уравнение (14.3.9) в виде

$$X^T\varphi(\delta)=0. \tag{14.3.11}$$

Мы видим, что уравнения $X^T\delta=0$ и $X^T\varphi(\delta)=0$ должны быть равносильными.

Будем считать это верным при любой матрице $X=X_{Nn}$ ранга n и при любом наборе чисел $(\delta_1, \ldots, \delta_N)$. Возьмем теперь матрицу вида

$$X=\left\|\begin{matrix}0 & \ldots\ldots & 0 & 0 & +1\\ 0 & \ldots\ldots & 0 & +1 & -\dfrac{x}{n-1}\\ 0 & \ldots\ldots 0 & +1 & -1 & -\dfrac{x}{n-1}\\ \ldots & \ldots & \ldots & \ldots & \ldots\\ 0 & \ldots\ -1 & 0 & \ldots & -\dfrac{x}{n-1}\\ \ldots & \ldots & \ldots & \ldots & \ldots\\ 0 & +1 & \ldots & \ldots & 0\\ +1 & -1 & \ldots\ 0 & \ldots & 0\\ -1 & 0 & \ldots\ 0 & \ldots & 0\end{matrix}\right\| \tag{14.3.12}$$

такую, что сумма элементов во всех столбцах, кроме последнего, равна нулю.

Рассмотрим также матрицу

$$\delta=\left\|\begin{matrix}\left.\begin{matrix}x\\ 1\\ \vdots\\ 1\end{matrix}\right\}n\\ \left.\begin{matrix}1\\ \vdots\\ 1\end{matrix}\right\}N-n\end{matrix}\right\|.$$

Здесь x — переменный аргумент.

Имеем: ранг $(X) = n$ при $x \neq 0$. Далее, очевидно, что $X^T\delta = 0$ при любом x; следовательно, должно быть $X^T\varphi(\delta) = 0$. Последняя строчка этого равенства дает

$$\varphi_1(x) - \frac{x}{n-1}\sum_{r=2}^{n}\varphi_r(1) = 0,$$

откуда

$$\varphi_1(x) = \alpha_1 x \tag{14.3.13}$$

при постоянном α_1. Аналогично получим

$$\varphi_r(x) = \alpha_r x \qquad (r = 1, 2, \ldots, N). \tag{14.3.14}$$

Далее, $\varphi_r(x) = \frac{f_r'}{f_r}$, так что получаем дифференциальное уравнение

$$\frac{f_r'}{f_r} = \alpha_r x.$$

Интегрируя, находим: $\ln f_r = c_0 x^2 + c_1$, где c_i — постоянные, зависящие только от r. Поэтому

$$f_r = \exp(c_0 x^2 + c_1) = c_2 e^{c_0 x^2}. \tag{14.3.15}$$

Но f_r должна быть вероятностной плотностью, так что

$$\int_{-\infty}^{\infty} f_r(x)\,dx = 1.$$

Отсюда следует, что $c_0 < 0$. Положим

$$c_0 = -\frac{1}{2\mu^2},$$

тогда из (14.3.15) находим

$$c_2 = \frac{1}{\sqrt{2\pi}\,\mu},$$

так что $f_r(x)$ — нормальная плотность $N(0, \mu)$. Так как погрешности Δ_r предполагались несмещенными и равноточными, $\mu = \sigma$ не должно зависеть от r.

Мы доказали нормальность f_r для равноточных погрешностей, отсюда она следует и для неравноточных погрешностей. В самом деле, из § 1 гл. VI мы знаем, что случай неравноточных погрешностей сводится на случай равноточных, если фундаментальное уравнение $L - \Delta = XA$ умножить слева на $P^{\frac{1}{2}}$, так что получится

$$P^{\frac{1}{2}}L - P^{\frac{1}{2}}\Delta = P^{\frac{1}{2}}XA.$$

При прежних условиях тогда имеет место нормальность вектора $P^{\frac{1}{2}}\Delta$, а стало быть, и вектора Δ.

Таким образом обосновывается утверждение, что из совместной эффективности оценок, получаемых по методу наименьших квадратов, при весьма широких условиях следует нормальность вектора погрешностей. Между методом наименьших квадратов и нормальным законом распределения существует глубокая связь.

§ 4. Ненормальный вектор погрешностей. Одна формула Гаусса. Теорема А. Н. Колмогорова, А. А. Петрова, Ю. М. Смирнова

Рассмотрим схему уравнивания с помощью элементов

$$y_r = a_1 x_{r1} + \ldots + a_n x_{rn} \qquad (r = 1, \ldots, N), \tag{14.4.1}$$

$$y_r = l_r - \Delta_r, \tag{14.4.2}$$

или, в матричном виде,

$$\begin{aligned} Y &= XA; \\ Y &= L - \Delta. \end{aligned} \qquad \text{ранг } (X) = n, \tag{14.4.3}$$

Мы сохраним условия и обозначения гл. VIII, но вектор погрешностей

$$\Delta = \begin{Vmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_N \end{Vmatrix} \tag{14.4.4}$$

будем считать имеющим независимые компоненты Δ_r, которые уже не предполагаются нормальными. Будем предполагать, что

$$E(\Delta_r) = 0; \qquad D(\Delta_r) = \frac{\sigma^2}{p_r}, \tag{14.4.5}$$

$$E(\Delta_r^4) = \frac{\mu_4}{p_r^2}, \tag{14.4.6}$$

где $p_r (r = 1, 2, \ldots, N)$ — известные числа (веса). Такой случай, как указано в § 1 гл. VI, легко сводится на случай равноточных измерений. Заметим, что для нормальных погрешностей Δ_r (14.4.6) вытекает из (14.4.5); для случая ненормальных Δ_r это уже не так. При замене наблюдений l_r на групповое выборочное среднее

$$\bar{l}_r = \frac{l_{r1} + \ldots + l_{rm}}{m}$$

группы независимых наблюдений имеем

$$D(\bar{l}_r) = \frac{\sigma^2}{m}, \quad \text{если} \quad D(l_{rj}) = \sigma^2 \qquad (j = 1, \ldots, m).$$

Пусть существуют центральные моменты четвертого порядка $\mu_4(l_{rj}) = \mu_4$. Легко подсчитать, что

$$\mu_4(\bar{l}_r) = \frac{3\sigma^4}{m^2} + \frac{\mu_4 - 3\sigma^4}{m^3}. \tag{14.4.7}$$

Для нормального закона имеем

$$\mu_4 = 3\sigma^4, \tag{14.4.8}$$

так что в этом случае

$$\mu_4(\bar{l}_r) = \frac{\mu_4}{m^2},$$

замена l_r на групповое среднее $\bar{l}_r$ придает наблюдениям вес $p_r = m$ (см. § 6 гл. VI); соблюдается также равенство (14.4.6). В общем случае ненормальных наблюдений l_{rj} при замене l_r на $\bar{l}_r$ будет выполняться равенство (14.4.7), отвечающее (14.4.5); (14.4.6), вообще говоря, не будет выполняться.

Мы знаем, что для ненормального вектора погрешностей Δ уже нельзя гарантировать описанные ранее экстремальные свойства получаемых по методу наименьших квадратов оценок, и, согласно § 3 этой главы, это отвечает существу дела. Как видно из § 2 гл. VII, в таком случае при условиях (14.4.5) можно гарантировать лишь минимальность дисперсий полученных оценок порознь, среди всех линейных несмещенных оценок измеряемых элементов и линейных функций от них. Но минимальность дисперсии и сама по себе есть условие типа наименьших квадратов, и подобные теоремы, хотя и нетривиальные, дают много менее того, что имеет место при нормальном векторе погрешностей Δ.

Надо еще заметить, что из решений нормальных уравнений

$$\tilde{A} = C^{-1}X^TPL$$

видно, что оценки $\tilde{a}_1, \ldots, \tilde{a}_n$ для параметров $a_1, \ldots, a_n$ получаются в виде линейных функций независимых наблюдений $l_1, \ldots, l_N$, и поэтому можно ожидать, согласно теореме Ляпунова (теорема 2.7.1), что эти оценки $\tilde{a}_1, \ldots, \tilde{a}_n$ будут приближенно нормальными даже при ненормальном векторе погрешностей Δ. В таком случае несмещенность и минимальность дисперсий $\tilde{a}_i$ среди линейных несмещенных оценок a_i будет иметь уже благоприятный вероятностный смысл, близкий к поясненному в § 5 гл. VI.

Вернемся к формулам (14.4.5) и (14.4.6). В случае нормальности вектора погрешностей Δ будет $\mu_4 = 3\sigma^4$ и

$$E(\Delta_r^4) = \frac{3\sigma^4}{p_r}. \tag{14.4.9}$$

Мы знаем, что в этом случае несмещенной оценкой для σ^2 будет $\tilde{s}^2 = \frac{[p\tilde{v}\tilde{v}]}{N-n}$ и что $\tilde{s}^2$ распределено как $\frac{\sigma^2}{N-n}\chi^2_{N-n}$. Так как

$$D\left(\chi^2_{N-n}\right) = 2(N-n)\sigma^4 \tag{14.4.10}$$

(см. § 6 гл. II), то

$$D\left(\tilde{s}^2\right) = \frac{1}{(N-n)^2} D\left([p\tilde{v}\tilde{v}]\right) = \frac{2\sigma^4}{N-n}. \tag{14.4.11}$$

Ввиду того, что для нормального вектора погрешностей Δ имеет место равенство $\mu_4 = 3\sigma^4$, (14.4.11) можно переписать так:

$$D\left(\tilde{s}^2\right) = \frac{\mu_4 - \sigma^4}{N-n}. \tag{14.4.12}$$

К. Ф. Гаусс ([7], § 39) указал обобщение формулы (14.4.12) для дисперсии $D\left(\tilde{s}^2\right)$ на случай ненормального вектора погрешностей при условиях (14.4.5) и (14.4.6) и (наиболее естественном) условии равноточности $[p_r = 1,\ (r = 1, \ldots, N)]$. Выведем эту формулу Гаусса в матричных обозначениях.

Теорема 14.4.1 (К. Ф. Гаусс). *В условиях* (14.4.5) *и* (14.4.6) *имеем*

$$D\left(\tilde{s}^2\right) = \frac{\mu_4 - \sigma^4}{N-n} - \frac{\mu_4 - 3\sigma^4}{(N-n)^2}(n - \Omega), \tag{14.4.13}$$

где (в наших обычных обозначениях)

$$\Omega = \sum_{r=1}^{N} \{U\}^2_{rr}, \tag{14.4.14}$$

$$U = P^{\frac{1}{2}} X C^{-1} X^T P^{\frac{1}{2}}. \tag{14.4.15}$$

Здесь $C = X^T P X$ — матрица нормальной системы уравнений.

Доказательство. Рассмотрим сперва случай равноточных измерений $P = E$; от него легко будет перейти к данному случаю, как пояснялось в § 1 гл. VI. Для этого случая имеем в обозначениях гл. VI [см. (6.6.2)]

$$\tilde{V} = (U - E)\Delta; \quad U = XC^{-1}X^T \tag{14.4.16}$$

При этом имеем, как выяснено в гл. VI,

$$U^2 = U; \quad U^T = U; \qquad \text{ранг } (U) = n. \tag{14.4.17}$$

Там же было выяснено, что симметрические матрицы U и $E - U$ являются полуопределенными. Этот факт весьма важен для дальнейшего.

Имеем далее

$$\tilde{s}^2 = \frac{[\tilde{v}\tilde{v}]}{N-n},$$

$$[\tilde{v}\tilde{v}] = \tilde{V}^T \tilde{V} = \Delta^T (U-E)^T (U-E)\,\Delta = \Delta^T (U-E)^2\,\Delta.$$

Далее,

$$(U-E)^2 = U^2 - 2U + E = U - 2U + E = E - U.$$

Таким образом,

$$[\tilde{v}\tilde{v}] = \Delta^T (E-U)\,\Delta. \qquad (14.4.18)$$

Положим $U = U_{NN} = \|u_{ij}\|$; тогда из (14.4.18) имеем

$$[\tilde{v}\tilde{v}] = \sum_{r=1}^{N} \Delta_r^2 - \sum_{i,\,j=1}^{N} u_{ij}\Delta_i\Delta_j. \qquad (14.4.19)$$

При этом $u_{ij} = u_{ji}$, так как $U^T = U$. Переписывая (14.4.19) так:

$$[\tilde{v}\tilde{v}] = \sum_{r=1}^{N} (1-u_{rr})\,\Delta_r^2 - 2 \sum_{\substack{i<j \\ i,\,j=1,\,\ldots,\,N}} u_{ij}\Delta_i\Delta_j, \qquad (14.4.20)$$

имеем, ввиду того что $E(\tilde{s}^2) = \sigma^2$,

$$D(\tilde{s}^2) = E(\tilde{s}^2 - \sigma^2)^2 = E(\tilde{s}^4) - 2\sigma^2 E\tilde{s}^2 + \sigma^4 =$$
$$= E(\tilde{s}^4) - 2\sigma^4 + \sigma^4 = E(\tilde{s}^4) - \sigma^4.$$

Таким образом, получаем

$$D(\tilde{s}^2) = \frac{1}{(N-n)^2}\,E([\tilde{v}\tilde{v}]^2) - \sigma^4. \qquad (14.4.21)$$

Найдем $E([\tilde{v}\tilde{v}]^2)$. Заметим, что погрешности Δ_i независимы в совокупности и $E(\Delta_i) = 0$. Это упрощает вычисления; мы должны возвести сумму (14.4.18) в квадрат и найти математическое ожидание этого квадрата. При этом появляются члены, содержащие произведения вида

$$\Delta_r^2\Delta_\mu\Delta_\nu \quad (\mu < \nu) \quad \text{и} \quad \Delta_i\Delta_j\Delta_\mu\Delta_\nu \quad (i<j,\ \mu<\nu).$$

Имеем далее $E\left(\Delta_r^2\Delta_\mu\Delta_\nu\right) = 0$, ибо произведение будет содержать член Δ_ρ ($\rho = \mu$ или ν, $\rho \neq r$) и потому обратится в нуль (см. гл. II). По тем же соображениям $E(\Delta_i\Delta_j\Delta_\mu\Delta_\nu)$ может быть не равным нулю только в случае, если $i = \mu$ и $j = \nu$. Случай $i = \nu$, $j = \mu$ не может иметь место, ибо по условию $i < j$ и $\mu < \nu$. Учитывая это, получаем из (14.4.20)

$$E([\tilde{v}\tilde{v}]^2) = \sum_{r,\,t=1}^{N} (1-u_{rr})(1-u_{tt})\,E\,\Delta_r^2\Delta_t^2 +$$
$$+ 4\sum_{i<j} (u_{ij})^2\,E\Delta_i^2\Delta_j^2. \qquad (14.4.22)$$

Имеем

$$E\left(\Delta_i^2\Delta_j^2\right)=E\left(\Delta_i^2\right)E\left(\Delta_j^2\right)=\sigma^4 \qquad (i<j) \tag{14.4.23}$$

Далее, $E\left(\Delta_r^2\Delta_t^2\right)=\sigma^4$ при $r\neq t$ и равно

$$E\left(\Delta_r^4\right)=\mu_4 \tag{14.4.24}$$

при $r=t$. Это позволяет записать (14.4.22) в виде

$$E\left([\tilde{v}\tilde{v}]^2\right)=(\mu_4-\sigma^4)\sum_{r=1}^{N}(1-u_{rr})^2+\sigma^4\left(\sum_{r=1}^{N}(1-u_{rr})\right)^2+$$
$$+2\sigma^4\sum_{i,\,j=1}^{N}(u_{ij})^2-2\sigma^4\sum_{i=1}^{N}(u_{ii})^2. \tag{14.4.25}$$

Докажем теперь равенства

$$\sum_{r=1}^{N}u_{rr}=n; \qquad \sum_{i,\,j=1}^{N}(u_{ij})^2=n. \tag{14.4.26}$$

Согласно (14.4.17) имеем

$$\sum_{r=1}^{N}u_{rr}=\mathrm{Sp}\,(U)=n. \tag{14.4.27}$$

Далее, рассмотрим выражение $\sum\limits_{i,\,j=1}^{N}u_{ij}^2$ — сумму квадратов всех элементов матрицы U. Докажем равенство

$$\sum_{i,\,j=1}^{N}u_{ij}^2=\mathrm{Sp}\,(UU^T). \tag{14.4.28}$$

В самом деле, $\mathrm{Sp}\,(UU^T)$ есть сумма диагональных элементов матрицы UU^T. А диагональные элементы ее, как видно из самого определения операции умножения матриц, суть суммы квадратов элементов строк матрицы U. Это доказывает (14.4.28). Далее, из (14.4.17) и (14.4.27) находим

$$U^T=U, \quad UU^T=U^2=U; \quad \mathrm{Sp}\,(UU^T)=\mathrm{Sp}\,(U)=n,$$

что и доказывает (14.4.26).

Не так просто вычисляется величина [см. (14.4.22)]

$$\sum_{i=1}^{N}(u_{ii})^2=\Omega.$$

Внеся (14.4.26) и (14.4.27) в (14.4.22), находим

$$E\left([\tilde{v}\tilde{v}]^2\right)=(\mu_4-\sigma^4)(N-2n+\Omega)+\sigma^4(N-n)^2+2\sigma^4 n-2\sigma^4\Omega.$$

Отсюда

$$E\left([\tilde{v}\tilde{v}]^2\right)=(\mu_4-\sigma^4)(N-n)-(\mu_4-3\sigma^4)(n-\Omega)+\sigma^4(N-n)^2.$$

Подставляя это в (14.4.21), приходим к (14.4.13), что и доказывает теорему для равноточных наблюдений.

В общем случае, согласно сказанному в гл. VI, § 1, достаточно заменить X на $P^{\frac{1}{2}}X$, L на $P^{\frac{1}{2}}L$, чтобы придти к равноточным наблюдениям. При этом $\tilde{V}$ заменяется на $P^{\frac{1}{2}}V$; $[\tilde{v}\tilde{v}]=\tilde{V}^T\tilde{V}$ на $\tilde{V}^TP\tilde{V}=[p\tilde{v}\tilde{v}]$, так что

$$\tilde{s}^2=\frac{[p\tilde{v}\tilde{v}]}{N-n}. \tag{14.4.29}$$

Матрица $C=X^TX$ заменяется на X^TPX и $U=X^TC^{-1}X$ на $U=P^{\frac{1}{2}}XC^{-1}X^TP^{\frac{1}{2}}$. Формула Гаусса (14.4.13) останется в силе. Все свойства матрицы U, разумеется, останутся в силе, так как U будет играть роль соответствующей матрицы для системы равноточных наблюдений. Эти свойства могут быть также проверены непосредственно.

Если вектор Δ нормален, то из формулы Гаусса (14.4.13) получается старая формула (14.4.12). В общем же случае Гауссом были даны оценки величины Ω снизу и сверху. Оценки Гаусса, которые приводят к соответствующим оценкам $D(\tilde{s}^2)$, имеют вид

$$\frac{n^2}{N}\leqslant\Omega\leqslant N. \tag{14.4.30}$$

Нижнюю оценку Гаусса, как мы узнаем далее, нельзя улучшить в общем случае, верхняя же оценка, как заметили А. Н. Колмогоров, А. А. Петров и Ю. М. Смирнов [23], весьма груба, и, как видно из (14.4.13), иногда может дать тривиальную отрицательную границу для $D(\tilde{s}^2)$ снизу (ясно, что $D(\tilde{s}^2)\geqslant 0$). Указанные три автора дали верхнюю оценку Ω, которую уже нельзя улучшить в общем случае, и получили неулучшаемую оценку $D(\tilde{s}^2)$ снизу.

Теорема 14.4.2 (**А. Н. Колмогоров, А. А. Петров, Ю. М. Смирнов** [23]). *Имеет место неулучшаемая в общем случае оценка*

$$\Omega\leqslant n, \tag{14.4.31}$$

так что при $\mu_4-3\sigma^4\geqslant 0$

$$\frac{1}{N-n}\left(\mu_4-\sigma^4-\frac{n}{N}(\mu_4-3\sigma^4)\right)\leqslant D(\tilde{s}^2)\leqslant\frac{\mu_4-\sigma^4}{N-n} \tag{14.4.32}$$

и при $\mu_4-3\sigma^4\leqslant 0$

$$\frac{\mu_4-\sigma^4}{N-n}\leqslant D(\tilde{s}^2)\leqslant\frac{1}{N-n}\left(\mu^4-\sigma^4+\frac{n}{N}(3\sigma^4-\mu_4)\right). \tag{14.4.33}$$

Заметим, что всегда $\mu_4\geqslant\sigma^4$, как это видно из (14.4.32), так что эти оценки никогда не бывают тривиальными.

Мы будем доказывать неравенства

$$\frac{n^2}{N} \leqslant \Omega \leqslant n, \qquad (14.4.34)$$

но не будем доказывать их неулучшаемости. Последняя выводится из одной алгебраической теоремы А. И. Мальцева (см. [23] и А. И. Мальцев [34]).

Доказательство. Надо доказать, что

$$\frac{n^2}{N} \leqslant \sum_{i=1}^{N} u_{ii}^2 \leqslant n. \qquad (14.4.35)$$

Сперва докажем, что

$$0 \leqslant u_{ii} \leqslant 1. \qquad (14.4.36)$$

В самом деле, если бы (14.4.36) не выполнялось, то хотя бы одна из матриц U и $E-U$ имела бы отрицательный диагональный элемент u_{ii} или $1-u_{ii}$. Но, как пояснялось выше, эти матрицы полуопределенные, и потому это невозможно, так как противоречит условиям полуопределенности (см. § 4 гл. 1). Таким образом, (14.4.36) имеет место. Отсюда выводим с помощью неравенства Коши — Буняковского [см. (1.1.1)]

$$N \sum_{i=1}^{N} u_{ii}^2 \geqslant \left(\sum_{i=1}^{N} u_{ii} \right)^2 = (\mathrm{Sp}\,(U))^2 = n^2;$$

$$\sum_{i=1}^{N} u_{ii}^2 \geqslant \frac{n^2}{N}. \qquad (14.4.37)$$

Далее, из (14.4.36) видим, что $0 \leqslant u_{ii}^2 \leqslant u_{ii}$, откуда

$$\sum_{i=1}^{N} u_{ii}^2 \leqslant \sum_{i=1}^{N} u_{ii} = \mathrm{Sp}\,(U) = n, \qquad (14.4.38)$$

что и доказывает (14.4.34). Далее, (14.4.30) и (14.4.31) получаются подстановкой неравенств (14.4.34) в формулу Гаусса (14.4.13). Верхняя оценка для Ω, данная самим Гауссом, отвечает использованию лишь того факта, что $0 \leqslant u_{ii} \leqslant 1$, так что $\sum_{i=1}^{N} u_{ii}^2 \leqslant N$.

§ 5. Метод обработки наблюдений Коши

Рассмотрим схему уравнивания по элементам. Для простоты и без нарушения общности положим $X_0 = 0$, так что система фундаментальных уравнений примет вид (в обозначениях гл. VIII)

$$Y = L - \Delta = XA,$$

или, подробнее,

$$y_r = l_r - \Delta_r = \sum_{j=1}^{N} x_{rj} a_j \qquad (r = 1,\ 2,\ \ldots,\ N). \qquad (14.5.1)$$

Наблюдения предполагаются равноточными. Мы видели, что при нормальности вектора погрешностей Δ метод наименьших квадратов имеет некоторые экстремальные свойства среди широкого класса других методов (см. § 5 гл. VI), так что его применение в этом смысле выгодно. Но в вычислительном отношении решение системы нормальных уравнений, несмотря на обычно применяемые упрощающие приемы (например, способы Гаусса и Гаусса — Дулиттла; см. § 12 гл. VI), все еще представляет затруднения. Поэтому иногда прибегают и к другим методам оценивания a_j по наблюдениям l_r $(r = 1,\ 2,\ \ldots,\ N)$ и фундаментальным уравнениям (14.5.1).

Составление нормальных уравнений и получение с их помощью оценок для параметров $(a_1,\ \ldots,\ a_n)$ есть, в конечном счете, метод такого комбинирования N уравнений с погрешностями избыточной системы (14.5.1), чтобы в результате можно было получить разрешимую систему линейных уравнений относительно n параметров $(a_1,\ \ldots,\ a_n)$. Таких методов может быть много, помимо метода наименьших квадратов, хотя, как пояснено выше, они не будут уже иметь экстремальных свойств, присущих методу наименьших квадратов. Один из таких методов был предложен О. Коши.

Здесь мы изложим вкратце его содержание и разберем некоторые простейшие вопросы, связанные с его эффективностью по сравнению с методом наименьших квадратов *).

Предписание О. Коши состоит в следующем. Сначала система (14.5.1) подготавливается, т. е. умножением каждого уравнения на $+1$ или -1 добиваются того, что $x_{1j} > 0$ $(j = 1,\ 2,\ \ldots,\ N)$. Предположим, что это уже сделано. Далее, все уравнения складываются, так что получается

$$(x_{11} + x_{21} + \ldots + x_{N1})\, a_1 + (x_{12} + x_{22} + \ldots + x_{N2})\, a_2 + \ldots +$$
$$+ (x_{1n} + x_{2n} + \ldots + x_{Nn})\, a_n = y_1 + \ldots + y_N. \qquad (14.5.2)$$

Будем придерживаться обозначений Коши, полагая

$$x_{1j} + x_{2j} + \ldots + x_{Nj} = Sx_j \qquad (j = 1,\ 2,\ \ldots,\ n). \qquad (14.5.3)$$

Имеем тогда

$$a_1 Sx_1 + \ldots + a_n Sx_n = Sy. \qquad (14.5.4)$$

*) Здесь используется дипломная работа Л. С. Бартеньевой [3] о методе Коши, защищенная в 1956 г. в Ленинградском государственном университете.

В силу предыдущего, $Sx_1 > 0$. Деля (14.5.4) на Sx_1, получаем

$$a_1 + a_2 \frac{Sx_2}{Sx_1} + \ldots + a_n \frac{Sx_n}{Sx_1} = \frac{Sy}{Sx_1}. \qquad (14.5.5)$$

Умножаем (14.5.5) соответственно на $x_{11}, x_{12}, \ldots, x_{1N}$ и вычитаем из полученных N уравнений уравнение системы (14.5.1); тогда члены, содержащие a_1, исчезают и получается система из N уравнений с $(n-1)$ неизвестными $a_2, \ldots, a_n$

$$\left(x_{12} - x_{11}\frac{Sx_2}{Sx_1}\right)a_2 + \ldots + \left(x_{nn} - x_{11}\frac{Sx_n}{Sx_1}\right)a_n = y_1 - x_{11}\frac{Sy}{Sx_1},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\left(x_{N2} - x_{N1}\frac{Sx_2}{Sx_1}\right)a_2 + \ldots + \left(x_{Nn} - x_{N1}\frac{Sx_n}{Sx_1}\right)a_n = y_N - x_{N1}\frac{Sy}{Sx_1}.$$

Далее, подготавливаем так же новую систему и производим с ней такие же операции. В конце концов, вообще говоря, получаем систему N уравнений с одним неизвестным a_n

$$x'_{rn} a_n = y'_r \qquad (r = 1, 2, \ldots, N); \qquad (14{,}5.6)$$

где $x'_{rn} \geqslant 0$ — получившиеся коэффициенты, а y'_r — линейные комбинации из y_q $(q = 1, 2, \ldots, N)$, получившиеся в процессе работы.

Далее, заменяем в каждом из y'_r неизвестные нам y_q на известные и наблюдаемые l_q. Результат такой замены обозначим l'_r. После этого находим a'_n из уравнений

$$a'_n = \frac{Sl'_r}{Sx'_r}. \qquad (14.5.7)$$

В предыдущих уравнениях подставляем a'_n на место a_n, l_q на место y_q, складываем их и находим a'_{n-1}, и т. д. пока не найдем оценку a'_1.

Иногда (на практике довольно часто) случается, что в процессе такой работы коэффициенты при $a_k, a_{k+1}, \ldots, a_n$ оказываются весьма малыми; тогда обычно соответствующие члены отбрасываются, а с оставшимися поступают указанным образом. Найдя соответствующие оценки a'_i, подставляют их в уравнение, а с оставшимися параметрами поступают, как ранее.

Мы видим, что метод Коши (если нет отбрасывания малых членов) приводит к линейным оценкам для $(a_1, \ldots, a_n)$: оценки $(a'_1, \ldots, a'_n)$ для этих параметров по методу Коши являются линейными функциями наблюдений $l_1, \ldots, l_N$. Несложно доказать и такую теорему:

Теорема 14.5.1. *Оценки $a'_1, \ldots, a'_n$ параметров $a_1, \ldots, a_n$ по методу Коши являются несмещенными, если не происходило отбрасывания малых членов.*

Доказательство. Оценки a_i' суть линейные функции наблюдений l_r, если не происходило отбрасывания малых членов. При замене l_r на y_r ($r=1, 2, \ldots, N$) и a_i' на a_i все употребляемые в методе Коши выражения обратятся в точные равенства. Среди этих выражений есть выражение вида

$$a_i' = L_i(l_1, \ldots, l_N) \qquad (i=1, 2, \ldots, n), \tag{14.5.8}$$

где L_i — линейная функция наблюдений, так что

$$E\left(L_i(l_1, \ldots, l_N)\right) = L_i\left(y_1, \ldots, y_N\right). \tag{14.5.9}$$

Но при замене l_r на y_r в правой части (14.5.8) a_i' можно заменить на a_i, т. е.

$$E\left(a_i'\right) = a_i,$$

что и требовалось доказать.

Если в процессе вычислений происходило отбрасывание малых членов, то получившиеся оценки, вообще говоря, будут приближенно несмещенными.

Перейдем теперь к вопросу об эффективности получаемых по методу Коши оценок; при этом будем предполагать, что не происходило отбрасывания малых членов, и ограничимся случаями одного и двух параметров a_1 и a_2 и равноточных наблюдений l_r.

В гл. VII (теореме 7.2.2) было доказано, что даже при ненормальности вектора случайных погрешностей ($\Delta_1, \ldots, \Delta_N$), а лишь при наличии у них дисперсий метод наименьших квадратов дает эффективные оценки параметров a_i среди всех методов оценивания, приводящих к линейным несмещенным оценкам. Так как метод Коши также приводит к таким оценкам, имеем

$$D\left(a_i'\right) \geqslant D\left(\tilde{a}_i\right) \qquad (i=1, 2, \ldots, n), \tag{14.5.10}$$

где через $\tilde{a}_i$, как и ранее, обозначим оценки для a_i по методу наименьших квадратов.

Правда, выражение качества оценки с помощью ее дисперсии есть, по существу, также подход к ней с точки зрения наименьших квадратов, но, как пояснялось в гл. V, такое выражение получает более ощутимый реальный смысл, если оценки a_i' и $\tilde{a}_i$ нормальны или приближенно нормальны. Если вектор погрешностей ($\Delta_1, \ldots, \Delta_N$) нормален, то $\tilde{a}_i$ и a_i' нормальны, как линейные функции от него. Даже если он ненормален (но Δ_i независимы), то при большом числе уравнений N сравнительно с числом параметров n оценки a_i' и $\tilde{a}_i$ будут, вообще говоря, приближенно нормальны в силу теоремы А. М. Ляпунова 2.7.1. Таким образом, неравенства (14.5.10) будут означать, что, вообще говоря, вероятности уклонений a_i' не более чем на заданное ε, будут больше у оценок $\tilde{a}_i$, чем у a_i'.

Перейдем к исследованию систем с одним и двумя неизвестными. Для системы с одним неизвестным

$$x_{r1}a_1 = y_r \qquad (r = 1, 2, \ldots, N) \qquad (x_{r1} \geqslant 0) \quad (14.5.11)$$

имеем оценки $a'_1 = \frac{Sl}{Sx_1}$ (метод Коши); $\tilde{a}_1 = \frac{[x_1 l]}{[x_1 x_1]}$ (метод наименьших квадратов).

Отсюда

$$D(a'_1) = \frac{N\sigma^2}{\left(\sum_{r=1}^{N} x_{r1}\right)^2}; \quad D(\tilde{a}_1) = \frac{\sigma^2}{[x_1 x_1]}. \qquad (14.5.12)$$

Отношение $\frac{D(\tilde{a}_1)}{D(a'_1)}$ называется *эффективностью оценки* a'_1. Имеем

$$e_1 = \frac{D(\tilde{a}_1)}{D(a'_1)} = \frac{\left(\sum_{r=1}^{N} x_{r1}\right)^2}{N \sum_{r=1}^{N} x_{r1}^2}. \qquad (14.5.13)$$

Это отношение должно быть меньше единицы согласно общей теории. В этом нетрудно убедиться и непосредственно, замечая, что в силу неравенства Буняковского — Коши (1.1.1)

$$\left(\sum x_{r1}\right)^2 \leqslant N \sum_{r=1}^{N} x_{r1}^2.$$

Знак равенства достигается только в случае, если все x_{r1} равны между собой, т. е. если имеются повторные наблюдения, отвечающие одному и тому же фундаментальному уравнению $x_{11}a_1 = y_1$. В этом случае попросту $a'_1 = \tilde{a}_1$, как видно из их выражений. Вместе с тем e_1 может быть сколь угодно малым. Если возьмем, например, $x_{r1} = \frac{1}{r}$ $(r = 1, 2, \ldots, N)$, то простой подсчет показывает, что при больших N имеем $e_1 \approx \frac{6(\ln N)^2}{\pi^2 N}$.

Для подобного случая метод наименьших квадратов значительно выгоднее метода Коши. Вообще (14.5.13) дает ориентировку в вопросе о том, во сколько раз теряется в точности при применении метода Коши.

Перейдем теперь к системе с двумя неизвестными. Пусть матрица X системы имеет вид

$$X = \begin{Vmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{N1} & x_{N2} \end{Vmatrix}; \quad \text{ранг } (X) = 2 \text{ и } x_{r1} \geqslant 0 \qquad (r = 1, 2, \ldots, N).$$

Полагая, как обычно, $x_j=(x_{1j}, x_{2j}, \ldots, x_{Nj})$, имеем, как и ранее,

$$\left.\begin{aligned} D(\tilde{a}_1) &= \frac{[x_2x_2]\,\sigma^2}{[x_1x_1]\,[x_2x_2]-[x_1x_2]^2}, \\ D(\tilde{a}_2) &= \frac{[x_1x_1]\,\sigma^2}{[x_1x_1]\,[x_2x_2]-[x_1x_2]^2}. \end{aligned}\right\} \qquad (14.5.14)$$

Найдем теперь оценки a_1' и a_2' для параметров a_1, a_2 по методу Коши. Имеем

$$x_{r1}a_1+x_{r2}a_2=y_r, \qquad (r=1,\ 2,\ \ldots,\ N), \qquad (14.5.15)$$
$$x_{r1}\geqslant 0 \qquad (r=1,\ 2,\ \ldots,\ N).$$

Для a_2 получаем уравнения

$$\left(x_{12}-x_{11}\frac{Sx_2}{Sx_1}\right)a_2=y_1-x_{11}\frac{Sy}{Sx_1},$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$\left(x_{N2}-x_{N1}\frac{Sx_2}{Sx_1}\right)a_2=y_N-x_{N1}\frac{Sy}{Sx_1}.$$

Пусть $\varepsilon_r=\pm 1$ — знак величины $x_{r2}-x_{r1}\frac{Sx_2}{Sx_1}$. В таком случае имеем

$$\sum \varepsilon_r\left(x_{r2}-x_{r1}\frac{Sx_2}{Sx_1}\right)a_2=\sum \varepsilon_r\left(y_r-x_{r1}\frac{Sy}{Sx_1}\right),$$

или

$$a_2\left(S(x_2\varepsilon)-S(x_1\varepsilon)\frac{Sx_2}{Sx_1}\right)=S(y\varepsilon)-S(x_1\varepsilon)\frac{Sy}{Sx_1}.$$

Отсюда для a_2 получаем оценку

$$a_2'=\frac{S(l\varepsilon)-S(x_1\varepsilon)\dfrac{Sl}{Sx_1}}{S(x_2\varepsilon)-S(x_1\varepsilon)\dfrac{Sx_2}{Sx_1}} \qquad (14.5.16)$$

и для a_1 — оценку

$$a_1'=\frac{Sl-a_2'Sx_2}{Sx_1}. \qquad (14.5.17)$$

Рассмотрим внимательнее оценку a_2'. Имеем

$$a_2'=\frac{1}{S(x_2\varepsilon)-S(x_1\varepsilon)\dfrac{Sx_2}{Sx_1}}\left(\sum_{r=1}^{N} l_r\left(\varepsilon_r-\frac{S(x_1\varepsilon)}{Sx_1}\right)\right), \qquad (14.5.18)$$

так что

$$D(a_2')=\frac{\sigma^2}{S(x_2\varepsilon)-S(x_1\varepsilon)\dfrac{Sx_2}{Sx_1}}\sum_{r=1}^{N}\left(\varepsilon_r-\frac{S(x_1\varepsilon)}{Sx_1}\right)^2. \qquad (14.5.19)$$

Таким образом, эффективность оценки a_2' на основании (14.5.14) имеет вид

$$e_1 = \frac{D(\tilde{a}_2)}{D(a_2')} = \frac{[x_1x_1][x_2x_2] - [x_1x_2]^2}{[x_1x_1]\left(S(x_2\varepsilon) - S(x_1\varepsilon)\frac{Sx_2}{Sx_1}\right)} \sum_{r=1}^{N}\left(\varepsilon_r - \frac{S(x_1\varepsilon)}{Sx_1}\right)^2. \quad (14.5.20)$$

На основании общей теории $0 \leqslant e_1 \leqslant 1$.

Для решения вопроса о том, какая потеря точности в процентном отношении происходит при применении метода Коши по сравнению с методом наименьших квадратов, надо знать матрицу X и найти по ней выражение (14.5.20). В практических задачах для случая строго равноточных наблюдений $D(a_2')$ может быть больше дисперсии $D(a_2)$ в 10 и более раз; стало быть, штандарты оценок могут отличаться в три и более раза. Тем не менее, результаты могут быть достаточно точными, а вычисления — весьма простыми, что заставляет во многих случаях предпочитать способ Коши способу наименьших квадратов.

Таблица 30

Результаты определения параметров a_1 и a_2

Дата (август 1955 г.)	Метод наименьших квадратов	Метод Коши
4—5	4,904 0,280	4,904 0,280
13—14	5,696 0,195	5,696 0,195
14—15	5,707 0,237	5,708 0,231
15—16	5,709 0,248	5,715 0,218
23—24	5,398 0,220	5,404 0,200
25—26	5,299 0,106	5,284 0,156
26—27	5,269 0,150	5,280 0,113
28—29	5,188 0,234	5,190 0,227

Л. С. Бартеньева [3] приводит данные по определению поправки часов из наблюдений звезд службы времени Пулковской обсерватории за август 1955 г. Имеют место уравнения

$$u + A_r a = (\alpha - T)\,r \qquad (r = 1,\ 2,\ \ldots,\ N),$$

где A_r, α взяты из звездных каталогов, T — наблюдаемый момент времени. Азимут инструмента a и поправка часов u — искомые величины.

Обозначим $u = a_1$; $a = a_2$. По ряду наблюдений получились результаты, приведенные в табл. 30. Они показывают, что точность наблюдений была удовлетворительной (σ^2 малым). Правда, уверенности в том, что наблюдения можно было считать равноточными, не было; при этом метод Коши, не использующий весов наблюдений, применялся пунктуально, а метод наименьших квадратов — в предположении равноточности, т. е. без учета весов. Таким образом, формула вида (14.5.20) для сравнения эффективности двух методов могла уже не иметь места.

ЛИТЕРАТУРА

1. Аббе (Abbe E.), Über die Gesetzmässigkeit in der Verteilung der Fehler bei Beobachtungsreihen. Werke, Bd. 2, Jena, 1906, S. 55—81.
2. Арлей Н. и Бух К. Р., Введение в теорию вероятностей и математическую статистику, ИЛ, 1951.
3. Бартеньева Л. С., Эффективность метода Коши (дипломная работа). ЛГУ, 1956.
4. Бернштейн С. Н., Теория вероятностей, изд. 4, Гостехиздат, 1946.
5. Вальд (Wald A.), The fitting of straight lines if both variables are subject to errors, Ann. Math. Statistics, vol. 11, № 3, 1940, 284—300.
6. Гантмахер Ф. Р., Теория матриц. Гостехиздат, 1953.
7. Гаусс (Gauss C. F.), Theoria motus corporum coelestium. Hamburg, 1809.
8. Гаусс (Gauss C. F.), Disquisitio de elementis ellipticis Palladis. 1810.
9. Гаусс (Gauss C. F.), Theoria combinationis observationum erroribus minimis obnoxiae, 1821.
10. Гельмерт (Helmert F.), Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate. Leipzig — Berlin, 1907.
11. Гнеденко Б. В., Курс теории вероятностей, изд. 2. Гостехиздат, 1954.
12. Грэббс (Grubbs F.), Sample criteria for testing outlying observations. Ann. Math. Statistics, vol. 21, № 1, 1950, 27—58.
13. Гузурбазар (Huzurbazar V. S.), The likelihood equation, consistency, and maxima of the likelihood function. Ann. Eugenics, vol. 14, 1948.
14. Дунин-Барковский И. В. и Смирнов Н. В., Теория вероятностей и математическая статистика в технике. Гостехиздат, 1955.
15. Дэвид (David F. N.), Probability theory for statistical methods. Cambridge, 1951.
16. Дюге (Dugué D.), Application des propriétés de la limite au sens du calcul des probabilités á l'étude de diverses questions de l'estimation. Journ. de l'Ec. Polytechnique, 3, № 4, 1937.
17. Зимовнов В. Н., Вопросы оценки точности результатов измерений. Геодезиздат, 1951.
18. Идельсон Н. И., Способ наименьших квадратов, изд. 2, Л., 1932.
19. Идельсон Н. И., Способ наименьших квадратов и теория математической обработки наблюдений. Геодезиздат, 1947.
20. Кендалл (Kendall M. G.), The advanced theory of statistics, vol. 1, 2, London, 1948.
21. Колмогоров А. Н., О логарифмически нормальном законе распределения размеров частиц при дроблении, ДАН СССР, т. 31, 1941, 99—101.
22. Колмогоров А. Н., К обоснованию метода наименьших квадратов, Успехи матем. наук, т. 1, вып. 1, 1946, 57—70.

23. Колмогоров А. Н., Петров А. А., Смирнов Ю. М., Одна формула Гаусса из теории метода наименьших квадратов. Изв. АН СССР, сер. матем., т. 11, № 6, 1947, 561—566.
24. Крамер (Cramér H.), Contributions to the theory of statistical estimation. Skandinavisk Aktuarietidskrift, vol. 29, 1946, 85—94.
25. Крамер (Cramér H.), Математические методы статистики, ИЛ, 1948.
26. Красовский Ф. Н., Курс геодезии, ч. I и II. ГИЗ, 1931.
27. Кризи (Creasy M. A.), Confidence limits for the gradient in the linear functional relationship. Journ. Roy. Stat. Soc., Ser. B, vol. 18, № 1, 1956, 65—69.
28. Курош А. Г., Курс высшей алгебры, изд. 5. Гостехиздат, 1956.
29. Лаплас (Laplace P. S.), Théorie analytique des probabilités. Paris, 1812.
30. Лежандр (Legendre A. M.), Nouvelles méthodes pour la détermination des orbites des comètes. Paris, 1806, Appendice sur la méthode des moindres carrés.
31. Лёв (Loeve M.), On sets of probability laws and their limit elements. Univ. Calif. Publ. in Statistics, 1950.
32. Линник Ю. В., Некоторые замечания к методу наименьших квадратов, с приложением к задачам прямых и обратных засечек. Теория вероятностей и ее применения, т. 2, вып. 3, 1957, 349—359.
33. Линник Ю. В. и Хусу А. П., Математико-статистическое описание неровностей профиля поверхности при шлифовании. Инженерный сборник АН СССР, т. XX, 1954, 154—159.
34. Мальцев А. И., Замечание к работе А. Н. Колмогорова, А. А. Петрова и Ю. М. Смирнова „Одна формула Гаусса из теории метода наименьших квадратов“. Изв. АН СССР, сер. матем., т. 11, № 6, 1947, 567—568.
35. Мальцев А. И., Основы линейной алгебры, изд. 2, Гостехиздат, 1956.
36. Марков А. А., Закон больших чисел и метод наименьших квадратов, (1898). Избр. труды. Изд. АН СССР, 1951, стр. 233—251.
37. Марков А. А., Исчисление вероятностей, изд. 4, ГИЗ, 1924.
38. Нейман Ю., Статистическая оценка как проблема классической теории вероятностей, Успехи матем. наук, вып. 10, 1944, 207—229.
39. Немчинов В. С., Полиномы Чебышева и математическая статистика, М., 1946.
40. Петров В. В., О методе наименьших квадратов и его экстремальных свойствах. Успехи матем. наук, т. 9, вып. 1, 1954, 41—62.
41. Рао (Rao C. R.), On the linear combination of observations and the general theory of least squares. Sankhya, vol. 7., p. 3, 1946, 237—256.
42. Рао (Rao C. R.), Advanced Statistical Methods in Biometric Research., N. Y., 1952.
43. Рао (Rao C. R.), Information and accuracy attainable in the estimation of statistical parameters, Bull. Calcutta Math. Soc., № 37, 1945.
44. Романовский В. И., Основные задачи теории ошибок, Гостехиздат, 1947.
45. Слуцкий Е. Е., Таблицы для вычисления неполной Г-функции и функции вероятности. Изд. АН СССР, 1950.
46. Смирнов В. И., Курс высшей математики, т. III, ч. 1, изд. 4, Гостехиздат, 1949.
47. Фихтенгольц Г. М., Курс дифференциального и интегрального исчисления, т. III, Гостехиздат, 1949.
48. Хальд А., Математическая статистика с техническими приложениями, ИЛ, 1956.

49. Хёфдинг (Hoeffding W.), A class of statistics with asymptotically normal distribution. Annals of Math. Statistics, vol. 19, № 3, 1948, 293—325.

50. Хирвонен (Hirvonen R. A.), Веса и весовые коэффициенты. Maanmittaus, т. 30, № 1—2, 1955, 10—23 (финск.).

51. Чеботарёв А. С., Способ наименьших квадратов с основами теории вероятностей, ОНТИ, 1936.

52. Чебышев П. Л., Об интерполировании по способу наименьших квадратов, 1859, соч., т. 1.

53. Чебышев П. Л., Sur l'intégration des différentielles irrationnels. Journ. de math., ser. I., t. 18, 1853.

54. Эзекиель (Ezekiel M.), Methods of correlation analysis, 2-nd Edition, N. Y., 1941.

55. Эйткен (Aitken A. C.), On least squares and linear combination of observations, Proc. Roy. Soc. Edin., vol. 55, 1935, 42—48.

56. Ющенко А. П., Способ наименьших квадратов. Изд. „Морской транспорт", 1956.

ПРИЛОЖЕНИЯ

Таблица I

t-распределение Стьюдента *)

k \ p	0,99	0,98	0,95	0,90	0,80	0,70	0,60
1	63,657	31,821	12,706	6,314	3,078	1,963	1,376
2	9,925	6,965	4,303	2,920	1,886	1,386	1,061
3	5,841	4,541	3,182	2,353	1,638	1,250	0,978
4	4,604	3,747	2,776	2,132	1,533	1,190	0,941
5	4,032	3,365	2,571	2,015	1,476	1,156	0,920
6	3,707	3,143	2,447	1,943	1,440	1,134	0,906
7	3,499	2,998	2,365	1,895	1,415	1,119	0,896
8	3,355	2,896	2,306	1,860	1,397	1,108	0,889
9	3,250	2,821	2,262	1,833	1,383	1,100	0,883
10	3,169	2,764	2,228	1,812	1,372	1,093	0,879
11	3,106	2,718	2,201	1,796	1,363	1,088	0,876
12	3,055	2,681	2,179	1,782	1,356	1,083	0,873
13	3,012	2,650	2,160	1,771	1,350	1,079	0,870
14	2,977	2,624	2,145	1,761	1,345	1,076	0,868
15	2,947	2,602	2,131	1,753	1,341	1,074	0,866
16	2,921	2,583	2,120	1,746	1,337	1,071	0,865
17	2,898	2,567	2,110	1,740	1,333	1,069	0,863
18	2,878	2,552	2,101	1,734	1,330	1,067	0,862
19	2,861	2,539	2,093	1,729	1,328	1,066	0,861
20	2,845	2,528	2,086	1,725	1,325	1,064	0,860
21	2,831	2,518	2,080	1,721	1,323	1,063	0,859
22	2,819	2,508	2,074	1,717	1,321	1,061	0,858
23	2,807	2,500	2,069	1,714	1,319	1,060	0,858
24	2,797	2,492	2,064	1,711	1,318	1,059	0,857
25	2,787	2,485	2,060	1,708	1,316	1,058	0,856
26	2,779	2,479	2,056	1,706	1,315	1,058	0,856
27	2,771	2,473	2,052	1,703	1,314	1,057	0,855
28	2,763	2,467	2,048	1,701	1,313	1,056	0,855
29	2,756	2,462	2,045	1,699	1,311	1,055	0,854
30	2,750	2,457	2,042	1,697	1,310	1,055	0,854
40	2,704	2,423	2,021	1,684	1,303	1,050	0,851
60	2,660	2,390	2,000	1,671	1,296	1,046	0,848
120	2,617	2,358	1,980	1,658	1,289	1,041	0,845
∞	2,576	2,326	1,960	1,645	1,282	1,036	0,842

*) В таблице приведены значения γ_p для величины t, имеющей распределение Стьюдента с k степенями свободы, определяемые из условия

$$P(|t| \leqslant \gamma_p) = p.$$

Иными словами, вероятность P для наблюденного значения t не превзойти по абсолютной величине γ_p равна p.

Таблица II

Пятипроцентные пределы γ для распределения Фишера $F_{k_1,\,k_2}(x)$ *)

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	16	20	24	30	50	100	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	252	253	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,41	19,43	19,44	19,45	19,46	19,47	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,74	8,69	8,66	8,64	8,62	8,58	8,56	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,84	5,80	5,77	5,74	5,70	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,68	4,60	4,56	4,53	4,50	4,44	4,40	4,36
6	5,99	5,14	4.76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,92	3,87	3,84	3,81	3,75	3,71	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,57	3,49	3,44	3,41	3,38	3,32	3,28	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,28	3,20	3,15	3,12	3,08	3,03	2,98	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,07	2,98	2,93	2,90	2,86	2,80	2,76	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,91	2,82	2,77	2,74	2,70	2,64	2,59	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,79	2,70	2,65	2,61	2,57	2,50	2,45	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,69	2,60	2,54	2,50	2,46	2,40	2,35	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,60	2,51	2,46	2,42	2,38	2,32	2,26	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,53	2,44	2,39	2,35	2,31	2,24	2,19	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,48	2,39	2,33	2,29	2,25	2,18	2,12	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,33	2,28	2.24	2,20	2,13	2,07	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,38	2,29	2,23	2,19	2,15	2,08	2,02	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,25	2,19	2,15	2,11	2,04	1,98	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,31	2,21	2,15	2,11	2,07	2,00	1,94	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,28	2,18	2,12	2,08	2,04	1,96	1,90	1,84
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,23	2,13	2,07	2,03	1,98	1,91	1,84	1,78
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,18	2,09	2,02	1,98	1,94	1,86	1,80	1,73
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,05	1,99	1,95	1,90	1,82	1,76	1,69
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,12	2,02	1,96	1,91	1,87	1,78	1,72	1,65
32	4,15	3,30	2,90	2,67	2,51	2,40	2,32	2,25	2,19	2,14	2,07	1,97	1,91	1,86	1,82	1,74	1,67	1,59
36	4,11	3,26	2,86	2,63	2,48	2,36	2,28	2,21	2,15	2,10	2,03	1,93	1,87	1,82	1,78	1,69	1,62	1,55
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,00	1,90	1,84	1,79	1,74	1,66	1,59	1,51
60	4,00	3,15	2,76	2,52	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,81	1,75	1,70	1,65	1,56	1,48	1,39
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,85	1,75	1,68	1,63	1,57	1,48	1,39	1,28
200	3,89	3,04	2,65	2,41	2,26	2,14	2,05	,98	1,92	1,87	1,80	1,69	1,62	1,57	1,52	1,42	1,32	1,19
∞	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,83	1,75	1,64	1,57	1,52	1,46	1,35	1,24	1,00

*) См. гл. II, § 6. В таблице приведены значения γ, удовлетворяющие соотношению $P(X>\gamma)=0{,}05$.

Таблица III

Нижние γ_1 и верхние γ_2 границы доверительного интервала

$$\gamma_1 s_1 < \sigma < \gamma_2 s_1 \qquad \left(s_1 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)^{*)}$$

k \ p	0,99		0,98		0,95		0,90	
	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2	γ_1	γ_2
1	0,356	159	0,388	79,8	0,446	31,9	0,510	15,9
2	0,434	14,1	0,466	9,97	0,521	6,28	0,578	4,40
3	0,483	6,47	0,514	5,11	0,566	3,73	0,620	2,92
4	0,519	4,39	0,549	3,67	0,599	2,87	0,649	2,37
5	0,546	3,48	0,576	3,00	0,624	2,45	0,672	2,090
6	0,569	2,98	0,597	2,62	0,644	2,202	0,690	1,916
7	0,588	2,66	0,616	2,377	0,661	2,035	0,705	1,797
8	0,604	2,440	0,631	2,205	0,675	1,916	0,718	1,711
9	0,618	2,277	0,644	2,076	0,688	1,826	0,729	1,645
10	0,630	2,154	0,656	1,977	0,699	1,755	0,739	1,593
11	0,641	2,056	0,667	1,898	0,708	1,698	0,748	1,550
12	0,651	1,976	0,677	1,833	0,717	1,651	0,755	1,515
13	0,660	1,910	0,685	1,779	0,725	1,611	0,762	1,485
14	0,669	1,854	0,693	1,733	0,732	1,577	0,769	1,460
15	0,676	1,806	0,700	1,694	0,739	1,548	0,775	1,437
16	0,683	1,764	0,707	1,659	0,745	1,522	0,780	1,418
17	0,690	1,727	0,713	1,629	0,750	1,499	0,785	1,400
18	0,696	1,695	0,719	1,602	0,756	1,479	0,790	1,385
19	0,702	1,666	0,725	1,578	0,760	1,460	0,794	1,370
20	0,707	1,640	0,730	1,556	0,765	1,444	0,798	1,358
21	0,712	1,617	0,734	1,536	0,769	1,429	0,802	1,346
22	0,717	1,595	0,739	1,519	0,773	1,416	0,805	1,335
23	0,722	1,576	0,743	1,502	0,777	1,402	0,809	1,326
24	0,726	1,558	0,747	1,487	0,781	1,391	0,812	1,316
25	0,730	1,541	0,751	1,473	0,784	1,380	0,815	1,308
26	0,734	1,526	0,755	1,460	0,788	1,371	0,818	1,300
27	0,737	1,512	0,758	1,448	0,791	1,361	0,820	1,293
28	0,741	1,499	0,762	1,436	0,794	1,352	0,823	1,286
29	0,744	1,487	0,765	1,426	0,796	1,344	0,825	1,279
30	0,748	1,475	0,768	1,417	0,799	1,337	0,828	1,274
40	0,774	1,390	0,792	1,344	0,821	1,279	0,847	1,228
50	0,793	1,336	0,810	1,297	0,837	1,243	0,861	1,199
60	0,808	1,299	0,824	1,265	0,849	1,217	0,871	1,179
70	0,820	1,272	0,835	1,241	0,858	1,198	0,879	1,163
80	0,829	1,250	0,844	1,222	0,866	1,183	0,886	1,151
90	0,838	1,233	0,852	1,207	0,873	1,171	0,892	1,141
100	0,845	1,219	0,858	1,195	0,878	1,161	0,897	1,133
200	0,887	1,15	0,897	1,13	0,912	1,11	0,925	1,09

*) Пояснения к таблице даны в примере 1 в конце гл. IV.

Таблица IV

Квантили распределения величины $v = \frac{x_{max} - \bar{x}}{s}$ или $v_1 = \frac{\bar{x} - x_{min}}{s}$

n \ p	0,10	0,05	0,025	0,01
3	1,406	1,412	1,414	1,414
4	1,645	1,689	1,710	1,723
5	1,791	1,869	1,917	1,955
6	1,894	1,996	2,067	2,130
7	1,974	2,093	2,182	2,265
8	2,041	2,172	2,273	2,374
9	2,097	2,237	2,349	2,464
10	2,146	2,294	2,414	2,540
11	2,190	2,343	2,470	2,606
12	2,229	2,387	2,519	2,663
13	2,264	2,426	2,562	2,714
14	2,297	2,461	2,602	2,759
15	2,326	2,493	2,638	2,800
16	2,354	2,523	2,670	2,837
17	2,380	2,551	2,701	2,871
18	2,404	2,577	2,728	2,903
19	2,426	2,600	2,754	2,932
20	2,447	2,623	2,778	2,959
21	2,467	2,644	2,801	2,984
22	2,486	2,664	2,823	3,008
23	2,504	2,683	2,843	3,030
24	2,520	2,701	2,862	3,051
25	2,537	2,717	2,880	3,071

Таблица V

Квантили r_p распределения величины $r = \frac{q^2}{s_1^2}$ *)

n \ p	0,001	0,01	0,05	n \ p	0,001	0,01	0,05
4	0,295	0,313	0,390	33	0,503	0,614	0,722
5	0,208	0,269	0,410	34	0,509	0,619	0,726
6	0,182	0,281	0,445	35	0,515	0,624	0,729
7	0,185	0,307	0,468	36	0,521	0,629	0,733
8	0,202	0,331	0,491	37	0,526	0,634	0,736
9	0,221	0,354	0,512	38	0,532	0,638	0,740
10	0,241	0,376	0,531	39	0,537	0,642	0,743
11	0,260	0,396	0,548	40	0,542	0,647	0,746
12	0,278	0,414	0,564	41	0,548	0,651	0,749
13	0,295	0,431	0,578	42	0,552	0,655	0,752
14	0,311	0,447	0,591	43	0,557	0,659	0,755
15	0,327	0,461	0,603	44	0,562	0,662	0,758
16	0,341	0,475	0,614	45	0,566	0,666	0,760
17	0,355	0,487	0,624	46	0,570	0,669	0,763
18	0,368	0,499	0,633	47	0,574	0,673	0,765
19	0,381	0,510	0,642	48	0,578	0,676	0,768
20	0,393	0,520	0,650	49	0,582	0,679	0,770
21	0,404	0,530	0,657	50	0,585	0,681	0,772
22	0,414	0,539	0,665	51	0,589	0,684	0,774
23	0,424	0,548	0,671	52	0,592	0,687	0,776
24	0,433	0,556	0,678	53	0,596	0,690	0,778
25	0,442	0,564	0,684	54	0,599	0,692	0,780
26	0,451	0,571	0,689	55	0,602	0,695	0,782
27	0,459	0,578	0,695	56	0,605	0,697	0,784
28	0,467	0,585	0,700	57	0,608	0,700	0,785
29	0,475	0,591	0,705	58	0,611	0,702	0,787
30	0,482	0,598	0,709	59	0,614	0,705	0,789
31	0,489	0,603	0,714	60	0,617	0,707	0,791
32	0,496	0,609	0,718				

*) Квантили вычислены по значениям, данным в таблице Харта, умножением их на $\frac{n-1}{2n}$ (см. B. J. Hart, Significance levels for the ratio of the mean square succesive difference to the variance, Ann. Math. Stat., vol. 13, 1942, 445—447).

Таблица VI

Значения $\gamma = \dfrac{1}{(1-p_0)^{\frac{2}{n-2}}} - 1$ (см. стр. 262)

n \ p_0	0,90	0,95	0,99	n \ p_0	0,90	0,95	0,99
3	99,0	399,	9998,	33	0,160	0,213	0,346
4	9,000	19,0	99,0	34	0,155	0,206	0,333
5	3,642	6,368	20,54	35	0,150	0,199	0,322
6	2,162	3,472	9,000	36	0,145	0,193	0,311
7	1,512	2,314	5,310	37	0,141	0,187	0,301
8	1,154	1,714	3,642	38	0,136	0,181	0,291
9	0,931	1,353	2,728	39	0,133	0,176	0,283
10	0,778	1,115	2,162	40	0,129	0,171	0,274
11	0,668	0,946	1,783	41	0,125	0,166	0,266
12	0,585	0,820	1,512	42	0,122	0,162	0,259
13	0,520	0,724	1,310	43	0,119	0,157	0,252
14	0,468	0,648	1,154	44	0,116	0,153	0,245
15	0,425	0,586	1,031	45	0,113	0,149	0,239
16	0,389	0,534	0,931	46	0,110	0,146	0,233
17	0,359	0,491	0,848	47	0,108	0,142	0,227
18	0,334	0,454	0,778	48	0,105	0,139	0,222
19	0,311	0,422	0,719	49	0,103	0,136	0,217
20	0,292	0,395	0,668	50	0,101	0,133	0,212
21	0,274	0,371	0,624	51	0,099	0,130	0,207
22	0,259	0,349	0,585	52	0,096	0,127	0,202
23	0,245	0,330	0,551	53	0,095	0,125	0,198
24	0,233	0,313	0,520	54	0,093	0,122	0,194
25	0,222	0,298	0,492	55	0,091	0,120	0,190
26	0,212	0,284	0,468	56	0,089	0,117	0,186
27	0,202	0,271	0,445	57	0,087	0,115	0,182
28	0,194	0,259	0,425	58	0,086	0,113	0,179
29	0,186	0,248	0,406	59	0,084	0,111	0,175
30	0,179	0,239	0,389	60 *)	0,083	0,109	0,172
31	0,172	0,229	0,374				
32	0,166	0,221	0,359				

*) При больших значениях n значения γ можно получать по приближенной формуле

$$\gamma \approx -\frac{2}{n-2}\ln(1-p_0).$$

Линник Юрий Владимирович

Метод наименьших квадратов
и основы математико-статистической теории обработки наблюдений

Редактор *В. В. Петров*

Техн. редактор *Р. Г. Польская* Корректор *Т. С. Петрова*

Сдано в набор 17/I 1958 г. Подписано к печати 18/VI 1958 г. Бумага $60 \times 92^1/_{16}$.
Физ. печ. л. 21. Т-03963 Условн. печ. л. 21. Уч.-изд. л. 20,29. Тираж 6000 экз. Цена 12 р. 15 к.
Заказ № 2725

Государственное издательство Физико-математической литературы.
Москва, В-71, Ленинский проспект, 15

Типография № 2 им. Евг. Соколовой УПП Ленсовнархоза. Ленинград, Измайловский пр., 29.

ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ

ФИЗМАТГИЗ

Москва, В-71, Ленинский проспект, 15

ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ

ФИЗМАТГИЗ

Москва, В-71, Ленинский проспект, 15.