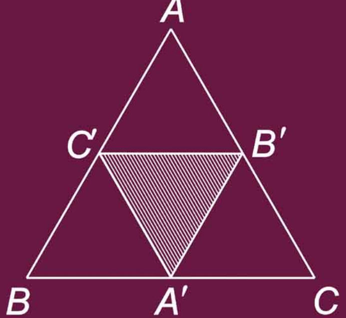
ЛЕКЦИИ ПО

R&C Dynamics

МАТЕМАТИЧЕСКОЙ ФИЗИКЕ

А. Пуанкаре

ТЕОРИЯ ВЕРОЯТНОСТЕЙ



COURSE DE LA FACULTÉ DES SCIENCES DE PARIS

COURS DE PHYSIQUE MATHEMATIQUE

CALCUL
DES
PROBABILITÉS

PAR

H. POINCARÉ

MEMBRE DE L'INSTITUT,

RÉDACTION DE

A. QUIQUET,

ANCIEN ÉLÉVE DE L'ÉCOLE NORMALE SUPERIEURE.

PARIS,

GAUTHIER-VILLARS, IMPRIMEUR-LIBRAIRE

DU BUREAU DES LONGITUDES, DE L'ÉCOLE POLYTECHNIQUE.

Quai des Grands-Augustins, 55

1912

А. Пуанкаре

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Перевод с французского
В. В. Шуликовской

Научный редактор
А. В. Летчиков

Редакция журнала
«Регулярная и хаотическая динамика»

Ижевск

1999

УДК 511.6

Пуанкаре А.

Теория вероятностей. — Ижевск: Ижевская республиканская типография, 1999, 280 с.

Книга является одной из частей курса лекций А. Пуанкаре. В ней рассмотрены как общие основы теории вероятностей, так и нетрадиционные вопросы, которые практически не содержатся ни в одном курсе. Рассмотрены различные приложения к физике, математике и механике.

Книга полезна широкому кругу читателей — физикам, математикам, историкам науки.

Оригинал-макет подготовлен в редакции журнала «Регулярная и хаотическая динамика»
http://www.rcd.com.ru

Содержание

Предисловие редактора перевода

Читателю предлагается перевод книги «Теория вероятностей» великого французского математика Анри Пуанкаре (1854–1912). Вышедшая в свет в 1912 году, эта книга стала учебным пособием для многих математиков, а также специалистов в различных областях науки, инструментом исследований которых стали теоретико-вероятностные методы. Очевидно, что в двадцатом веке теория вероятностей существенно продвинулась вперед как в научном, так и в методическом направлениях. Пожалуй, не имеет смысла перечислять огромное количество учебников, в которых изложены современные концепции теории вероятностей. В каждом из них можно найти точные определения и свойства математических понятий вероятности, случайной величины, ее числовых характеристик — математического ожидания и дисперсии. Однако, увлекаясь формально-логическим описанием понятий теории вероятностей, авторы книг забывают о природе появления случайности, о приложении тех теоретических результатов, которые ими получены. В противоположность этим учебникам А. Пуанкаре в своей книге не дает точных определений некоторых понятий, оставляет без строгих доказательств некоторые результаты, останавливаясь в основном на их идейном обосновании и подтверждении в приложениях. Такой подход дает возможность изучения теории вероятностей, не требующего овладения тяжелым математическим аппаратом, что делает эту книгу доступной для достаточно широкого круга читателя.

Не менее важными являются философские воззрения Анри Пуанкаре, изложенные во введении этой книги. Развитие теории вероятностей как математической науки, изучающей закономерности случайных явлений, в конце девятнадцатого века было предопределено серьезными требованиями со стороны естествознания и общественной практики. У многих исследователей того времени возникал вопрос о фундаментальной натурфилософской роли теории вероятностей. Более того, признание объективности статистических закономерностей поднимало вопрос об изучении условий их возникновения. Подавляющее большинство ученых того времени исходили из лапласовской картины мира. Таким образом, возникла потребность построения такой картины мира, в которой объективные статистические законы, математические понятия вероятности и случайной величины не противоречили бы основным принципам концепции лапласовского детерминизма.

Решая эту задачу, А. Пуанкаре как философ твердо придерживается детерминизма. Для него случай есть производная от неопределенности в силу ограниченности знания данных. Анализируя реальную практику науки, А. Пуанкаре выделяет три основные причины, которые требуют использование понятия объективной случайности. Во-первых, это ограниченная точность измерения, не дающая возможность точно угадать результат эксперимента в тех ситуациях, когда малые возмущения вызывают большие следствия. Во-вторых, сложность построенной модели эксперимента, когда результат определяет такое большое количество различных причин, что исследователю становится невозможным достоверно определить его. В этом случае, упрощая модель, исследователь вносит неопределенность в эксперимент. В-третьих, то, что А. Пуанкаре называет «невежеством». Такого рода причины возникают в тех ситуациях, когда исследователь в силу ограниченности своих знаний не может предопределить результаты своего эксперимента. А. Пуанкаре считал, что его рассуждения обосновывают наличие случайности в мире, в котором ничто не происходит произвольно, подчиняясь действию однозначных законов. Он видел, что наука все больше обращается к вероятностно-статистическим идеям, при этом, не вступая в противоречие с представлениями об однозначной детерминированности явлений действительности.

Понимая важность развития теории вероятностей как математической науки, А. Пуанкаре в своей книге привлекает глубокий аналитический аппарат для решения задач. Здесь можно найти методы моментов и характеристических функций при доказательстве центральной предельной теоремы, методы наименьших квадратов и наибольшего правдоподобия при статистической оценке неизвестных параметров, вычисление вероятностей методом производящих функций. Стоит выделить оригинальное построение изложения материала в книге и большое количество задач из других разделов математики таких, как теория чисел, алгебра, математическая физика, небесная механика, функциональный анализ, дифференциальная геометрия.

Конечно, всю книгу нельзя назвать легкой для чтения. Введение и первые несколько глав под силу прочитать любому интересующемуся теорией вероятностей читателю, поскольку они не содержат сколько-нибудь серьезного математического аппарата. Последующие главы более трудны для чтения и требуют знания основных понятий математического анализа, в частности, дифференциального и интегрального исчислений. Велико было искушение для редактора дать историческую и научную оценку каждого факта, приведенного в книге. Однако в этом случае пришлось бы переписать всю книгу. Редактор и переводчик книги решили отказаться от этого и в целом постарались сохранить стиль изложения А. Пуанкаре, чтобы читатель мог самостоятельно оценить изящность и верность рассуждений великого математика.

Несомненно, эта книга откроет для многих читателей, в том числе и математиков, новый взгляд на теорию вероятностей, на ее историю и на роль А. Пуанкаре в ее развитии.

А. В. Лётчиков

Введение[1]

I. Случай

«Не будет ли чрезмерной смелостью рассуждать о законах случая? Не является ли случай противоположностью любого закона?» Так пишет Бертран в начале своей «Теории вероятностей». Вероятность противоположна определенности; это то, о чем не знают; следовательно, может показаться, что вероятность нельзя вычислить. Вот как минимум одно очевидное противоречие, о котором уже много написано.

Кроме того, что такое случай? Древние различали явления, которые, по-видимому, повиновались гармоническим законам, установленным единожды и навсегда, и явления, приписываемые случаю; именно те, которые нельзя было предвидеть, т. к. они не подчинялись никакому закону. Ни в одной области точные законы не определяли всего, они лишь очерчивали пределы, в которых дозволялось пребывать случаю. В рамках этой концепции слово «случай» имело точный, объективный смысл: то, что являлось случаем для одного, оставалось случаем и для другого, и для обоих сразу.

Но мы больше не придерживаемся этой концепции; мы сделались абсолютными детерминистами, такими, которые, желая сохранить за человеком права свободного арбитра, позволяют детерминизму по меньшей мере безраздельно править в неорганическом мире. Любое явление, каким бы малым оно ни было, имеет причину, и бесконечно мощный разум, бесконечно хорошо знающий законы природы, смог бы предвидеть все, от начала веков. Если бы подобный разум существовал, мы не могли бы играть с ним ни в какую азартную игру, мы бы всегда проигрывали.

Действительно, для него слово «случай» не имело бы смысла, точнее, случая не существовало бы. Только по причине нашей слабости и нашего невежества случай существует для нас. Даже если не покидать пределы нашего слабого человечества, то, что является случаем

[1] Это введение взято из главы «Случай» моей работы «Наука и метод» («Science et Méthode», Flammarion). — *Прим. авт.*

для невежды, уже не будет таковым для ученого. Случай — не более чем мера нашего невежества. Случайные явления — по определению те, чьих законов мы не знаем.

Но будет ли это определение вполне удовлетворять нас? Когда первые халдейские пастухи следили за движением звезд, они еще не были знакомы с законами астрономии; пришли ли они к выводу, что звезды движутся случайно? Если современный физик изучает новое явление и открывает соответствующий закон во вторник, сказал бы он в понедельник, что это явление случайно? Более того, предсказывая какое-то явление, не ссылаются ли часто на то, что Бертран называет законом случая? Например, в кинетической теории газа мы находим известные законы Мариотта и Гей–Люссака, обязанные гипотезе, что скорость молекул газа изменяется нерегулярно, т. е. случайно. Все физики говорят, что данные законы оказались бы куда сложнее, если бы скорости подчинялись бы какому-нибудь элементарно простому закону, если бы молекулы были, так сказать, *организованы*, если бы они подчинялись какой-нибудь дисциплине. Только благодаря случаю, только благодаря нашему невежеству мы можем установить эти законы; но если слово «случай» — лишь простой синоним невежества, то что это значит? Надо ли это понимать следующим образом?

«Вы требуете от меня предсказать некоторые еще не произошедшие явления. Если бы, к несчастью, я знал их законы, то я не смог бы решить вашу задачу без неосуществимых вычислений, и мне пришлось бы отказать вам; но так как я имею счастье не знать их, я вам сейчас же отвечу. И, что самое необычное, мой ответ будет верен.»

Итак, надо признать, что слово «случай» не совпадает с тем именем, которое мы даем нашему невежеству, что среди явлений, чьи причины нам неизвестны, мы должны различать явления случайные, о которых мы предварительно узнаем в теории вероятностей, и неслучайные, о которых мы не можем сказать ничего, т. к. мы еще не определили законы, которые бы управляли ими. Очевидно, что для первых, собственно случайных явлений, сведения, поставляемые теорией вероятностей, не перестанут быть верными в тот день, когда мы лучше изучим явления.

Директор страховой компании, занимающийся страхованием жизни, не знает, когда умрет каждый из его клиентов, но он ведет расчеты на основе теории вероятностей и закона больших чисел, и не ошибается, и успешно распределяет дивиденды среди своих акционеров. Эти дивиденды не исчезли бы, если бы однажды какой-нибудь слишком проницательный и болтливый медик решил с ведома полиции проинформировать директора о шансах на жизнь его застрахованных. Этот медик рассеял бы неведение директора, но не имел бы никакого влияния на дивиденды, которые, очевидно, не являются результатом этого неведения.

II. Определение случая

Чтобы найти наилучшее определение случая, мы рассмотрим некоторые события из тех, что принято считать случайными, события, к которым кажется применимой теория вероятностей; затем мы выясним, каковы их общие свойства.

Первый пример, который мы выберем, будет связан с неустойчивым равновесием; мы хорошо знаем, что если конус поставить на вершину, то он сразу же упадет, но мы не знаем, на какой бок; кажется, что это может решить только случай. Если бы конус был совершенно симметричным, если бы его ось была совершенно вертикальной, если бы он не подчинялся никакой другой силе, кроме силы тяжести, он бы вообще не упал. Но малейший дефект симметрии заставит его легко наклониться в одну или другую сторону, а как только конус наклонится, он совершенно упадет на эту же сторону, как бы ни был мал этот наклон. Даже если симметрия совершенна, одно очень легкое колебание, одно дуновение ветра сможет наклонить его на угол в несколько секунд; этого будет достаточно, чтобы вызвать падение, и направление падения совпадет с направлением начального отклонения.

Очень мелкая, ускользающая от нас причина вызывает значительное следствие, которое мы не можем не заметить; тогда мы говорим, что этим следствием мы обязаны случаю. Если бы мы точно знали законы природы и состояние Вселенной в начальный момент времени, мы бы могли точно предсказать состояние этой же самой Вселенной в позднейший момент. Но даже если бы законы природы больше не были секретом для нас, мы все равно не смогли бы знать состояние мира иначе, чем *приблизительно.* Если это позволяет нам предвидеть дальнейшую ситуацию *с тем же самым приближением*, мы получаем все, что нужно, и говорим, что данное явление было предвидено, что оно управляется некоторыми законами; но так бывает не всегда, может случиться, что небольшие различия в начальных условиях порождают очень большую разницу в конечных явлениях; маленькая ошибка в первых производит большую ошибку в последних. Предсказание становится невозможным, и мы имеем случайное событие.

Наш второй пример будет очень похож на первый, мы позаимствуем его из метеорологии. Почему метеорологам с таким трудом удается предсказать погоду более-менее точно? Почему нам кажется, что ливни и даже бури приходят случайно, так что многие люди находят совершенно естественным молиться о дожде или ясной погоде, в то время как они сочли бы смешным просить в молитве о затмении? Мы видим, что наибольшие потрясения происходят главным образом в тех областях, где атмосфера находится в неустойчивом равновесии. Метеорологи хорошо знают, что это равновесие неустойчиво, что где-то должен возникнуть циклон, но они совершенно неспособны сказать, где именно; одной десятой градуса в какой-нибудь точке больше или меньше — и циклон разражается здесь, а не там, и производит опустошение в местности, которую должен был бы пощадить. Если бы мы вычислили эту десятую долю градуса, мы бы смогли узнать о циклоне заранее, но наблюдения не были ни достаточно точны, ни достаточно интенсивны, и из-за этого все кажется обязанным вмешательству случая. Здесь мы снова находим тот же контраст между мелкой, незаметной для наблюдателя причиной и значительными следствиями, которые иногда оборачиваются ужасными бедствиями.

Перейдем к другому примеру, к распределению малых планет в поясе зодиака. Значения их начальной долготы могли быть любыми, но их средние скорости были различны, и эти планеты находились в движении настолько долго, что сейчас мы можем считать их распределенными вдоль зодиака случайно. Очень маленькие начальные различия в расстоянии до Солнца или в средних скоростях дали огромную разницу в их нынешней долготе; действительно, излишек в одну тысячную секунды при среднем перемещении за день даст одну секунду за три года, один градус за десять тысяч лет, полную окружность за три или четыре миллиона лет, а что будет по истечении времени, прошедшего с тех пор, как малые планеты оторвались от лапласова облака? Вот еще один пример малой причины и большого следствия, точнее, маленьких различий в причине и больших различий в следствии.

Игра в рулетку уведет нас немного в сторону, т. к. она непохожа на предыдущий пример. Предположим, что стрелка, которую можно поворачивать вокруг некоторой оси, находится на циферблате, разделенном на 100 секторов, красных и черных попеременно. Если стрелка останавливается на красном секторе, партия выиграна, если нет — проиграна. Очевидно, все зависит от начального импульса, который мы сообщим стрелке. Я предполагаю, что стрелка сделает 10 или 20 оборотов, но она может остановиться более или менее быстро в зависимости от того, толкнул я более или менее сильно. Достаточно, чтобы импульс изменился всего на одну или две тысячных, — и моя стрелка остановится на черном секторе, или на следующем, который окажется красным. Мускульное чувство не в состоянии оценить эти различия, они ускользнули бы даже от очень тонких инструментов. Итак, я не могу предвидеть, что сделает стрелка, которую я только что толкнул, вот почему мое сердце бьется, и я полагаюсь на случай. Разница в причине неощутима, а разница в следствии имеет для меня очень большое значение, т. к. речь идет обо всей моей ставке.

III

Поговорим теперь о других примерах, в которых появятся немного иные свойства. Займемся сначала кинетической теорией газа. Как следует представлять себе заполненный газом сосуд? Бесчисленные молекулы, наделенные большими скоростями, пересекают его во всех направлениях; каждое мгновение они ударяются о стенку сосуда или друг о друга, и эти удары происходят при очень разных условиях. В таком случае нас особенно интересует не малость причин, а их сложность. Кроме того, здесь присутствует и первая особенность, она тоже играет важную роль. Если бы одна молекула отклонилась от своей траектории вправо или влево на очень маленькую величину, сравнимую с радиусом взаимодействия молекул газа, она бы избежала удара или испытала бы его в других условиях, и направление ее скорости после удара изменилось бы на 90° или, возможно, 180°.

И это еще не все. Мы только что видели, что достаточно отклонить молекулу на бесконечно малую величину до удара, чтобы после удара она отклонилась на величину конечную. Если же молекула испытывает два последовательных удара, то перед первым из них достаточно будет отклонить эту молекулу на бесконечно малую величину второго порядка, чтобы после первого столкновения мы имели бесконечно малую первого порядка, а после второго столкновения — конечную величину. Но молекула не подвергается всего лишь двум ударам, она испытывает очень большое число ударов за секунду. И если после первого удара величина отклонения умножается на какое-то очень большое число A, то после n ударов она умножится на A^n; отклонение окажется очень большим не только из-за того, что A велико, т. е. из-за того, что малые причины приводят к большим следствиям, но и потому что показатель степени n велик, т. е. потому что удары очень многочисленны и причины этого сложны.

Перейдем ко второму примеру; почему во время ливня капли дождя кажутся нам распределенными случайно? Конечно, из-за сложности причин, определяющих их образование. Представим себе ионы, рассеянные в атмосфере; долгое время на них действуют постоянно меняющиеся потоки воздуха, они вовлекаются в вихри очень малых размеров, и их конечное распределение не имеет ничего общего с начальным. Вдруг температура снижается, водяные пары конденсируются, и каждый из этих ионов становится центром одной дождевой капли. Для того чтобы знать, каким окажется распределение этих капель и сколько их упадет на каждую мостовую, мало изучить начальное распределение ионов, требуется подсчитать воздействия тысячи вихрей воздуха, мелких и капризных.

Произойдет то же самое, если мы будем рассматривать частицы пыли, взвешенные в воде; сосуд полон течений, их законы нам неизвестны, мы знаем только то, что они очень сложны; в конце какого-то промежутка времени частицы будут распределены случайно, т. е. равномерно, по всему сосуду, причем именно из-за сложности этих течений. Если бы все они повиновались какому-то простому закону, если бы, например, сосуд вращался и течения циркулировали бы вокруг его оси, описывая круги, то картина была бы совсем другой, т. к. каждая частица сохранила бы свою начальную высоту и начальное расстояние до оси вращения.

Можно прийти к тому же результату для смеси двух жидкостей или двух тонко размолотых порошков. В качестве более грубого примера мы можем рассмотреть то, что происходит при тасовании игральных карт. При каждом движении карты подвергаются некоторой перестановке (аналогичной тем, которые составляют предмет теории подстановок). Какая из этих перестановок осуществится? Я утверждаю, что вероятность, с которой должна произойти именно данная перестановка (например, такая, которая помещает на n-ое место карту, изначально занимавшую место $\varphi(n)$), эта вероятность зависит от привычек игрока. Но если игрок будет тасовать карты достаточно долго, произойдет большое число последовательных перестановок; в итоге конечный порядок карт будет определяться только случаем; я хочу сказать, что все возможные порядки будут равновероятны. Такой результат возникает из-за большого числа последовательных перестановок, т. е. из-за сложности явления.

Наконец, несколько слов о теории ошибок. Здесь причины и сложны, и многочисленны. Сколько ловушек подстерегает наблюдателя, даже вооруженного наилучшими инструментами? Надо по возможности замечать наиболее грубые из этих ловушек и избегать их. К наиболее грубым относятся прежде всего те, которые порождают систематические ошибки. Но даже если мы допустим, что нам удалось избежать их, остается много мелких ловушек, которые, при скоплении их последствий, могут стать опасными. Именно отсюда проистекают случайные ошибки; мы приписываем их случаю, т. к. их причины слишком сложны. Здесь снова имеются малые причины, но каждая из них приводит лишь к незначительным следствиям; только из-за их объединения и из-за большого количества этих причин их следствия становятся угрожающими.

IV

Можно принять еще и третью точку зрения, которая не так важна, как две предыдущие, и на которой я бы настаивал меньше. Когда мы стараемся предвидеть какое-то событие и изучаем все, предшествующее ему, мы должны исследовать предыдущую ситуацию; но мы не можем сделать это во всей Вселенной и удовлетворяемся знанием того, что происходит по соседству с тем местом, где должно случиться наше событие, или того, что, по-видимому, имеет какое-то отношение к этому событию. Исследование не может быть полным, здесь надо уметь выбирать. Однако может случиться так, что мы оставим в стороне обстоятельства, которые изначально покажутся нам совершенно чуждыми ожидаемому событию, которым мы никогда бы не подумали приписывать какое-то влияние на него и которые, между тем, вопреки всем ожиданиям, сыграют важную роль.

Человек идет по улице по каким-то своим делам; кто-то, находящийся в курсе этих дел, мог бы сказать, по какой причине он вышел в этот час, почему он идет по улице. На крыше работает кровельщик; нанявший его предприниматель мог бы в определенной мере предвидеть то, что он сделает. Но прохожий почти не думает о кровельщике, как и кровельщик о прохожем; кажется, что они принадлежат двум совершенно чуждым друг другу мирам. И все-таки кровельщик роняет черепицу, которая убивает прохожего, и, не колеблясь, говорит, что это случай.

Наша слабость не позволяет нам объять всю Вселенную полностью, мы вынуждены делить ее на части. Мы пытаемся сделать это как можно менее искусственно, и тем не менее время от времени случается, что две таких части взаимодействуют между собой. И тогда последствия этого взаимодействия кажутся нам обязанными случаю.

Будет ли это третьим способом понять случай? Не всегда. В действительности он восходит к первому или второму способам. Всякий раз, когда два мира, в общем чужие друг для друга, воздействуют друг на друга некоторым образом, законы этого воздействия обязательно будут очень сложны; с другой стороны, хватило бы очень маленького изменения начальных условий в этих двух мирах, чтобы их взаимодействие вообще не имело места. Как мало бы потребовалось, чтобы прохожий прошел одной секундой позже или чтобы кровельщик уронил свою черепицу секундой раньше!

V. Законы случая

Все вышесказанное еще не объясняет нам, почему случай подчиняется каким-то законам. Достаточно ли нам найти незначительные или сложные причины, чтобы предвидеть если не то, каковы окажутся результаты *в каждом случае*, то, как минимум, каковы они будут *в среднем*? Чтобы ответить на этот вопрос, лучше всего вернуться к какому-нибудь примеру из рассмотренных ранее.

Я начну с примера с рулеткой. Как уже было сказано, место, где остановится стрелка, будет зависеть от сообщенного ей начального импульса. С какой вероятностью данный импульс имеет ту или иную величину? Я ничего не знаю об этом, но для меня естественней всего допустить, что эта вероятность представляется непрерывной аналитической функцией. Вероятность того, что импульс заключен между a и $a+\varepsilon$ будет, таким образом, практически равна вероятности того, что он заключен между $a+\varepsilon$ и $a+2\varepsilon$ *при условии, что* ε *очень мало.* Это общее свойство всех аналитических функций. Малые изменения функции пропорциональны малым изменениям переменной.

Но мы предположили, что очень малой разницы в импульсе будет достаточно для изменения цвета сектора, на котором остановится стрелка. От a до $a+\varepsilon$ он красный, от $a+\varepsilon$ до $a+2\varepsilon$ — черный; вероятность получить красный сектор совпадет с аналогичной вероятностью для следующего за ним черного; поэтому полная вероятность красного равна полной вероятности черного.

В этой задаче нам должна быть задана аналитическая функция, представляющая собой вероятность определенного начального импульса. Но теорема останется верной, какой бы ни была эта функция, т. к. результат здесь зависит от общего свойства всех аналитических функций. Это означает, что мы не нуждаемся ни в каких данных.

Все, о чем мы говорили в случае с рулеткой, так же можно применить и к примеру с малыми планетами. Мы можем рассматривать зодиак как гигантскую рулетку, на которую Создатель бросил очень большое число маленьких шаров и сообщил им различные начальные импульсы, меняя их при этом по какому-то закону. Нынешнее распределение планет равномерно и не зависит от этого закона по тем же причинам, что и в предыдущем случае. Итак, понятно, почему происходящее повинуется законам случая, когда маленькие различия в причинах достаточны для возникновения большой разницы в следствиях. Можно считать, что вероятности этих маленьких различий пропорциональны самим различиям, и именно потому, что эти различия малы, а малые приращения непрерывной функции пропорциональны приращениям аргумента.

Перейдем к совершенно другому примеру, в котором повсюду возникает сложность причин; я предполагаю, что какой-то игрок тасует карты. Каждый раз он меняет порядок карт, а изменить его можно многими способами. Для простоты изложения рассмотрим только три карты. Карты, которые до перестановки занимали места 1 2 3 соответственно, после перестановки смогут занять места

$$1\,2\,3,\quad 2\,3\,1,\quad 3\,1\,2,\quad 3\,2\,1,\quad 1\,3\,2,\quad 2\,1\,3.$$

Каждая из этих шести гипотез возможна, и они имеют вероятности

$$p_1,\, p_2,\, p_3,\, p_4,\, p_5,\, p_6$$

соответственно.

Сумма этих шести чисел равна 1, и это все, что мы знаем; все шесть вероятностей, естественно, зависят от привычек игрока, с которыми мы незнакомы.

При второй и следующих перестановках все начнется сначала, в тех же условиях; я хочу сказать, что p_4, например, всегда будет равна вероятности того, что три карты, занимавшие после n-й и перед $(n+1)$-й перестановками места 1 2 3 соответственно, эти три карты займут места 3 2 1 после $(n+1)$-й перестановки. И это правило будет верным для любого числа n, т. к. привычки игрока, его манера тасовать карты, останутся теми же.

Но если количество перестановок очень велико, то карты, занимавшие перед первой перестановкой места 1 2 3, после последней перестановки смогут занять места

$$1\,2\,3, \quad 2\,3\,1, \quad 3\,1\,2, \quad 3\,2\,1, \quad 1\,3\,2, \quad 2\,1\,3,$$

и вероятность этих шести гипотез будет практически одной и той же и равной $\frac{1}{6}$; этот результат останется верным для любых $p_1, \ldots, p_6$, которые нам неизвестны. Большое число перестановок, т. е. сложность причин, приводит к равенству вероятностей.

Точно так же можно рассуждать в случае, когда имеется более трех карт; но даже для трех карт рассуждения будут слишком сложны, и я удовольствуюсь случаем всего лишь двух карт[1]. У нас есть только две гипотезы

$$1\,2, \qquad 2\,1$$

с вероятностями p_1 и $p_2 = 1 - p_1$. Пусть есть n перестановок; я выигрываю 1 франк, если в конце карты займут первоначальный порядок, и проигрываю 1 франк, если в конце они поменяются местами. Итак, мое математическое ожидание равно

$$(p_1 - p_2)^n.$$

Разность $p_1 - p_2$, разумеется, меньше 1; таким образом, при очень больших n мое ожидание будет нулевым, и нам не нужно знать p_1 и p_2, чтобы выяснить, что игра честная. И все-таки здесь имеется одно исключение, когда одно из чисел p_1 и p_2 равно 1, а второе нулю. *Но*

[1] Более полные вычисления смотрите в главе «Различные вопросы». — *Прим. авт.*

этот случай невероятен, т. к. тогда наши начальные гипотезы были бы слишком просты. Эти соображения применимы не только к тасованию карт, но и ко всем перемешиваниям порошков и жидкостей, и даже молекул в кинетической теории газа. Возвращаясь к этой теории, представим себе газ, чьи молекулы не могут сталкиваться друг с другом, но отражаются при ударе от стенки сосуда, в который заключен газ. Если форма сосуда достаточно сложна, то распределение молекул и их скоростей скоро станет равномерным. Но этого не случится, если сосуд сферический или имеет форму прямоугольного параллелепипеда. Почему? Потому что в первом случае постоянным останется расстояние от центра сферы до произвольной траектории; во втором случае постоянной будет абсолютная величина угла траектории со сторонами параллелепипеда.

Итак, видно, что следует понимать под *слишком простыми* условиями; это условия, которые сохраняют что-то неизменным, позволяют существовать какому-то инварианту. Будут ли дифференциальные уравнения в какой-то задаче слишком просты для применения законов случая? На первый взгляд этот вопрос кажется лишенным точного смысла; но теперь мы знаем, о чем речь. Уравнения слишком просты, если они что-то сохраняют, допускают одинаковый интеграл; понятно, что если какое-то начальное условие остается неизменным, то конечная ситуация уже не сможет не зависеть от начальной.

Перейдем, наконец, к теории ошибок. Мы не знаем, из-за чего происходят случайные ошибки, и именно из-за нашего незнания мы знаем, что они будут подчиняться закону Гаусса. Таков парадокс. Его можно объяснить примерно так же, как в предыдущих случаях. Нам нужно знать только одно: что ошибки очень многочисленны, что они очень малы, что каждая из них может быть как отрицательной, так и положительной. Какая кривая изображает вероятность каждой из них? Ничего не зная, мы предполагаем только то, что эта кривая симметрична. Тогда можно доказать, что итоговая ошибка подчиняется закону Гаусса, и этот итоговый закон не зависит от неизвестных нам частных. Так еще раз простота результата порождается сложностью самих данных.

VI

Мы постарались дать определение случая, и теперь следует задать себе один вопрос. Если мы смогли определить случай, то будет ли он иметь объективный характер?

Мы можем спрашивать себя об этом. Я говорил о причинах очень малых или очень сложных. Но разве очень малое для одного не может быть большим для другого и то, что кажется очень сложным одному, не может показаться простым другому? Частично я уже ответил на этот вопрос, т. к. выше было *совершенно точно* указано, в каком случае дифференциальные уравнения становятся слишком просты для применения законов случая. Но эту проблему следует изучить немного более подробно, т. к. здесь можно принять еще и другие точки зрения.

Что означает выражение «слишком мало»? Для того, чтобы понять это, достаточно сослаться на слова, сказанные выше. Различие очень мало, интервал очень мал, когда в пределах этого интервала вероятность остается практически постоянной. А почему эта вероятность должна рассматриваться как постоянная на данном интервале? Потому что мы допускаем, что вероятностный закон представлен непрерывной кривой, и не только непрерывной в аналитическом смысле слова, но и *практически* непрерывной, как уже объяснялось выше. Это означает, что у кривой не только нет точек разрыва, у нее также не будет слишком крутых или слишком резких подъемов и спадов.

Но что позволяет нам выдвигать такую гипотезу? Как было сказано выше, причина в том, что от начала веков в мире есть сложные законы, не прекращающие действовать все еще в одном и том же направлении, и эти законы постоянно заставляют мир двигаться к однообразию, так что он никогда бы не смог вернуться назад. Вот почему постепенно понижаются подъемы и уменьшаются спады, и из-за этого наши вероятностные кривые возрастают и убывают очень медленно. За миллиарды миллиардов веков будет сделан еще один шаг к единообразию, и интервалы возрастания и убывания кривой растянутся еще в десять раз: средний участок нашей кривой станет в десять раз больше. И тогда та длина, которая не кажется нам очень малой, т. к. на нашей кривой дугу такой длины нельзя считать прямолинейной, в ту эпоху станет, наоборот, рассматриваться как очень маленькая, т. к. крутизна подъема уменьшится в десять раз и дугу такой длины можно будет практически заменить прямой линией.

Итак, слова об очень малом остаются относительными; но они относительны не для одного и того же человека, а для современного состояния мира. Эти слова существенно изменятся, когда мир станет еще более однообразным, так что все еще больше смешается. Но тогда, без сомнения, люди не смогут больше жить, им придется освободить место для других существ; должен ли я сказать, лучших или худших? В том смысле, что наш критерий остается верным для всех людей, он остается объективным.

С другой стороны, что означает выражение «очень сложное»? Я уже дал одно решение этой проблемы, то, на которое я сослался в начале нашего параграфа, но есть и другие. Мы говорили, что из-за сложных причин возникает все более и более тесное перемешивание, но через какое время это перемешивание удовлетворит нас? Когда накопится достаточно сложностей? Когда карты будут достаточно перетасованы? Если мы смешиваем два порошка, белый и голубой, то наступает момент, когда цвет смеси нам кажется одинаковым, но это происходит из-за несовершенства наших чувств; цвет уже кажется одинаковым дальнозоркому, т. к. он вынужден смотреть издалека, но он все еще неодинаков для близорукого. И когда цвет станет одним и тем же для всех взглядов, предел еще можно будет отодвинуть с помощью инструментов. Ни у одного человека нет шансов различить бесконечное разнообразие, которое, если права кинетическая теория, скрывается под однообразной внешностью газа. Однако, если принять идеи Гуи о броуновском движении, не окажется ли, что микроскоп способен продемонстрировать нам что-то аналогичное?

Этот новый критерий так же относителен, как и первый, и если он сохраняет объективный характер, то от того, что все люди имеют практически одни и те же органы чувств, а возможности их инструментов ограничены, да и используются они не иначе как в исключительных случаях.

VII. Вероятность в гуманитарных науках

То же самое происходит и в гуманитарных науках, особенно в истории. Историк вынужден делать выбор между событиями изучаемой эпохи; он рассказывает только о тех из них, которые кажутся ему наиболее важными. Так, например, он довольствуется изложением лишь наиболее значительных событий XVI века и самых заметных фактов из истории XVII века. Если первых оказывается достаточно для объяснения вторых, говорят, что эти события «соответствуют законам истории». Но если причиной какого-то крупного события XVII века оказывается мелкое происшествие XVI века, о котором не сообщил ни один историк, которым пренебрег весь мир, тогда говорят, что это событие обязано своим происхождением случаю; слово «случай» имеет тот же смысл, что и в точных науках; оно означает, что малые причины произвели большое следствие.

Самый большой случай — это рождение какого-нибудь великого человека. Только случайно может произойти встреча двух половых клеток разного пола, содержащих, каждая со своей стороны, в точности те таинственные элементы, чье взаимодействие должно произвести гения. Все согласны, что эти элементы являются очень редкими, а их встреча должна быть еще более редкой. Как мало требуется для того, чтобы отклонить несущий эти элементы сперматозоид с его пути; достаточно отклонить его на десятую долю миллиметра — и Наполеон не родился бы, и судьбы континента были бы иными. Ни один пример не может лучше заставить нас понять настоящий характер случая.

Еще одно слово о парадоксах, возникающих при применении теории вероятностей в гуманитарных науках. Было доказано, что ни одна палата никогда не содержала ни одного депутата от оппозиции или, по крайней мере, такое событие было бы настолько невероятным, что можно было бы не колеблясь ставить на обратное и спорить на миллион против одного су. Кондорсе постарался вычислить, сколько понадобится присяжных для того, чтобы судебные ошибки сделались практически невозможными. Если бы результаты этого вычисления были использованы, мы бы, конечно, испытали то же разочарование, которое ждет нас в споре, когда мы доверяемся подсчетам, говорящим, что оппозиция никогда не имеет своих представителей в палате.

Законы случая неприменимы к таким вопросам. Даже если правосудие не всегда совершается по достойным мотивам, то оно хотя бы не доверяет методу Бридуа; возможно, напрасно, т. к. лишь случайные судебные ошибки могли бы защитить нас от системы Кондорсе.

Что здесь можно сказать? Мы склонны приписывать случаю события этой природы, т. к. их причины темны; но это не настоящий случай. Конечно, причины неизвестны нам и даже сложны, но они сложны недостаточно, т. к. сохраняют кое-что неизменным; мы видели, что именно это отличает «слишком простые» причины. Когда люди сталкиваются друг с другом, они уже не ведут себя случайно и независимо друг от друга; они воздействуют друг на друга. Многочисленные причины приходят в действие, беспокоят людей, увлекают их направо и налево, но есть вещь, которую эти причины не могут разрушить, это наши привычки Панургова стада[1]. Они сохраняются в неизменности.

VIII. Различные размышления

Может возникнуть и много других вопросов; я бы хотел затронуть некоторые из них, прежде чем выделить те, которыми я займусь более подробно. Когда мы констатируем какой-то простой результат, например, находим круглое число, мы говорим, что подобный результат не может быть случайным, и пытаемся объяснить его причины неслучайным образом. Действительно, существует лишь очень слабая вероятность, что среди 10 000 чисел случай выберет круглое, например, 10 000; это всего лишь один шанс из 10 000. Но имеется не более одного шанса из 10 000 выбрать другое, не важно какое число; однако такой результат не удивит нас и не отнимет желания приписать его случаю; только из-за того, что он кажется менее поразительным.

Будет ли это просто нашей иллюзией, или есть примеры, когда такой способ ви́дения оправдан? Надо надеяться, что да, т. к. иначе все научные исследования стали бы невозможными. Что мы делаем, когда хотим проверить какую-нибудь гипотезу? Мы не можем проконтролировать все ее следствия, т. к. их будет бесконечно много; мы удовлетворяемся проверкой только некоторых из них, и, если это нам удается, мы объявляем, что гипотеза подтвердилась, т. к. такой успех нельзя объяснить случаем. И этот способ лежит в основе всех рассуждений. Сейчас я не смогу полностью обосновать его, т. к. это заняло бы у меня слишком много времени; но я могу, по меньшей мере, сказать следующее: у нас есть две гипотезы, т. к. наш результат может иметь либо одну простую причину, либо совокупность сложных, которые мы называем случаем. Мы считаем естественным предполагать, что первая гипотеза должна приводить к простому результату, и если мы получаем простой результат, например, круглое число, то нам кажется более

---

[1] Толпа бессмысленных подражателей (из романа Ф. Рабле «Гаргантюа и Пантагрюэль»). — *Прим. перев.*

правдоподобным приписать его простой причине, которая приводит к этому результату почти точно, чем случаю, который может дать нам его только с одним шансом из 10 000; но в действительности у простой причины не больше шансов привести к этому результату.

В дальнейшем мы увидим, почему в таблице логарифмов десятичные знаки кажутся распределенными соответственно законам случая. Тот же вопрос можно задать о десятичных знаках числа π. Но этот случай будет более тонким.

Допустим, мы знаем, что число π находится между 3 и 4, и рассматриваем первые n десятичных знаков; возьмем целую часть числа $10^n(\pi - 3)$, она лежит между 0 и 10^n; из 10^n целых чисел, заключенных в этих пределах, выбираем случайным образом одно; находим, сколько у него будет цифр 7 и 5 и насколько первое число превышает второе. Пусть это превышение равно e. Предположим, что среди наших 10^n чисел есть N таких, у которых отношение $\frac{e}{n}$ меньше ε.

Закон больших чисел говорит нам, что $N \cdot 10^{-n}$ стремится к 1, когда n бесконечно растет. Итак, если мы выбираем число, лежащее в этих пределах *случайно*, то вероятность того, что превышение e и все аналогичные превышения будут сравнительно малы, т. е. вероятность того, что десятичные знаки будут распределены равномерно, случайным образом, окажется очень близка к вероятности достоверного события.

Но почему мы имеем право рассуждать так, как если бы число π выбиралось случайно? Если бы вместо π мы рассмотрели простую рациональную дробь, чья запись в десятичной форме порождает периодическую дробь, то результат вовсе бы не был верен.

По-видимому, число π кажется нам выбранным случайно, т. к. его происхождение сложно, и мы бессознательно рассуждаем об этом числе так же, как о следствиях, возникающих из совокупности сложных причин.

Глава 1
Определение вероятностей

1. Почти невозможно дать удовлетворительное определение *вероятности.* Обычно говорят, что вероятность события равна отношению числа исходов, благоприятных для этого события, к числу всех возможных исходов.

Так, если первое число равно n, а второе N, то вероятность будет равна $\frac{n}{N}$; в известных случаях это определение не вызывает никаких сложностей. Вероятность вытащить короля из колоды в 32 карты равна $^4\!/_{32}$, т. к. полное число всех возможных исходов, т. е. карт, равно 32, и среди этих карт есть четыре короля; получается, что $N = 32$, $n = 4$. Когда бросают кость, вероятность получить 4 равна $\frac{1}{6}$, т. к. $N = 6$ и $n = 1$, потому что кость имеет шесть граней, из которых только одна содержит 4. Из урны, содержащей n белых и p черных шаров, вытаскивают один шар; вероятность того, что он окажется белым, равна

$$\frac{n}{n+p}.$$

2. Рассмотрим немного более сложный пример. Две урны, неразличимые внешне, содержат: первая — n белых и p черных, вторая — n' белых и p' черных шаров. Пусть надо вытащить один шар; спрашивается, какова вероятность вытащить белый. Можно было бы сказать, что общее число исходов равно $n + n' + p + p'$, а вероятность равна $\frac{n+n'}{n+n'+p+p'}$. В то же время можно сказать, что сначала могут представиться два исхода, т. е. выбор первой или второй урны; вероятность предпочесть первую урну равна $\frac{1}{2}$, вторую — тоже $\frac{1}{2}$, т. к. имеется столько же шансов протянуть руку к одной урне, как и ко второй. Если я опустил руку в первую урну, то вероятность того, что, выбрав первую урну, я вытащу белый шар, равна $\frac{n}{n+p}$; аналогичная вероятность для второй урны равна $\frac{1}{2} \cdot \frac{n'}{n'+p'}$.

Сумма $\frac{1}{2} \cdot \frac{n}{n+p} + \frac{1}{2} \cdot \frac{n'}{n'+p'}$ и будет правильной оценкой требуемой вероятности, и две найденные оценки совпадут только в частном случае, когда $\frac{n}{n+p} = \frac{n'}{n'+p'}$, т. е. $\frac{n}{p} = \frac{n'}{p'}$. Откуда вытекает это расхождение? Дело в том, что $n + n' + p + p'$ исходов не будут *одинаково* вероятны.

Так, предположим, что в первой урне шаров в два раза больше:

$$n' + p' = \frac{1}{2}(n+p).$$

Вероятность того, что я выну один *данный* шар из этой урны, равна $\frac{1}{2(n+p)}$; а вероятность вынуть его из второй урны будет равна $\frac{1}{n+p}$.

Таким образом, в определение вероятности надо добавить: при условии, что все исходы *одинаково* правдоподобны(1).

Приведем два других примера, принадлежащих Бертрану.

3. Задача о трех шкатулках. В каждой из трех одинаковых шкатулок A, B, C есть по два ящичка, α и β; каждый ящичек в шкатулке A содержит золотую монету, в B — серебрянную, а в C один из ящиков содержит золотую монету, другой — серебряную:

	A	B	C
α	золото	серебро	золото
β	золото	серебро	серебро

Чему равна вероятность того, что, открыв случайным образом один из шести ящиков, мы найдем золотую монету? Шесть исходов $A\alpha$, $A\beta$, $B\alpha$, $B\beta$, $C\alpha$, $C\beta$ будут равновероятны; из этих шести исходов три: $A\alpha$, $A\beta$ и $C\alpha$ — благоприятны для обнаружения золотой монеты. Итак, вероятность равна $\frac{1}{2}$.

Если мы выбираем одну из шкатулок случайно, то вероятность выбрать C совпадет с $\frac{1}{3}$.

Я открываю наугад одну из шкатулок и нахожу там золотую монету; какова вероятность того, что вторая монета будет серебряной?

Мне попалась либо шкатулка C, либо шкатулка A; в первом случае вторая монета будет серебряной, во втором — золотой. Кажется, что вероятность равна $\frac{1}{2}$. Но это заключение неверно.

Перед тем как открыть шкатулку, я знал, что найду золотую или серебряную монету с одинаковой вероятностью, т. е. с вероятностью $\frac{1}{2}$; между тем я могу найти золотую монету в трех случаях, $A\alpha$, $A\beta$, $C\alpha$, и из этих трех исходов лишь один $C\alpha$ благоприятствует обнаружению серебряной монеты во втором ящике.

Когда мы в первый раз оценили вероятность как $\frac{1}{2}$, два рассмотренных исхода не были равновероятны: исход A соответствовал $A\alpha$ и $A\beta$ и был вдвое вероятнее, чем исход C, который соответствовал только $C\alpha$.

4. Задача об игре в шары. Два одинаково искусных игрока, Пьер и Поль, играют в шары; Пьер должен бросить два шара, Поль — один, и победит тот из них, чей шар окажется ближе всего к цели.

Чему равна вероятность того, что выиграет Поль?

Пусть A и B — шары Пьера, C — шар Поля; если упорядочить шары по их близости к цели, то мы получим шесть возможных событий:

$$ABC, \quad BCA, \quad CAB, \quad ACB, \quad CBA, \quad BAC.$$

Эти шесть событий равновероятны; количество событий, обеспечивающих победу Пьера, равно четырем, два события обеспечивают победу Поля; итак, Поль выигрывает с вероятностью $\frac{1}{3}$.

Можно было бы рассуждать иначе: шар Пьера A может оказаться дальше от цели, чем C, или наоборот.

$$A > C \quad \text{или} \quad A < C.$$

То же верно для шара B:

$$B > C \qquad B < C.$$

Итак, возможны четыре случая:

$$\begin{gathered} A > C \quad \text{и} \quad B > C, \\ A < C \quad \text{и} \quad B > C, \\ A > C \quad \text{и} \quad B < C, \\ A < C \quad \text{и} \quad B < C. \end{gathered}$$

Только один исход, первый, благоприятен для Поля, т. к. его шар в этом случае ближе к цели, чем A и B; вероятность получается равной $\frac{1}{4}$.

Но эти четыре исхода не равновероятны.

$A > C$ и $B > C$ соответствуют двум сочетаниям CAB, CBA,
$A < C$ и $B > C$ соответствуют одному сочетанию ACB,
$A > C$ и $B < C$ соответствуют одному сочетанию BCA,
$A < C$ и $B < C$ соответствуют двум сочетаниям ABC, BAC.

5. Таким образом, полное определение вероятности включает в себя некоторое заявление о принципах, по которым распознаются все равновероятные исходы. Математическое определение здесь невозможно; нам придется в каждом случае заключить некоторое *соглашение* и заявлять, что мы будем рассматривать какие-то исходы как равновероятные. Эти соглашения не будут совершенно произвольными, но они ускользают от математического ума, и математик не должен их изучать, после того как они уже приняты.

Следовательно, всякая вероятностная задача решается в два этапа: первый, так сказать, метафизический, обосновывает то или иное соглашение; второй, математический, применяет к этим соглашениям какие-то численные законы(2).

6. Мы будем группировать вопросы, которыми собираемся заниматься, по-разному в зависимости от разных точек зрения; сначала рассмотрим все вопросы с точки зрения возможных исходов.

К первой группе мы отнесем все те задачи, где число возможных исходов конечно и не превышает определенные пределы; это задачи об азартных играх и простые задачи комбинаторного анализа.

Ко второй группе относятся задачи, в которых число исходов остается конечным, но становится очень большим; здесь можно получить лишь приближенные оценки вероятности, используя закон больших чисел, теорему Бернулли и т. д. Эти задачи составляют предмет статистики.

В третьей категории число возможных исходов бесконечно.

Так, если мы бросаем иглу на листок бумаги, на котором начерчены параллельные линии, то вероятность, с который игла пересечет одну из этих линий, зависит от бесконечного числа возможных исходов.

Именно в этом последнем случае особенно важно очень тщательно определить предварительные соглашения.

Пусть, например, известно, что дробное или иррациональное число x заключено между 0 и 1, и требуется узнать вероятность того, что x лежит между 0 и $\frac{1}{2}$; количество возможных исходов бесконечно. Хочется сказать, что эта вероятность равна $\frac{1}{2}$; однако точно так же можно сказать, что если $0 < x < \frac{1}{2}$, то его квадрат, обозначим его y, лежит между 0 и $\frac{1}{4}$.

Если $x^2 = y$, а x лежит между 0 и 1, то получаем, что $0 < y < 1$. Благоприятными будут все случаи, когда $0 < y < \frac{1}{4}$; если мы поделим интервал, заключенный между 0 и 1, на четыре равные части, то вероятность того, что y лежит между 0 и $\frac{1}{4}$, будет равна $\frac{1}{4}$.

Однако было бы грубой ошибкой оценивать в $\frac{1}{4}$ вероятность того, что x лежит между 0 и $\frac{1}{2}$.

Действительно, в первый раз мы считали равновероятными две гипотезы:

$$x_0 < x < x_0 + \varepsilon \quad \text{и} \quad x_1 < x < x_1 + \varepsilon,$$

где ε — одно и то же; во второй раз мы считали равновероятными гипотезы

$$x_0^2 < x^2 < x_0^2 + \varepsilon \quad \text{и} \quad x_1^2 < x^2 < x_1^2 + \varepsilon,$$

и эти два соглашения противоречат друг другу.

Здесь x — произвольная константа; выше, в задаче о бросании иглы, имелись три произвольные константы: координаты середины иглы и ее направление. В других вероятностных задачах еще больше произвольных констант или даже произвольных законов. Например, функция $y = f(x)$ может казаться более вероятной, чем другая такая же функция; это происходит при интерполяции. Такие задачи мы отнесем к четвертой группе[(3)].

7. Примем другую точку зрения.

Вопрос о вероятности возникает только из-за нашего невежества: если бы у нас были все данные о проблеме, то не имело бы места ничего, кроме определенности. С другой стороны, наше невежество не должно быть полным, иначе мы бы вообще не смогли ничего оценить. Поэтому можно классифицировать вероятностные задачи в зависимости от того, больше или меньше глубина нашего невежества.

Например, вероятность того, что шестой десятичный знак произвольного числа из таблицы логарифмов равен шести, a priori совпадает с $\frac{1}{10}$; в действительности все условия в этой задаче определены, и если бы мы хотели потратить время на подсчеты, мы бы узнали эту вероятность точно. То же самое получается при интерполяции, вычислении интегралов по методу Котеса или Гаусса и т. д.

В физических задачах наше невежество становится больше; речь идет о предвидении события, т. е. о каком-то будущем явлении, которое, с одной стороны, зависит от явления предыдущего, с другой стороны, — от закона, объединяющего предыдущие с последующим. Может случится так, что мы знаем этот закон, но не предыдущее явление; с какой вероятностью произойдет последующее?

Допустим, например, что мы знаем закон движения молекул; если бы мы точно знали их начальное расположение, мы бы смогли сказать, где они будут находиться в данный момент; вероятность, с которой эти молекулы займут определенное конечное положение, будет зависеть от вероятности, которую мы *по договоренности* припишем той или иной начальной позиции. В каждом случае необходима какая-то отдельная гипотеза.

Так, когда ищут вероятность того, что кометы имеют эллиптические орбиты, надо заключить соглашение, предполагать, что на большом расстоянии от Солнца эти небесные тела распределены в пространстве равномерно, так же как и направления их скоростей.

Другой аналогичный вопрос: будут ли случайны пробелы, возникающие в ряду малых планет? Их начальное расположение опять неизвестно, но астроном знает закон их движения. Какие соглашения относительно их начальных позиций следует выбрать в этом случае?

Сложно сделать это, не впадая в произвол. Однако есть гипотезы, которые кажутся совершенно невероятными: нельзя допустить, что начальные скорости таковы, что все орбиты имеют один и тот же эксцентриситет.

С другой стороны, может случиться так, что какие-то результаты окажутся в определенной мере независимыми от принятого закона, соединяющего предыдущее с последующим. Рассмотрим очень большое число малых планет с различными средними скоростями; их радиус-векторы, дли́ны, начальные скорости распределены по какому-то закону. Через очень большой промежуток времени эти малые планеты распределятся равномерно по всем долготам. Их число будет одинаковым в равных секторах.

8. Наконец, может случиться так, что в других задачах наше невежество окажется еще большим, и сам закон ускользнет от нас. В этом случае определить вероятность практически невозможно. Если, например, x — неизвестная функция от t, то мы не знаем точно, какую вероятность надо изначально приписать величине x_0, чтобы найти

$$\int_{t_0}^{t_1} x\,dt.$$

Тогда мы часто позволяем себе положиться на некоторое смутное чувство, которое очень сильно́ в нас; мы не умеем оправдать его, но без него в любом случае невозможна никакая наука. Самые точные законы возникают из отдельных экспериментов, чьи результаты следует обобщать. Когда Кеплер выводил свои законы из наблюдений Тихо Браге, кто-нибудь мог возразить ему: «Тихо Браге не всегда смотрел на небо, не могло ли случится так, что, когда он его не наблюдал, закон, который вы ищете, изменялся?»

Конечно, Кеплер нашел бы это возражение смешным, он бы ответил: «Эта гипотеза совершенно неправдоподобна». Тем самым он бы апелировал к этому плохо определенному чувству вероятности.

9. Задачи на условную вероятность[1] будут более тонкими, чем задачи о вероятности следствий.

Когда мы только что решали задачу об урне, мы знали, что в ней было n белых и p черных шаров; когда мы искали вероятность вытащить белый шар, причина была известна: причиной была урна с n белыми и p черными шарами.

Но можно поставить обратную задачу, когда я знаю, что в урне всего $n + p$ шаров, но не знаю, как они распределены. Я вытаскиваю черный шар; чему равна вероятность того, что черных шаров больше, чем белых? Это и есть условная вероятность.

Мы постоянно проводим параллели с физикой; законы известны нам только по их следствиям, которые мы наблюдаем. Пытаться вывести из следствий законы, касающиеся причин, и означает решить задачу на условные вероятности.

---

[1]В оригинале используется термин Probabileté des causes, т. е. вероятность причин. — *Прим. перев.*

10. Не настаивая больше на метафизической стороне вероятностных вопросов, с единственной целью побудить к размышлениям на эту тему, я замечу еще, что вероятность может быть субъективной. Можно иметь какие-то личные основания верить, что одна гипотеза вероятнее другой.

Вероятность может быть объективной, например, в статистике, когда ищут вероятное число людей, которые умрут в течение года; однако это число немного скачет. В каких пределах наши предвидения подтверждаются? И почему?

Это немного таинственно, недоступно математику.

Как бы то ни было, в математическом изложении теории вероятностей я буду придерживаться порядка, обозначенного выше.

Я начну с задач, в которых число возможных исходов конечно; затем, в случае очень большого числа исходов, я изучу теорему Бернулли и ее следствия, условные вероятности, задачи, где возникают произвольные константы; в предположении, что число исходов бесконечно, я изложу теорию ошибок, очень важный раздел; наконец, я покажу, как можно находить законы, т. е. произвольные функции.

Глава 2
Сложение и умножение вероятностей

11. Вычисление вероятностей основывается на двух теоремах: теореме сложения вероятностей и теореме умножения вероятностей.

Имея два события A и B, можно ставить различные вероятностные задачи в зависимости от того, сколько событий должно произойти: одно, оба или ни одного.

Либо A и B происходят оба — гипотеза, которую я назову AB.

Либо A происходит, B нет — гипотеза, которую я назову AB'.

Либо A не происходит, B происходит — гипотеза, которую я назову $A'B$.

Либо ни A, ни B не происходят — гипотеза, которую я назову $A'B'$.

Предполагаем, что AB реализуется в α различных исходах, AB' реализуется в β различных исходах, $A'B$ реализуется в γ различных исходах, $A'B'$ реализуется в δ различных исходах. Общее число возможных исходов равно $\alpha + \beta + \gamma + \delta$; заранее предполагается, что исходы равновероятны.

Рассмотрим различные вероятности. Вероятность того, что A произойдет, равна

$$p_1 = \frac{\alpha + \beta}{\alpha + \beta + \gamma + \delta}, \qquad (A)$$

где благоприятны исходы AB и AB'.

Вероятность того, что B произойдет, равна

$$p_2 = \frac{\alpha + \gamma}{\alpha + \beta + \gamma + \delta}, \qquad (B)$$

а вероятность того, что произойдет как минимум одно из двух событий, равна

$$p_3 = \frac{\alpha + \beta + \gamma}{\alpha + \beta + \gamma + \delta}, \qquad (A \text{ или } B)$$

где благоприятны три первых исхода AB, AB' и $A'B$. Вероятность того, что оба события произойдут, равна

$$p_4 = \frac{\alpha}{\alpha+\beta+\gamma+\delta}, \qquad (A \text{ и } B)$$

где благоприятна только одна гипотеза AB. Еще мы должны рассмотреть вероятность того, что A произойдет, если произошло B:

$$p_5 = \frac{\alpha}{\alpha+\gamma}, \qquad (A, \text{ если } B).$$

Мы заранее знаем, что B произошло, поэтому число возможных исходов сократилось, как и число благоприятных.

Вероятность того, что A произойдет, если B не произошло, равна

$$p_6 = \frac{\beta}{\beta+\delta}. \qquad (A, \text{ если } B'),$$

где $\beta+\delta$ — число возможных исходов.

Вероятность того, что B произойдет, если произошло A, равна

$$p_7 = \frac{\alpha}{\alpha+\beta}. \qquad (B, \text{ если } A)$$

Вероятность того, что B произойдет, если известно, что A не произошло, равна

$$p_8 = \frac{\gamma}{\gamma+\delta}. \qquad (B, \text{ если } A')$$

12. Вышеупомянутые теоремы сводятся к простым тождествам.

Рассмотрим p_1, p_2, p_3, p_4. Имеем

$$p_1 + p_2 = p_3 + p_4,$$

$$p_4 = \frac{\alpha}{\alpha+\beta+\gamma+\delta} = \frac{\alpha}{\alpha+\gamma} \cdot \frac{\alpha+\gamma}{\alpha+\beta+\gamma+\delta} = p_2 \cdot p_5;$$

аналогично $p_4 = p_1 \cdot p_7$.

Итак, сумма вероятностей того, что произойдет A, и того, что произойдет B, равна сумме вероятностей того, что произойдет хотя бы одно из них, и того, что оба они произойдут:

$$(A) + (B) = (A \text{ или } B) + (A \text{ и } B)^1.$$

[1]Здесь и далее имеется в виду сумма и умножение соответствующих формул.

Вероятность того, что A и B оба произойдут, равна вероятности того, что произойдет B, умноженной на вероятность того, что произойдет A, когда известно, что B произошло.

Или, наоборот, она равна вероятности того, что A произойдет, умноженной на вероятность того, что B произойдет в предположении, что должно произойти A

$$(A \text{ и } B) = (B) \cdot (A, \text{ если } B) = (A) \cdot (B, \text{ если } A).$$

13. Предположим, в частности, что $\alpha = 0$, откуда $p_4 = 0$; тогда

$$p_1 + p_2 = p_3.$$

Когда два события не могут произойти одновременно, вероятности A и B дают в сумме вероятность того, что какое-то одно из них произойдет.

Так, когда событие может произойти двумя разными способами, причем эти два способа не могут встретится одновременно, вероятность наступления данного события равна сумме вероятности того, что оно произойдет первым способом, и того, что оно произойдет вторым способом.

Это и есть теорема *сложения вероятностей.*

14. Может случиться, что $p_5 = p_1$. Тогда

$$\frac{\alpha}{\alpha+\gamma} = \frac{\alpha+\beta}{\alpha+\beta+\gamma+\delta},$$

откуда

$$\frac{\alpha+\gamma}{\alpha} = \frac{\alpha+\beta+\gamma+\delta}{\alpha+\beta},$$

$$1 + \frac{\gamma}{\alpha} = 1 + \frac{\gamma+\delta}{\alpha+\beta},$$

$$\frac{\alpha+\beta}{\alpha} = \frac{\gamma+\delta}{\gamma},$$

$$1 + \frac{\beta}{\alpha} = 1 + \frac{\delta}{\gamma},$$

$$\frac{\alpha}{\beta} = \frac{\gamma}{\delta}.$$

Когда выполняется это последнее условие, имеем $p_1 = p_5$; так же $p_1 = p_6$, если поменять местами α с β и γ с δ; кроме того, $p_2 = p_7$, так как p_1 меняется на p_2, а p_5 на p_7; и аналогично $p_2 = p_8$.

Итак,

$$p_1 = p_5 = p_6, \qquad p_2 = p_7 = p_8.$$

Иначе говоря, вероятность того, что A произойдет, остается одинаковой, известно ли, что B произошло, или известно, что B не произошло; другими словами, вероятность A не зависит от B.

Говорят, что эти два события независимы.

Из $p_5 = p_1$ получаем

$$p_4 = p_1 \cdot p_2;$$

вероятность того, что оба независимых события произойдут, равна произведению вероятностей каждого из них.

Это теорема об *умножении вероятностей.*

15. Из колоды в 32 карты одновременно вытаскивают 2 штуки.

Вероятность того, что первая карта — король, равна

$$p_1 = \frac{1}{8};$$

вероятность того, что вторая карта — король, равна

$$p_2 = \frac{1}{8};$$

вероятность того, что обе вытащенные карты — короли, равна

$$p_4 = \frac{4 \times 3}{32 \times 31}.$$

Среди всех размещений карт по две, то есть 32×31, ищем те, которые благоприятствуют нашему событию: их будет 4×3, так как в колоде 4 короля, и из них можно сформировать столько же размещений по 2, сколько из 4 букв. Вероятность того, что среди 2 карт будет как минимум один король, равна

$$p_3 = p_1 + p_2 - p_4 = \frac{8 \times 31 - 12}{32 \times 31}.$$

Нельзя говорить, что вероятность наличия хотя бы одного короля равна удвоенной вероятности того, что есть один король.

Урна содержит K шаров, занумерованных от 1 до K. Если искать вероятность появления № 1 при одновременном доставании двух шаров, то № 1 может стоять либо на первом шаре, либо на втором; эти два события несовместны и полная вероятность равна

$$p_1 + p_2 = \frac{2}{K}.$$

Вернемся к королям из колоды карт. Будут ли наши события независимы? Неверно, что $p_4 = p_1 \cdot p_2$; но $p_4 = p_1 \cdot p_7$.

Обращаясь к значению p_7, видим, что если первое событие A произошло, то остаются только 3 короля и 31 карта, и вероятность наступления B равна

$$p_7 = \frac{3}{31}.$$

Итак,

$$p_4 = \frac{4 \times 3}{32 \times 31}.$$

Другой пример независимых событий: я бросаю две игральные кости; какова вероятность того, что на каждой выпадет 6?

Вероятность того, что на первой выпадет 6, равна $\frac{1}{6}$. Вероятность того, что на второй выпадет 6, равна $\frac{1}{6}$.

Вероятность того, что на обеих выпадет 6, равна $\frac{1}{36}$, так как эти два события независимы.

16. Условие независимости не всегда так же очевидно, и теорему умножения можно применить неправильно, что и случается довольно часто.

Допустим, я ищу закон распределения попаданий при стрельбе из пистолета; у меня нет данных ни о стрелке, ни о пистолете. Получается вопрос в стиле: «Каков возраст капитана?»

Тем не менее возьмем две оси координат с началом в центре мишени: пусть x и y — прямоугольные, а ρ и ω — полярные координаты точки M.

Задача остается неопределенной, даже если мы считаем вероятность попаданий одинаковой во всех направлениях.

Вероятность того, что M находится в малом элементе поверхности $d\sigma$, можно обозначить как

$$f(x, y)\, d\sigma.$$

Следует определить $f(x;\,y)$; чтобы вероятность оставалась одинаковой во всех направлениях, эта функция должна зависеть от ρ; тогда эта вероятность запишется как $f(\rho)\,d\sigma$. Найдем вероятность того, что абсцисса точки попадания лежит между x и $x+dx$; она представляется как

$$\varphi(x)\,dx.$$

Так же вероятность того, что ордината точки попадания лежит между y и $y+dy$, представляется как

$$\psi(y)\,dy.$$

Но чтобы вероятность оставалась одной и той же во всех направлениях, следует предположить, что φ и ψ равны; тогда во втором случае имеем $\varphi(y)\,dy$.

Теперь проведем неправильное рассуждение: найдем вероятность того, что M находится в малом прямоугольнике со сторонами dx и dy. Должны одновременно произойти два события:

1° абсцисса лежит между x и $x+dx$;

2° ордината лежит между y и $y+dy$.

В силу теоремы об умножении вероятностей искомая вероятность будет

$$\varphi(x)\cdot\varphi(y)\,dx\,dy.$$

С другой стороны, эта вероятность выражается через $f(\rho)\,d\sigma$; итак, имеем

$$\varphi(x)\cdot\varphi(y)=f(\rho).$$

Возьмем производные от логарифмов обеих частей по переменной x, помня, что $\rho^2=x^2+y^2$; получим

$$\frac{\varphi'(x)}{\varphi(x)}=\frac{f'(\rho)}{f(\rho)}\frac{x}{\rho}.$$

Итак,

$$\frac{\varphi'(x)}{x\cdot\varphi(x)}=\frac{f'(\rho)}{\rho\cdot f(\rho)},$$

и, аналогично,

$$\frac{\varphi'(y)}{y\cdot\varphi(y)}=\frac{f'(\rho)}{\rho\cdot f(\rho)}.$$

Левые части каждого из этих уравнений зависят от x и y соответственно; они равны между собой, следовательно, это константы.

$$\frac{\varphi'(x)}{x \cdot \varphi(x)} = h,$$

$$\ln \varphi(x) = \frac{hx^2}{2} + \ln C,$$

$$\varphi(x) = C \cdot e^{\frac{hx^2}{2}}.$$

Это рассуждение неверно: мы применили теорему об умножении вероятностей, то есть предположили, что события независимы; иначе говоря, положение пули относительно оси абсцисс не зависит от ее положения относительно оси ординат.

Опишем четыре одинаковых относительно $d\sigma$ области вокруг четырех вершин A, B, C, D прямоугольника со сторонами, параллельными осям координат. Назовем α, β, γ, δ соответствующие вероятности того, что M попадет в каждый из этих элементов поверхности.

Я предположил, что ордината пули была одинакова для B и D, расположенных на параллели к оси OX, что ее абсцисса одинакова для A и B, расположенных на одной параллели к оси Oy; иначе говоря, что $\frac{\alpha}{\beta} = \frac{\gamma}{\delta}$, но эта гипотеза совершенно беспочвенна(4).

17. Максвелл допустил ту же ошибку в теории газа. Рассмотрим газ, образованный очень большим количеством молекул, наделенных разными скоростями; как будут распределены эти скорости у разных молекул?

Выберем трехмерную прямоугольную систему координат и поместим в ее начало вектор, представляющий полярный радиус, широту и долготу скорости молекулы. Оценим вероятность того, что конечная точка M этого вектора находится в малом элементе объема $d\tau$.

Если я естественным образом предполагаю, что скорости практически одинаковы во всех направлениях, то эта вероятность будет представляться как

$$f(r)\, d\tau.$$

Вероятность того, что первая координата лежит между x и $x + dx$, запишется как $\varphi(x)\, dx$; вероятность того, что вторая координата лежит между y и $y+dy$, запишется как $\varphi(y)\,dy$; вероятность того, что третья координата лежит между z и $z+dz$, запишется как $\varphi(z)\,dz$.

Вероятность того, что M находится в малом параллелепипеде со сторонами, параллельными осям координат, равна $f(r)\,dx\,dy\,dz$. Если бы теорема об умножении вероятностей была применима, мы бы сразу получили

$$f(r) = \varphi(x)\varphi(y)\varphi(z),$$

а это неправильно.

18. Задача о выборах. Эта задача допускает изящное решение, принадлежащее г. Андре́. Имеются два кандидата, A и B; хорошо осведомленный избиратель заранее знает, что A будет иметь m, а B — n голосов, где m больше n. Требуется найти вероятность того, что A будет иметь большинство в течение всей процедуры подсчета бюллетеней.

Чтобы оценить число возможных исходов, заметим, что m бюллетеней A и n бюллетеней B могут появляться в таком количестве разных порядков, сколько есть перестановок с повторениями из m букв A и n букв B, то есть

$$\frac{(m+n)!}{m!\,n!}.$$

Я разделю возможные исходы на три группы.

В первую я помещу те исходы, когда A имеет большинство в начале и сохраняет его все время, пусть здесь имеется N_1 благоприятных исходов.

Во вторую группу я помещу все исходы, когда первый бюллетень принадлежит B; то есть A теряет большинство в начале; здесь N_2 благоприятных исходов.

В третью я помещу все исходы, когда A имеет большинство в начале, но затем теряет его, чтобы восстановить к концу; здесь N_3 благоприятных исходов.

Имеем

$$N_1 + N_2 + N_3 = \frac{(m+n)}{m!\,n!},$$

то есть надо подсчитать $\dfrac{N_1}{N_1+N_2+N_3}$.

Оцениваем N_2: первый вскрытый бюллетень подан за B; уберем его, и останется m бюллетеней A и $(n-1)$ бюллетеней B. Число возможных исходов равно

$$\frac{(m+n-1)!}{m!\,(n-1)!},$$

что дает нам величину N_2.

Я хочу показать, что $N_3 = N_2$.

Лемма 1. ** Предположим, что на выборах голоса разделились поровну: подано q бюллетеней за A и q — за B. Допустим также, что в начале A имеет большинство голосов и сохраняет его до последнего бюллетеня, а потом теряет, так как в конце голоса делятся поровну; последний бюллетень, таким образом, содержит имя B. Будем вскрывать их в обратном порядке, тогда B потеряет большинство только с последним бюллетенем.*

В какой-то определенный момент голосования вскрыто a бюллетеней A и b бюллетеней B и известно, что $a > b$, так как большинство голосов у A. Остается вскрыть $q-a$ бюллетеней A и $q-b$ бюллетеней B.

При обратном порядке голосование уже показало бы, что

$$q - b > q - a,$$

т.е. у B было бы большинство.

Доказав лемму, вернемся к нашей задаче.

Рассмотрим комбинацию α из третьей группы:

$$AABAB|BABAA. \tag{α}$$

Здесь A имеет преимущество вплоть до черты, затем, со следующим бюллетенем, теряет его в первый раз.

Пусть слева от черты у A есть q, у B есть $q-1$ бюллетеней, т.е. всего $2q-1$ бюллетеней.

Рассмотрим другую комбинацию, которую назовем производной от α:

$$BABAA|AABAB.$$

Она получается, если взять из α последовательно бюллетени с местами

$$2q,\ 2q+1,\ 2q+2,\ \ldots,\ m+n,\ 1,\ 2,\ \ldots,\ 2q-1,$$

т.е. перенести налево от черты то, что было справа, и наоборот.

Если $2q$-й бюллетень будет по определению первым, где A потеряет большинство, то каждая комбинация α будет иметь единственную производную комбинацию.

Теперь рассмотрим комбинацию β из второй группы; она начинается с B:

$$BABAAABA; \qquad (\beta)$$

к концу большинство имеет A.

Образуем комбинацию β' следующим способом: сначала первый бюллетень с именем B, затем последний бюллетень из β, затем предпоследний и т. д., то есть бюллетени β в обратном порядке вплоть до второго:

$$BABAAABA; \qquad (\beta')$$

Ясно, что сперва B имеет большинство голосов, но к концу теряет его.

Предположим, что $2p$-й бюллетень в первый раз отнимает его у B большинство. В нашем примере он второй по счету. Комбинация β' служит для определения числа p: в нашем примере это единица.

Бюллетень, занимающий в β' $2p$-е место, в β занимает $(m + n + 2 - 2p)$-е место.

Я помещу в β черту перед членом, занимающим это место:

$$BABAAAB|A. \qquad (\beta)$$

Рассмотрим, наконец, следующую комбинацию, которую я обозначу γ и назову производной от β; я начну с бюллетеней, стоявших справа от черты (в примере он только один), а затем возьму те, что были слева:

$$A|BABAAAB. \qquad (\gamma)$$

Бюллетени β взяты в порядке

$$m + n + 2 - 2p, \ldots, m + n, 1, 2, \ldots, m + n - 2p, m + n + 1 - 2p.$$

Я заявляю, что эта производная всегда принадлежит к третьей группе.

Во-вторых, $(m + n + 2 - 2p)$-й бюллетень должен быть A, т. к. в β' он отнимет у B большинство; итак, γ начинается с A.

Кандидат A не будет все время сохранять большинство. Действительно, первые $2p$ бюллетеней γ — это первые $2p$ бюллетеней β', взятые в обратном порядке; а согласно предположению, после вскрытия этих $2p$ бюллетеней оба кандидата имеют равное число голосов.

Итак, в γ кандидат A теряет большинство, и γ принадлежит к третьей группе.

Таким образом, каждая комбинация из второй группы имеет, причем единственную, производную, принадлежащую к третьей группе.

Я заявляю, что если для какой-то комбинации α из третьей группы я образую ее производную β, а затем — производную от β, то я вернусь к α.

Покажем, что q в α соответствует p в β.

Действительно, образуем β':

$$BBABAA|AABA. \tag{β'}$$

Если я беру $2q$ первых бюллетеней из β', то это в точности $2q$ первых бюллетеней из α, вскрытых в обратном порядке, и по лемме B потеряет большинство только к концу; с другой стороны, мы знаем, что в β' кандидат B потеряет большинство только с $2p$-ым бюллетенем; итак,

$$p = q,$$

и производная от β будет α.

Если я составляю комбинацию β, я получаю из нее производную γ; этой же производной от β будет α.

Можно сказать, что комбинации второй группы сопряжены с комбинациями из третьей так, что каждая комбинация будет производной от своей сопряженной из другой группы.

Итак,

$$N_3 = N_2;$$

$$N_2 + N_3 = 2 \cdot \frac{(m+n-1)!}{m!\,(n-1)!},$$

$$\frac{N_2 + N_3}{N_1 + N_2 + N_3} = \frac{2n}{m+n}.$$

Это вероятность того, что A не всегда имеет большинство, а

$$\frac{N_1}{N_1 + N_2 + N_3} = 1 - \frac{2n}{m+n} = \frac{m-n}{m+n}.$$

19. Задача об игральных костях. Бросают n костей и спрашивают, какова вероятность того, что общее число очков равно K.

Предположим сначала, что костей только две; для каждой из них возможны шесть разных исходов, а вместе они дадут тридцать шесть комбинаций:

$$\begin{array}{lll}
1\,1 & & \\
1\,2 & 2\,1 & \\
1\,3 & 2\,2 & 3\,1 \\
1\,4 & 2\,3 & 3\,2 \\
1\,5 & 2\,4 & \cdot\,\cdot \\
1\,6 & 2\,5 & \cdot\,\cdot \\
 & 2\,6 & \cdot\,\cdot \\
 & & \cdot\,\cdot
\end{array}$$

Из этих комбинаций:

одна соответствует числу	$K = 2$,
2 соответствуют числу	$K = 3$,
3 соответствуют числу	$K = 4$,
....	...
1 соответствует числу	$K = 12$.

Вероятность получить 2 равна $\frac{1}{36}$,

Вероятность получить 3 равна $\frac{2}{36}$,

Вероятность получить 4 равна $\frac{3}{36}$,

....

Займемся общей задачей об n костях, здесь полное число возможных исходов 6^n; действительно, пусть $\alpha_1, \alpha_2, \ldots, \alpha_n$ — одна из комбинаций, каждое из чисел α может принимать шесть значений: $1, 2, 3, \ldots, 6$; тогда искомое число — это количество сочетаний с повторениями из 6 букв по n, то есть 6^n.

Общее число очков должно равняться изначально заданному K,

$$\alpha_1 + \alpha_2 + \ldots + \alpha_n = K.$$

Рассмотрим один из 6^n возможных исходов и поставим ему в соответствие одночлен

$$t_1^{\alpha_1} \cdot t_2^{\alpha_2} \cdot \ldots \cdot t_n^{\alpha_n}.$$

Составим из этих одночленов сумму П, меняя $\alpha_1, \alpha_2, \ldots, \alpha_n$ от 1 до 6.

$$\Pi = \sum t_1^{\alpha_1} \cdot t_2^{\alpha_2} \cdot \ldots \cdot t_n^{\alpha_n},$$

что можно переписать как

$$\Pi = (t_1 + t_1^2 + \ldots + t_1^6)(t_2 + t_2^2 + \ldots + t_2^6) \ldots (t_n + t_2^n + \ldots + t_n^6).$$

Это произведение n сомножителей; полагаем в нем

$$t_1 = t_2 = \ldots = t_n = t.$$

Одночлен примет вид

$$t^{\alpha_1+\alpha_2+\ldots+\alpha_n} = t^K,$$

а многочлен П вырождается в

$$(t + t^n + \ldots + t^6)^n.$$

Пусть N — число благоприятных исходов; имеем N одночленов, равных t^K, и их сумма равна Nt^K, и, если составить сумму $\sum Nt^K$ по всем возможным значениям K, то

$$\sum Nt^K = \Pi.$$

Искомая вероятность равна $\frac{N}{6^n}$[1].

Величину N легко сосчитать.

Это n-я степень суммы членов геометрической прогрессии:

$$\Pi = \left(\frac{t - t^7}{1 - t}\right)^n = (t - t^7)^n \cdot (1 - t)^{-n};$$

$(t - t^7)^n$ и $(1 - t)^{-n}$ можно разложить по формуле бинома; найдя произведение двух разложений, я получу коэффициент при t^K, то есть N.

Вернемся к случаю двух костей. Получаем $(t - t^7)^2(1 - t)^{-2}$ и

$$(t - t^7)^2 = t^2 - 2t^8 + t^{14},$$
$$(1 - t)^{-2} = 1 + 2t + 3t^2 + \ldots .$$

[1] Метод вычисления вероятностей таким образом сейчас известен как метод производьных функций.

Оценим коэффициент при t^K в произведении этих двух разложений. Во-первых, K не превосходит 12; затем мы рассмотрим два случая: когда число очков лежит от 2 до 7 и когда оно лежит от 8 до 12.

Если число K равно, самое большее, 7, т. е. $K \leqslant 7$, то в первом разложении не надо брать ни t^8, ни t^{14}, и придется рассматривать только

$$t^2 + 2t^3 + 3t^4 + \ldots + 6t^7.$$

Итак, для значений 2, 3, 4, ... , 7 величина N соответственно равна 1, 2, 3, ... , 6.

Если K равно или превосходит 8, то не надо рассматривать t^{14}, и придется сосчитать только

$$t^2(1 + 2t + \ldots) - 2t^8(1 + 2t + \ldots).$$

Коэффициент при t^K в первом одночлене будет $K - 1$.

Коэффициент при t^8 во втором одночлене будет — 2; при t^9 — 4; и т. д.; при t^K — $2(K - 7)$.

Итак,

$$N = K - 1 - 2(K - 7) = 13 - K.$$

Для $n > 2$ выражения окажутся сложнее.

20. Задача о лотерее. В урне имеется μ шаров, занумерованных от 1 до μ; из них достают n штук; какова вероятность того, что среди них будет K заранее отмеченных шаров?

Пусть n извлеченных шаров имеют разные номера, принимающие значения от 1 до μ.

Если порядок извлечения шаров учитывается, то возможные исходы совпадают по количеству с размещениями из μ букв по n:

$$\frac{\mu!}{(\mu - n)!}.$$

Когда остается $\mu - i + 1$ шаров, у каждого из них столько же шансов быть извлеченным из урны, как и у других; мы предполагаем все выборки равновероятными.

Если не учитывать порядок, то число возможных исходов не превосходит числа сочетаний из μ букв по n,

$$\frac{\mu!}{(\mu - n)!\, n!}$$

Я больше не считаю разными гипотезы, которые отличаются только порядком извлечения шаров. Будут ли все сочетания, как и размещения, равновероятны? Да, т. к. каждому соответствует $n!$ размещений.

Число благоприятных исходов равно числу тех сочетаний, в которые входят K отмеченных шаров; за их вычетом в урне остается $\mu - K$ шаров. Тогда число благоприятных исходов равно числу сочетаний из $\mu - K$ букв по $n - K$:

$$\frac{(\mu - K)!}{(\mu - n)!\,(n - K)!}.$$

Вероятность получить K отмеченных номеров в лотерее равна, таким образом,

$$\frac{(\mu - K)!\,n!}{\mu!\,(n - K)!}.$$

21. Задача об игре с выходящим. Три игрока A, B, C играют по следующим правилам: A и B играют вместе; C не играет. Проигравший выходит и его замещает C. После каждой партии проигравшего вновь заменяют. Игра заканчивается, когда один игрок выигрывает два раза подряд.

Естественно предположить, что игра азартная, и вероятность выиграть партию равна $\frac{1}{2}$ для каждого игрока.

Например, можно получить:

1-я партия AB; A выиграл;

2-я партия AC; если A выиграл, он победил; если C выиграл, возвращается B;

3-я партия BC; если C выиграл, он победитель; если B выиграл, возвращается A;

4-я партия BA; и так далее.

Допустим, что A выиграл первую партию. Требуется найти вероятность победы для каждого из игроков. Пусть эти вероятности равны x, y, z для A, B, C соответственно.

Сначала возможны две гипотезы. Если A выигрывает вторую партию, он победит, и вероятности победы для трех игроков станут $1, 0, 0$.

Если A проигрывает, он уступает место C, B находится в условиях A, C — в условиях B; вероятности станут z, x, y.

Применим теоремы сложения и умножения вероятностей.

Игрок A может победить при двух гипотезах, исключающих друг друга:

1° выиграв рассматриваемую партию;

2° проиграв ее.

Итак, вероятность победы для A

$$x = \frac{1}{2} \times 1 + \frac{1}{2} \times z.$$

Игрок B может победить только одним способом: A проигрывает рассматриваемую партию, и затем B становится победителем.

$$y = \frac{1}{2}x.$$

Аналогично для C вероятность будет

$$z = \frac{1}{2}y.$$

Отсюда $x = \frac{4}{7}$; $y = \frac{2}{7}$; $z = \frac{1}{7}$.

Заметим, что $x + y + z = 1$; когда есть много возможных событий, и таких, что одно и только одно из них непременно наступит, сумма их вероятностей равна 1, но здесь мы не имеем в точности этот случай, т. к. партия может продолжаться бесконечно. Здесь наша сумма равна 1, потому что вероятность бесконечно долгой партии равна 0.

Мы предположили, что A выиграл первую партию. Перейдем к началу игры; перед первой партией C вне игры.

Есть две гипотезы: выиграет A или выиграет B.

Если A, то B выходит, C вступает в игру, и вероятности для каждого игрока станут

$$\frac{4}{7}; \quad \frac{1}{7}; \quad \frac{2}{7}.$$

Если выиграет B, они станут $\frac{1}{7}$, $\frac{4}{7}$, $\frac{2}{7}$.

Игрок A может победить, хоть выиграв первую партию, хоть проиграв. Вероятность первой гипотезы получается умножением вероятности для A выиграть первую партию и вероятности победить после этого, то есть $\frac{1}{2} \times \frac{4}{7}$. Вероятность второй гипотезы аналогично равна $\frac{1}{2} \times \frac{1}{7}$.

Теперь складываем вероятности:
Для A, $\frac{1}{2} \times \frac{4}{7} + \frac{1}{2} \times \frac{1}{7} = \frac{5}{14}$;
Для B точно также, $\frac{5}{14}$;
Для C, без дальнейших вычислений, $\frac{4}{14}$.[5]

Глава 3

Математическое ожидание

22. Пусть в какой-то определенный момент игры игрок имеет вероятность выиграть, равную p; ставка равна α.

По определению математическое ожидание равно $p\alpha$. Принято считать, что если α — проигрыш, то $p\alpha$ отрицательно.

Если разные гипотезы с соответствующими вероятностями p_1, $p_2, \ldots, p_n$ приводят к выигрышам $\alpha_1, \alpha_2, \ldots, \alpha_n$ соответственно, то определением математического ожидания будет сумма

$$p_1\alpha_1 + p_2\alpha_2 + \ldots + p_n\alpha_n.$$

Игра справедливая, когда математическое ожидание одинаково для всех игроков.

Чтобы упростить решение некоторых вопросов, часто договариваются о введении фиктивных игроков, и оценивают их математическое ожидание.

Пусть есть два события, A и B.

Вероятность наступления A равна p_1.

Вероятность наступления B равна p_2.

Вероятность наступления A или B равна p_3.

Вероятность наступления A и B равна p_4.

Здесь

$$p_1 + p_2 = p_3 + p_4.$$

Если эти два события несовместны, получим $p_4 \equiv 0$, т. е. теорему о сложении вероятностей; но я буду предполагать, что это не так. Пусть выигрыш составляет 1 франк, если происходит A; точно также 1 франк, если происходит B.

Полное математическое ожидание при рассмотрении этих двух событий будет $p_1 = p_2$.

Итак, совместны события или нет, полное математическое ожидание будет суммой частичных.

Если бы произошли оба события, игрок бы получил 2 франка.

Более сложный случай: возможно определенное количество событий $A_1, A_2, \ldots, A_q$; вероятность того, что A_i произойдет, равна p_i, вероятность того, что одновременно произойдут A_i и A_k, равна p_{ik}, вероятность того, что одновременно произойдут A_i, A_j и A_k, равна p_{ilk}.

Игроку обещают заплатить 1 франк за каждое произошедшее событие; если их будет n, он получит n франков. Его полное математическое ожидание будет суммой тех, которые обеспечены каждым событием, т. е. $\sum p_i$.

Пусть ему обещают столько франков, сколько будет сочетаний по два в ряду событий. Если происходят 2 события, он получает 1 франк; если происходят 3 события A, B, C, — 3 франка, т. к. есть три сочетания AB, BC, CA'; если n событий — $\dfrac{n\cdot(n-1)}{2}$ франков.

Когда происходят два события A_i и A_k, эта комбинация обеспечивает ему 1 франк, математическое ожидание равно p_{ik}, а полное математическое ожидание —

$$\sum p_{ik}.$$

Если происходят n событий и игроку дают 1 франк за группу из трех, он получит $\dfrac{n(n-1)(n-2)}{1\cdot 2\cdot 3}$ франков, и его математическое ожидание будет

$$\sum p_{ijk}.^{1}$$

Теперь будем играть по следующим правилам: я плачу 1 франк за произошедшее событие; игрок платит мне 1 франк за сочетание двух событий; я ему плачу 1 франк за сочетание трех событий; он мне — 1 франк за сочетание четырех событий и т. д.

События...	1,	2,	3,	...	
Выигрыш игрока	1,	−1,	1,	−1,	...

Его полный выигрыш

$$\sigma = n - \frac{n(n-1)}{1\cdot 2} + \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3} - \ldots,$$

[1] В подлиннике стояли ошибочные формулы:

$$\frac{n(n-1)}{2}\sum p_{ik} \quad \text{и} \quad \frac{n(n-1)(n-2)}{1\cdot 2\cdot 3}\sum p_{ijk}$$ — *Прим. ред.*

если осуществляется n событий.

Если $n = 0$, $\sigma = 0$.

В общем случае получаем

$$1 - \sigma = 1 - n + \frac{n(n-1)}{1 \cdot 2} - \frac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3} + \ldots = (1-1)^n.$$

Итак, при $n > 0$ $\sigma = 1$.

Таким образом, когда ни одно событие не происходит, игрок не получает ничего, когда событий n, он всегда получает 1 франк.

Его математическое ожидание

$$P = \sum p_i - \sum p_{ik} + \sum p_{ijk} - \ldots .$$

Теперь можно сформулировать обобщение теоремы сложения вероятностей: вероятность того, что произойдет хотя бы одно событие, равна

$$\sum p_i - \sum p_{ik} + \sum p_{ijk} - \ldots .$$

23. Задача о совпадении. В урне есть μ шаров, занумерованных от 1 до μ; я вытаскиваю их один за другим, пока урна не опустеет. Совпадение происходит, если в i-й раз я вытаскиваю шар с номером i.

Найдем вероятность того, что будет как минимум одно совпадение.

Во-первых, вероятность совпадения в i-й раз равна $\frac{1}{\mu}$. Действительно, исходов будет столько же, сколько есть перестановок из μ букв, т. е. $\mu!$ Сколько из них благоприятных? Это те, у которых i-й шар стоит на i-м месте; здесь я могу переставлять $\mu - 1$ остальных шаров, значит, будет $(\mu - 1)!$ благоприятных исходов. Вероятность равна

$$p_i = \frac{(\mu - 1)!}{\mu!} = \frac{1}{\mu}.$$

Ищем вероятность совпадения в i-й и k-й раз. Два шара занимают определенное место: если мы переставляем $\mu - 2$ других, то видим, что число благоприятных исходов равно $(\mu - 2)!$; вероятность равна

$$p_{ik} = \frac{1}{\mu(\mu - 1)}.$$

Аналогично

$$p_{ijk} = \frac{1}{\mu(\mu-1)(\mu-2)}.$$

Все p_i равны; всего может быть μ совпадений, поэтому

$$\sum p_i = 1.$$

Чтобы вычислить $\sum p_{ik}$, замечаем, что возможное количество двойных совпадений равно $\frac{\mu(\mu-1)}{1\cdot 2}$.

$$\frac{\mu(\mu-1)}{1\cdot 2}\cdot p_{ik} = \sum p_{ik} = \frac{1}{1\cdot 2}.$$

Аналогично

$$\frac{\mu(\mu-1)(\mu-2)(\mu-2)}{1\cdot 2\cdot 3}p_{ijk} = \sum p_{ijk} = \frac{1}{1\cdot 2\cdot 3}.$$

Если я обещаю игроку столько франков, сколько произошло простых совпадений, его математическое ожидание будет 1; оно будет равно $\frac{1}{2}$, если он получит 1 франк за сочетание двух совпадений; $\frac{1}{6}$, если он получит 1 франк за сочетание трех совпадений, и т. д. А с какой вероятностью произойдет хотя бы одно совпадение? Это то, что мы только что называли P.

$$P = \frac{1}{1!} - \frac{1}{2!} + \frac{1}{3!} - \ldots \pm \frac{1}{\mu!},$$

последний член соответствует случаю μ одновременных совпадений. Слагаемые в P — это μ первых членов в разложении $1 - e^{-1}$, а этот ряд сходится с большой скоростью. Погрешность настолько же мала, насколько велико μ, и для $\mu = 20$ она меньше $\frac{1}{20!}$, т. е. несущественна.

Искомая вероятность равна $1 - e^{-1}$.

24. Вероятные значения. Пусть p_1 — вероятность того, что какая-то данная величина a равна a_1.

Пусть p_2 — вероятность того, что эта величина a равна a_2.

....

Пусть p_n — вероятность того, что эта величина a равна a_n.

Вероятное значение[1] a равно по определению

$$a_1 p_2 + a_2 p_2 + \ldots + a_n p_n;$$

это математическое ожидание игрока, которому обещали сумму, равную a.

Вероятное значение a^2 ни в коем случае не совпадает с квадратом вероятного значения a. По определению вероятное значение a^2 равно $\sum a_i^2 p_i$, тогда как квадрат вероятного значения a равен $\left(\sum a_i p_i\right)^2$.

Пусть b — вероятное значение для a^2, c — для a.

Имеем

$$b - c^2 = \sum p_i \sum a_i^2 p_i - \left(\sum a_i p_i\right)^2,$$

т. к. $\sum p_i = 1$.

Мы перепишем правую часть с помощью тождества

$$\sum X^2 \cdot \sum X'^2 - \left(\sum XX'\right)^2 = \frac{1}{2} \sum (XY' - YX')^2,$$

где все Y — это X, взятые с другими индексами.

Полагаем

$$X = \sqrt{p_i},$$
$$X' = a_i \cdot \sqrt{p_i},$$

откуда $XX' = a_i p_i$ и

$$\sum X^2 = \sum p_i,$$
$$\sum X'^2 = \sum a_i^2 p_i,$$
$$\sum XX'^2 = \sum a_i p_i.$$

Следовательно,

$$b - c^2 = \frac{1}{2} \sum (\sqrt{p_i} a_k \sqrt{p_k} - \sqrt{p_k} a_i \sqrt{p_i})^2,$$
$$b - c^2 = \frac{1}{2} \sum p_i p_k (a_k - a_i)^2.$$

[1]Везде дальше под вероятным значением А. Пуанкаре понимает математическое ожидание величины a. — *Прим. ред.*

Числа p_i и p_k положительны, поэтому $b - c^2$ больше нуля и вероятное значение квадрата a всегда больше квадрата вероятного значения a, исключая случай $a_k = a_i$, т. е. когда a может принимать только одно значение.

Ясно, что вероятное значение суммы двух выражений равно сумме вероятных значений этих выражений. Напротив, вероятное значение произведения двух выражений, вообще говоря, не равно произведению их вероятных значений. Все, что мы только что говорили о вероятном значении квадрата, вполне это доказывает.

Пусть, однако, f и φ — две ***не зависящие*** друг от друга функции.

Пусть p_i — вероятность того, что $f = f_i$, а q_k — того, что $\varphi = \varphi_k$.

Так как эти две функции независимы, вероятность того, что $f = f_i$ и одновременно $\varphi = \varphi_k$, будет $p_i q_k$, в силу теоремы об умножении вероятностей. Тогда вероятное значение произведения $f\varphi$ равно

$$\sum p_i q_k f_i \varphi_k = \left(\sum p_i f_i\right)\left(q_k \varphi_k\right),$$

то есть произведению вероятных значений f и φ.[6]

25. Пусть игроку обещают (как обычно, вынимая без возвращения шары из урны, содержащей μ шаров) платить 1 франк за каждый (локальный) максимум в списке, который получится, если записывать номера шаров в порядке их появления. Чему равно математическое ожидание?

Предположим, максимум есть на i-м шаге; мы вытащили три шара, a, b, c, $(i-1)$-й, i-й, $(i+1)$-й, и, т. к. это максимум, $a < b > c$.

Есть $\mu!$ возможных исходов. Не трогая остальные шары, я переставляю a, b, c между собой; возможны 6 комбинаций, из которых две, a, b, c и c, b, a, будут благоприятны.

В этой группе вероятность максимума равна, таким образом, $\frac{1}{3}$. Между тем есть $\frac{\mu!}{6}$ таких групп, соответствующих i-му шагу, и *для этого шага* математическое ожидание равно $\frac{1}{3}$.

Полное математическое ожидание будет суммой частичных; с другой стороны, если нет специальной договоренности, а я ее не предполагаю, ни первый, ни последний шаг не могут привести к выплате.

Итак, полное математическое ожидание равно $\frac{\mu - 2}{3}$.

26. Пусть у каждого из n игроков есть игральная кость и каждый делает ставку в 1 франк; тот, кому выпадает самое большое число очков, заберет n франков; и если самое большое число выпадет у нескольких игроков, они поделят банк между собой.

Пусть у первого игрока A выпало K очков; каково в этот момент его математическое ожидание?

Вероятность того, что у какого-то другого определенного игрока выпало K очков, равна $\frac{1}{6}$; меньше, чем K очков, выпало с вероятностью $\frac{K-1}{6}$.

Какова вероятность того, что A делит банк с $i-1$ данным игроком? Эти $i-1$ игроков должны иметь K очков, $n-i$ остальных — меньше, чем K. Искомая вероятность равна произведению вероятностей, $\left(\frac{1}{6}\right)^{i-1}$ на $\left(\frac{K-1}{6}\right)^{n-i}$.

Итак, вероятность того, что A делит банк с $i-1$ данным игроком равна $\frac{(K-1)^{n-i}}{6^{n-1}}$. Из $i-1$ можно составить столько же групп, сколько есть сочетаний из $n-1$ буквы по $i-1$, т. е. $\frac{(n-1)!}{(i-1)!(n-i)!}$.

Каждое из этих сочетаний дает игроку A найденную выше вероятность, а соответствующий выигрыш равен $\frac{n}{i}$; математическое ожидание A равно

$$\frac{(n-1)!}{(i-1)!(n-i)!}\cdot\frac{(K-1)^{n-i}}{6^{n-1}}\cdot\frac{n}{i}=\frac{N!}{i!(n-i)!}\cdot\frac{(K-1)^{n-i}}{6^{n-1}}.$$

Следует найти сумму этих математических ожиданий от $i=1$ до $i=n$, в последнем случае банк делится поровну между всеми. Так как

$$\sum\frac{n!}{i!(n-i)!}\frac{(K-1)^{n-i}}{6^{n-1}}$$

есть разложение бинома

$$\frac{(1+(K-1))^n}{6^{n-1}},$$

за исключением члена, соответствующего $i=0$, т. е. $\frac{(K-1)^n}{6^{n-1}}$, математическое ожидание равно

$$\frac{K^n - (K-1)^n}{6^{n-1}}.$$

27. Санкт-Петербургский парадокс. Теория математического ожидания привела к возникновению этого знаменитого парадокса. Поль бросает монету; если выпадает орел, он платит Пьеру 1 франк и партия заканчивается; если решка, монету бросают снова. Если во второй раз выпадает орел, Пьер получает 2 франка и партия заканчивается; если решка — начинается снова. В третий раз либо Пьер получит 4 франка, либо партия продолжится, и т. д. Если решка выпадает n раз подряд, а в $(n+1)$-й раз будет орел, Поль платит 2^n франков.

Какую сумму должен отдать Пьер Полю в начале игры, чтобы игра была справедливой? Иначе говоря, каково математическое ожидание выигрыша Пьера?

Вероятность выпадения орла в первый раз равна $\frac{1}{2}$, соответствующее ожидание тоже $\frac{1}{2}$.

Вероятность выпадения решки в первый раз, а орла — во второй равна произведению вероятностей, т. е. $\frac{1}{4}$, т. к. это независимые события. Математическое ожидание равно $\frac{1}{4} \times 2 = \frac{1}{2}$.

В третий раз оно равно $\frac{1}{8} \times 4 = \frac{1}{2}$.

В n-й раз — $\frac{1}{2^{n+1}} \times 2^n = \frac{1}{2}$.

Все члены ряда равны $\frac{1}{2}$; математическое ожидание выигрыша Пьера бесконечно: он никогда не оплатит слишком дорогое право на участие в игре.

Этот парадокс старались объяснить разными способами. Говорили, что Поль не бесконечно богат: пусть, например, его состояние больше 2^p, но меньше 2^{p+1}; если в $(p+1)$-й раз выпадает орел, Поль должен 2^p франков и сможет заплатить; но если орел выпадет в следующий раз, Поль будет должен 2^{p+1} франков и окажется банкротом. Итак, Пьер может получить только 2^p франков; его математическое ожидание становится

$$\frac{1}{2} + \frac{1}{4} \times 2 + \frac{1}{8} \times 4 + \ldots + \frac{1}{2^{p+1}} \cdot 2^p + \frac{1}{2^{p+2}} \cdot 2^p + \frac{1}{2^{p+3}} \cdot 2^p + \ldots$$

Ряд превращается в

$$\frac{1}{2}+\frac{1}{2}+\ldots+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots;$$

$p+1$ слагаемых превращаются в $\frac{1}{2}$, остальные дают в сумме $\frac{1}{2}$; математическое ожидание будет равно

$$\frac{p+1}{2}+\frac{1}{2}=\frac{p+2}{2}.$$

Если состояние Поля равно, например, одному миллиарду, можно считать $p = 30$, и математическое ожидание Пьера будет $\frac{32}{2} = 16$. Видно, что оно заметно уменьшилось.

Еще можно сказать, что удовольствие выиграть 1000 франков больше для того, у кого пока ничего нет, чем для миллионера; что удовольствие удвоить состояние не зависит от размеров состояния.

Если кто-то обладает состоянием x, удовольствие от выигрыша суммы h измеряется как

$$\log\frac{x+h}{x}.$$

Математическое ожидание заменяется на моральное.

Если p_h — вероятность получить выигрыш h, то моральное ожидание равно

$$\sum p_h \log\frac{x+h}{x},$$

или

$$\frac{1}{2}\log\frac{x+1}{x}+\frac{1}{4}\log\frac{x+2}{x}+\ldots+\frac{1}{2^{n+1}}\log\frac{x+2^n}{x}+\ldots$$

для Санкт-Петербургского парадокса. Это ряд, очевидно, сходится.

28. Разорение игрока. Два игрока, у одного из которых, A, есть m франков, а у другого, B, — n франков, играют партию на 1 франк и продолжают игру, *пока один из них не разорится.* Вероятность наступления этого события будет функцией от m и n, т. е. $\varphi(m, n)$, и если сумма состояний, $m+n=s$, — константа, то φ будет функцией от s и n, т. е. от n. Назовем $\varphi(n)$ вероятность того, что в конце концов разорится B; здесь мы будем предполагать, что условия игры несправедливы.

Если A в каждой партии выигрывает с вероятностью p, то B — с вероятностью $1-p$.

Пусть начинается новая партия. Имеются две гипотезы: выигрывает A, и у B останется $(n-1)$ франк; выиграет B, и у него будет $(n+1)$ франк.

Функция $\varphi(n)$ выражает вероятность того, что B, проиграв эту партию, в конце разорится, т. е. $p\cdot\varphi(n-1)$, а также вероятность того, что B, выиграв эту же самую партию, все равно под конец разорится, т. е. $(1-p)\cdot\varphi(n+1)$.

$$\varphi(n)=p\cdot\varphi(n-1)+(1-p)\cdot\varphi(n+1). \qquad \text{(I)}$$

Это рекурентное соотношение и служит для определения $\varphi(n)$. Кроме того, надо знать граничные условия.

Если $n=0$, то B уже разорен; если $n=s$, $m=0$, то у A ничего нет. Итак, $\varphi(0)=1$, $\varphi(s)=0$.

29. Решим более общее рекурентное уравнение

$$A_k\cdot\varphi(n+K)+A_{k-1}\cdot\varphi(n+K-1)+\ldots+A_1\cdot\varphi(n+1)+A_0\cdot\varphi(n)=0,$$

где все A — постоянные коэффициенты. Это уравнение в конечных разностях, линейное, с постоянными коэффициентами, по решению напоминающее линейные дифференциальные уравнения с постоянными коэффициентами.

Предположим, что найдено K решений $\varphi_1(n)$, $\varphi_2(n)$, ... , $\varphi_k(n)$ таких, что

$$\sum_q A_q\cdot\varphi_i(n+q)=0.$$

Еще одно решение получим, положив

$$\varphi(n)=\alpha_1\varphi_1(n)+\alpha_2\cdot\varphi_2(n)+\ldots+\alpha_k\varphi_k(n).$$

Действительно, умножим предыдущую $\sum$ на α_i и суммируем полученные выражения по i.

$$\sum_i\alpha_i\sum_q A_q\cdot\varphi_i(n+q)=\sum_q A_q\sum_i\alpha_i\cdot\varphi_i(n+q)=0.$$

Если φ_1, φ_2, ... , φ_k линейно независимы, их линейная комбинация будет общим решением. Действительно, предположим, что это не так, тогда

$$\varphi(n) = \alpha_1 \cdot \varphi_1 + \alpha_2\varphi_2 + \ldots + \alpha_k \cdot \varphi_k + \psi.$$

Я хочу выбрать α_1, α_2, ... , α_k так, чтобы выполнялась следующая система:

$$\begin{aligned}
\varphi(0) &= \alpha_1 \cdot \varphi_1(0) + \alpha_2\varphi_2(0) + \ldots + \alpha_k \cdot \varphi_k(0),\\
\varphi(1) &= \alpha_1 \cdot \varphi_1(1) + \alpha_2\varphi_2(1) + \ldots + \alpha_k \cdot \varphi_k(1),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\varphi(K-1) &= \alpha_1 \cdot \varphi_1(K-1) + \alpha_2\varphi_2(K-1) + \ldots + \alpha_k \cdot \varphi_k(K-1),
\end{aligned}$$

Эти K линейных уравнений однозначно определяют $\alpha_1, \alpha_2, \ldots, \alpha_K$ при условии, что определитель системы отличен от нуля, что имеет место, когда $\varphi_1, \varphi_2, \ldots, \varphi_K$ линейно независимы. В этом случае $\psi(0) = \psi(1) = \ldots = \psi(K-1) = 0$, а также верно следующее соотношение:

$$A_k \cdot \psi(n+K) + A_{k-1} \cdot \psi(n+K-1) + \ldots + A_1\psi(n+1) + A_0\psi(n) = 0.$$

Если $n = 0$, все члены, кроме $A_K \cdot \psi(K)$, обнуляются; тогда $\psi(K)$ — нуль.

Если $n = 1$, $\psi(K+1)$ — нуль, и т. д.

Итак, $\psi(n)$ тождественный нуль, и $\varphi(n)$ сводится к $\alpha_1\varphi_1 + \alpha_2\varphi_2 + \ldots + \alpha_k\varphi_k$. Значит, достаточно знать K линейно независимых частных интегралов, чтобы найти общий.

Для нахождения K линейно независимых интегралов я полагаю $\varphi(n) = \beta^n$. Тогда

$$A_K \cdot \beta^{n+K} + A_{K-1} \cdot \beta^{n+K-1} + \ldots + A_0 \cdot \beta^n = 0$$

или

$$A_K\beta^K + A_{K-1} \cdot \beta^{K-1} + \ldots + A_0 = 0,$$

откуда имеем K частных значений β и, следовательно, K частных интегралов.

Существует исключение — случай, когда уравнение относительно β имеет кратные корни, например, двукратный корень $\beta_1 = \beta_2$;

меняя коэффициенты непрерывным образом, мы действительно можем получить два равных корня. Тогда у нас больше не будет K решений.

Пусть β_1^n и β_2^n — решения; $\dfrac{\beta_1^n - \beta_2^n}{\beta_1 - \beta_2}$ — их линейная комбинация и, следовательно, тоже решение. Когда β_1 стремится к β_2, предел, по соображениям непрерывности, все еще будет решением. Этот предел получается дифференцированием по β_1 обоих членов отношения, что дает $\dfrac{n \cdot \beta^{n-1}}{1}$; тогда $n \cdot \beta^{n-1}$ или, если угодно, $n \cdot \beta^n$ будет новым решением.

При трехкратном корне получим, кроме того, $n^2 \cdot \beta_1^n$ и т. д.

30. Применим это правило к нашей задаче, т. е. к уравнению (I) параграфа 28.

Полагаем $\varphi(n) = \beta^n$; получается

$$\beta^n = p \cdot \beta^{n-1} + (1-p) \cdot \beta^{n+1}$$

или

$$\beta = p + (1-p)\beta^2.$$

Это уравнение второго порядка имеет очевидный корень $\beta = 1$; другой корень равен $\dfrac{p}{1-p}$; именно это значение мы будем в дальнейшем называть β. Два решения β^n и 1 дают общее решение $\varphi(n)$:

$$\varphi(n) = a \cdot \beta^n + b.$$

Граничные условия задают нам произвольные константы:

$$1 = a + b,$$
$$0 = a\beta^s + b;$$

откуда

$$a = \frac{1}{1-\beta^s}, \qquad b = \frac{-\beta^s}{1-\beta^s},$$

и

$$\varphi(n) = \frac{\beta^n - \beta^s}{1-\beta^s}.$$

Это выражение не определено при $\beta = 1$.

Когда $\beta = 1$, $p = \frac{1}{2}$; игра будет справедливой. Здесь мы найдем значение $\varphi(n)$ по правилу Лопиталя:

$$\varphi(n) = \frac{n \cdot \beta^{n-1} - s \cdot \beta^{s-1}}{- s \cdot \beta^{s-1}} = \frac{s-n}{s} = \frac{m}{s}.$$ [1]

Необходимо рассмотреть три случая.

1° $\beta > 1$. Игра выгодна для A, $p > 1-p$,

$$\varphi(n) = \frac{\beta^s - \beta^n}{\beta^s - 1}.$$

2° $\beta = 1$. Игра справедливая, $p = 1-p$,

$$\varphi(n) = \frac{s-n}{s}.$$

3° $\beta < 1$. Игра выгодна для B, $p < 1-p$,

$$\varphi(n) = \frac{\beta^n - \beta^s}{1 - \beta^s}.$$

Вероятность разорения B равна $\dfrac{\beta^s - \beta^n}{\beta^s - 1}$.

Чтобы найти вероятность разорения A, меняем местами p и $1-p$; β меняется на $\frac{1}{\beta}$, а n на m. Отсюда находим

$$\frac{\beta^{-s} - \beta^{-m}}{b^{-s} - 1},$$

то есть

$$\frac{1 - \beta^n}{1 - \beta^s}.$$

31. Сумма двух вероятностей

$$\frac{\beta^n - \beta^s}{\beta^s - 1} + \frac{1 - \beta^n}{1 - \beta^s}$$

---

[0]В этом случае 1 является корнем уравнения кратности 2.

равна единице, что не было очевидно *a priori*. Действительно, вероятность того, что партия продолжается бесконечно долгое время, может быть положительна.

Предположим, что s очень велико. Когда $\beta > 1$, β^s очень велико и будет главным членом в $\varphi(n)$; предел $\varphi(n)$ равен 1. При $\beta = 1$ ее предел тоже 1. При $\beta < 1$ β^s стремится к 0 при возрастании s, и предел $\varphi(n)$ равен β^n.

Заключение. Итак, если s велико, а n конечно, мы наверняка разоримся при справедливой или выгодной для противника игре. Но если игра выгодна для нас, вероятность разорения тем меньше, чем больше наше состояние.

Профессиональный игрок, банкир, играет с целым светом, т. е. с бесконечно богатым противником, но условия игры обеспечивают ему преимущество. Напротив, вкладчик, который будет играть бесконечно, *наверняка разорится.*

32. Ж. Бертран вычислил *вероятный момент его разорения.*

Для банкира $\beta < 1$, вероятность банкротства равна β^n; β — это соотношение шансов, благоприятных для вкладчика, к шансам банкира; предположим, $\beta = \frac{19}{20}$, т. е. $\beta = 1 - \frac{1}{20}$.

Пусть n — состояние банкира, если за единицу принять ставку в каждой партии; для этой ставки имеется максимум, поэтому для n тоже есть определенный максимум; пусть $n = 1000$.

Вероятность банкротства равна

$$\left(1 - \frac{1}{20}\right)^{1000} = \left(\left(1 - \frac{1}{20}\right)^{20}\right)^{50},$$

или близка к e^{-50}, очень малой величине.

На этом мы закончим изучение примеров, возникающих в простых задачах комбинаторного анализа.

Глава 4

Теорема Бернулли

33. Теперь мы займемся вопросами, связанными с формулой Стирлинга, теоремой Бернулли и вероятностями событий, возникающих при повторении испытаний.

Предположим, что два противоположных события A и B имеют вероятности p и q соответственно. При каждом испытании обязательно происходит одно из них, и они не могут произойти одновременно; итак,

$$p + q = 1,$$

сумма вероятностей равна вероятности достоверного события.

Испытание повторяют m раз; в каждом испытании происходит одно из двух событий. Так, при бросании кости событием A может быть выпадение 6, событием B — любого другого числа очков; $p = \frac{1}{6}$; $q = \frac{5}{6}$.

Событие A, как и B, произойдет какое-то определенное число раз. Нас интересует вероятность того, что A произойдет α, B — $m - \alpha$ раз.

Предполагаем, что вероятность остается одной и той же при каждом испытании. Для игральной кости вероятность получить 6 очков всегда равна $\frac{1}{6}$. Напротив, для колоды в 32 карты вероятность вытащить короля равна $\frac{1}{8}$ при первом испытании, $\frac{4}{31}$ или $\frac{3}{31}$ при втором, в зависимости от того, вытащили короля в первый раз или нет.

34. Сначала найдем вероятность того, что события следуют в каком-то определенном порядке

$$AABAABBAB;$$

вероятности каждого из событий будут

$$ppqppqqpq,$$

а произведение вероятностей — вероятность того, что эти события происходят одновременно, — равна

$$p^5 \cdot q^4.$$

В общем, вероятность того, что в определенной последовательности произойдут α событий A и $m - \alpha$ событий B, равна

$$p^\alpha \cdot q^{m-\alpha}.$$

Она не зависит от порядка наступления событий.

35. Если нужно, чтобы m испытаний, взятых ***в каком угодно порядке,*** приводили к α событиям A и $m-\alpha$ событиям B, то в силу правила сложения вероятностей искомая вероятность будет суммой стольких слагаемых, равных $p^\alpha \cdot q^{m-\alpha}$, сколько есть перестановок с повторением из α букв A и $m - \alpha$ букв B; это число равно

$$\frac{m!}{\alpha!\,(m-\alpha)!}.$$

Вероятность того, что события наступят в каком угодно порядке, равна

$$u_\alpha = \frac{m!}{\alpha!\,(m-\alpha)!} p^\alpha \cdot q^{m-\alpha};$$

это один из членов разложения $(p+q)^m$.

Если я просуммирую все выражения, которые можно получить, полагая α равным $0, 1, \ldots, m$, я найду

$$\sum u_\alpha = (p+q)^m = 1.$$

Сумма вероятностей всех возможных исходов должна равняться единице, так как достоверно известно, что один и только один из этих исходов произойдет.

36. Которая из этих вероятностей будет самой большой?

Я буду вычислять отношение последующего члена к предыдущему;

$$u_{\alpha+1} = \frac{m}{(\alpha+1)!\,(m-\alpha-1)!} \cdot p^{\alpha+1} q^{m-\alpha-1},$$

$$\frac{u_{\alpha+1}}{u_\alpha} = \frac{\alpha!}{(\alpha+1)!} \frac{(m-\alpha)!}{(m-\alpha-1)!} p \cdot q^{-1} = \frac{m-\alpha}{\alpha+1} \cdot \frac{p}{q}.$$

Так же можно заменить α на $\alpha - 1$:

$$\frac{u_\alpha}{u_{\alpha-1}} = \frac{m-\alpha+1}{\alpha} \cdot \frac{p}{q}.$$

Для того чтобы u_α было максимальным из всех вероятностей, необходимо, чтобы

$$u_{\alpha+1} < u_\alpha > u_{\alpha-1}.$$

Итак,

$$\frac{u_{\alpha+1}}{u_\alpha} < 1 \quad \text{и} \quad \frac{u_\alpha}{u_{\alpha-1}} > 1,$$

то есть

$$\frac{m-\alpha}{\alpha+1} \cdot \frac{p}{q} < 1 \quad \text{и} \quad \frac{m-\alpha+1}{\alpha} \cdot \frac{p}{q} > 1.$$

Эти неравенства можно переписать как

$$\begin{aligned} &(m-\alpha)p < (\alpha+1)q \quad &\text{и} \quad &(m-\alpha+1)p > \alpha q,\\ &mp - \alpha p < \alpha q + q \quad &\text{и} \quad &mp - \alpha p + p > \alpha q. \end{aligned}$$

Так как

$$\alpha p + \alpha q = \alpha(p+q) = \alpha,$$

получаем

$$mp < \alpha + q \quad \text{и} \quad mp > \alpha - p.$$

Мы приходим к неравенствам

$$mp + p > \alpha > mp - q.$$

Отсюда находим верхний и нижний пределы α. Разница между ними равна $p + q = 1$; значит, α лежит между двумя числами, вообще говоря, дробными, которые отличаются на единицу, и, так как α — целое число, эти два предела полностью определяют его.

Исключение составляет случай целого $mp + p$; тогда $mp - q$ тоже целое. Здесь возможны колебания при выборе α, так как два последовательных члена в разложении $(p+q)^m$ равны между собой.

Если m очень велико, отношение $\frac{\alpha}{m}$ лежит между $p + \frac{1}{m}$ и $p - \frac{1}{m}$; то есть $\frac{\alpha}{m}$ близкó к p.

Это одна из формулировок *теоремы Бернулли.*

Если я выбрал α так, чтобы u_α было как можно бóльшим, отношение количества событий A к количеству событий B будет почти совпадать с отношением p и q.

37. Какова вероятность того, что α удалено от mp на расстояние h? Пусть

$$\alpha - mp = h.$$

Я называю h уклонением и буду искать вероятное значение его абсолютной величины, так же как и вероятное значение его квадрата.

Займемся оценкой вероятного значения h, вероятного значения $|h|$ —модуля h — и вероятного значения h^2.

Я буду рассматривать вероятное значение какой-то величины M,[1] т. е. $\sum M \cdot u_\alpha$.

$$\sum M \cdot u_\alpha = \sum M \cdot \frac{m!}{\alpha!\,(m-\alpha)!} p^\alpha \cdot q^{m-\alpha};$$

вторая часть равенства — многочлен с целыми коэффициентами, однородный, имеющий степень m относительно p и q; я обозначу его $F(p;\,q)$.

Попытаемся выразить вероятное значение произведения $M\alpha$, т. е.

$$\sum M \cdot \alpha \frac{m!}{\alpha!(m-\alpha)!} p^\alpha \cdot q^{m-\alpha}.$$

Мы перешли от одного выражения к другому, умножая на α все слагаемые; дифференцируя $p^\alpha \cdot q^{m-\alpha}$ по p, мы бы имели $\alpha \cdot p^{\alpha-1} \cdot q^{m-\alpha}$; итак, вероятное значение $M\alpha$ равно

$$\frac{p \cdot dF}{dp}.$$

Числа p и q не будут независимы, так как их сумма равна 1. Мы взяли производную так, словно они независимы, ведь мы дифференцировали по p, считая q константой. Более того, M может зависеть от p, потому что h зависит от p.

Я избегаю этой путаницы следующим образом.

Вместо p и q я ввожу две вспомогательные переменные, x и y, и рассматриваю функцию

$$F(x,\,y) = \sum M \frac{m!}{\alpha!\,(m-\alpha)!} x^\alpha \cdot y^{m-\alpha}.$$

[1] Величина M зависит от α. — ***Прим. ред.***

Даже если M зависит от p, я не заменяю p и q на x и y; тогда вероятное значение M равно $F(p;\, q)$, а

$$\sum M\alpha \frac{m!}{\alpha!\,(m-\alpha)!} x^{\alpha} \cdot y^{m-\alpha}$$

равно $x \cdot \dfrac{dF}{dx}$.

Мы заменим x и y на p и q после дифференцирования.

38. Применим предыдущие рассуждения к нашей задаче.

Пусть сначала $M = 1$;

$$F(x,\, y) = \sum \frac{m!}{\alpha!\,(m-\alpha)!} x^{\alpha} \cdot y^{m-\alpha} = (x+y)^m;$$

если теперь $x = p,\ y = q$, то

$$F(p,\, q) = (p+q)^m = 1,$$

и действительно, вероятное значение единицы равно 1.

Чтобы найти вероятное значение α, я дифференцирую $F(x,\, y)$ по x и умножаю на x, что дает мне $mx \cdot (x+y)^{m-1}$; затем я полагаю $x = p$, $y = q$. Вероятное значение α равно mp.

Чтобы найти вероятное значение α^2, я дифференцирую член $mx \cdot (x+y)^{m-1}$ по x, затем умножаю на x; это дает мне сначала

$$m(x+y)^{m-1} + m(m-1)x(x+y)^{m-2},$$

затем

$$mx(x+y)^{m-1} + m(m-1)x^2(x+y)^{m-2}.$$

Полагая $x = p$ и $y = q$, получаю выражение для вероятного значения α^2:

$$mp + m^2p^2 - mp^2.$$

Найдем теперь вероятные значения h и h^2. Вероятное значение h равно вероятному значению α минус вероятное значение mp, т. е. $mp - mp = 0$.

Итак, вероятное значение уклонения равно 0.

Величина h^2 равна

$$h^2 = \alpha^2 - 2mp\alpha + m^2p^2;$$

тогда ее вероятное значение будет равно

$$(mp - m^2p^2 - mp^2) - 2mp \cdot mp + m^2p^2,$$

или

$$mp(1 - p),$$

то есть mpq.

Вероятное значение квадрата h равно mpq.

Заодно мы убедились, что оно больше квадрата вероятного значения h, равного нулю.

39. Перейдем к вероятному значению модуля h; найдем сначала, каким будет математическое ожидание для игрока, которому платят единичную сумму, если уклонение положительно, и 0, если оно отрицательно[1].

В сумме $\sum u_\alpha$ надо ограничиваться теми слагаемыми, чье уклонение положительно. Пусть u_β — последний член $\sum u_\alpha$, для которого уклонение положительно; имеем

$$\beta > mp \quad \text{и} \quad \beta - 1 < mp,$$

и математическое ожидание для нашего игрока будет равно

$$p^m + \frac{m}{1}p^{m-1}q + \ldots + \frac{m!}{\beta!\,(m-\beta)!}p^\beta \cdot q^{m-\beta},$$

то есть $F(p;\, q)$, если положить

$$F(x,\, y) = x^m + \frac{m}{1}x^{m-1}y + \ldots + \frac{m!}{\beta!\,(m-\beta)!}x^\beta \cdot y^{m-\beta}.$$

Если теперь предположить, что игроку предложили сумму α, его математическое ожидание будет $x \cdot \dfrac{dF}{dx}$ (заменяем $x = p,\ y = q$ после дифференцирования). Наконец, вероятное значение функции, равной h для $h > 0$ и 0 для $h < 0$, будет математическим ожиданием для игрока, которому предлагают $\alpha - mp$, если $\alpha > mp$; т. е.

$$x \cdot \frac{dF}{dx} - mpF.$$

[1]Фактически здесь вычисляется математическое ожидание индикатора события, состоящего в том, что уклонение положительно. — *Прим. ред.*

Функция F — однородный многочлен степени m от x и y; тогда

$$mF = x\frac{dF}{dx} + y\frac{dF}{dy}.$$

Вышеупомянутое математическое ожидание превращается в

$$x\frac{dF}{dx} - px\frac{dF}{dx} - py\frac{dF}{dy}$$

или

$$xq\frac{dF}{dx} - yp\frac{dF}{dy},$$

если заменить $x = p$, $y = q$ после дифференцирования. Имеем сначала

$$pq\left(\frac{dF}{dx} - \frac{dF}{dy}\right);$$

с другой стороны,

$$\frac{dF}{dx} = m \cdot x^{m-1} + m(m-1)x^{m-2} \cdot y + \ldots + \frac{m!}{\beta!(m-\beta)!}\beta x^{\beta-1} \cdot y^{m-\beta},$$

$$\frac{dF}{dy} = m \cdot x^{m-1} + m(m-1)x^{m-2} \cdot y + \ldots$$

В сумме $\frac{dF}{dx}$ на одно слагаемое больше, это последнее слагаемое и представляет собой $\frac{dF}{dx} - \frac{dF}{dy}$, оно равно

$$\frac{m!}{\beta!\,(m-\beta)!} \cdot p^{\beta} \cdot q^{m-\beta} \cdot \frac{\beta}{p}.$$

Итак, мы получаем математическое ожидание нашего игрока:

$$p \cdot q \cdot \frac{m!}{\beta!\,(m-\beta)!} \cdot p^{\beta} \cdot q^{m-\beta} \cdot \frac{\beta}{p} = \frac{m!}{\beta!\,(m-\beta)!} \cdot p^{\beta} \cdot q^{m-\beta} \cdot \beta \cdot q.$$

Это произведение βq и u_{β}, последнего слагаемого, соответствующего положительному уклонению.

40. Пусть игроку предлагают сумму, равную абсолютной величине уклонения: пусть E — его математическое ожидание при условии, что ему должны платить только за положительное уклонение, E' — при условии, что ему должны платить только за отрицательное уклонение.

Вероятное значение h равно $E - E'$.

Так как вероятное значение h равно нулю, то

$$E - E' = 0,$$

и вероятное значение $|h|$ равно $2E$.

41. Итак, вероятное значение уклонения h, взятого с учетом знака, равно нулю; вероятное значение h^2 равно mpq, для $|h|$ оно равно $2E$ или

$$2\beta q \frac{m!}{\beta!\,(m-\beta)!} \cdot p^{\beta} \cdot q^{m-\beta}.$$

Величина β соответствует последнему слагаемому, чье уклонение положительно, она мало отличается от mp.

Это выражение можно приблизительно оценить как

$$2mpq \frac{m!}{\beta!\,(m-\beta)!} \cdot p^{\beta} \cdot q^{m-\beta}.$$

Оно гораздо больше всех остальных, но, вообще говоря, довольно мало.

Вероятное значение $|h|$ намного меньше mpq. Согласно уже сделанному замечанию, квадрат вероятного значения $|h|$ меньше вероятного значения h^2: вероятное значение $|h|$ достоверно меньше, чем $\sqrt{mpq}$.

Глава 5
Применение формулы Стирлинга

42. Сейчас я покажу, как можно найти приблизительное значение максимального слагаемого u_α, вероятное значение $|h|$ и т. д.

Вычисление этих приблизительных величин связано с *формулой Стирлинга.*

Имеем
$$n! = \int\limits_0^\infty x^n \cdot e^{-n}\, dx = \Gamma(n+1),$$

то есть функцию Эйлера; этот интеграл сохраняет смысл и при положительном, но нецелом n. Если положить

$$\Gamma(n+1) = n^n \cdot e^{-n} \cdot \sqrt{2\pi n},^1$$

то отношение левой части равенства к правой стремится к единице, когда n неограниченно возрастает.

Я не буду приводить доказательства в общем случае и ограничусь тем, что касается целого n.

Формула Стирлинга служит для вычисления факториала целого числа
$$n! = n^n \cdot e^{-n} \cdot \sqrt{2\pi n};$$

это *асимптотическая* формула. Абсолютная погрешность, которую допускают, беря правую часть равенства вместо левой, неограниченно возрастает вместе с n, но относительная погрешность стремится к нулю.

Из этого следует, что абсолютная погрешность логарифма от $n!$ стремится к 0.

Сначала я могу написать

$$n! = n^n e^{-n} \sqrt{n} \cdot F(n);$$

---

[1] Равенство в этой формуле понимается как эквивалентность правой и левой частей при $n \to +\infty$.

я покажу, что $F(n)$ стремится к конечному пределу C при бесконечно большом n; так что
$$n! = C \cdot n^n e^{-n}\sqrt{n}.$$

43. Рассмотрим произведение
$$\frac{F(2)}{F(1)} \cdot \frac{F(3)}{F(2)} \cdots \frac{F(n+1)}{F(n)}.$$

Это бесконечное произведение сходится, или, что то же самое, ряд с общим членом
$$\log\frac{F(n+1)}{F(n)} \text{ — сходящийся}^1.$$

Имеем
$$(n+1)! = (n+1)^{n+1} \cdot e^{-(n+1)}\sqrt{n+1} \cdot F(n+1),$$
$$\frac{F(n+1)}{F(n)} = \frac{(n+1)!}{n!} \cdot \frac{n^n e^{-n}\sqrt{n}}{(n+1)^{n+1} \cdot e^{-(n+1)}\sqrt{n+1}},$$
$$\frac{F(n+1)}{F(n)} = \frac{n+1}{(n+1)^{n+1}} \cdot n^n \cdot e \cdot \sqrt{\frac{n}{n+1}} = e\left(\frac{n}{n+1}\right)^{n+\frac{1}{2}},$$
$$\log\frac{F(n+1)}{F(n)} = 1 - \left(n+\frac{1}{2}\right) \cdot \log\left(1+\frac{1}{n}\right).$$

Правая часть преобразуется в ряд с помощью разложения $\log(1+x)$:
$$1 - \left(n+\frac{1}{2}\right)\left(\frac{1}{n} - \frac{1}{2n^2} + \frac{1}{3n^3} - \ldots\right)$$
или
$$1 - 1 + \frac{1}{2n} - \frac{1}{3n^2} + \ldots - \frac{1}{2n} + \frac{1}{4n^2} - \ldots,$$
то есть
$$-\frac{1}{12n^2} + \ldots.$$

---

[1]Здесь и далее рассматривается только функция натурального логарифма. — *Прим. ред.*

Если я назову u_n первый член этого последнего ряда,

$$u_n = -\frac{1}{12n^2},$$

то

$$\lim n^2 u_n = -\frac{1}{12}.$$

По признаку Гаусса ряд сходится.

44. Остается вычислить величину C.

Воспользуемся формулой Валлиса:

$$\frac{\pi}{2} = \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\ldots\cdot\frac{2n}{2n-1}\cdot\frac{2n}{2n+1}\cdot\ldots;$$

отсюда следует, что при n, бóльших любой наперед заданной величины,

$$\sqrt{\frac{\pi}{2}} = \frac{2\cdot 4\cdot\ldots\cdot 2n}{1\cdot 3\cdot\ldots\cdot(2n-1)\sqrt{2n+1}}.$$

В числителе стоят первые n четных чисел, в знаменателе — первые n нечетных; я умножаю верх и низ на $2\cdot 4\cdot\ldots\cdot 2n$; получается

$$\frac{(2\cdot 4\cdot\ldots\cdot 2n)^2}{(2n)!\sqrt{2n+1}} = \frac{2^{2n}\cdot(n!)^2}{(2n)!\sqrt{2n+1}}.$$

При бесконечно больших n эта дробь имеет своим пределом $\sqrt{\frac{\pi}{2}}$. Если мы заменим здесь факториалы их приблизительными значениями при очень больших n, то дробь превратится в

$$\frac{2^{2n}\cdot n^{2n}\cdot e^{-2n}\cdot C^2\cdot n}{(2n)^{2n}\cdot e^{-2n}\cdot C\cdot\sqrt{2n}\cdot\sqrt{2n+1}}$$

или

$$C\sqrt{\frac{n^2}{2n(2n+1)}};$$

и, когда n неограниченно растет, станет равной $\frac{c}{2}$.

Этот предел должен совпадать с предыдущим, $\frac{\pi}{2}$, отсюда

$$C = \sqrt{2\pi}.$$

45. Асимптотическая оценка u_α. Мы уже видели, что когда два противоположных события, A и B, имеют вероятности p и q соответственно, такие, что $p+q=1$, вероятность того, что из m событий произойдут α равных A и $m-\alpha$ равных B, будет

$$u_\alpha = \frac{m!}{\alpha!\,(m-\alpha)!}\,p^\alpha \cdot q^{m-\alpha}.$$

Вычисляем приблизительное значение u_α, предполагая m очень большим, и, более того,

$$\alpha = mp + \lambda\sqrt{m},$$

где

$$\frac{\lambda\sqrt{m}}{mp} < 1.$$

Число $\lambda\sqrt{m}$ очень велико, если m очень велико, но мы будем считать λ конечным, и, следовательно, очень малой — относительную погрешность, когда вместо α берется mp; то есть $\frac{\alpha}{mp}$ близко к 1,

$$\frac{\alpha}{mp} = 1 + \frac{\lambda}{p\sqrt{m}}.$$

В качестве следствия имеем

$$m - \alpha = mq - \lambda\sqrt{m},$$

так как $p+q=1$.

В выражении u_α я заменяю каждый факториал его значением, вычисленным с помощью формулы Стирлинга:

$$u_\alpha = \frac{m^m e^{-m}\sqrt{2\pi m}p^\alpha q^{m-\alpha}}{\alpha^\alpha \cdot e^{-\alpha}\sqrt{2\pi\alpha}(m-\alpha)^{m-\alpha}\cdot e^{-(m-\alpha)}\sqrt{2\pi(m-\alpha)}},$$

$$u_\alpha = \frac{m^m p^\alpha q^{m-\alpha}}{\alpha^\alpha\cdot(m-\alpha)^{m-\alpha}}\sqrt{\frac{m}{2\pi\alpha(m-\alpha)}}.$$

Объединяем члены с показателем α и с показателем $m-\alpha$:

$$u_\alpha = \left(\frac{mp}{\alpha}\right)^\alpha \cdot \left(\frac{mq}{m-\alpha}\right)^{m-\alpha}\sqrt{\frac{m}{2\pi\alpha(m-\alpha)}}.$$

Между тем

$$\frac{m}{\alpha(m-\alpha)}=\frac{m}{(mp+\lambda\sqrt{m})(mq-\lambda\sqrt{m})},$$

$\frac{\alpha}{mp}$ стремится к единице, как и $\frac{m-\alpha}{mq}$; корень, входящий в u_α, имеет пределом

$$\sqrt{\frac{m}{2\pi\, mp\, mq}} \quad \text{или} \quad \frac{1}{\sqrt{2\pi\, mpq}}.$$

Мы можем записать

$$\log u_\alpha = \log\frac{1}{\sqrt{2\pi\, mpq}} - \alpha\log\frac{\alpha}{mp} - (m-\alpha)\log\frac{m-\alpha}{mq}$$

или

$$\log u_\alpha = \log\frac{1}{\sqrt{2\pi\, mpq}} - (mp+\lambda\sqrt{m})\log\left(1+\frac{\lambda}{p\sqrt{m}}\right) -$$
$$-(mq-\lambda\sqrt{m})\log\left(1-\frac{\lambda}{q\sqrt{m}}\right).$$

При очень большом m мы можем разложить логарифмы от $1+\frac{\lambda}{p\sqrt{m}}$ и от $1-\frac{\lambda}{q\sqrt{m}}$ по формуле разложения $\log(1+x)$. Итак,

$$(mp+\lambda\sqrt{m})\log\left(1+\frac{\lambda}{p\sqrt{m}}\right)$$

становится

$$(mp+\lambda\sqrt{m})\left(\frac{\lambda}{p\sqrt{m}}-\frac{\lambda^2}{2p^2m}+\frac{\lambda^3}{3p^3m\sqrt{m}}-\ldots\right).$$

Сейчас я ищу асимптотическую оценку u_α, т. е. такую величину, чтобы отношение u_α к этой величине стремилось к 1, когда m неограниченно возрастает; тогда в предыдущем произведении я смогу пренебречь всеми членами, стремящимися к 0 и содержащими m или $\sqrt{m}$ в знаменателе.

Останется

$$\lambda\sqrt{m}-\frac{\lambda^2}{2p}+\frac{\lambda^2}{p} \quad \text{или} \quad \lambda\sqrt{m}+\frac{\lambda^2}{2p}.$$

Отсюда я выведу значение произведения

$$(mq - \lambda\sqrt{m}) \cdot \log\left(1 - \frac{\lambda}{q\sqrt{m}}\right),$$

меняя в предыдущем результате λ на $-\lambda$, p на q; сумма этих произведений равна

$$\left(\lambda\sqrt{m} + \frac{\lambda^2}{2p}\right) + \left(-\lambda\sqrt{m} + \frac{\lambda^2}{2q}\right)$$

или $\frac{\lambda^2}{2}\left(\frac{1}{p} + \frac{1}{q}\right)$, то есть $\frac{\lambda^2}{2pq}$.

Итак,

$$\log u_\alpha = \log\frac{1}{\sqrt{2\pi\, mpq}} - \frac{\lambda^2}{2pq}.$$

Заменяя логарифмы самими числами, получаем приблизительное значение u_α:

$$u_\alpha = \frac{e^{-\frac{\lambda^2}{2pq}}}{\sqrt{2\pi\, mpq}}.$$

Когда m бесконечно растет, отношение u_α к предыдущему выражению стремится к единице.

46. Сначала я прослежу, что происходит с максимальным членом.

Максимальное u_α получается, если придать α значение, мало отличающееся от mp; итак, λ равно нулю, и значение максимального выражения будет $\frac{1}{\sqrt{2\pi\, mpq}}$.

Это выражение убывает при росте m. Нельзя считать, что при бесконечно больших m искомая вероятность приближается к вероятности достоверного события; напротив, она стремится к нулю.

Это и есть теорема Бернулли, формулировку которой мы сейчас уточним.

47. Какова вероятность того, что λ заключено между λ и $\lambda + d\lambda$?

Я считаю $d\lambda$ очень малым; однако $d\lambda\sqrt{m}$ — целое число, что означает, что λ имеет порядок $\frac{1}{\sqrt{m}}$.

Если я придаю λ очень малое приращение, значение экспоненты не изменится и u_α будет практически константой.

Искомая вероятность — это сумма членов, у которых α меняется от α до $\alpha + K$, α и $\alpha + K$ определяются как

$$\alpha = mp + \lambda\sqrt{m},$$
$$\alpha + K = mp + (\lambda + d\lambda)\sqrt{m},$$

то есть

$$K = d\lambda\sqrt{m}.$$

Число α должно находиться в таких пределах, чтобы

$$\lambda + d\lambda \geqslant \frac{\alpha - mp}{\sqrt{m}} > \lambda,$$

то есть должно равняться одному из чисел

$$\alpha + 1,\ \alpha + 2,\ \ldots,\ \alpha + K.$$

Сумма вероятностей будет равна

$$u_{\alpha+1} + u_{\alpha+2} + \ldots + u_{\alpha+K}.$$

Есть K слагаемых, практически равных u_α; искомая вероятность равна

$$\frac{K \cdot e^{-\frac{\lambda^2}{2pq}}}{\sqrt{2\pi\, mpq}};$$

заменяя K на $d\lambda\sqrt{m}$, получим

$$\frac{d\lambda e^{-\frac{\lambda^2}{2pq}}}{\sqrt{2\pi\, mpq}}.$$

48. Пусть $h^2 = \dfrac{1}{2pq}$, тогда мы возвращаемся к рассмотрению следующего выражения:

$$\frac{h \cdot dx \cdot e^{-h^2x^2}}{\sqrt{\pi}}.$$

Это выражение представляет собой вероятность того, что некая величина x лежит между x и $x + dx$; вероятность того, что она лежит между x_0 и x_1, будет равна

$$\int\limits_{x_0}^{x_1} \frac{h \cdot dx \cdot e^{-h^2x^2}}{\sqrt{\pi}};$$

соответственно, вероятность того, что x меняется от $-\infty$ до $+\infty$, равна

$$\int_{-\infty}^{+\infty} \frac{h \cdot dx \cdot e^{-h^2x^2}}{\sqrt{\pi}}.$$

Обозначая $hx = y$, превращаем этот последний интеграл в

$$\int_{-\infty}^{+\infty} \frac{dy\, e^{-y^2}}{\sqrt{\pi}}.$$

Это известный интеграл, чья величина равна 1.

49. Остановимся на некоторых следствиях этих вычислений. Вероятность того, что λ лежит между $-\infty$ и $+\infty$, равна 1, что кажется тавтологией. Но это заключение не было таким уж очевидным: формула, которой мы пользовались, была приближенной и верной только при λ малых по сравнению с $\sqrt{m}$.

Пусть сначала

$$\alpha = mp + \lambda\sqrt{m};$$

вероятность того, что λ лежит между λ_0 и λ_1, стремится, согласно предыдущим рассуждениям, к

$$\int_{\lambda_0}^{\lambda_1} \frac{d\lambda\, e^{-\frac{\lambda^2}{2pq}}}{\sqrt{2\pi pq}},$$

когда m неограниченно растет. С другой стороны, если λ_0 и λ_1 бесконечно удалены друг от друга, интеграл стремится к единице.

Теперь считаем

$$\alpha > mp(1 - \varepsilon)$$

и

$$\alpha < mp(1 + \varepsilon).$$

Пусть $F(\varepsilon, m)$ — вероятность того, что так оно и есть. Я заявляю, что для любого данного ε могу взять m настолько большим, чтобы разность $1 - F(\varepsilon, m)$ была меньше заданной величины η. Выбираем сначала число λ настолько большое, чтобы разность

$$1 - \int\limits_{-\lambda}^{+\lambda} \frac{d\lambda}{\sqrt{2\pi pq}} \cdot e^{-\frac{\lambda^2}{2pq}}$$

была меньше, чем $\frac{\eta}{2}$. Это возможно, т. к. интеграл стремится к 1 при бесконечно больших λ. Я возьму m настолько большим,

1° чтобы

$$\lambda < \varepsilon p\sqrt{m},$$

откуда

$$F(\varepsilon, m) > F\left(\frac{\lambda}{p\sqrt{m}}, m\right);$$

2° чтобы разность

$$\left| F\left(\frac{\lambda}{p\sqrt{m}}, m\right) - \int\limits_{-\lambda}^{+\lambda} \frac{d\lambda}{\sqrt{2\pi pq}} \cdot e^{-\frac{\lambda^2}{2pq}} \right| < \frac{\eta}{2};$$

что возможно, т. к. для фиксированного λ предел вероятности

$$F\left(\frac{\lambda}{p\sqrt{m}}, m\right) \quad \text{при} \quad m = \infty$$

представляется через интеграл $\int\limits_{-\lambda}^{+\lambda}$.

Итак, получим

$$1 - F(\varepsilon, m) < \eta.$$

Иначе говоря, *вероятность того, что α будет находиться между $mp(1-\varepsilon)$ и $mp(1+\varepsilon)$, как бы ма́ло ни было ε, стремится к единице, когда m бесконечно растет*(7).

50. Можно спросить, каково вероятное значение x^n, равное по определению

$$\int\limits_{-\infty}^{+\infty} \frac{hx^n\,dx}{\sqrt{\pi}} e^{-h^2x^2}.$$

Если n нечетно, этот интеграл равен нулю.

Если n четно, он совпадает с

$$2\int\limits_{0}^{\infty} \frac{hx^n\,dx}{\sqrt{\pi}} e^{-h^2x^2}.$$

Мы нашли вероятное значение абсолютной величины уклонения; найдем вероятное значение абсолютной величины x^n, то есть $|x^n|$. Это

$$\int\limits_{-\infty}^{+\infty} \frac{h\cdot|x^n|\,dx}{\sqrt{\pi}} e^{-h^2x^2}.$$

Функция под знаком $\int$ четная; тогда этот интеграл равен

$$2\int\limits_{0}^{+\infty} \frac{h\cdot|x^n|\,dx}{\sqrt{\pi}} e^{-h^2x^2}.$$

51. Итак, во всех случаях мы возвращаемся к одному и тому же интегралу, а сам он сводится к интегралам Эйлера.

Считаем

$$h^2x^2 = y,$$

откуда

$$h\,dx = \frac{1}{2}y^{-\frac{1}{2}}\,dy.$$

Вышеуказанный интеграл превращается в

$$2\int\limits_{0}^{+\infty} \frac{1}{2}\frac{dy\,y^{-\frac{1}{2}}\cdot y^{\frac{n}{2}}}{h^n\cdot\sqrt{\pi}} e^{-y}$$

или

$$\frac{1}{h^n\sqrt{\pi}}\int\limits_0^{+\infty} dy\cdot y^{\frac{n-1}{2}}e^{-y}.$$

Это интеграл Эйлера:

$$\frac{\Gamma\left(\frac{n+1}{2}\right)}{h^n\cdot\sqrt{\pi}}.$$

Если n четно и равно 2μ, то

$$\Gamma\left(\frac{n+1}{2}\right)=\Gamma\left(\frac{1}{2}+\mu\right)=\Gamma\left(\frac{1}{2}\right)\cdot\frac{1}{2}\cdot\frac{3}{2}\cdot\ldots\cdot\left(\mu-\frac{1}{2}\right),$$

и, так как

$$\Gamma\left(\frac{1}{2}\right)=\sqrt{\pi},$$

вероятное значение $|x^n|$ равно

$$\frac{1}{h^n}\cdot\frac{1\cdot 3\cdot\ldots\cdot(2\mu-1)}{2^\mu}.$$

Если n нечетно и равно $2\mu+1$, вероятное значение $|x^n|$ равно

$$\frac{\Gamma(\mu+1)}{h^n\sqrt{\pi}}$$

или

$$\frac{\mu!}{h^n\sqrt{\pi}}.$$

Полагаем $n=0$; вероятное значение $|1|$ равно 1.

Полагаем $n=2$; вероятное значение $|x^2|$ равно $\frac{1}{2h^2}$.

Мы нашли вероятное значение $(\alpha-mp)^2$ и получили mpq. Теперь мы найдем вероятное значение λ^2, т. е. квадрата от

$$\frac{\alpha-mp}{\sqrt{m}};$$

оно должно совпадать с pq.

Это проверяется без труда, т. к.

$$h^2 = \frac{1}{2pq}.$$

Вероятное значение $|x|$ равно

$$\frac{1}{h\sqrt{\pi}}.$$

Мы знаем вероятное значение $|\alpha - mp|$; это произведение $2mpq$ и максимального члена суммы $\sum u_\alpha$, то есть $2mpq\, u_\alpha$, где α мало отличается от mp.

Чтобы найти этот максимальный член, достаточно положить $\lambda = 0$; получаем

$$\frac{1}{\sqrt{2\pi\, mpq}}.$$

Итак, вероятное значение $|\alpha - mp|$ равно

$$\frac{2mpq}{\sqrt{2\pi\, mpq}},$$

то есть $\dfrac{\sqrt{2mpq}}{\sqrt{\pi}}$.

Отсюда выводится вероятное значение $|\lambda|$: деля на $\sqrt{m}$, находим

$$\sqrt{\frac{2pq}{\pi}} = \frac{1}{h\sqrt{\pi}}.$$

52. Мы нашли асимптотическую оценку члена

$$u_\alpha = \frac{m!}{\alpha!\,(m-\alpha)!}\, p^\alpha q^{m-\alpha},$$

считая $\alpha = mp + \lambda\sqrt{m}$, m очень велико.

Эта асимптотическая оценка u_α равна

$$\frac{e^{\frac{-\lambda^2}{2pq}}}{\sqrt{2\pi\, mpq}}.$$

Теперь будем искать асимптотическую оценку u_α для $\alpha = m\varepsilon$.

Сравнивая два выражения для α, получаем

$$\lambda = (\varepsilon - p)\sqrt{m}.$$

Возникает искушение принять за асимптотическую оценку выражение

$$\frac{e^{-\frac{m(\varepsilon-p)^2}{2pq}}}{\sqrt{2\pi\, mpq}};$$

но это выражение не будет точным.

Точное значение u_α равно

$$u_\alpha = \frac{m!}{\alpha!\,(m-\alpha)!}p^\alpha \cdot q^{m-\alpha}.$$

Если m и α очень велики, асимптотическое выражение u_α равно

$$\frac{m^m e^{-m}\sqrt{2\pi m}}{\alpha^\alpha \cdot e^{-\alpha}\sqrt{2\pi\alpha}(m-\alpha)^{m-\alpha}e^{-(m-\alpha)}\sqrt{2\pi(m-\alpha)}}p^\alpha \cdot q^{m-\alpha}.$$

Отношение этих двух выражений для u_α стремится к единице всегда, когда m и $m-\alpha$ неограниченно растут, а

$$\alpha = \varepsilon m,$$

где ε стремится к конечной величине.

Я хочу упростить асимптотическое выражение u_α:

$$\frac{m^m}{\alpha^\alpha(m-\alpha)^{m-\alpha}}\sqrt{\frac{m}{2\pi\alpha(m-\alpha)}}p^\alpha \cdot q^{m-\alpha}.$$

Пусть $\alpha = \varepsilon m$; я полагаю

$$m - \alpha = \varepsilon' m,$$

откуда $\varepsilon' = 1 - \varepsilon$.

Асимптотическое выражение принимает вид

$$\frac{m^m}{(m\varepsilon)^{m\varepsilon}\cdot(m\varepsilon')^{m\varepsilon'}}\sqrt{\frac{1}{2\pi\, m\varepsilon\varepsilon'}}p^{m\varepsilon}q^{m\varepsilon'},$$

и, так как $m^m = m^{m\varepsilon} \cdot m^{m\varepsilon'}$,

$$\left(\frac{mp}{m\varepsilon}\right)^{m\varepsilon} \cdot \left(\frac{mq}{m\varepsilon'}\right)^{m\varepsilon'} \sqrt{\frac{1}{2\pi\, m\varepsilon\varepsilon'}}.$$

Пусть

$$A = \left(\frac{p}{\varepsilon}\right)^{\varepsilon} \cdot \left(\frac{q}{\varepsilon'}\right)^{\varepsilon'};$$

тогда

$$u_\alpha = \frac{A^m}{\sqrt{2\pi\, m\varepsilon\varepsilon'}}.$$

53. Я заявляю, что A всегда меньше 1. Оставляя ε и ε' константами, я буду изменять p и q, сохраняя связывающее их соотношение $p + q = 1$.

Чему равен максимум A?

Этот максимум достигается, когда $p^\varepsilon \cdot q^{\varepsilon'}$ максимально, т. е. когда p и q пропорциональны своим степеням:

$$\frac{p}{\varepsilon} = \frac{q}{\varepsilon'} = \frac{p+q}{\varepsilon+\varepsilon'} = 1.$$

Итак, максимум достигается при

$$p = \varepsilon, \qquad q = \varepsilon',$$

и тогда $A = 1$.

Следующее рассуждение позволит нам лучше понять поведение A.

54. Я буду считать p и q константами и менять ε и ε'. Для этого я рассмотрю

$$\log\frac{1}{A} = \varepsilon\log\frac{\varepsilon}{p} + \varepsilon'\log\frac{\varepsilon'}{q}.$$

Как изменяется это число? Возьмем его полный дифференциал

$$d\varepsilon \cdot \log\frac{\varepsilon}{p} + d\varepsilon' \cdot \log\frac{\varepsilon'}{q} + d\varepsilon + d\varepsilon';$$

он должен обратиться в нуль в точке максимума. Но ε и ε' не являются независимыми переменными. Имеем

$$d\varepsilon + d\varepsilon' = 0.$$

Отсюда

$$\log \frac{\varepsilon}{p} = \log \frac{\varepsilon'}{q},$$

$$\frac{\varepsilon}{p} = \frac{\varepsilon'}{q} = \frac{\varepsilon + \varepsilon'}{p+q} = 1.$$

Максимум достигается при $\varepsilon = p$, $\varepsilon' = q$ и этот максимум равен единице.

Как меняется A? При $\varepsilon = 0$ A равно q; при $\varepsilon = 1$ A равно p; таким образом, A начинает возрастать с q, при $\varepsilon = p$ достигает 1, а затем убывает до p.

55. Формула, дающая приблизительную оценку $n!$, имеет вид

$$n! = n^n \cdot e^{-n} \cdot \sqrt{n} \cdot F(n).$$

Функция $F(n)$ стремится к пределу $\sqrt{2\pi}$ при неограниченно растущем n, а

$$(n+1)! = (n+1)^{n+1} e^{-(n+1)} \sqrt{n+1} F(n+1).$$

Отсюда получается

$$\frac{F(n+1)}{F(n)} = \left(\frac{n}{n+1}\right)^n \cdot e \sqrt{\frac{n}{n+1}},$$

$$\log \frac{F(n+1)}{F(n)} = 1 - \left(n + \frac{1}{2}\right) \log \left(1 + \frac{1}{n}\right).$$

Хочется знать, возрастает или убывает $F(n)$ при росте n, т. е. будет логарифм в левой части положительным или отрицательным.

Я делю правую часть на $n + \frac{1}{2}$; остается

$$\frac{1}{n + \frac{1}{2}} - \log \left(1 + \frac{1}{n}\right).$$

Я полагаю $n = \frac{1}{x}$ и рассматриваю функцию

$$\varphi(x) = \frac{1}{\frac{1}{x} + \frac{1}{2}} - \log(1+x),$$

то есть

$$\varphi(x) = \frac{2x}{x+2} - \log(1+x).$$

Положительно или отрицательно $\varphi(x)$? Переменная x меняется от 0 до 1.

При $x = 0$

$$\varphi(0) = 0;$$

при $x = 1$

$$\varphi(1) = \frac{2}{3} - \log 2.$$

Так как $\log 2 = 0{,}69\ldots$, $\varphi(1)$ отрицательно. Надо выяснить, обращается ли в нуль производная

$$\varphi'(x) = \frac{2(x+2) - 2x}{(x+2)^2} - \frac{1}{1+x},$$

$$\varphi'(x) = \frac{4}{(x+2)^2} - \frac{1}{1+x} = \frac{4(x+1) - (x+2)^2}{(x+1)(x+2)^2}.$$

Знаменатель всегда положителен, числитель равен $-x^2$. Итак, $\varphi'(x)$ всегда отрицательна, следовательно, $\varphi(x)$ убывает; значит, $\varphi(x)$ остается отрицательной и

$$F(n+1) < F(n).$$

Если $n = 1$, то $n! = 1$ и $1 = e^{-1}F(1)$, то есть $F(1) = e$. Также известно, что

$$F(\infty) = \sqrt{2\pi}.$$

Величина $F(n)$ всегда убывает, но это убывание не очень велико, т. к.

$$e = 2{,}8\ldots, \quad \text{а} \quad \sqrt{2\pi} = 2{,}5\ldots$$

56. Займемся описанием величины u_α,

$$u_\alpha = \frac{m^m}{\alpha^\alpha(m-\alpha)^{m-\alpha}} p^\alpha \cdot q^{m-\alpha} \sqrt{\frac{m}{\alpha(m-\alpha)}} \frac{F(m)}{F(\alpha) \cdot F(m-\alpha)}.$$

Речь идет о поиске верхнего предела этого выражения. Во-первых,

$$F(m) < F(m-\alpha),$$

откуда

$$\frac{F(m)}{F(\alpha)F(m-\alpha)} < \frac{1}{F(\alpha)}.$$

Но само $\frac{1}{F(\alpha)}$ меньше, чем $\frac{1}{\sqrt{2\pi}}$; асимптотическая оценка u_α и есть его верхний предел.

57. Какова вероятность того, что α меньше, чем εm? Эта вероятность, обозначим ее П, равна

$$\Pi = u_0 + u_1 + \ldots + u_\beta,$$

где $\beta < \varepsilon m$, $\beta + 1 \geqslant \varepsilon m$.

Я предполагаю

$$\varepsilon < p.$$

Пусть сначала

$$\beta + 1 = \varepsilon m;$$

u_0, u_1, ... , u_β возрастают.

$$\Pi < u_{\beta+1} \cdot (\beta + 1).$$

У нас есть верхний предел $u_{\beta+1}$; итак, если $\beta + 1 = \varepsilon m$, то

$$\Pi < \frac{A^m \varepsilon m}{\sqrt{2\pi\, m\varepsilon\varepsilon'}},$$

или

$$\Pi < \frac{A^m \sqrt{\varepsilon m}}{\sqrt{2\pi\varepsilon'}}.$$

Пусть теперь

$$\beta + 1 > \varepsilon m.$$

Тогда

$$\Pi < u_\beta(\beta + 1).$$

Ищем верхний предел u_β; я полагаю

$$A = \varphi(\varepsilon).$$

Если бы $\frac{\beta}{m}$ равнялось ε, получилось бы

$$u_\alpha < \left[\varphi\left(\frac{\beta}{m}\right)\right]^m \sqrt{\frac{m}{2\pi\beta(m-\beta)}}.$$

Между тем A — это функция от ε, растущая при росте ε вплоть до $\varepsilon = p$, и $\varphi(\varepsilon) > \varphi\left(\frac{\beta}{m}\right)$.

Речь идет о верхнем пределе, поэтому

$$u_\beta < A^m \sqrt{\frac{m}{2\pi\beta(m-\beta)}}.$$

Кроме того, β превосходит $\varepsilon m - 1$,

$$\beta > \varepsilon m - 1, \qquad m - \beta > \varepsilon' m - 1.$$

Тогда

$$u_\beta < A^m \sqrt{\frac{m}{2\pi(\varepsilon m - 1)(\varepsilon' m - 1)}}.$$

Чтобы вернуться к вероятности Π, меньшей $u_\beta \cdot (\beta + 1)$, заметим, что $\beta + 1$ само меньше $\varepsilon m + 1$; мы приходим, наконец, к формуле, немного более сложной, чем для целого εm: если εm целое, то $\Pi < A^m \sqrt{\frac{\varepsilon m}{2\pi\varepsilon'}}$, если εm нецелое, то $\Pi < A^m \sqrt{\frac{(\varepsilon m + 1)^2 m}{2\pi(\varepsilon m - 1)(\varepsilon' m - 1)}}$.

Итак:

вероятность того, что α меньше εm, если ε меньше p, никогда не превосходит одной из только что вычисленных величин.

Данная вероятность стремится к нулю, когда m бесконечно растет, при условии, что $\varepsilon < p$. Это теорема Бернулли, которую можно сформулировать по-другому: вероятность того, что α лежит между $mp(1-\theta)$ и $mp(1+\theta)$, стремится к единице, если θ постоянно, а m растет.

Глава 6
Закон Гаусса и повторение испытаний

58. Мы положили

$$\alpha = mp + \lambda\sqrt{m}$$

и нашли вероятность того, что λ лежит между двумя пределами λ_0 и λ_1; эта вероятность при очень больших m представляется следующим интегралом:

$$\int\limits_{\lambda_0}^{\lambda_1} \sqrt{\frac{1}{2\pi\, pq}} e^{-\frac{\lambda^2}{2pq}}\, d\lambda.$$

Теперь мы готовы исследовать, что происходит, когда вероятность того, что x лежит между x_0 и x_1, представляется интегралом:

$$\int\limits_{x_0}^{x_1} \frac{h}{\sqrt{\pi}} e^{-h^2x^2}\, dx.$$

Для краткости я буду называть вероятностный закон *нормальным*, если значение вероятности представлено данным интегралом.

Я предполагаю, что x положительно; вероятность этого равна

$$\int\limits_{0}^{\infty} \frac{h}{\sqrt{\pi}} e^{-h^2x^2}\, dx,$$

то есть $\frac{1}{2}$.

Если я рассматриваю

$$\int\limits_{0}^{x_0} \frac{h}{\sqrt{\pi}} e^{-h^2x^2}\, dx,$$

то этот интеграл будет постоянно расти, когда x_0 менятся от 0 до $+\infty$, потому что все его элементы положительны. В частности, он достигает значения $\frac{1}{4}$.

Величина x_0, при которой интеграл равен $\frac{1}{4}$, называется ***вероятным уклонением.*** Модуль $|x|$ достигает и не достигает этой величины с одинаковой вероятностью.

Число x_0 пропорционально $\frac{1}{h}$, а его точную величину можно найти в таблицах значений данного интеграла.

Пусть x — некоторая величина с нормальным вероятностным законом.

Вероятное значение $|x|$ равно

$$\frac{1}{h\sqrt{\pi}}.$$

Вероятное значение x^2 равно

$$\frac{1}{2h^2}.$$

Вероятностный закон αx также нормален. Пусть $\alpha x = x'$; вероятностный закон x' равен

$$\int\limits_{x'_0}^{x'_1} \frac{h}{\alpha\sqrt{\pi}} e^{-h^2 \frac{x'^2}{\alpha^2}}\, dx',$$

и достаточно положить $h' = \frac{h}{\alpha}$.

Вероятное значение квадрата αx равно

$$\frac{\alpha^2}{2h^2}.$$

59. Предположим, что для x вероятность находиться между x_0 и x_1 выражена интегралом

$$\int\limits_{x_0}^{x_1} \frac{h}{\sqrt{\pi}} e^{-h^2x^2}\, dx,$$

а для y вероятность находиться между y_0 и y_1 выражена интегралом

$$\int_{y_0}^{y_1} \frac{h'}{\sqrt{\pi}} e^{-h'^2 y^2}\, dy.$$

Предположим, кроме того, что величины x и y независимы; это условие можно переформулировать следующим образом: вероятность того, что первая из них лежит между x_0 и x_1 не зависит от вероятности того, что вторая из них лежит между y_0 и y_1.

Вероятное значение x^2 равно $\frac{1}{2h^2}$; у y^2 оно равно $\frac{1}{2h'^2}$.

С какой вероятностью точка с координатами x и y окажется внутри некоторой области?

Займемся сначала случаем прямоугольной области.

Вероятность того, что $(x,\, y)$ попадет внутрь прямоугольника, т. е. что одновременно выполняются две системы неравенств

$$x_0 < x < x_1, \qquad y_0 < y < y_1,$$

представляется следующим двойным интегралом, взятым по всему прямоугольнику:

$$\iint \frac{hh'}{\pi} e^{-h^2x^2 - h'^2y^2}\, dx\, dy.$$

Если область произвольная, я разделяю ее на бесконечно малые прямоугольники. Полная вероятность будет суммой двойных интегралов, взятых по этим элементарным прямоугольникам; в конечном счете она совпадет с двойным интегралом, взятым по всем элементам области.

60. Теперь предполагаем, что

$$x + y = z.$$

Вероятность того, что z лежит между z и $z + dz$, совпадает с вероятностью того, что точка $(x,\, y)$ заключена между двумя бесконечно близкими параллельными прямыми; искомая вероятность будет равна вероятности попасть в эту бесконечно малую область.

Я хочу разбить данную бесконечно малую область на части.

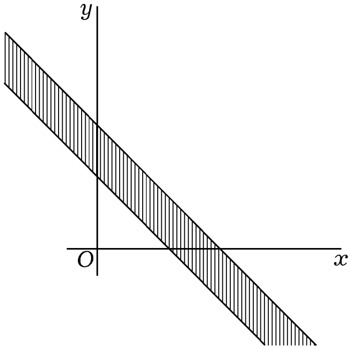


Рис. 1

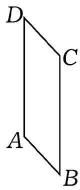


Рис. 2

Для этого я разделю ось абсцисс на бесконечно много частей и в точках деления проведу параллели к оси ординат; так я получу бесконечно много малых параллелограммов; какова площадь одного из них, $ABCD$?

Точки A и B удовлетворяют уравнению

$$x + y = z;$$

координаты A равны x и $z - x$, координаты B равны $x + dx$ и $z - x - dx$.

Точки C и D удовлетворяют уравнению

$$x + y = z + dz;$$

координаты D — x и $z + dz - x$; координаты C — $x + dx$ и $z + dz - x - dx$.

Площадь параллелограмма равна $dx\,dz$.

Двойной интеграл будет суммой интегралов, взятых по каждому параллелограмму:

$$dx\,dz\,\frac{hh'}{\pi}e^{-h^2x^2-h'^2(z-x)^2}.$$

При первом интегрировании y и z считаются константами, а x меняется от $-\infty$ до $+\infty$.

Интеграл равен

$$dz\,\frac{hh'}{\pi}\int\limits_{-\infty}^{+\infty} e^{-h^2x^2-h'^2(z-x)^2}\,dx.$$

Полагаем
$$P = h^2x^2 + {h'}^2(x-z)^2,$$
то есть
$$P = (h^2 + {h'}^2)x^2 - 2{h'}^2xz + {h'}^2z^2$$
или
$$P = (ax - b)^2 + c,$$
где
$$a = \sqrt{h^2 + {h'}^2},$$
$$b = \frac{{h'}^2 z}{a},$$
$$c = {h'}^2z^2 - b^2.$$
Мы будем оценивать
$$\int e^{-P}\,dx$$
или
$$\int e^{-(ax-b)^2-c}\,dx.$$

Пусть $ax - b = \xi$; этот интеграл равен
$$e^{-c}\int e^{-\xi^2}\frac{d\xi}{a},$$
наконец, т. к. ξ вместе с x меняется от $-\infty$ до $+\infty$, получим
$$e^{-c}\frac{\sqrt{\pi}}{a}.$$

Искомая вероятность того, что z лежит между z и $z + dz$, равна
$$dz\frac{hh'}{\sqrt{\pi}}\cdot\frac{e^{-c}}{a}.$$

С другой стороны,
$$c = {h'}^2z^2 - \frac{{h'}^4z^2}{h^2 + {h'}^2} = \frac{h^2{h'}^2z^2}{h^2 + {h'}^2}.$$

Итак, требуемая вероятность равна

$$dz\frac{hh'}{\sqrt{\pi}}\cdot\frac{e^{-z^2\frac{h^2h'^2}{h^2+h'^2}}}{\sqrt{h^2+h'^2}}.$$

Вероятностный закон нормален(8).

Вероятное значение z^2 или $(x+y)^2$ будет равно

$$\frac{1}{2\left(\frac{h^2h'^2}{h^2+h'^2}\right)} \quad \text{или} \quad \frac{1}{2h^2}+\frac{1}{2h'^2},$$

то есть вероятное значение $(x+y)^2$ — это сумма вероятных значений x^2 и y^2.

Вероятное значения $2xy$ здесь действительно равно нулю, но мы бы не получили этого a priori, если бы вероятностный закон не был нормальным.

Данное элегантное рассуждение принадлежит г. Д'Оканю.

61. Задача о повторении испытаний. Есть два противоположных события, A и B, с вероятностями p и q соответственно. Пусть, например, урна содержит μ белых и ν черных шаров,

$$p=\frac{\mu}{\mu+\nu}, \qquad q=\frac{\nu}{\mu+\nu};$$

из нее вытаскивают очень большое число шаров, равное m, каждый раз возвращая вынутый шар в урну. Если вытащено α белых шаров, то, по теореме Бернулли, $\frac{\alpha}{mp}$ скорее всего мало отличается от единицы.

Вероятное значение λ^2 будет равно pq.[1]

Изменим немного условия так, чтобы не один только случай определял наступление событий. Рассмотрим две урны, первая содержит μ белых и ν черных шаров, вторая — μ' белых и ν' черных, и договоримся вытаскивать шары по очереди то из одной урны, то из другой.

---

[1]Здесь, как и ранее, $\lambda=\dfrac{\alpha-mp}{\sqrt{m}}$ есть уклонение. — *Прим. ред.*

После очень большого числа попыток, равного m, будут вынуты α белых и $(m-\alpha)$ черных шаров. Итак, $\frac{\alpha}{m}$ будет очень близко к p, которое здесь равно

$$\frac{1}{2}\frac{\mu}{\mu+\nu}+\frac{1}{2}\frac{\mu'}{\mu'+\nu'}.$$

Но сохранится ли закон уклонений? В данном случае это невозможно.

Предположим, что в первой урне есть только белые, во второй — только черные шары: мы вытащим $\frac{m}{2}$ белых и $\frac{m}{2}$ черных шаров. Тогда $\frac{\alpha}{m}$ будет равно $\frac{1}{2}$, а уклонение — нулю, причем p тоже равно $\frac{1}{2}$. Вероятное значение λ^2 — нуль; а должно быть равно $\frac{1}{4}$, т. к.

$$pq=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}.$$

Мы видим, что уклонение не сохраняется.

62. Я хочу показать, что при вмешательстве какой-то другой причины, кроме случая, вероятное уклонение (или вероятное значение λ^2) будет меньше, чем если бы действовал один только случай.

Пусть m испытаний образуют две категории, в одной βm, в другой $\beta' m$ испытаний, и

$$\beta+\beta'=1.$$

Предполагаем, что два события A и B имеют, соответственно, вероятности p и q в первой категрии, p' и q' во второй.

Событие A присутствует α раз в первой категории, α' раз во второй; событие B присутствует $\beta m-\alpha$ и $\beta' m-\alpha'$ раз.

Полное число испытаний, благоприятных для A, равно $\alpha+\alpha'$; α очень близко к βmp, а α' — к $\beta' mp'$; $\frac{\alpha}{\beta mp}$ и $\frac{\alpha'}{\beta' mp'}$ очень мало уклоняются от единицы; $\alpha+\alpha'$ очень близко к $\beta mp+\beta' mp'$, т. е. уклонение имеет порядок $\sqrt{m}$, и эта серия событий будет очень похожа на ту, что появляется при простом повторении испытаний, где вероятности A и B были бы

$$\beta p+\beta' p' \quad \text{и} \quad \beta q+\beta' q'$$

соответственно.

Ищем законы уклонений. Я хочу положить

$$\alpha = \beta m p + \lambda\sqrt{\beta m},$$
$$\alpha' = \beta' m p' + \lambda'\sqrt{\beta' m}.$$

При полном испытании выполнялось бы равенство:

$$\alpha + \alpha' = m(\beta p + \beta' p') + \lambda''\sqrt{m},$$

где $\beta p + \beta' p'$ — вероятность A во всей серии испытаний.

Составим вероятные значения λ^2, λ'^2, λ''^2, которые обозначим (λ^2), (λ'^2), (λ''^2).

Получаем

$$(\lambda^2) = pq, \qquad (\lambda'^2) = p'q',$$

а так же, *если бы действовал только случай*,

$$(\lambda''^2) = (\beta p + \beta' p')(\beta q + \beta' q').$$

Ищем настоящее значение (λ''^2). Имеем

$$\lambda''\sqrt{m} = \lambda\sqrt{\beta m} + \lambda'\sqrt{\beta' m}$$

или

$$\lambda'' = \lambda\sqrt{\beta} + \lambda'\sqrt{\beta'}.$$

Два события являются независимыми: вероятность того, что λ заключено в двух данных пределах, не зависит от вероятности того, что λ' заключено в двух данных пределах. Вероятностные законы $\lambda\sqrt{\beta}$ и $\lambda'\sqrt{\beta'}$ нормальны, и вероятный закон их суммы $\lambda\sqrt{\beta} + \lambda'\sqrt{\beta'}$ тоже будет нормальным.

Вероятное значение λ''^2 будет равно

$$(\lambda''^2) = \beta(\lambda^2) + \beta'(\lambda'^2) = \beta pq + \beta' p'q'.$$

Это и есть выражение для вероятного значения квадрата уклонения.

Сравниваем две найденные величины; их разность равна

$$(\beta p + \beta' p')(\beta q + \beta' q') - (\beta pq + \beta' p'q')$$

или, если вспомнить, что $\beta + \beta' = 1$,

$$(\beta p + \beta' p')(\beta q + \beta' q') - (\beta pq + \beta' p'q')(\beta + \beta'),$$

то есть

$$\beta\beta'(pq' + p'q - pq - p'q')$$

или

$$\beta\beta'(p - p')(q - q').$$

Но

$$p - p' = q - q'.$$

Тогда рассмотренная разность положительна, и

$$(\lambda''^2) < (\beta p + \beta' p')(\beta q + \beta' q').$$

Это свойство используется в статистике. Пусть результаты наблюдений собраны в таблицу, и хочется узнать, случайны различия в наблюдениях или здесь действует не только случай.

Для какого-то определенного количества наблюдений сравнивают отношение числа наступлений события A и события B.

Пусть N и N' — полное количество наступлений A и B, тогда

$$p = \frac{N}{N + N'}; \qquad q = \frac{N}{N + N'}.$$

Теперь разделим таблицу на несколько серий; в каждой из них полное число испытаний равно m, и

$$\alpha = pm + \lambda\sqrt{m}$$

— число исходов, благоприятных для A. Для каждой серии вычисляется λ; затем находится среднее арифметическое для λ^2; если это среднее равно pq, то действует только случай; если оно меньше pq, то действует не зависящая от случая причина.

Глава 7

Непрерывная вероятность

63. Парадокс Бертрана. До сих пор, даже если нам приходилось иметь дело с очень большим числом возможных исходов, это число оставалось конечным. Иногда мы рассматривали предельные переходы и заменяли $\sum$ на $\int$.

Теперь мы займемся задачами, в которых число возможных исходов становится бесконечным.

Требуется дать точное определение этих исходов, и парадокс Бертрана предоставляет нам прекрасное свидетельство того, какого рода ошибки могут возникнуть в этих задачах; речь идет о следующем вопросе.

С какой вероятностью хорда некоей данной окружности будет больше стороны правильного треугольника, вписанного в эту окружность?

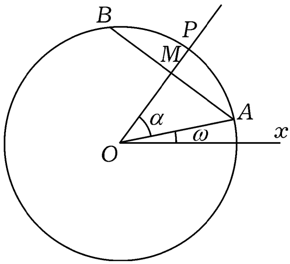


Рис. 3

Ж. Бертран решает эту задачу двумя способами и получает совершенно противоположные результаты.

Пусть AB — хорда; возьмем за единицу радиус OA; полярные координаты точки A будут 1 и ω.

Пусть α — угол AOM, OM — перпендикуляр, восставленный из центра хорды, P — точка пересечения этого перпендикуляра с кривой, а M — середина хорды.

Угол POx или θ, равен $\alpha + \omega$.

64. Первое рассуждение. Неважно, на каком участке окружности находится точка A. Вероятность того, что ω лежит между ω_0 и ω_1, пропорциональна разности $\frac{\omega_1 - \omega_0}{2\pi}$. Зафиксировав точку A, можно провести хорду во всех возможных направлениях, т. е., выбрав A, я могу придать α все возможные значения от 0 до $\frac{\pi}{2}$.

Вероятность того, что α будет находиться между α_0 и α_1, пропорциональна $\alpha_1 - \alpha_0$. Если бы AB была стороной вписанного равностороннего треугольника, то α было бы равно $60°$.

Так как α может принять любое из значений от $0°$ до $90°$, вероятность того, что хорда окажется больше стороны треугольника, равна

$$\frac{90° - 60°}{90° - 0°} = \frac{1}{3}.$$

65. Второе рассуждение. Хорду можно провести в каком угодно направлении. Вероятность того, что θ лежит между θ_0 и θ_1, пропорциональна $\theta_1 - \theta_0$.

Однажды зафиксировав это направление, я провожу OP: отрезок AB будет определен, когда я выберу точку M, т. е. расстояние $OM = \rho = \cos\alpha$.

Расстояние ρ может принимать любое значение от 0 до 1; следует считать, что вероятность попасть в промежуток от ρ_0 до ρ_1 попорциональна $\rho_1 - \rho_0$.

Хорда больше стороны треугольника, если OM лежит от 0 до $\frac{1}{2}$.

Итак, вероятность равна $\frac{1}{2}$.

Откуда берется это противоречие? В этих двух случаях мы принимаем разные гипотезы; мы определили вероятность двумя разными способами.

66. Пусть в общем случае требуется определить вероятность того, что какое-то число x лежит между x_0 и x_1; в принципе, мы можем сказать, что ничего о нем не знаем.

Эта вероятность должна зависеть от x_0 и x_1, т. е. она будет какой-нибудь функцией вида $P(x_0, x_1)$.

Если мы ищем вероятность того, что x лежит между x_0 и x_2, а

$$x_0 < x_1 < x_2,$$

то в силу принципа сложения вероятностей эта вероятность равна

$$P(x_0, x_2) = P(x_0, x_1) + P(x_1, x_2).$$

Если

$$x_2 = x_1 + dx,$$

то имеем

$$P(x_0, x_2) - P(x_0, x_1) = P(x_1, x_1 + dx_1).$$

Эта вероятность будет бесконечно мала и после деления на dx_1 будет зависеть только от x_1.

Итак, во всех случаях получаем

$$P(x_0, x_1) = \int_{x_0}^{x_1} \varphi(x)\, dx.$$

Но *нам неизвестна природа* $\varphi(x)$, эта функция остается произвольной; мы должны *задать* ее изначально по особой, имеющей какой-то смысл договоренности.

Точно так же вероятность того, что точка (x, y) лежит внутри данной области, равна

$$\iint \varphi(x, y)\, dx\, dy,$$

где двойной интеграл берется по всей области, но $\varphi(x, y)$ нам неизвестна.

Математик не имеет преимуществ при выборе этой гипотезы; но как только гипотеза выбрана, он должен внимательно следить, чтобы не принять другую, ей противоречащую.

67. Мы можем иметь много параметров $x_1, x_2, \ldots, x_p$. Интеграл порядка p,

$$\int \varphi(x_1, x_2, \ldots, x_p)\, dx_1\, dx_2 \ldots dx_p,$$

будет задавать вероятность того, что параметры x удовлетворяют некоторым условиям, *если функция* φ *определена*; надо всего лишь взять интеграл по всем значениям x, удовлетворяющим данным условиям. Но это определение имеет смысл, только когда функция φ задана по какой-то предварительной *договоренности.*

Предположим, что делается замена переменных $x_1, x_2, \ldots, x_p$ на $y_1, y_2, \ldots, y_p$; интеграл преобразуется в

$$\int \varphi \frac{\partial(x_1, x_2, \ldots, x_p)}{\partial(y_1, y_2, \ldots, y_p)}\, dy_1\, dy_2 \ldots dy_p$$

с помощью якобиана, или функционального определителя x по y.

Этот новый интеграл полностью определен; интегрирование проводится по тем y, которые соответствуют прежним x; они будут известны, т. к. мы знаем соотношения, связывающие x и y.

68. Применим все вышесказанное к парадоксу Бертрана.

В первом случае переменными были ω и α, во втором — θ и ρ.

В первый раз вероятность того, что ω попало в промежуток от ω_0 до ω_1, была пропорциональна $\omega_1 - \omega_0$; вероятность того, что α лежит от α_0 до α_1, была пропорциональна $\alpha_1 - \alpha_0$.

Эта вероятность представлялась как интеграл

$$\iint d\omega\, d\alpha,$$

взятый по всем парам значений ω и α, удовлетворявшим требуемым условиям.

Во второй раз мы предположили, что θ и ρ могут принимать все возможные значения с одинаковой вероятностью, и мы представили искомую вероятность как

$$\iint d\theta\, d\rho.$$

Как мы уже непосредственно установили, эти две гипотезы не совпадают. Ищем якобиан; имеем

$$d\rho = -\sin\alpha\, d\alpha, \qquad d\theta = d\omega + d\alpha.$$

Определитель

$$\begin{vmatrix} 0 & -\sin\alpha \\ 1 & 1 \end{vmatrix}$$

равен $\sin\alpha$. Итак, второй интеграл равен

$$\iint \sin\alpha\, d\omega\, d\alpha$$

и не совпадает с первым.

69. Другой пример. Рассмотрим на чертеже прямую AB; ее тангенциальные координаты равны $\frac{1}{a}$ и $\frac{1}{b}$. Будем искать вероятность того, что эта прямая займет какое-то определенное положение. Вероятность того, что a и b принимают все значения в определенных пределах, может быть, по первому соглашению, представлена как

$$\iint da\,db,$$

где $b = a\ \operatorname{tg}\omega$, если ω — угол между AB и AO.

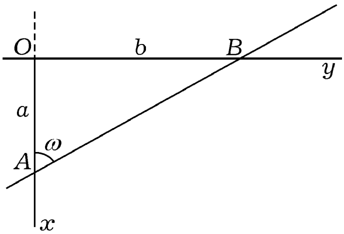


Рис. 4

Точно так же можно сказать, что ω принимает все возможные значения; откуда вероятность равна

$$\iint da\,d\omega.$$

Эти два интеграла не одно и то же, и второе соглашение, которое при поверхностном изучении могло бы показаться таким же обоснованным, как и первое, противоречит ему; действительно, якобиан равен

$$\frac{a}{\cos^2\omega},$$

и

$$\iint da\,db = \iint \frac{a}{\cos^2\omega}\,da\,d\omega.$$

70. Задача о разрезании палки. Палку длины 1 делят на три части x, y, z.

$$x + y + z = 1.$$

Мы допускаем, что для x вероятность попасть в промежуток от x до $x+$ $+\,dx$ по определению пропорциональна dx; для промежутка от x_0 до x_1 она будет пропорциональна $x_1 - x_0$.

Точно так же для y вероятность попасть в промежуток от y_0 до y_1 будет по определению пропорциональна $y_1 - y_0$.

Тогда вероятность того, что x и y удовлетворяют определенным условиям, выражается через интеграл

$$\iint dx\,dy,$$

взятый по всем значениям x и y, удовлетворяющим этим условиям.

Аналогично можно было бы предположить, что эта вероятность равна

$$\iint dy\,dz,$$

или

$$\iint dz\,dx,$$

так как в качестве переменных можно взять x и z или y и z.

Здесь все три определения эквивалентны; имеем

$$\iint dz\,dx = \iint dx\,dy\,\frac{\partial(z,\,x)}{\partial(x,\,y)},$$

а функциональный определитель в точности равен 1, т. к.

$$x = x,$$
$$z = 1 - x - y.$$

71. Какова вероятность того, что x, y и z образуют треугольник?

Нарисуем равносторонний треугольник с высотой 1: из точки M, внутренней для этого треугольника, опускаем перпендикуляры к трем сторонам. Полученная таким образом сумма их длин будет равна высоте треугольника, т. е. 1; они будут соответствовать трем частям палки, т. е. x, y, z.

Точку M можно рассматривать как определяющую способ деления палки: какова вероятность того, что эта точка попадет внутрь какой-то определенной области?

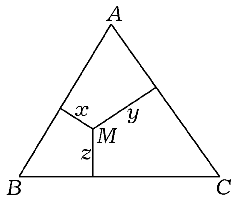


Рис. 5

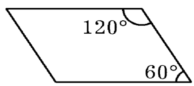


Рис. 6

Вероятность того, что x лежит между x и $x+dx$, а y лежит между y и $y+dy$, пропорциональна $dx\,dy$. Точка M будет лежать между двумя прямыми, параллельными BC, проведенными на расстояниях x и $x+dx$ от BC, и двумя прямыми, параллельными к AC, лежащими на расстояниях y и $y+dy$ от AC. Построенный таким образом параллелограмм имеет углы 60° и 120°, а его площадь равна

$$\frac{dx\,dy}{\sin 60^\circ}.$$

В этом случае вероятность будет пропорциональна площади параллелограмма; и вообще она будет пропорциональна площади рассмотренной области.

Так как точка M должна лежать внутри треугольника ABC, вероятность того, что M попадет внутрь какой-то определенной области, будет равна отношению площадей этой области и всего треугольника.

Соединим прямыми точки A', B' и C' — середины сторон треугольника. Для того чтобы из x, y и z можно было составить треугольник, M должна лежать внутри $A'B'C'$; если точка M лежит на одной из сторон $A'B'C'$, выполняется одно из следующих уравнений:

$$z = x + y, \qquad x = y + z, \qquad y = z + x;$$

если точка M находится вне $A'B'C'$, одна из трех величин x, y, z больше суммы двух других.

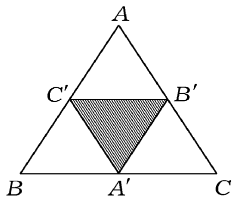


Рис. 7

Вероятность того, что из x, y, z можно составить треугольник, равна $\frac{1}{4}$.

72. Задача о бросании иглы. На листе бумаги нарисовано некоторое количество параллельных прямых, отстоящих друг от друга на одинаковое расстояние; это общее расстояние равно d, и на бумагу случайным образом бросают иглу, чья длина тоже равна d.

С какой вероятностью эта игла пересекает одну из прямых?

Можно поставить и более общую задачу.

Зафиксированы две координатные оси OX, OY и фигура F, неподвижно соединенная с этими осями.

Пусть, с другой стороны, есть подвижные оси OX и OY и фигура F_1 постоянной формы, но неподвижно соединенная с этими осями и перемещающаяся вместе с ними. Определяем положение подвижной фигуры по отношению к неподвижным осям.

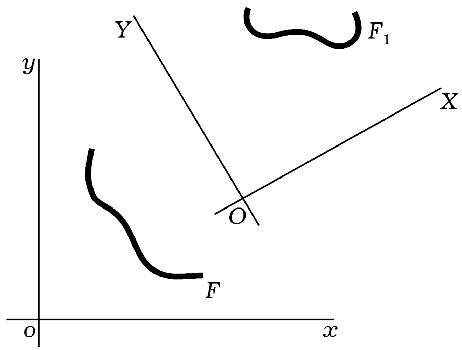


Рис. 8

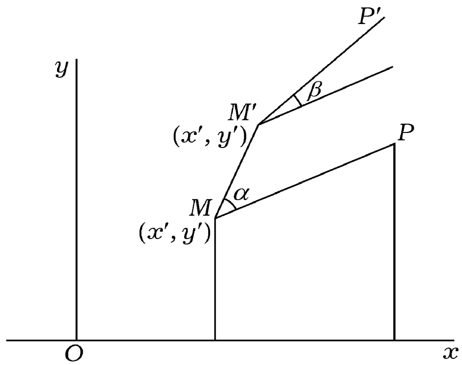


Рис. 9

Пусть M — одна из ее точек, MP — прямая, проходящая через M и неподвижно соединенная с фигурой F_1: достаточно определить положение MP по отношению к XOY. Это положение определено координатами x, y точки M и углом ω между MP и OX.

Вероятность того, что F удовлетворяет некоторым условиям, пропорциональна

$$\iiint dx\,dy\,d\omega.$$

Чтобы обосновать это определение, я покажу, что оно сохраняется, если взять другую точку M' на подвижной фигуре и другую прямую $M'P'$.

Прямая $M'P'$ будет соединена с MP.

Пусть l — длина MM', α — угол MM' с MP, а β — угол между $M'P'$ и MP: l, α, β — константы.

Прямая $M'P'$ определена относительно системы координат XOY координатами x', y' точки M' и углом ω' между $M'P'$ и осью OX.

Если по первому определению вероятность того, что подвижная фигура удовлетворяет некоторым условиям, равна $\iiint dx\,dy\,d\omega$, то эта вероятность также будет равна

$$\iiint dx'\,dy'\,d\omega'.$$

Действительно, якобиан равен единице; имеем отображение

$$\begin{aligned} x' &= x + l\cos(\omega+\alpha),\\ y' &= y + l\sin(\omega+\alpha),\\ \omega' &= \omega+\beta, \end{aligned}$$

и якобиан $\dfrac{\partial(x',\, y',\, \omega')}{\partial(x,\, y,\, \omega)}$ равен

$$\begin{vmatrix} 1 & 0 & -l\sin(\omega+\alpha)\\ 0 & 1 & l\cos(\omega+\alpha)\\ 0 & 0 & 1 \end{vmatrix},$$

то есть 1.

Итак, вероятностный закон одинаков, каким бы ни был выбор прямой MP.

73. Если я рассматриваю две фигуры φ, φ', равные между собой и соединенные с подвижными осями координат, то вероятность того, что φ' удовлетворяет каким-то условиям, совпадает с вероятностью того, что φ удовлетворяет тем же условиям. Рассмотрим прямую MP, соединенную с φ и другую прямую, $M'P'$, чье расположение относительно φ' совпадает с расположением MP относительно φ; пусть x, y, ω, с одной стороны, и x', y', ω', с другой стороны, — величины, определяющие положение этих двух прямых.

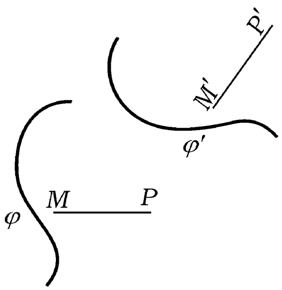


Рис. 10

Тогда и положение фигуры φ определено с помощью x, y и ω, а фигуры φ' — с помощью x', y', ω'. Для того чтобы φ удовлетворяла каким-то условиям, x, y, ω должны удовлетворять каким-то неравенствам. Для того чтобы φ' удовлетворяла тем же условиям, x', y', ω' должны удовлетворять тем же неравенствам. При разнице в обозначениях мы, тем не менее, получим один и тот же интеграл.

Итак, значение вероятности одинаково в обоих случаях.

74. Требуется найти вероятность, с которой ограниченный отрезок MP пересекает параллельные прямые в задаче о бросании иглы. Если второй отрезок $M'P'$, той же длины, неподвижно соединен с MP, то вероятность пересечения с параллелями будет для него той же самой.

Если вместо MP рассмотреть вдвое бо́льший отрезок MQ, то вероятность удвоится$^{(9)}$, т. к. этот отрезок состоит из двух частей, равных MP, назовем их MN и NQ, где N — середина MQ.

Предположим, игроку обещают столько франков, сколько точек пересечения с параллельными прямыми[1] будет у отрезка. Математическое ожидание при игре с MQ будет равно удвоенному ожиданию при игре с MN, т. к. оно совпадет с суммой ожидания для MN и ожидания для NQ. В общем, оно будет пропорционально длине отрезка.

Если NQ не будет продолжением MN, математическое ожидание все равно удвоится. Итак, математическое ожидание пропорционально общей длине линии, прямая она или ломаная, или, если пойти еще дальше, *какую бы форму она не имела.*

Если обещано столько франков, сколько точек пересечения имеет с параллельными прямыми какая-то кривая, математическое ожидание будет также пропорционально длине этой кривой.

Если кривая — это окружность диаметра d, то ее длина будет равна πd; в этом случае всегда есть две точки пересечения и математическое ожидание равно 2. Для кривой длины s это ожидание будет $\frac{2s}{\pi d}$; для прямой длины d — $\frac{2}{\pi}$.$^{(10)}$

[1] Можно договориться, что если один из концов MN попадает на прямую, это считается полупересечением. — *Прим. авт.*

75. Вернемся к парадоксу Бертрана, связанному с вероятностью того, что хорда некоторой окружности будет меньше стороны вписанного в нее правильного треугольника.

Проведем окружность C', концентрическую к первой, C, так, чтобы радиус C' был в два раза меньше. Случайным образом помещаем на чертеж отрезок. Если мы примем соглашения, сделанные только что в задаче о бросании иглы, то будет ли вероятность зависеть от новой, третьей гипотезы, или от одной из двух, изученных ранее?

Я могу считать отрезок фиксированным, а окружности — подвижными. Вероятность того, что одна из окружностей пересекает отрезок, пропорциональна ее длине; тогда вероятность того, что C' пересекает отрезок, равна отношению длин двух окружностей, т.е. $\frac{1}{2}$. Так мы возвращаемся к одной из гипотез Бертрана.

Глава 8
Различные приложения

76. Рассмотрим еще одну задачу, аналогичную задаче о бросании иглы.

На сфере S нарисована подвижная фигура; какова вероятность того, что эта фигура удовлетворяет некоторым условиям[1]?

Во-первых, как определить ее расположение?

Пусть p_0 — начальное, p_1 — конечное положение подвижной фигуры; из p_0 в p_1 можно перейти с помощью подходящего поворота, определенного осью и углом вращения.

Пусть α, β, γ — направляющие косинусы оси, а 2θ — угол вращения; полагаем

$$\lambda = \cos\theta, \quad \mu = \alpha\sin\theta, \quad \nu = \beta\sin\theta, \quad \rho = \gamma\sin\theta$$

и считаем λ, μ, ν, ρ переменными.

Они связаны соотношением

$$\lambda^2 + \mu^2 + \nu^2 + \rho^2 = 1.$$

Это соотношение сохраняется, если поменять знаки у λ, μ, ν, ρ, и нам достаточно знать любые три из этих величин.

Пусть вероятность представляется с помощью тройного интеграла

$$\int \frac{d\mu\, d\nu\, d\rho}{\lambda};$$

точно также ее можно записать через

$$\int \frac{d\lambda\, d\nu\, d\rho}{\mu};$$

[1] Здесь и далее рассматривается сфера радиуса 1. — *Прим. ред.*

Действительно, найдем якобиан этих новых переменных λ, ν, ρ по старым, т.е. μ, ν, ρ, считая λ определенной как функцию от μ, ν, ρ.

$$\lambda\, d\lambda = -\mu\, d\mu - \nu\, d\nu - \rho\, d\rho.$$

Якобиан равен

$$\begin{vmatrix} -\dfrac{\mu}{\lambda} & -\dfrac{\nu}{\lambda} & -\dfrac{\rho}{\lambda} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = -\frac{\mu}{\lambda}.$$

Итак, при замене переменных первый из тройных интегралов с точностью до знака переходит во второй после умножения на $\dfrac{\lambda}{\mu}$. Следовательно,

$$\int \frac{d\mu\, d\nu\, d\rho}{\lambda} \quad \text{и} \quad \int \frac{d\mu\, d\nu\, d\rho}{\mu}$$

дадут одинаковое определение вероятности. Это определение можно обосновать следующим образом: рассмотрим сферу

$$x^2 + y^2 + z^2 = 1.$$

Предполагаем, что для произвольной точки сферы вероятность попасть внутрь некоторой сферической области пропорциональна площади этой области. Данная площадь совпадает с интегралом

$$\int \frac{dx\, dy}{\cos \widehat{n\Sigma}} = \int \frac{dx\, dy}{z},$$

где $\cos \widehat{n\Sigma}$ — третий направляющий косинус нормали к сфере в данной точке.

В нашей задаче мы приняли совершенно аналогичную гипотезу, т. к.

$$\lambda^2 + \mu^2 + \nu^2 + \rho^2 = 1$$

было бы уравнением сферы в четырехмерном пространстве

77. Предварительно я займусь начальным положением p_0. Поворот λ, μ, ν, ρ зависит не только от p_1, но и от выбора начального положения.

Я хочу показать, что вероятность останется прежней, если вместо начальной позиции p_0 рассмотреть другую, p_0'.

Поворот от p_0' к p_1 определяется с помощью λ', μ', ν', ρ', а вероятность — через

$$\int \frac{d\mu'\,d\nu'\,d\rho'}{\lambda'}.$$

Я заявляю, что она будет пропорциональна предыдущей.

Пусть l, m, n, r определяют поворот от p_0' к p_0, тогда поворот λ', μ', ν', ρ' будет результатом двух других. Формулы композиции поворотов известны:

$$\begin{aligned}
\lambda' &= \lambda l - \mu m - \nu n - \rho r,\\
\mu' &= \lambda m + \mu l - \nu r + \rho n,\\
\nu' &= \lambda n + \mu r + \nu l - \rho m,\\
\rho' &= \lambda r - \mu n + \nu m + \rho l.
\end{aligned}$$

Теперь нужно вычислить якобиан λ', μ', ν', ρ' по λ, μ, ν, ρ. Я полагаю

$$\sigma = \sqrt{\lambda^2 + \mu^2 + \nu^2 + \rho^2}.$$

В нашей задаче $\sigma = 1$, но в принципе я могу придать λ, μ, ν, ρ любые значения. Аналогично

$$\sigma' = \sqrt{\lambda'^2 + \mu'^2 + \nu'^2 + \rho'^2}.$$

Каковы бы ни были эти величины, имеем

$$\sigma' = \sigma\sqrt{l^2 + m^2 + n^2 + r^2},$$

и если l, m, n, r — данные константы, то

$$\sigma' = \sigma.$$

Требуется вычислить

$$\frac{\partial(\mu', \nu', \rho')}{\partial(\mu, \nu, \rho)}$$

при условии $\sigma = 1$, т. е.

$$\begin{vmatrix}
\dfrac{\partial\mu'}{\partial\mu} & \dfrac{\partial\mu'}{\partial\nu} & \dfrac{\partial\mu'}{\partial\rho}\\[2mm]
\dfrac{\partial\nu'}{\partial\mu} & \dfrac{\partial\nu'}{\partial\nu} & \dfrac{\partial\nu'}{\partial\rho}\\[2mm]
\dfrac{\partial\rho'}{\partial\mu} & \dfrac{\partial\rho'}{\partial\nu} & \dfrac{\partial\rho'}{\partial\rho}
\end{vmatrix}.$$

Но я хочу довести число переменных до 4 рассматривать, с одной стороны, σ, μ, ν, ρ, с другой стороны, σ', μ', ν', ρ'. Так как $\sigma' = \sigma$, частные производные от σ будут 1, 0, 0, 0:

$$\begin{vmatrix} 1 & 0 & 0 & 0 \\ \dfrac{\partial\mu'}{\partial\sigma} & \dfrac{\partial\mu'}{\partial\mu} & \dfrac{\partial\mu'}{\partial\nu} & \dfrac{\partial\mu'}{\partial\rho} \\ \dfrac{\partial\nu'}{\partial\sigma} & \dfrac{\partial\nu'}{\partial\mu} & \dfrac{\partial\nu'}{\partial\nu} & \dfrac{\partial\nu'}{\partial\rho} \\ \dfrac{\partial\rho'}{\partial\sigma} & \dfrac{\partial\rho'}{\partial\mu} & \dfrac{\partial\rho'}{\partial\nu} & \dfrac{\partial\rho'}{\partial\rho} \end{vmatrix}.$$

Это тот же самый определитель, т. е. теперь можно написать, что

$$\frac{\partial(\mu', \nu', \rho')}{\partial(\mu, \nu, \rho)} = \frac{\partial(\sigma, \mu', \nu', \rho')}{\partial(\sigma', \mu, \nu, \rho)} =$$

$$= \frac{\partial(\sigma, \mu', \nu', \rho')}{\partial(\lambda', \mu', \nu', \rho')} \cdot \frac{\partial(\lambda', \mu', \nu', \rho')}{\partial(\lambda, \mu, \nu, \rho)} \cdot \frac{\partial(\lambda, \mu, \nu, \rho)}{\partial(\sigma, \mu, \nu, \rho)}.$$

Вычислим по очереди три последних определителя. Первый равен

$$\begin{vmatrix} \dfrac{\partial\sigma'}{\partial\lambda'} & \dfrac{\partial\sigma'}{\partial\mu'} & \dfrac{\partial\sigma'}{\partial\nu'} & \dfrac{\partial\sigma'}{\partial\rho'} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = \frac{\partial\sigma'}{\partial\lambda'} = \frac{\lambda'}{\sigma'}.$$

Второй равен

$$\begin{vmatrix} l & -m & -n & -r \\ m & l & -r & n \\ n & r & l & -m \\ r & -n & m & l \end{vmatrix} = (l^2 + m^2 + n^2 + r^2)^2 = 1.$$

Третий равен

$$\frac{\partial(\lambda, \mu, \nu, \rho)}{\partial(\sigma, \mu, \nu, \rho)} = \frac{1}{\dfrac{\partial(\sigma, \mu, \nu, \rho)}{\partial(\lambda, \mu, \nu, \rho)}} = \frac{1}{\frac{\lambda}{\sigma}} = \frac{\sigma}{\lambda}.$$

Итак, произведение всех трех равно $\frac{\lambda'}{\lambda}$, и интеграл

$$\int \frac{d\mu'\, d\nu'\, d\rho'}{\lambda'}$$

превращается в

$$\int \frac{d\mu\, d\nu\, d\rho}{\lambda'} \cdot \frac{\lambda'}{\lambda}$$

или

$$\int \frac{d\mu\, d\nu\, d\rho}{\lambda}.$$

Следовательно, вероятность остается одной и той же, каким бы ни было начальное положение.

78. Требуется найти вероятность того, что наша фигура p_0 удовлетворяет некоторым условиям.

Если рассмотреть другую фигуру p_0', равную первой и неподвижно соединенную с ней, то можно точно также потребовать, чтобы эта вторая фигура удовлетворяла каким-то условиям. Пусть тогда λ, μ, ν, ρ — параметры вращения, переводящего p_0 в p_1, а λ', μ', ν', ρ' — параметры вращения, переводящего p_0' в p_1. Вероятность того, что p_0 перейдет в p_1, удовлетворяющую каким-то условиям, представляется как

$$\int \frac{d\mu\, d\nu\, d\rho}{\lambda},$$

где параметры λ, μ, ν, ρ должны удовлетворять некоторым неравенствам.

Вероятность того, что p_0' переходит в p_1, удовлетворяющую *тем же* условиям, представляется как

$$\int \frac{d\mu'\, d\nu'\, d\rho'}{\lambda'}.$$

где λ', μ', ν', ρ' должны удовлетворять *тем же* неравенствам. Интегралы, отличающиеся только обозначениями, равны, и вероятность остается прежней.

Итак, вероятности того, что две подвижные фигуры, равные и соединенные друг с другом, удовлетворяют одному и тому же условию, равны между собой.

79. Выберем другое рассуждение, без λ, μ, ν, ρ.

Я определяю положение точки M подвижной фигуры ее координатами x, y, z, а положение дуги большой окружности MP — углом ω, который MP образует с MA, если MA — дуга большой окружности, проходящей через фиксированную точку A.

Вероятность запишется как

$$\int \Psi\, dx\, dy\, d\omega,$$

где ψ — функция от x, y, z и ω.

Взяв

$$dx\, dy = z\, d\sigma,$$

где $d\sigma$ — элемент поверхности сферы, получаем интеграл

$$\int \Phi\, d\sigma\, d\omega.$$

Этот интеграл нужно взять по всем σ — элементам поверхности сферы — и по всем значениям угла ω, удовлетворяющим данным условиям. Возвращаясь к переменным x и y, запишем

$$\int \Phi \frac{dx\, dy}{z}\, d\omega.$$

80. Я заявляю, что вид Φ остается неизменным, каким бы ни было положение точки A.

Рассмотрим другую фиксированную точку, B, и пусть ω' — угол между MP и MB, β — угол между MB и MA.

$$\omega' = \omega + \beta.$$

Вместо x, y, ω сделаем переменными x, y, ω'. Функциональный определитель $\dfrac{\partial(x,\, y,\, \omega')}{\partial(x,\, y,\, \omega)}$ равен

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \dfrac{\partial\beta}{\partial x} & \dfrac{\partial\beta}{\partial y} & 1 \end{vmatrix} = 1.$$

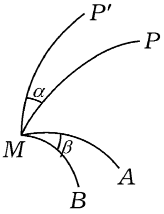


Рис. 11

81. Функция Φ не зависит от ω.

Действительно, рассмотрим другую дугу большой окружности, MP', и пусть ω' — угол между MP' и MA, α — угол между MP' и MP.

$$\omega'' = \omega + \alpha.$$

Вероятностный закон не изменится. Интеграл

$$\int \Phi \, d\sigma \, d\omega$$

выражает вероятность того, что MP удовлетворяет некоторым условиям. Интеграл

$$\int \Phi \, d\sigma \, d\omega''$$

будет выражать вероятность того, что MP' удовлетворяют тем же самым условиям. Эти две вероятности должны совпадать, значит, Φ не будет зависеть от ω.

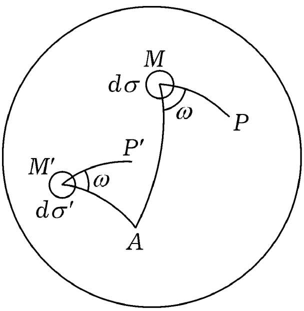


Рис. 12

82. Более того, Φ не зависит от x и y.

Действительно, пусть $d\sigma$ и $d\sigma'$ — два элемента поверхности сферы, равные между собой. Пусть l, m, n, r — параметры вращения, переводящего $d\sigma$ в $d\sigma'$. Пусть также λ, μ, ν, ρ — параметры вращения, переводящего подвижную фигуру в такое положение, чтобы M оказалась внутри $d\sigma$, а λ', μ', ν', ρ' — параметры вращения, переводящего M внутрь $d\sigma'$.

Пусть, кроме того, x, y — координаты центра масс $d\sigma$, а x', y' — центра масс $d\sigma'$.

Параметры λ', μ', ν', ρ' будут линейными функциями от переменных λ, μ, ν, ρ, с одной стороны, и l, m, n, r — с другой стороны; действительно, для доказательства этого факта надо всего лишь обратиться к формуле композиции вращений, приведенной выше. Между тем, как и раньше, мы имеем

$$\int \frac{d\mu \, d\nu \, d\rho}{\lambda} = \int \frac{d\mu' \, d\nu' \, d\rho'}{\lambda'}.$$

Вероятность того, что M лежит внутри $d\sigma$, равна

$$\int_0^{2\pi} d\omega \int \Phi\, d\sigma = 2\pi\Phi(x,\, y)\, d\sigma = \int \frac{d\mu\, d\nu\, d\rho}{\lambda},$$

параметры λ, μ, ν, ρ должны удовлетворять неравенствам, выражающим тот факт, что соответствующее вращение перемещает M внутрь $d\sigma$; интеграл берется по всем значениям параметров, удовлетворяющим этим условиям.

Вероятность того, что M лежит внутри $d\sigma'$, равна

$$2\pi\Phi(x',\, y')\, d\sigma' = \int \frac{d\mu'\, d\nu'\, d\rho'}{\lambda'}.$$

Итак,

$$\Phi(x,\, y)\, d\sigma = \Phi(x',\, y')\, d\sigma'$$

или, так как $d\sigma = d\sigma'$

$$\Phi(x,\, y) = \Phi(x',\, y'),$$

интеграл брался по всем значениям μ', ν', ρ' таким, чтобы вращение λ', μ', ν', ρ' переводило M внутрь $d\sigma'$; или, что то же самое, таким, чтобы вращение λ, μ, ν, ρ переводило M внутрь $d\sigma$.

83. Мы записали вероятностный закон в другой форме, но такое определение вероятности будет совпадать с прежним, когда мы брали в качестве переменных λ, μ, ν, ρ. В итоге вероятность того, что M попадет внутрь некоторой области $d\sigma$, и, в то же время, MP образует с MA угол ω, равна

$$\int \Phi\, d\sigma\, d\omega,$$

где Φ — константа, которую надо определить.

Возьмем интеграл по всем элементам сферы; угол ω меняется от 0 до 2π, а σ — от 0 до 4π. Значение интеграла будет равно

$$4\pi\Phi \times 2\pi = 8\pi^2\Phi;$$

но это вероятность, которая должна совпадать с 1. Итак,

$$\Phi = \frac{1}{8\pi^2}$$

и $\displaystyle\int \frac{d\sigma\, d\omega}{8\pi^2}$ или $\displaystyle\int \frac{dx\, dy\, d\omega}{8\pi^2 z}$ — искомый вероятностный закон.

84. Пусть на сфере есть неподвижная и движущая кривые; игроку обещают столько франков, сколько найдется точек пересечения; каково математическое ожидание игрока?

Оно пропорционально произведению длин кривых.

85. Если рассмотреть подвижную фигуру φ и две фиксированные, φ_1 и φ_2, то вероятность того, что φ займет заданное положение относительно φ_1, совпадет с вероятностью того, что φ займет то же положение относительно φ_2.

Иначе говоря, пусть λ, μ, ν, ρ определяют поворот, переводящий φ_1 в какое-то положение φ'; введем новые переменные λ', μ', ν', ρ', определяющие поворот, который переводит φ_2 в то же положение φ'; мы обнаружим, что вероятностные законы одинаковы, и получим, как и выше,

$$\int \frac{d\mu\, d\nu\, d\rho}{\lambda} = \int \frac{d\mu'\, d\nu'\, d\rho'}{\lambda'}.$$

Поворот $\lambda', \mu', \nu', \rho'$ — результат двух других, λ, μ, ν, ρ и l, m, n, r; этот последний, переводящий φ_2 в φ_1, можно считать заданным, и вычисление функционального определителя μ, ν, ρ по μ', ν', ρ' проводится также, как и в предыдущий раз.

Установив это, я откажусь от параметров λ, μ, ν, ρ, чья геометрическая интерпретация довольно сложна, и мы вернемся к переменным x, y и ω из параграфа 79.

86. Пусть M — точка подвижной фигуры на сферической поверхности, имеющая координаты x, y, z; а MP — дуга большой окружности, принадлежащая подвижной фигуре и составляющая угол ω с дугой другой большой окружности, MA, проходящей через другую фиксированную точку A.

1° С какой вероятностью M лежит внутри $d\sigma$, когда ω меняется от 0 до 2π? Эта вероятность равна

$$\int \frac{d\sigma\, d\omega}{8\pi^2} = \frac{d\sigma}{4\pi}.$$

2° С какой вероятностью подвижная окружность на сфере пересечется с фиксированной?

Пусть P — полюс фиксированной окружности, A — точка пересечения с плоскостью чертежа, а θ — угол POA.

Пусть P', A', θ' — аналогичные характеристики подвижной окружности.

Пусть φ — угол POP'.

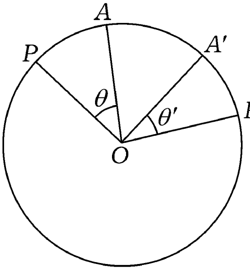


Рис. 13

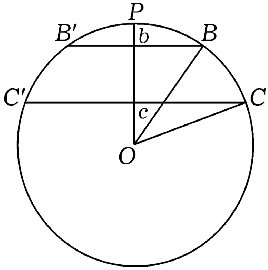


Рис. 14

Необходимое и достаточное условие пересечения — одновременное выполнение неравенств

$$\varphi < \theta + \theta',$$
$$\varphi > \theta - \theta'$$

в предположении, что $\theta > \theta'$.

Обозначим $BCB'C'$ область, в которой должен находиться полюс P' подвижной окружности; в проекции на плоскость чертежа две малые окружности, ограничивающие эту область, превращаются в отрезки BB', CC'; угол POB будет равен $\theta - \theta'$, угол POC — $\theta + \theta'$.

Искомая вероятность будет пропорциональна высоте bc(11)

$$bc = Ob - Oc = \cos(\theta - \theta') - \cos(\theta + \theta'),$$
$$bc = 2\sin\theta\sin\theta'.$$

87. Можно сформулировать и более общую задачу. Пусть даны:

1° n дуг фиксированных больших окружностей, c_1, c_2, … , c_n; все дуги равны между собой и имеют длину l;

2° n' дуг подвижных больших окружностей, c'_1, c'_2, … , c'_n той же длины l, соединенных друг с другом.

Я ищу точки пересечения подвижных дуг с неподвижными и обещаю заплатить столько франков, сколько есть точек пересечения.

Математическое ожидание игрока будет пропорционально произведению nn'.

Как только что было показано, вероятность пересечения c'_1 с c_1 совпадает с такой же вероятностью для c'_2 и c_1, и т. д.

Точно также она остается неизменной для c'_1 и c_2 и т. д.

Математическое ожидание составит столько франков, сколько сочетаний можно образовать из первых n дуг с n' вторыми. Предположим также, что длина этих последних равна l' и отлична от l: математическое ожидание будет равно $nn'll'$.

Если рассматривать две ломаные линии, составленные из дуг больших окружностей, то математическое ожидание по-прежнему будет пропорционально их длинам, т. к. если удвоить один из элементов, то соответствующее математическое ожидание тоже удвоится.

При предельном переходе этот вывод все еще будет верным, и вообще математическое ожидание будет пропорционально s и s' — длинам кривых. Пусть оно равно Kss'.

Ищем константу K.

Возьмем две больших окружности: их длина одинакова и равна 2π, они пересекаются в двух точках; математическое ожидание равно $K \times 4\pi^2$ и, т. к. оно равно 2,

$$K = \frac{1}{2\pi^2}.$$

Для произвольных окружностей математическое ожидание равно $\frac{ss'}{2\pi^2}$. Если окружности пересекаются, то точек пересечения будет две[1]. Между тем

$$s = 2\pi \sin\theta,$$
$$s' = 2\pi \sin\theta'.$$

Ожидание равно

$$\frac{2\pi \sin\theta \times 2\pi \sin\theta'}{2\pi^2} = 2\pi \sin\theta \sin\theta'.$$

Выше мы получили тот же результат другим способом.

[1]Вероятность того, что окружности касаются (т. е. имеют одну точку пересечения), равна нулю. — *Прим. ред.*

88. Пусть на небесной сфере случайным образом размещены N звезд.

Будем платить игроку один франк за каждую пару звезд p и p' такую, что угловое расстояние между ними меньше γ. Каково математическое ожидание игрока?

Пусть звезда p' должна попасть внутрь некоторой области. Поверхность этой области будет пропорциональна $\sin^2 \frac{\gamma}{2}$. Для $\gamma = \pi$ эта поверхность совпадет со всей сферой. Итак, вероятность равна

$$\frac{\sin^2 \frac{\gamma}{2}}{\sin^2 \frac{\pi}{2}} = \sin^2 \frac{\gamma}{2}.$$

Так как всего звезд N, они могут образовать $\frac{N(N-1)}{2}$ пар. Математическое ожидание равно

$$\frac{N(N-1)}{2} \cdot \sin^2 \frac{\gamma}{2}.$$

89. Рассмотрим *механическую систему*, чьи уравнения имеют гамильтонов вид; n переменных, $x_1, x_2, \ldots, x_n$, определяют положение системы; еще n переменных, $y_1, y_2, \ldots, y_n$ — ее скорости.

Пусть задана функция F, зависящая от x и y; уравнения имеют вид[1]

$$\frac{dx_i}{dt} = \frac{dF}{dy_i}; \qquad \frac{dy_i}{dt} = -\frac{dF}{dx_i}.$$

Нам известна F, т. е. закон движения, но неизвестны начальные условия.

Обозначим $x_1^0, x_2^0, \ldots, x_n^0$ и $y_1^0, y_2^0, \ldots, y_n^0$ значения переменных в момент $t = 0$.

Какова вероятность того, что эти переменные имеют какие-то определенные значения в данный момент t?

Если мне дан вероятностный закон, по которому переменные принимают вышеуказанные начальные значения, задача становится однозначно определенной. Я предполагаю, что этот вероятностный закон для начальных значений задан.

[1]Такую функцию принято называть гамильтонианом. — *Прим. ред.*

90. Я считаю, что вероятность пропорциональна интегралу

$$\int K\, dx_1^0\, dx_2^0 \ldots dx_n^0\, dy_1^0\, dy_2^0 \ldots dy_n^0,$$

где K — некоторая константа.

Можно предположить, что о начальных величинах ничего неизвестно; допустим, мы знаем только, что F лежит между F_1 и F_2; так как функция $F = \mathrm{const}$ будет интегралом уравнения движения (первый интеграл движения), F останется между F_1 и F_2, если начальное значение F лежало в этих пределах.

Предыдущий интеграл, взятый по всем значениям переменных, удовлетворяющим неравенствам

$$F_1 < F < F_2,$$

будет равен 1.

Если вероятностный закон верен для начальных значений переменных, он будет верен и для их конечных значений.

Достаточно знать, что функциональный определитель конечных значений по начальным равен единице.

Пусть $x', \ldots, y', \ldots$ — значения $x, \ldots, y, \ldots$ в момент t'; $x, \ldots, y, \ldots$ — их значения в момент t. Достаточно доказать наше предложение только для близких t и t'. Пусть

$$t' = t + \varepsilon.$$

Для простоты я рассмотрю случай двух переменных x и двух y.

$$x_1' = x_1 + \varepsilon \frac{dx_1}{dt} = x_1 + \varepsilon \frac{dF}{dy_1},$$

$$x_2' = x_2 + \varepsilon \frac{dF}{dy_2};$$

и

$$y_1' = y_1 + \varepsilon \frac{dy_1}{dt} = y_1 - \varepsilon \frac{dF}{dx_1},$$

$$y_2' = y_2 - \varepsilon \frac{dF}{dx_2}.$$

Функциональный определитель равен

$$\begin{vmatrix} 1+\varepsilon\dfrac{d^2F}{dx_1\,dy_1} & \varepsilon\dfrac{d^2F}{dx_2\,dy_1} & \varepsilon\dfrac{d^2F}{dy_1^2} & \varepsilon\dfrac{d^2F}{dx_1\,dy_2} \\ \varepsilon\dfrac{d^2F}{dx_1\,dy_2} & 1+\varepsilon\dfrac{d^2F}{dx_2\,dy_2} & \varepsilon\dfrac{d^2F}{dy_2\,dy_1} & \varepsilon\dfrac{d^2F}{dy_2^2} \\ -\varepsilon\dfrac{d^2F}{dx_1^2} & \varepsilon\dfrac{d^2F}{dx_1\,dx_2} & 1-\varepsilon\dfrac{d^2F}{dx_1\,dy_1} & -\varepsilon\dfrac{d^2F}{dx_1\,dy_2} \\ \varepsilon\dfrac{d^2F}{dx_2\,dx_1} & -\varepsilon\dfrac{d^2F}{dx_2^2} & -\varepsilon\dfrac{d^2F}{dx_2\,dx_1} & 1-\varepsilon\dfrac{d^2F}{dx_2\,dy_2} \end{vmatrix}.$$

Я буду раскладывать его, пренебрегая квадратом ε. Все элементы определителя, кроме диагональных, — бесконечно малые первого порядка. Поэтому как минимум второй порядок будут иметь все слагаемые, кроме входящих в

$$\left(1+\varepsilon\frac{d^2F}{dx_1\,dy_1}\right)\left(1+\varepsilon\frac{d^2F}{dx_2\,dy_2}\right)\left(1-\varepsilon\frac{d^2F}{dx_1\,dy_1}\right)\left(1-\varepsilon\frac{d^2F}{dx_2\,dy_2}\right)$$

или

$$\left[1-\varepsilon^2\left(\frac{d^2F}{dx_1\,dy_1}\right)^2\right]\left[1-\varepsilon^2\left(\frac{d^2F}{dx_2\,dy_2}\right)^2\right],$$

то есть определитель равен 1 с точностью до ε^2.

91. Из всего предыдущего вытекает, *что нужно уделить очень большое внимание выбору подходящего вероятностного закона.*

Вероятность того, что x лежит между x_0 и x_1, выражается через интеграл

$$\int\limits_{x_0}^{x_1}\varphi(x)\,dx;$$

при нахождении вероятностного закона нам придется выдвигать различные гипотезы о виде функции φ, но мы всегда будем считать ее непрерывной.

В общем случае вероятность того, что x удовлетворяет какому-то заданному условию, будет зависеть от выбора φ; однако, это происходит не всегда, и *решение некоторых задач не зависит от выбора вероятностного закона.*

Пример 1. Вероятность того, что x иррационально, всегда будет равна 1, как бы ни была выбрана непрерывная функция φ, а вероятность рационального x всегда будет бесконечно малой.

92. Пример 2. Барабан разделен на очень много равных частей, попеременно черных и красных; будем вращать его с большой скоростью. Когда он остановится, одна из частей окажется напротив фиксированной стрелки: с какой вероятностью эта часть будет красной или черной?

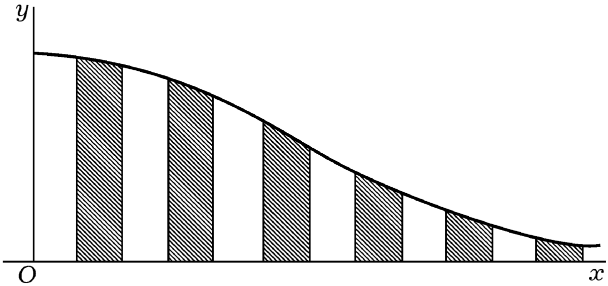


Рис. 15

Для полного решения задачи требуется знать произвольную функцию φ; решение будет зависеть от импульса, начальной угловой скорости. Вероятность того, что эта скорость попадет в интервал от ω_0 до ω_1, равна

$$\int\limits_{\omega_0}^{\omega_1} \varphi(\omega)\, d\omega,$$

где функция φ совершенно неизвестна.

С другой стороны, полный угол поворота барабана равен θ. Вероятность того, что θ лежит между θ_0 и θ_1, равна

$$\int\limits_{\theta_0}^{\theta_1} f(\theta)\, d\theta.$$

Мы точно так же ничего не знаем о $f(\theta)$. Тем не менее вероятность того, что нужный участок будет красным, всегда очень близка к $\frac{1}{2}$; таким образом она не зависит от f.

Я предполагаю, что каждому делению на барабане соответствует угол ε; разделим ось абсцисс на части, равные ε, а в точках деления проведем вертикальные прямые до пересечения с кривой

$$y = f(\theta).$$

Так как в точках деления цвет меняется, я заштрихую области, соответствующие, например, красным делениям.

Искомая вероятность равна отношению заштрихованной площади к полной.

Какова ни была форма кривой, при бесконечно растущем числе делений это отношение будет стремится к $\frac{1}{2}$.

Действительно, пусть A — максимальный угол, на который может повернуться барабан, т. е. $\theta < A$. Пусть функция $f(\theta)$ непрерывна и дифференцируема. Пусть, кроме того, ее производная не превышает некоторого максимума M:

$$|f'(\theta)| < M.$$

Я делю A на n равных частей; пусть ε — одна из них. Имеем

$$\varepsilon = \frac{A}{n}.$$

Рассмотрим две соседние области: разность их площадей меньше, чем $\varepsilon(\mu - \mu')$, где μ и μ' обозначают соответственно максимум и минимум $f(\theta)$ на этом интервале. Но $(\mu - \mu')$ меньше, чем $2M\varepsilon$: разность двух площадей меньше, чем $2M\varepsilon^2$.

Так как заштрихованных областей $\frac{n}{2}$, мы найдем разность площади двух полных областей, если умножим результат на $\frac{n}{2}$, что даст нам $M\varepsilon^2 n$ или $MA\varepsilon$.

Разность площади двух областей стремится к нулю вместе с ε, и вероятность оказывается равной $\frac{1}{2}$.

Если бы мы совершенно ничего не знали про φ или f, мы бы ничего не смогли вычислить; предпринимать вычисления можно только при наличии некоторых сведений о них. Но здесь нам оказалось достаточно знать, что у f есть ограниченная производная.

93. Пример 3. Рассмотрим большое число *планет* с практически круговыми орбитами. Пусть a — средняя скорость одной из этих планет, b — ее долгота в данный момент времени, который мы будем считать начальным. В момент t ее долгота равна

$$l = at + b.$$

Пусть вероятность того, что a и b удовлетворяют определенным условиям, равна

$$\int \varphi(a,\, b)\, da\, db.$$

Я заявляю, что в конце очень большого промежутка времени планеты будут одинаково распределены по всем знакам Зодиака.

Таким образом, вероятность того, что l лежит в данных пределах, не будет зависеть от φ.

Ищем вероятное значение функции e^{iml}: если m отлично от нуля, вероятное значение будет стремиться к нулю при неограниченном возрастании l. Это вероятное значение можно записать как

$$\iint e^{im(at+b)} \varphi(a,\, b)\, da\, db.$$

Интегрируя по частям, получаем

$$\int \frac{e^{im(at+b)}}{imt} \varphi\, db - \iint \frac{e^{im(at+b)}}{imt} \frac{d\varphi}{da}\, da\, db.$$

Если мы только предположим, что φ непрерывна, оба верхних выражения будут стремиться к нулю.

Я хочу найти вероятное значение какой-нибудь периодической функции $f(l)$. Формула Фурье дает нам

$$f(l) = \sum A_m e^{iml}.$$

Вероятное значение каждого из членов этого ряда, исключая константу A_0, равно 0. Итак, вероятное значение $f(l)$ равно

$$A_0 = \frac{1}{2\pi} \int\limits_0^{2\pi} f(l)\, dl.$$

Предположим, что

$$0 < l_0 < l_1 < 2\pi;$$

вероятность того, что l лежит между l_0 и l_1, равна

$$\frac{1}{2\pi}\int\limits_{l_0}^{l_1} dl = \frac{l_1 - l_0}{2\pi}.$$

Это вероятное значение функции $f(l)$, равной 1, если l лежит между l_0 и l_1, и 0 иначе.

При каких угодно t и φ вероятность будет практически пропорциональна $l_1 - l_0$; распределение планет будет равномерным.

Глава 9
Условные вероятности

94. Теперь мы займемся вопросами, связанными с так называемой условной вероятностью.

До сих пор мы решали задачи следующего содержания: если известно, что действует такая-то причина, то какова вероятность такого-то следствия?

Существуют обратные задачи: если известно, что произошло такое-то следствие, какова вероятность того, что действовала такая-то причина?

Эти задачи схожи с задачей о двух урнах, в первой из которых белых шаров намного больше, чем во второй: известно, что вынули белый шар, но неизвестно, из какой урны; гораздо больше оснований считать, что урна была первой, а не второй.

Чтобы дать более точные определения, необходимо принять особое соглашение, как в начале обсуждения любого вероятностного вопроса.

Когда число возможных исходов сравнивают с числом благоприятных, надо следить, чтобы все исходы были равновероятны. Соглашение, основанное на равной вероятности каких-то исходов, всегда остается в довольно большой степени произвольным.

Из колоды в 32 карты вытащили карту, известно, что это картинка; какова вероятность того, что это король?

До того как событие произошло, вероятность равнялась отношению числа королей к общему числу возможных исходов, т. е. $\frac{4}{32}$ или $\frac{1}{8}$. После того как событие произошло, число благоприятных исходов по прежнему 4; число возможных исходов меньше, оно равно количеству картинок, т. е. 12. Вероятность стала $\frac{4}{12}$ или $\frac{1}{3}$; она увеличилась.

95. Формула Байеса. Пусть на событие могут повлиять n разных причин, $C_1, C_2, \ldots, C_n$; вероятность того, что причина C_i, если в событии участвовала она, приведет к исходу A, равна p_i.

Если бы мы знали, что воздействует именно C_1, мы бы могли утверждать, что вероятность A равна p_1.

Предполагается, что две причины не могут действовать одновременно.

До того как событие произошло, каждая из этих причин a priori имела какую-то вероятность, которую я считаю известной: для C_i эта вероятность равнялась $\bar{\omega}_i$.

Случилось событие A: какова вероятность того, что это следствие причины C_i?

Занумеруем все возможные и вероятные исходы и для наглядности рассмотрим конкретный пример.

Пусть в каждой из M урн есть Q шаров; всего MQ шаров или MQ возможных исходов, которые я считаю равновероятными.

Урны делятся на группы $C_1, C_2, \ldots, C_n$. Урн вида C_1 будет $\bar{\omega}_1 M$; вида C_2 — $\bar{\omega}_2 M$; ... ; вида C_n — $\bar{\omega}_n M$. A priori вероятность того, что воздействовать будет причина C_n, равна

$$\frac{\bar{\omega}_n \cdot M}{M} = \bar{\omega}_n.$$

В урнах лежат белые или черные шары. Пусть, например, событие A наступает, если вынут белый шар. Вероятность вынуть белый шар из урн первого вида равна p_1.

В урнах вида C_1 лежит $p_1 Q$ белых шаров; в урнах вида C_2 — $p_2 Q$, ... , в урнах вида C_n — $p_n Q$.

Вынули белый шар; требуется найти вероятность того, что выбранная урна имела вид C_i. Число благоприятных исходов равно числу белых шаров в урнах вида C_i, т. е. $\bar{\omega}_i p_i MQ$.

Полное число всех возможных исходов совпадает с числом всех шаров, т. е.

$$\bar{\omega}_1 p_1 MQ + \bar{\omega}_2 p_2 MQ + \ldots + \bar{\omega}_n p_n MQ.$$

Отношение этих двух чисел и будет, по определению, искомой вероятностью

$$\frac{\bar{\omega}_i p_i}{\bar{\omega}_1 p_1 + \bar{\omega}_2 p_2 + \ldots + \bar{\omega}_n p_n}.$$

96. Итак, можно сказать: вероятность того, что действовала причина C_i, после того как, подействовав, она привела к событию A, равна произведению вероятностей.

Во-первых, причина C_i должна действовать, а вероятность этого a priori равна $\bar{\omega}_i$; затем, воздействовав, она приводит к событию A с вероятностью p_i. Произведение вероятностей равно $\bar{\omega}_i p_i$.

Но вопрос поставлен иначе.

Требуется, чтобы событие произошло, и только затем оно должно быть приписано причине C_i. Это снова произведение вероятностей.

Вероятность того, что событие происходит, равна

$$\bar{\omega}_1 p_1 + \bar{\omega}_2 p_2 + \ldots + \bar{\omega}_n p_n;$$

пусть (*если известно, что событие произошло*) вероятность того, что оно — результат причины C_i, равна x, тогда произведение вероятностей того, что событие произошло, и того, что оно — результат причины C_i, равно

$$x \cdot (\bar{\omega}_1 p_1 + \bar{\omega}_2 p_2 + \ldots + \bar{\omega}_n p_n),$$

откуда

$$x = \frac{\bar{\omega}_i p_i}{\bar{\omega}_1 p_1 + \bar{\omega}_2 p_2 + \ldots + \bar{\omega}_n p_n}.$$

Или иначе, возьмем формулу

$$(B)(A, \text{ если } B) = (A)(B, \text{ если } A),$$

приведенную в параграфе 12. Пишем

$$(C_i)(A, \text{ если } C_i) = (A)(C_i, \text{ если } A).$$

(C_i) — вероятность причины a priori, когда неизвестно, произошло ли следствие A; это $\bar{\omega}_i$.

$(A$, если $C_i)$ — вероятность следствия, если известно, что действовала причина C_i; это p_i.

$(C_i$, если $A)$ — вероятность причины a posteriori, когда известно, что A произошло; это x.

Вероятность (A) — константа, не зависящая от C_i.

Предыдущее равенство показывает, что $(C_i$, если $A)$ пропорционально произведению $(C_i)(A$, если $C_i)$; т.е. x пропорционально $p_i\bar{\omega}_i$.

97. Пусть в начале игры противник тянет карту и вытаскивает короля; какова вероятность того, что это шулер?

Пусть $\bar{\omega}_1$ — априорная вероятность того, что это не шулер; $\bar{\omega}_2$ — что шулер. В первом случае вероятность вытащить короля равна $\frac{1}{8}$, во втором — 1. A posteriori вероятность того, что мы имеем дело с шулером, равна

$$\frac{\bar{\omega}_2}{\dfrac{\bar{\omega}_1}{8} + \bar{\omega}_2}.$$

Если предположить, что $\bar{\omega}_2 = \bar{\omega}_1$, т. е. мы ничего не знаем о честности противника, вероятность равна $\frac{8}{9}$. Это очень большое число.

К счастью, в большинстве случаев можно a priori считать, что

$$\bar{\omega}_2 < \bar{\omega}_1.$$

98. В одной урне, чье содержимое неизвестно, лежит N шаров; мы вынимаем шар μ раз, каждый раз возвращая вынутый шар в урну, и нам попадаются только белые шары. Какова вероятность того, что в урне n белых шаров?

Пусть $\bar{\omega}_n$ — априорная вероятность того, что в урне n белых шаров, а p_n — вероятность того, что вытащили μ белых.

$$p_n = \left(\frac{n}{N}\right)^{\mu}.$$

Когда μ шаров вынуто, вероятность того, что в урне содержится n белых шаров, задается вышеуказанной формулой, и после сокращения на множитель $\left(\frac{1}{N}\right)^{\mu}$, общий для обеих частей дроби, она равна

$$\frac{\bar{\omega}_n n^{\mu}}{\bar{\omega}_1 \cdot 1^{\mu} + \bar{\omega}_1 \cdot 2^{\mu} + \ldots + \bar{\omega}_1 \cdot N^{\mu}}.$$

99. Требуется знать a priori числа $\bar{\omega}_1, \bar{\omega}_2, \ldots, \bar{\omega}_N$, о которых можно выдвигать множество гипотез.

Предположим, например, что любой состав шаров в урне равновероятен, т. е. все $\bar{\omega}_i$ равны. Каждое из них совпадет с $\frac{1}{N+1}$, т. к. существует $N+1$ возможность, включая случай, когда нет ни одного белого шара. Предыдущая дробь превращается в

$$\frac{n^\mu}{1^\mu + 2^\mu + \ldots + N^\mu}.$$

Во второй раз предположим, что в урну последовательно кладется N шаров, одни белые, другие черные, и каждый раз цвет определяется жребием.

Вероятность положить белый шар каждый раз равна $\frac{1}{2}$, а вероятность того, что в конце n из N шаров в урне будут белыми, оценивается по формуле

$$\frac{m!}{\alpha!\,(m-\alpha)!} p^\alpha \cdot q^{m-\alpha},$$

где надо положить

$$m = N, \qquad \alpha = n, \qquad p = q = \frac{1}{2},$$

что дает

$$\frac{N!}{n!\,(N-n)!} \left(\frac{1}{2}\right)^N.$$

Так, $\bar{\omega}_n$ — априорная вероятность того, что в урне было n белых шаров, — будет пропорциональна биномиальному коэффициенту, а в выражении апостериорной вероятности того же события нам придется приравнивать все $\bar{\omega}$ к этим различным коэффициентам:

$$\bar{\omega}_n = \frac{N!}{n!\,(N-n)!}.$$

100. Результат будет совершенно не похож на предыдущий.

Пусть N очень велико;

$$1^\mu + 2^\mu + \ldots + N^\mu$$

— многочлен от N степени $\mu+1$, в котором я могу оставить только член с наибольшей степенью, $\frac{N^{\mu+1}}{\mu+1}$. Тогда при первой гипотезе вероятность станет

$$\frac{n^\mu(\mu+1)}{N^{\mu+1}},$$

и, например, при $\mu = 2$ она равна $\dfrac{3n^2}{N^3}$.

При второй гипотезе оценим сперва $\bar{\omega}_n n^\mu$, если μ по-прежнему равно 2.

$$\bar{\omega}_n \cdot n^2 = n^2 \frac{N!}{n!\,(N-n)!}.$$

Затем оценим знаменатель

$$\bar{\omega}_1 \cdot 1^\mu + \bar{\omega}_2 \cdot 2^\mu + \ldots + \bar{\omega}_N \cdot N^\mu.$$

Для этого оценим выражение

$$1 + e^x \cdot \bar{\omega}_1 + e^{2x} \cdot \bar{\omega}_2 + \ldots + e^{Nx} \cdot \bar{\omega}_N,$$

это не что иное, как разложение

$$(1+e^x)^N.$$

Я дважды дифференцирую по x:

$$1^2 \cdot e^x \cdot \bar{\omega}_1 + 2^2 \cdot e^{2x} \cdot \bar{\omega}_2 + \ldots + N^2 \cdot e^{Nx} \cdot \bar{\omega}_N.$$

Чтобы найти искомый знаменатель при $\mu = 2$, достаточно положить $x = 0$.

Итак, этот знаменатель равен удвоенному коэффициенту при x^2 в разложении $(1+e^x)^N$ по степеням x; ограничимся членами второго порядка:

$$(1+e^x)^n = \left(2 + x + \frac{x^2}{2}\right)^N = 2^N\left(1 + \frac{x}{2} + \frac{x^2}{4}\right)^N,$$

то есть

$$2^N\left[1 + \frac{Nx}{2} + \frac{N(N-1)}{8}x^2 + \frac{Nx^2}{4}\right].$$

В скобках слагаемым со старшей степенью числа N будет $\dfrac{N^2}{8}x^2$. Искомый знаменатель приблизительно равен удвоенному значению $2^N \cdot \dfrac{B^2}{8}$, т. е $2^{N-2}N^2$.

Таким образом, при второй гипотезе вероятность того, что в урне лежат n белых шаров, равна

$$\frac{N!}{n!\,(N-n)!}\cdot\frac{n^2}{N^2}\cdot\frac{1}{2^{N-2}};$$

она оказывается гораздо меньше, чем при первой гипотезе.

Действительно, доказывая теорему Бернулли, мы видели, что

$$\frac{N!}{n!\,(N-n)!}\cdot\frac{1}{2^N}$$

очень мало, за исключением случая, когда n и $N-n$ практически равны p и q, т. е. $\frac{1}{2}$.

101. Два игрока в шахматы сыграли $n+m$ партий: первый выиграл n, второй — m раз. Если $n>m$, то следует предположить, что первый игрок сильнее.

При розыгрыше новой партии у первого игрока шансов выиграть больше.

Кто-то, хорошо знающий игроков, мог бы представить вероятность выигрыша для первого игрока числом p. Но я, который никогда не видел игру этих партнеров, ничего не знаю о p; я попытаюсь найти p теоретически.

Вероятность того, что p лежит между p_0 и p_1, можно представить как

$$\int\limits_{p_0}^{p_1} f(p)\,dp,$$

где функция $f(p)$ неизвестна.

Вероятность того, что p лежит между p и $p+dp$ равна a priori $f(p)\,dp$; она соответствует $\bar{\omega}_i$.

Вероятности p_i соответствует выражение

$$\frac{(n+m)!}{n!\,m!}p^n q^m.$$

В данном случае причиной будет тот факт, что первый игрок выигрывает с вероятностью p.

Вероятность того, что первый игрок выиграет n партий, если p — вероятность выиграть каждую из $n+m$ партий равна

$$\frac{(n+m)!}{n!\,m!}p^n q^m$$

Чему равна a posteriori вероятность того, что p лежит между p и $p+dp$?

Здесь $\bar{\omega}_i p_i$, если заменить q на $1-p$, будет равно

$$p^n(1-p)^m\frac{(n+m)!}{n!\,m!}f(p)\,dp.$$

Сумма $\bar{\omega}_i p_i$ равна

$$\int_0^1 p^n(1-p)^m\frac{(n+m)!}{n!\,m!}f(p)\,dp;$$

этот интеграл надо взять от 0 до 1, т. к. вероятность p, очевидно, лежит в этих пределах.

За недостатком сведений обычно выдвигают гипотезу, что $f(p)=1$.

Тогда интеграл оценивается очень просто: он принимает вид

$$\frac{(n+m)!}{n!\,m!}\int_0^1 p^n(1-p)^m\,dp.$$

Его можно свести к интегралу Эйлера первого рода:

$$B(n+1,\,m+1)=\frac{\Gamma(n+1)\cdot\Gamma(m+1)}{\Gamma(n+m+2)};$$

здесь все Г-функции — факториалы, и это выражение — не что иное, как

$$\frac{n!\,m!}{(n+m+1)!}.$$

Тогда интеграл, соответствующий сумме всех $\bar{\omega}_i p_i$ попросту равен $\dfrac{1}{n+m+1}$, и апостериорная вероятность того, что p лежит между p и $p+dp$, равна

$$\varphi(p)\,dp=\frac{(n+m+1)!}{n!\,m!}p^n(1-p)^m\,dp.$$

102. Какой станет вероятность того, что этот игрок выиграет следующую партию? Теперь ее легко найти. Вероятность того, что p лежит между p и $p + dp$, равна $\varphi(p)\,dp$; если это условие выполнено, то вероятность выиграть следующую партию равна p; в силу правила умножения вероятностей объединение этих двух условий имеет вероятность $p\varphi(p)\,dp$.

Проинтегрируем данное выражение от 0 до 1:

$$\int_0^1 p\varphi(p)\,dp.$$

Если заменить $\varphi(p)$ ее значением, получится

$$\frac{(n+m+1)!}{n!\,m!}\int_0^1 p^{n+1}(1-p)^m\,dp.$$

Это снова интеграл Эйлера, и мы приходим к выражению

$$\frac{(n+m+1)!}{n!\,m!}\cdot\frac{(n+1)!\,m!}{(n+m+2)!}$$

или

$$\frac{n+1}{n+m+2}.$$

Если бы я применил то же рассуждение к азартной игре, я не имел бы права считать $f(p) = 1$. Действительно, a priori p должно совпадать с $\frac{1}{2}$. Тогда $f(p)$ должна быть бесконечной при $p = \frac{1}{2}$.

103. Пусть N — *количество всех малых планет*; среди них есть определенное число M известных. В течение года наблюдают n планет, m из которых известны.

Требуется найти вероятное значение N. Это значение не должно сильно отличаться от $\frac{Mn}{m}$, но такой оценки недостаточно: надо учесть возможную разницу между настоящим и вероятным числом планет.

Мы будем действовать так: в первую очередь предположим, что нам известна вероятность, с которой в течение года можно наблюдать реально существующую планету; пусть эта вероятность равна p; допустим, что она одинакова для известных и неизвестных планет.

Так как мы наблюдали n планет, вероятное значение N должно, на первый взгляд, совпадать с $\frac{n}{p}$; но невозможно, чтобы это совпадение было совершенно точным: числа 1, 2, ... , N имеют собственные вероятности, которые я назову $\bar{\omega}_1, \bar{\omega}_2, \ldots, \bar{\omega}_N$, и вероятное значение N равно

$$\bar{\omega}_1 + 2\bar{\omega}_2 + \ldots + N\bar{\omega}_N.$$

Если предполагать, что p дано, а все числа 1, 2, ... , N равновероятны, мы придем, как будет показано, к вероятному значению $\frac{n+q}{p}$.

Решив этот первый вопрос, займемся другой задачей; мы предположили, что p известно, теперь откажемся от этого допущения и определим вероятное значение N как функцию m и M, что даст нам результат, очень близкий к $\frac{Mn}{m}$, как мы и предвидели выше.

104. Итак, я назову $\bar{\omega}_N$ априорную вероятность того, что количество планет равно N; p_N — вероятность того, что в год наблюдают n планет, если всего их N.

Апостериорная вероятность — это условная вероятность; она выражается как

$$\frac{\bar{\omega}_N \cdot p_N}{\sum \bar{\omega}_N p_N}.$$

В качестве первой гипотезы насчет $\bar{\omega}$ предположим, что все они равны; предыдущая формула упрощается:

$$\frac{p_N}{\sum p_N}.$$

Чтобы вычислить p_N, применим теорему о повторении испытаний. Вероятность того, что из N планет за год наблюдают n равна

$$p_N = \frac{N!}{n!(N-n)!} p^n \cdot q^{N-n},$$

если p — вероятность наблюдать любую реально существующую планету; q — вероятность не наблюдать ее.

Здесь n — постоянная величина; это наименьшее возможное N. Пусть N неограниченно растет,

$$\sum p_N = p^N + \frac{n+1}{1} p^n q + \frac{(n+1)(n+2)}{1 \cdot 2} p^n q^2 + \ldots =$$
$$= p^n (1-q)^{-(n+1)} = F(p, q).$$

Если я воспользуюсь соотношением $p = 1 - q$, то

$$\sum p_N = p^n \cdot p^{-(n+1)} = \frac{1}{p}.$$

Если гипотеза об $\bar{\omega}$ верна, то вероятность того, что всего есть N планет, равна $p \cdot p_N$.

105. Вероятное значение N равно

$$\frac{\sum N_{p_N}}{\sum p_N}.$$

Немного проще будет найти вероятное значение $N - n$, т. е.

$$\frac{\sum (N-n)p_N}{\sum p_N}.$$

Пусть

$$A = \frac{N!}{n!(N-n)!};$$

тогда

$$\sum p_N = \sum Ap^n \cdot q^{N-n},$$

$$\sum (N-n)p_N = \sum Ap^n \cdot (N-n)q^{N-n}.$$

Для оценки второго члена достаточно продифференцировать $F(p,\, q)$ по q и результат умножить на q:

$$q\frac{dF}{dq} = p^n q(n+1)(1-q)^{-(n+2)};$$

после дифференцирования заменим $1 - q$ на p. Остается

$$\frac{(n+1)q}{p^2}.$$

В силу предыдущих рассуждений, это выражение равно $\sum Ap^n(N - - n)q^{N-n}$. Так как, с другой стороны,

$$\sum p_N = \frac{1}{p},$$

вероятное значение $N - n$ равно

$$\frac{\dfrac{(n+1)q}{p^2}}{\dfrac{1}{p}} = \frac{(n+1)q}{p}.$$

Следовательно, вероятное значение N равно

$$\frac{(n+1)q}{p} + n = \frac{n+q}{p}.$$

Эта величина мало отличается от $\frac{n}{p}$ как и предполагалось. Действительно,

$$\frac{n+q}{p} = \frac{n+1}{p} - 1.$$

106. Теперь будем считать p неизвестным; допустим, нам хочется узнать, с какой вероятностью p лежит между p и $p + dp$.

Априорная вероятность этого события, $\bar{\omega}_i$, будет равна

$$\bar{\omega}_i = f(p)\,dp,$$

где $f(p)$ — неизвестная функция от p.

Пусть p_i — вероятность того, что при определенном p происходит данное событие, т. е. из M планет наблюдают m.

$$p_i = \frac{M!}{m!\,(M-m)!} p^m q^{M-m}.$$

Все возможные значения p лежат от 0 до 1. Тогда

$$\frac{\bar{\omega}_i p_i}{\sum \bar{\omega}_i p_i} = \frac{\dfrac{M!}{m!\,(M-m)!} p^m q^{M-m} f(p)\,dp}{\dfrac{M!}{m!\,(M-m)!} \int\limits_0^1 p^m q^{M-m} f(p)\,dp}.$$

Каково вероятное значение N? Умножим числитель на $\frac{n+q}{p}$ и проинтегрируем от 0 до 1. Вероятное значение N равно

$$\overline{N} = \frac{\int\limits_0^1 p^m q^{M-m} \frac{n+q}{p} f(p)\, dp}{\int\limits_0^1 p^m q^{M-m} f(p)\, dp}.$$

Результат зависит от $f(p)$; предположим, что эта функция равна 1, а также заменим $\frac{n+q}{p}$ на $\frac{n+1}{p} - 1$.

$$\overline{N} = \frac{\int\limits_0^1 p^m q^{M-m} \frac{n+1}{p}\, dp - \int\limits_0^1 p^m q^{M-m}\, dp}{\int\limits_0^1 p^m q^{M-m}\, dp}.$$

$$\overline{N} = \frac{(n+1)\int\limits_0^1 p^{m-1} q^{M-m}\, dp}{\int\limits_0^1 p^m q^{M-m}\, dp} - 1.$$

Положим

$$\overline{N} = (n+1)J - 1,$$

где J — отношение двух эйлеровых интегралов.

$$J = \frac{\dfrac{\Gamma(m)\cdot\Gamma(M-m+1)}{\Gamma(M+1)}}{\dfrac{\Gamma(m+1)\cdot\Gamma(M-m+1)}{\Gamma(M+2)}} = \frac{\Gamma(M+2)\Gamma(m)}{\Gamma(M+1)\Gamma(m+1)},$$

$$J = \frac{M+1}{m},$$

и, следовательно,

$$\overline{N} = \frac{(n+1)(M+1)}{m} - 1.$$

Эта величина очень близка к $\frac{Mn}{m}$, как мы и заявляли заранее.

Глава 10
Теория ошибок и среднее арифметическое

107. Я предполагаю, что произведены различные измерения

$$x_1,\, x_2,\, \ldots\,,\, x_n$$

одной и той же величины; с какой вероятностью ее настоящее значение лежит между z и $z + dz$?

Следует ввести закон распределения ошибок. Я предположу, что истинное значение измеряемой величины равно z; какова вероятность того, что результат наблюдения попадет в интервал от x_1 до $x_1 + dx_1$? В любом случае я могу представить эту вероятность как $dx_1 \cdot \varphi(x_1,\, z)$.

Если мы договорились принять это выражение за закон распределения ошибок, то с какой вероятностью z попадет в интервал от z до $z + dz$?

Эта задача на условные вероятности, и мы будем вычислять

$$\frac{\bar{\omega}_i p_i}{\sum \bar{\omega}_i p_i}.$$

Здесь $\bar{\omega}_i$ — априорная вероятность того, что z лежит между z и $z + dz$; эта вероятность представляется как

$$\bar{\omega}_i = \psi(z)\, dz,$$

где ψ — какая-то функция, зависящая от наших сведений о z. Число p_i — это вероятность того, что для наблюдаемой величины, равной z, результаты наблюдений попадут в интервалы от x_1 до $x_1 + dx_1$, от x_2 до $x_2 + dx_2$, ... , от x_n до $x_n + dx_n$. Вероятности этих событий равны

$$dx_1 \cdot \varphi(x_1,\, z),\; dx_2 \cdot \varphi(x_2,\, z),\; \ldots\,,\; dx_n \cdot \varphi(x_n,\, z)$$

соответственно.

Но p_i — вероятность одновременного наступления этих событий; так как все они независимы, p_i равна произведению вероятностей

$$p_i = dx_1 \cdot dx_2 \ldots dx_n \varphi(x_1,\, z)\varphi(x_2,\, z) \ldots \varphi(x_n,\, z).$$

Числитель искомой апостериорной вероятности равен

$$dz\, dx_1\, dx_2 \ldots dx_n\, \psi(z)\varphi(x_1,\, z)\varphi(x_2,\, z) \ldots \varphi(x_n,\, z).$$

Чтобы найти знаменатель $\sum \bar{\omega}_i p_i$, надо проинтегрировать это выражение по z. При делении константы $dx_1\, dx_2 \ldots dx_n$ исчезнут, и останется

$$\frac{dz \cdot \psi(z)\varphi(x_1,\, z)\varphi(x_2,\, z) \ldots \varphi(x_n,\, z)}{\int\limits_{-\infty}^{+\infty} dz \cdot \psi(z)\varphi(x_1,\, z)\varphi(x_2,\, z) \ldots \varphi(x_n,\, z)}.$$

108. Этот результат нам мало что дал бы, если бы у нас не было никаких данных о φ и ψ. Но насчет φ была выдвинута гипотеза, которая называется законом распределения ошибок.

Этот закон не получают с помощью строгих рассуждений; доказательства, которые можно было бы привести, окажутся, кроме всего прочего, грубыми; некоторые опираются на утверждение, что вероятность уклонения пропорциональна самому уклонению. Тем не менее все верят в этот закон. Как мне однажды сказал г. Липпманн, потому что экспериментаторы думают, что это математическое утверждение, а математики — что это результат экспериментов.

Вот как пришел к этому закону Гаусс.

Когда мы ищем, какое значение лучше всего придать z, нам ничего не остается, кроме как взять среднее от $x_1,\, x_2,\, \ldots,\, x_n$, конечно, если нет никаких соображений для обоснования другого выбора. Тогда нужен закон распределения ошибок, соответствующий этому правилу. Гаусс ищет, какой должна быть φ, *чтобы среднее значение оказалось наиболее вероятным.*

109. Если dz — константа, вероятность того, что z лежит в интегрвале dz, равна

$$\psi(z)\varphi(x_1,\, z)\varphi(x_2,\, z) \ldots \varphi(x_n,\, z)\, dz$$

Наиболее вероятным значением будет то, при котором эта функция достигает максимума. Предположим, что максимум достигается при среднем z.

Сначала Гаусс приравнивает ψ к 1, затем полагает, что $\varphi(x_1, z)$ имеет вид $\varphi(z - x_1)$. Какой тогда должна быть функция φ, чтобы максимум

$$\varphi(z - x_1)\varphi(z - x_2) \ldots \varphi(z - x_n)$$

достигался при данном z? Приравняем к нулю логарифмическую производную предыдущего выражения по z:

$$\frac{\varphi'(z - x_1)}{\varphi(z - x_1)} + \frac{\varphi'(z - x_2)}{\varphi(z - x_2)} + \ldots + \frac{\varphi'(z - x_n)}{\varphi(z - x_n)} = 0.$$

Я обозначу

$$\frac{\varphi'(z - x_1)}{\varphi(z - x_1)} = F(x_1).$$

Уравнение, которое надо проверить, превратится в

$$F(x_1) + F(x_1) + \ldots + F(x_n) = 0.$$

Это условие должно выполняться всякий раз, когда

$$x_1 + x_2 + \ldots + x_n = nz.$$

Я придам $x_1, x_2, \ldots, x_n$ приращения $dx_1, dx_2, \ldots, dx_n$; если z остается константой, сумма всех x должна тоже остаться константой, и мы получаем

$$F'(x_1)dx_1 + F'(x_2)dx_2 + \ldots + F'(x_n)dx_n = 0,$$
$$dx_1 + dx_2 + \ldots + dx_n = 0.$$

Эти два уравнения должны быть равносильны, откуда

$$F'(x_1) = F'(x_2) = \ldots = F'(x_n),$$

то есть $F'(x_1)$ — константа, которую я обозначу a.

$$F(x_1) = a(z - x_1) + b,$$

и

$$\log \varphi(z - x_1) = \frac{a(z - x_1)^2}{2} + b(z - x_1) + c.$$

Определим константы a, b, c.

$$F(x_1) + F(x_2) + \ldots + F(x_n) = \sum a(z - x_1) + nb = 0,$$
$$x_1 + x_2 + \ldots + x_n - nz = -\sum (z - x_1) = 0.$$

Так как эти два уравнения должны быть равносильны, получаем

$$b = 0,$$

и можно написать, что

$$\varphi(z - x_1) = e^c \cdot e^{\frac{a(z-x_1)^2}{2}};$$

c находится из условия

$$\int_{-\infty}^{+\infty} \varphi(z - x_1)\, dx_1 = 1.$$

Считая

$$-a = 2h,$$
$$z - x_1 = y,$$

находим

$$\varphi(y) = \sqrt{\frac{h}{\pi}} e^{-hy^2}.$$

110. Ж. Бертран выдвигает следующие возражения.

Функция φ взята в виде $\varphi(z - x_1)$, тогда как на самом деле она должна быть $\varphi(z, x_1)$. Более того, мы предположили, что $\psi(z) = 1$, а этого нельзя утверждать a priori.

Другое возражение: будет ли среднее *наиболее вероятным значением* или *вероятным значением*? Это не одно и то же.

Предположим, что какая-то величина x может принять значения

$$1,\ 2,\ \ldots,\ n-1 \quad \text{или} \quad n,$$

и каждое из них имеет вероятность

$$p_1,\ p_2,\ \ldots,\ p_n,$$

так что

$$p_1 + p_2 + \ldots + p_n = 1.$$

Вероятное значение x по определению равно

$$\bar{x} = p_1 + 2p_2 + \ldots + np_n.$$

Наиболее вероятное значение x будет соответствовать максимальному из чисел p.

В случае задачи об ошибках вероятное значение z совпадет с отношением

$$\frac{\int\limits_{-\infty}^{+\infty} dz \cdot z \cdot \psi(z)\varphi(x_1, z)\varphi(x_2, z)\ldots\varphi(x_n, z)}{\int\limits_{-\infty}^{+\infty} dz \cdot \psi(z)\varphi(x_1, z)\varphi(x_2, z)\ldots\varphi(x_n, z)}.$$

Бертран говорит, что Гаусс должен был искать функцию φ не из условия, что среднее будет наиболее вероятным значением z, а из условия, что среднее будет вероятным значением z.

111. Можно попытаться избавиться от принятых нами гипотез, что $\varphi(x_1, z)$ имеет форму $\varphi(z - x_1)$, а $\psi(z)$ равно 1; можно спросить какой вид нужно придать этим двум функциям, чтобы среднее арифметическое $x_1, x_2, \ldots, x_n$ было наиболее вероятным значением z.

Иначе говоря, при этом среднем арифметическом должен достигаться максимум

$$\psi(z)\varphi(x_1, z)\varphi(x_2, z)\ldots\varphi(x_n, z).$$ (12)

В точке максимума логарифмическая производная равна нулю, т. е., если обозначить

$$\frac{\varphi'(x_1, z)}{\varphi(x_1, z)} = F(x_1, z),$$

$$\frac{\psi'(z)}{\psi(z)} = \chi,$$

должно выполняться

$$F(x_1, z) + F(x_2, z) + \ldots + F(x_n, z) + \chi = 0.$$

Этому равенству должно удовлетворять значение z, определенное из уравнения

$$x_1 + x_2 + \ldots + x_n = nz.$$

Я придам $x_1, x_2, \ldots, x_n$ приращения $dx_1, dx_2, \ldots, dx_n$. Я предполагаю, что z не меняется и последнее равенство по-прежнему верно; тогда χ — константа, и

$$\frac{dF(x_1)}{dx_1}dx_1 + \frac{dF(x_2)}{dx_2}dx_2 + \ldots + \frac{dF(x_n)}{dx_n}dx_n = 0,$$
$$dx_1 + dx_2 + \ldots + dx_n = 0.$$

Это имеет место только при

$$\frac{dF(x_1)}{dx_1} = \frac{dF(x_2)}{dx_2} = \ldots = \frac{dF(x_n)}{dx_n}.$$

Итак,

$$\frac{dF}{dx_1} = A',$$

где A' зависит только от z; и

$$F = A'x_1 + B',$$

B' тоже зависит только от z.

Необходимое нам условие принимает вид

$$A'(x_1 + x_2 + \ldots + x_n) + nB' + \chi = 0,$$

то есть

$$n(A'z + B') + \chi = 0.$$

Это равенство должно быть верным, какими бы ни были z и n; тогда

$$\chi = 0,$$

то есть $\psi(z)$ — константа, и

$$A'z + B' = 0.$$

Вот какими стали бы рассуждения Гаусса, если бы мы захотели провести их, принимая во внимание первое замечание Бертрана.

112. Из $F(x_1, z) = A'x_1 + B'$ легко вывести, что

$$\log \varphi(x_1, z) = Ax_1 + B + \log \theta(x_1).$$

Здесь $\log \theta(x_1)$ является функцией только от x_1; A и B — функции от z, причем производные A' и B' таковы, что

$$A'z + B' = 0.$$

Итак,

$$\varphi(x_1, z) = \theta(x_1) e^{Ax_1 + B}.$$

Таким был бы результат, если не накладывать других условий, кроме постулата Гаусса о среднем значении.

В формуле участвуют две произвольные функции, θ и A; B связано с A некоторым соотношением.

113. Гауссу было сделано и другое замечание. Величина, которую надо было принять за z, не наиболее вероятное значение, а вероятное значение. Действительно, наиболее вероятное значение — это то, которое соответствует самому большому p; оно может отличаться от всех других, тогда как эти другие могут быть очень близки между собой и очень мало отличаться от истинного. Они не влияют на наиболее вероятное значение, но вносят наибольший вклад в вероятное значение, равное по определению

$$x_1 p_1 + x_2 p_2 + \ldots + x_n p_n.$$

Вероятное значение z равно

$$\frac{\int\limits_{-\infty}^{+\infty} z\psi(z)\varphi(x_1, z)\varphi(x_2, z)\ldots\varphi(x_n, z)\,dz}{\int\limits_{-\infty}^{+\infty} \psi(z)\varphi(x_1, z)\varphi(x_2, z)\ldots\varphi(x_n, z)\,dz}.$$

(Две величины под знаком $\int$ отличаются только множителем z в числителе.)

Надо выбрать ψ и φ так, чтобы это вероятное значение было средним:

$$\frac{x_1 + x_2 + \ldots + x_n}{n}.$$

114. Я предположу, что p наблюдений дали результат x_1; p других — результат x_2; ... ; наконец, p последних — результат x_n. Число p везде одно и то же, я считаю его очень большим.

В двух интегралах будет p множителей, равных $\varphi(x_1, z)$, p множителей, равных $\varphi(x_2, z)$, ... , p множителей, равных $\varphi(x_n, z)$.

Я обозначу

$$\Phi = \varphi(x_1, z)\varphi(x_2, z)\ldots\varphi(x_n, z).$$

Речь идет о проверке равенства

$$\frac{\int\limits_{-\infty}^{+\infty} z\psi(z)\Phi^p\,dz}{\int\limits_{-\infty}^{+\infty} \psi(z)\Phi^p\,dz} = \frac{x_1 + x_2 + \ldots + x_n}{n}.$$

Это равенство должно иметь место при сколь угодно большом p.

Если бы вместо двух $\int$ мы имели отношение двух $\sum$, нам бы пришлось рассмотреть дробь

$$\frac{a_1X_1^p + a_2X_2^p + \ldots + a_nX_n^p}{b_1X_1^p + b_2X_2^p + \ldots + b_nX_n^p},$$

где X_1, X_2, ... , X_n были бы функциями от x_1, x_2, ... , x_n, а a_1, a_2, ... , a_n, b_1, b_2, ... , b_n — функциями от z.

Я считаю все величины X положительными. Каков предел этого отношения, когда p бесконечно растет? Пусть X_i — наибольшее из всех X: пределом будет $\frac{a_i}{b_i}$.

Действительно, это соотношение можно записать как

$$\frac{\sum a_k(\frac{X_k}{X_i})^p}{\sum b_k(\frac{X_k}{X_i})^p}.$$

Все дроби $\frac{X_k}{X_i}$ меньше 1, кроме одной, той, что соответствут $k = i$. Итак, когда p бесконечно растет, данное отношение имеет своим пределом $\frac{a_i}{b_i}$.

Распространим этот результат на интегралы

$$\int \varphi_1(z)\Phi^p\,dz \quad \text{и} \quad \int \varphi_2(z)\Phi^p\,dz;$$

Здесь $\varphi_1(z)$ играет ту же роль, что a_i, $\varphi_2(z)$ — ту же, что b_i. Каким будет предел отношения этих интегралов? Пусть z_0 — величина, при которой Φ достигает максимума. Предел будет равен

$$\frac{\varphi_1(z_0)}{\varphi_2(z_0)},$$

то есть

$$\frac{z_0\psi(z_0)}{\psi(z_0)} = z_0.$$

Это z_0 и должно совпадать со средним арифметическим.

115. Мы возвращаемся к предыдущему вопросу: Φ должна достигать максимума, когда z равна среднему арифметическому. Мы знаем, при каких условиях это имеет место:

$$\varphi(x_1,\, z) = \theta(x_1)e^{Ax_1+B},$$

где A' и B' — производные функций A и B по z — связаны соотношением

$$A'z + B' = 0.$$

Когда предполагается, что φ зависит только от разности $z - x_1$, ее логарифмическая производная по z,

$$A'x_1 + B',$$

должна иметь первую степень относительно $z - x_1$; тогда

$$\varphi(z - x_1) = Ce^{A'(z-x_1)^2}.$$

Здесь A' и $C = \theta(x_1)$ — константы.

В общем случае, когда не предполагается, что φ зависит только от $z - x_1$, в $\varphi(x_1,\, z)$ остается определить три функции: во-первых, $\psi(z)$, которую теперь нельзя объявить константой, как было сделано при вычислении наиболее вероятного значения; во-вторых, $\theta(x_1)$, в третьих, A. Что касается B, то есть соотношение, связывающее его с A.

116. Займемся немного более полным определением этих произвольных функций.

Я предположу, что p наблюдений дают результат x_1: их среднее арифметическое будет x_1; тогда

$$\Phi = \theta(x_1)e^{Ax_1+B},$$

и должно выполняться равенство

$$\int z\psi(z)\Phi^p\,dz = x_1\int \psi(z)\Phi^p\,dz;$$

откуда

$$\theta^p\int (z-x_1)\psi(z)e^{p(Ax_1+B)}\,dz = 0.$$

Это соотношение должно быть верным, какими бы ни были p и x_1.

Число θ^p можно было вынести из-под знака $\int$, т. к. оно не содержит z, и мы не сможем определить θ с помощью этого равенства.

117. Ищем $\psi(z)$. Мы знаем, что $Ax_1 + B$ — функция от z, достигающая максимума при $z = x_1$; пусть этот максимум равен u_0^2. Я могу положить

$$Ax_1 + B = u_0^2 - u^2;$$

где u вещественнозначно. Так же

$$\int\limits_{x_1}^{z} (z-x_1)\cdot\psi(z)\,dz$$

— интеграл, который всегда положителен и обращается в нуль только при $z = x_1$; тогда я смогу считать его равным v^2, откуда

$$Ax_1 + B = u_0^2 - u^2,$$
$$(z-x_1)\psi(z)\,dz = 2v\,dv.$$

Чтобы окончательно определить u и v, надо уточнить их знак, т. к. пока мы определили только u^2 и v^2. Мы будем считать u и v положительными, если z больше x_1, и отрицательными иначе; итак, u и v всегда одного знака.

Кроме того,

$$u_0^2 = A(x_1)\cdot x_1 + B(x_1).$$

Наш интеграл с точностью до константы равен

$$\int 2v\,dv\,e^{-pu^2}.$$

Я могу считать, что v выражается как функция от u,

$$2v\,dv = f(u)\,du,$$

и тогда

$$\int f(u)e^{-pu^2}\,du$$

должен быть равен нулю при всех p, если пределы интегрирования равны $-\infty$ и $+\infty$.

118. Это может случиться, только когда $f(u)$ — нечетная функция. Меняя u на $-u$, получим

$$\int f(-u)e^{-pu^2}du = 0,$$

откуда

$$\int [f(u) + f(-u)]e^{-pu^2}\,du = 0.$$

Это соотношение верно при сколь угодно больших p.

Если $f(u)$ нечетно, то

$$f(u) + f(-u) = 0.$$

Если $f(u)$ не является нечетной, я разложу ее в ряд по возрастающим степеням u. Интеграл не сможет равняться 0 при любом p. Действительно,

$$f(u) + f(-u) = \alpha u^{2n} + \beta u^{2n+2} + \ldots$$

Я положу

$$u\sqrt{p} = \xi;$$

интеграл превратится в

$$\int \left(\frac{\alpha\xi^{2n}}{p^n} + \frac{\beta\xi^{2n+2}}{p^{n+1}} + \ldots\right) e^{-\xi^2}\frac{d_\xi}{p^{\frac{1}{2}}},$$

и он должен быть тождественно равен нулю.

Если мы умножим все на $p^{n+\frac{1}{2}}$, первое слагаемое не будет содержать p, а другие будут; интеграл с множителем, близким к $p^{n+\frac{1}{2}}$, при очень больших p почти совпадет с

$$\alpha \int\limits_{-\infty}^{+\infty} e^{-\xi^2} \xi^{2n} \, d\xi,$$

а это не нуль.

Итак, $f(u)$ должна быть нечетной функцией от u.

119. Мы полагали

$$(z - x_1)\psi(z) \, dz = f(u) \, du.$$

Дифференцируем, считая x константой, другое уравнение для u:

$$Ax_1 + B = u_0^2 - u^2;$$

получаем

$$dz(A'x_1 + B') = -2u \, du;$$

или, помня о соотношении

$$A'z + B' = 0,$$
$$A'x_1 + B' = A'x_1 - A'z = -A'(z - x_1).$$

Итак,

$$A(z - x_1) \, dz = 2u \, du,$$

откуда

$$\frac{\psi(z)}{A'} = \frac{f(u)}{2u}.$$

Функция $\dfrac{f(u)}{u}$ — четная, значит, $\dfrac{\psi(z)}{A'}$ не должна меняться при замене u на $-u$.

Между тем $\dfrac{\psi(z)}{A'}$ — функция не от u, а от z, не зависящая от x; я заявляю, что она должна вырождаться в константу.

Действительно, я рассмотрю два каких-то значения z, равных z_1 и z_2, при которых A и B равны A_1 и B_1, A_2 и B_2 соответственно; я выберу x_1 так, что

$$A_1x_1 + B_1 = A_2x_2 + B_2;$$

значение $u_0^2 - u^2$ будет одним и тем же при z_1 и z_2.

Тогда $\dfrac{f(u)}{u}$, которая зависит только от u^2, тоже будет одинаковой, и

$$\frac{\psi(z_1)}{A'(z_1)} = \frac{\psi(z_2)}{A'(z_2)}.$$

Следовательно, $\dfrac{\psi(z)}{A'}$ — константа.

120. Таким образом, наиболее общий способ удовлетворить постулату Гаусса (измененному согласно возражению Ж. Бертрана насчет того, что среднее совпадает с вероятным значением) приводит к

$$\varphi(x_1,\, z) = \theta(x_1) e^{-\int \psi(z)(z-x_1)\,dz}.$$

121. Рассмотрим

$$\int\limits_{-\infty}^{+\infty} \Pi \left(z - \frac{x_1 + x_2 + \ldots + x_n}{n} \right),$$

где

$$\Pi = \psi(z)\varphi(x_1,\, z)\varphi(x_2,\, z)\ldots\varphi(x_n,\, z)dz.$$

Я заявляю, что этот интеграл равен нулю, т. е. среднее равно вероятному значению.

Я обозначу

$$\frac{x_1 + x_2 + \ldots + x_n}{n} = x,$$

и

$$\varphi(x_1,\, z)\varphi(x_2,\, z)\ldots\varphi(x_n,\, z) = \theta(x_1)\theta(x_2)\ldots\theta(x_n)e^p.$$

Показатель p можно записать как

$$p = -\int \psi(z)[(z - x_1) + (z - x_2) + \ldots + (z - x_n)]\,dz,$$

то есть

$$p = -n \int \psi(z)(z - x)\,dz.$$

Теперь остается показать, что нулю равен следующий интеграл:

$$\int \theta(x_1)\theta(x_2)\ldots\theta(x_n)(z - x)\psi(z)e^{-n\int \psi(z)(z-x)dz}dz.$$

Если мы положим

$$\int (z-x)\psi(z)\,dz = u^2,$$

откуда

$$(z-x)\psi(z)\,dz = 2u\,du,$$

то данный интеграл примет вид

$$2\theta(x_1)\theta(x_2)\ldots\theta(x_n)\int ue^{-nu^2}du.$$

Он равен нулю, если u изменяется от $-\infty$ до $+\infty$.

122. Итак, функция φ зависит от ψ, а ψ зависит от сведений, которые мы можем иметь a priori о поведении z.

Кроме того, φ зависит от умения наблюдателя и априорной вероятности его ошибки.

Нет никаких оснований считать, что эти две априорные вероятности зависят друг от друга. Следовательно, единственной разумной гипотезой при нахождении закона Гаусса будет предположить, что $\psi = 1$.

123. Остается $\theta(x_1)$. Ничто не заставляет считать эту функцию равной 1. Например, известно, что некоторые наблюдения, вроде астрономических или меридиональных, подвержены особым ошибкам, которые называются *десятичными.*

Когда измеряют какую-то величину, когда снимают показания приборов, результат оценивается до какого-то порядка, и найденное число будет самым близким, с точностью до этого порядка, к величине, которую хотели узнать.

Между тем замечено, что каждый наблюдатель кажется расположенным к каким-то конкретным десятичным знакам; этот факт можно выразить аналитически, сказав, что $\theta(x_1)$ периодично и достигает максимума на этих десятичных знаках.

124. Какое мнение теперь должно сложиться о постулате Гаусса? Сказать, что он принят всеми, не значит обосновать его, т. к. может оказаться, что ни у кого нет достаточных сведений о том, что такое закон распределения ошибок.

Если бы мы применили те же рассуждения к z^2,

$$z^2 = \frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n},$$

то приемлемое значение z было бы

$$z = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}}.$$

Такое может случиться в серии измерений квадрата неизвестной величины, а по постулату Гаусса надо брать среднее из n найденных с помощью наблюдений величин.

Ж. Бертран приводит пример со стрелкой, показывающей квадрат измеряемого угла. Должен ли он брать среднее по показаниям стрелки, т. е. среднее квадратов углов, или надо взять среднее самих углов? Ни одно из этих решений не было бы разумным. При измерении каждого угла совершаются две ошибки:

1° ошибка в показаниях прибора, и вероятная ошибка в показаниях прибора была бы равна среднему значению угла;

2° ошибка отсчета, а вероятная ошибка отсчета была бы равна среднему значению квадрата угла.

125. Теперь правило среднего кажется лишенным смысла. Однако почему же оно почти никогда не подводит нас? Почему брать среднее оказывается правильным? ***Прежде всего потому, что ошибки очень малы.***

Если вместо z я измеряю $f(z)$ и применяю к $f(z)$ постулат Гаусса,

$$f(z) = \frac{f(x_1) + f(x_2) + \ldots + f(x_n)}{n},$$

то из-за того, что x_1 очень близко к z, я получу

$$f(x_1)f(z) + (x_1 - z)f'(z);$$

то же для $x_2, \ldots, x_n$. Отсюда я выведу

$$f(z) = f(z) + \frac{\sum(x_1 - z)f'(z)}{n},$$

то есть

$$\sum(x_1 - z) = 0$$

или $nz = x_1 + x_2 + \ldots + x_n$.

Итак, мы приходим к одному и тому же результату, измеряем ли мы непосредственно величину z или какую-нибудь функцию $f(z)$ от этой величины.

126. С другой стороны, будет ли правильно ограничиваться нахождением среднего? Будет ли этот принцип неоспорим?

Если в n наблюдениях $n-1$ результатов оказались очень близки друг к другу, а n-й очень далек от них, следует ли брать среднее? Результат бы очень отличался от центра тяжести первых $n-1$ наблюдений, $n-1$ хороших наблюдений. Какие-то экспериментаторы получили отклонение в n-й раз: это, можно сказать, случай, это плохой результат.

Но тогда мы возьмем величину, которая уже не будет средней; нашлись основания отбросить постулат.

Когда принимается закон Гаусса, вероятная ошибка в среднем равна $\frac{1}{\sqrt{n}}$; так что при увеличении числа наблюдений мы будем получать все бо́льшую точность. Однако, измеряя один метр миллион миллионов раз без верньера[1], мы никогда не оценим его с точностью до миллиметра, до микрона.

Впрочем, это объяснимо: при наблюдении очень малых величин нельзя отвечать ни за что; у гипотезы, что ошибка лежит в промежутке от 0 до 1 микрона нет никаких преимуществ перед гипотезой, что ошибка лежит в промежутке от 1 до 2 микронов.

127. В XI главе я, кроме того, установлю следующую теорему: *Когда берется среднее, квадрат допущенной ошибки равен*

$$\left(z - \frac{x_1 + x_2 + \ldots + x_n}{n}\right)^2.$$

Гаусс показывает, что вероятное значение этого выражения стремится к нулю, когда n неограниченно возрастает, каков бы ни был закон распределения ошибок.

Этот факт обосновывает выбор среднего: оно становится все более и более вероятным по мере роста n, не будучи самым вероятным.

Но этот способ оправдать выбор среднего *независимо от закона распределения ошибок*, является, так сказать, опровержением рассуждений Гаусса, изложенных выше, т. к. эти рассуждения должны были установить, что какой-то очень частный закон — *единственный*, способный оправдать одинаковое использование среднего на практике.

[1] Верньер — механическое устройство, позволяющее осуществлять точную настройку измерительных приборов. — *Прим. ред.*

Достаточно странно, что этим опровержением мы обязаны самому Гауссу.

Глава 11
Обоснование закона Гаусса

128. Мы будем рассуждать в следующем порядке:

1° найдем, как записывается вероятное значение p-й степени ошибки, если верен закон Гаусса;

2° покажем, что такое выражение вероятного значения существует только для этого закона;

3° найдем выражение вероятного значения для произвольного вероятностного закона;

4° снова найдем это выражение в случае, когда результирующая ошибка равна сумме многих частичных ошибок, не зависящих друг от друга;

5° найдем, каким станет это выражение, когда частичные ошибки очень многочисленны и очень малы;

6° получив то же выражение, что и для закона Гаусса, сделаем вывод, что закон Гаусса должен быть верен всякий раз, когда результирующая ошибка складывается из очень малых, очень многочисленных и независимых частичных;

7° получим тот же результат другим способом.

129. Пусть измеряется величина z. Вероятность того, что результат измерения лежит между x_1 и $x_1 + dx_1$, можно представить как

$$\varphi(x_1,\, z)\, dx_1.$$

Как уже было сказано, Гаусс предполагает, что φ зависит только от $z - x_1$; тогда вероятность будет равна $\varphi(z - x_1)dx_1$; более того, он предполагает, что систематических ошибок нет, т. е. φ — четная функция и не меняется при замене $x_1 - z$ на $z - x_1$.

Пусть y_1 — ошибка;

$$y_1 = x_1 - z.$$

Вероятность равна $\varphi(y_1)dy_1$. Нам надо будет рассмотреть вероятное значение y_1 и, в более общем случае, y_1^p; оно будет равно

$$\int\limits_{-\infty}^{+\infty} \varphi(y_1)y_1^p\,dy_1.$$

Так как φ — четная функция, этот интеграл равен нулю при нечетных p.

130. Можно сделать два наблюдения, y_1 и y_2, и рассматривать вероятное значение какой-нибудь функции от y_1 и y_2, например, $y_1^{m_1}y_2^{m_2}$.

Пусть $\varphi(y_1)$ — вероятность того, что первая ошибка лежит между y_1 и y_1+dy_1; $\varphi(y_2)$ — вероятность того, что вторая ошибка лежит между y_2 и y_2+dy_2.

Вероятное значение $y_1^{m_1}y_2^{m_2}$ по определению равно

$$\iint y_1^{m_1}y_2^{m_2}\varphi(y_1)\varphi(y_2)\,dy_1\,dy_2,$$

интегралы по y_1 и по y_2 взяты от $-\infty$ до $+\infty$.

Так как функция под знаком $\iint$ равна произведению функции от y_1 на функцию от y_2, а все пределы интегрирования постоянны, наш двойной интеграл будет произведением двух простых:

$$\int y_1^m\varphi(y_1)\,dy_1 \int y_2^m\varphi(y_2)\,dy_2.$$

Равенство интегралов показывает, что вероятное значение произведения равно произведению вероятных значений.

Конечно при этом два множителя должны быть различны: вероятное значение y_1^4, например, не будет квадратом вероятного значения y_1^2; но вероятное значение $y_1^2y_2^2$ равно произведению вероятных значений y_1^2 и y_2^2.

131. Пусть m_2 — нуль. В интеграле

$$\int\limits_{-\infty}^{+\infty} \varphi(y_2)\,dy_2$$

мы должны получить единичное вероятное значение, т. е. 1; между тем очевидно, что

$$\int\limits_{-\infty}^{+\infty} \varphi(y_2)\, dy_2 = 1,$$

так как этот интеграл выражает вероятность того, что y_2 лежит между $-\infty$ и $+\infty$, т. е. вероятность достоверного события.

132. Если произведено много наблюдений $y_1, y_2, \ldots, y_n$, то вероятное значение какой-то функции ψ от этих наблюдений равно

$$\int \varphi(y_1)\varphi(y_2)\ldots\varphi(y_n)\psi(y_1, y_2, \ldots, y_n)\, dy_1\, dy_2 \ldots dy_n.$$

Если функция ψ нечетна, симметричные части интеграла будут попарно равны и противоположны по знаку; такой интеграл равен нулю.

133. Я возвращаюсь к гипотезе Гаусса:

$$\varphi(y) = \sqrt{\frac{h}{\pi}} e^{-hy^2}.$$

Тогда вероятное значение y^{2p+1} равно нулю; ищем вероятное значение y^{2p}: по определению это

$$\int\limits_{-\infty}^{+\infty} \sqrt{\frac{h}{\pi}} e^{-hy^2} y^{2p} dy.$$

Сначала заметим, что

$$\int\limits_{-\infty}^{+\infty} \sqrt{\frac{h}{\pi}} e^{-hy^2} dy = 1$$

или

$$\int e^{-hy^2} dy = \sqrt{\pi} \cdot h^{-\frac{1}{2}}.$$

Я дифференцирую это равенство p раз по h:

$$\int e^{-hy^2} (-y^2)^p dy = \sqrt{\pi} h^{-p-\frac{1}{2}} \left(-\frac{1}{2}\right) \left(-\frac{3}{2}\right) \left(-\frac{5}{2}\right) \cdots \left(-\frac{2p-1}{2}\right).$$

Знак «$-$» повторяется p раз справа и слева; следовательно, он исчезает:

$$\int e^{-hy^2} y^{2p} dy = \sqrt{\pi} h^{-p-\frac{1}{2}} \frac{1 \cdot 3 \cdot 5 \ldots (2p-1)}{2^p}.$$

С другой стороны,

$$2 \cdot 4 \cdot 6 \ldots 2p = 2^p \cdot p!,$$

$$\int e^{-hy^2} y^{2p} dy = \sqrt{\pi} h^{-p-\frac{1}{2}} \frac{1}{2^{2p}} \frac{(2p)!}{p!}.$$

Вероятное значение y^{2p} есть произведение этого интеграла и $\sqrt{\frac{h}{\pi}}$; итак,

$$\overline{y^{2p}} = \frac{1}{h^p} \frac{(2p)!}{p!\, 2^{2p}}.$$

134. Этот результат получен при гипотезе Гаусса:

$$\varphi(y) = \sqrt{\frac{h}{\pi}}\, e^{-hy^2};$$

возникает вопрос: будет ли закон Гаусса единственным, дающим такой результат? Да, он единственен.

Ищем вероятное значение

$$e^{-n(y_0-y)^2}.$$

По определению оно равно

$$\int\limits_{-\infty}^{+\infty} \sqrt{\frac{h}{\pi}} e^{-hy^2} e^{-n(y_0-y)^2} dy.$$

Этот интеграл можно вычислить непосредственно; также можно разложить $e^{-n(y_0-y)^2}$ в ряд, сходящийся при всех значениях y,

$$e^{-n(y_0-y)^2} = \sum A_p y^p.$$

Отсюда

$$\int\limits_{-\infty}^{+\infty} \sqrt{\frac{h}{\pi}} e^{-hy^2} e^{-n(y_0-y)^2} dy = \sum A_{2p} \overline{y^{2p}}.$$

Точно также и для другой функции φ

$$\int\limits_{-\infty}^{+\infty} \varphi(y) e^{-n(y_0-y)^2} dy = \sum A_{2p} \overline{y^{2p}}.$$

По нашей гипотезе средние значения y^{2p} одинаковы в обоих случаях. Тогда отношение двух интегралов равно 1.

Так как это отношение остается одним и тем же, каким бы ни было n, воспользуемся доказанной теоремой.

Предел отношения

$$\frac{\int f_1(y)\varphi^n(y)\,dy}{\int f_2(y)\varphi^n(y)\,dy}$$

при бесконечно больших n равен

$$\frac{f_1(y_0)}{f_2(y_0)},$$

если в пределах интегрирования $\varphi(y)$ достигает максимума при $y = y_0$.

В нашем случае $e^{-n(y_0-y)^2}$ достигает максимума при $y = y_0$; предел отношения двух интегралов равен

$$\frac{\sqrt{\frac{h}{\pi}}\, e^{-hy_0^2}}{\varphi(y_0)}.$$

Так как это отношение всегда остается равным 1,

$$\varphi(y_0) = \sqrt{\frac{h}{\pi}}\, e^{-hy_0^2},$$

где y_0 было взято совершенно произвольно, т. е. закон Гаусса — единственный, при котором вероятное значение y^{2p} совпадает с приведенным выше.$^{(13)}$

135. Предположим, что имеется какой-то закон распределения ошибок.

Пусть было сделано n наблюдений, которые дали n индивидуальных ошибок $y_1, y_2, \ldots, y_n$. Возьмем среднее от этих наблюдений: мы допускаем ошибку, равную среднему от индивидуальных ошибок:

$$\frac{y_1 + y_2 + \ldots + y_n}{n}.$$

Гаусс предполагал вычислять вероятное значение квадрата этой ошибки; по определению это интеграл

$$\int \varphi(y_1)\varphi(y_2)\ldots\varphi(y_n)\left(\frac{y_1+y_2+\ldots+y_n}{n}\right)^2 dy_1\, dy_2\, \ldots\, dy_n.$$

Я разложу квадрат:

$$\left(\frac{y_1+y_2+\ldots+y_n}{n}\right)^2 = \frac{\sum y_1^2}{n^2} + \frac{2\sum y_1 y_2}{n^2}.$$

Искомое вероятное значение равно

$$\frac{\sum \overline{y_1^2}}{n^2} + \frac{2\sum \overline{y_1 y_2}}{n^2}.$$

Вероятное значение произведения $y_1 y_2$ равно $\bar{y}_1 \times \bar{y}_2$; так как функции y_1 и y_2 нечетны, $\bar{y}_1$ и $\bar{y}_2$ равны нулю. Остается

$$\frac{\bar{y}_1+\bar{y}_2+\ldots+\bar{y}_n}{n}.$$

Интеграл сводится к сумме n интегралов; но в каждом из них участвует одна и та же функция

$$\varphi(y_1)\varphi(y_2)\ldots\varphi(y_n),$$

умноженная на $y_1^2\, dy_1$, или $y_2^2\, dy_2$, ... , или $y_n^2\, dy_n$. Это один и тот же интеграл с точностью до обозначений.

В итоге вероятное значение квадрата ошибки равно

$$\frac{n\bar{y}_1^2}{n^2} = \frac{\bar{y}_1^2}{n}.$$

Таким образом, вероятное значение квадрата допущенной ошибки равно квадрату вероятного значения индивидуальной ошибки, деленному на n.

Этого свойства достаточно для обоснования метода средних; это свойство выполнено, каким бы ни был закон распределения ошибок.

В любом случае, как мы видели в предыдущей главе, только для закона Гаусса среднее будет наиболее вероятной величиной.

Для любого другого вероятностного закона среднее будет *все более и более вероятным*, когда наблюдений проводится все больше и больше, но оно так и не станет наиболее вероятным значением.

136. Ищем вероятное значение

$$\left(\frac{y_1 + y_2 + \ldots + y_n}{n}\right)^{2p+1};$$

это нечетная функция, которую мы возводим в нечетную степень: вероятное значение должно равняться нулю.

Ищем вероятное значение

$$\left(\frac{y_1 + y_2 + \ldots + y_n}{n}\right)^{2p}.$$

По формуле, обобщающей формулу бинома,

$$(y_1 + y_2 + \ldots + y_n)^{2p} = \sum \frac{(2p)!}{\alpha_1!\,\alpha_2!\,\ldots\alpha_\mu!} y_1^{\alpha_1} y_2^{\alpha_2} \ldots y_\mu^{\alpha_\mu},$$

где

$$\alpha_1 + \alpha_2 + \ldots + \alpha_\mu = 2p.$$

Надо взять среднее значение каждого слагаемого и разделить его на n^{2p}. Среднее значение выражений, у которых один из показателей α нечетный, равно нулю. Для того чтобы это выражение оказалось отличным от нуля, все показатели должны быть четными.

137. В качестве примера возьмем

$$(y_1 + y_2 + \ldots + y_n)^4.$$

Все α могут быть четными только в двух случаях:

1° $\alpha_1 = 4$, остальные нули;

2° $\alpha_1 = \alpha_2 = 2$, что приводит к сумме

$$\sum y_1^4 + 6 \sum y_1^2 y_2^2 + R.$$

Здесь R — сумма членов с нулевым средним значением; коэффициент при $\sum y_1^2 y_2^2$ равен $\frac{4!}{2!\,2!} = 6$.

Рассмотрим еще

$$(y_1 + y_2 + \ldots + y_n)^6.$$

Все α могут быть четными только в трех случаях:

1° $\alpha_1 = 6$;
2° $\alpha_1 = 4,\ \alpha_2 = 2$;
3° $\alpha_1 = \alpha_2 = \alpha_3 = 2$.

Если R — сумма членов с нулевым средним значением, то

$$(y_1 + y_2 + \ldots + y_n)^6 = \sum y_1^6 + 15 \sum y_1^4 y_2^2 + 90 \sum y_1^2 y_2^2 y_3^2 + R.$$

Коэффициент при $\sum y_1^4 y_2^2$ равен $\frac{6!}{4!\,2!} = 15$; коэффициент при $\sum y_1^2 y_2^2 y_3^2$ равен $\frac{6!}{2!\,2!\,2!} = 90$.

Условимся обозначать M_p среднее значение y_1^p. Во-первых, среднее значение $(y_1 + y_2 + \ldots + y_n)^4$ равно

$$\overline{(y_1 + y_2 + \ldots + y_n)^4} = nM_4 + 6\frac{n(n-1)}{2}M_2^2.$$

Действительно, в $\sum y_1^4$ все члены имеют одно и то же среднее значение, а их n штук; в $\sum y_1^2 y_2^2$ все члены также имеют одинаковое среднее, и их $\frac{n(n-1)}{2}$ штук.

Среднее значение $(y_1 + y_2 + \ldots + y_n)^6$ будет равно

$$\overline{(y_1 + y_2 + \ldots + y_n)^6} = nM_6 + 15n(n-1)M_4M_2 + 90\frac{n(n-1)(n-2)}{6}M_3^2.$$

В $\sum y_1^4 y_2^2$ столько слагаемых, сколько есть *размещений* из n букв по 2; тогда коэффициент при $M_4 M_2$ равен

$$15n(n-1).$$

138. Можно продолжить вычисления для других значений $2p$; выражения станут все более и более сложными.

Примем во внимание, что n очень велико; в правой части будут слагаемые с n, n^2, n^3 и т. д.

При приближенных вычислениях можно рассматривать только слагаемые с самой высокой степенью n. Для $2p = 4$ это $3n^2M_3^2$; для $2p = 6$ это $15n^3M_2^3$. Итак, средняя ошибка

$$y = \frac{y_1 + y_2 + \ldots + y_n}{n},$$

возведенная в четвертую степень, имеет вероятное значение

$$\bar{y}^4 = 3\left(\frac{M_2}{n}\right)^2,$$

для шестой степени получаем

$$\bar{y}^6 = 15\left(\frac{M_2}{n}\right)^3.$$

Найдем это вероятное значение в общем случае:

$$(y_1 + y_2 + \ldots + y_n)^{2p} = \sum\left[\frac{(2p)!}{\alpha_1!\,\alpha_2!\,\ldots\alpha_\mu!}\sum y_1^{\alpha_1} y_2^{\alpha_2}\ldots y_\mu^{\alpha_\mu}\right].$$

Во второй $\sum$ переставляются только индексы при y.

Сохраним лишь те слагаемые, у которых все α четны; вероятное значение остальных равно нулю. Получаем

$$\overline{(y_1 + y_2 + \ldots + y_n)^{2p}} = \sum\frac{(2p)!}{\alpha_1!\,\alpha_2!\,\ldots\alpha_\mu!} N M_{\alpha_1} M_{\alpha_2}\ldots M_{\alpha_\mu}.$$

Действительно, у всех слагаемых во второй $\sum$,

$$\sum y_1^{\alpha_1} y_2^{\alpha_2}\ldots y_\mu^{\alpha_\mu},$$

одно и то же вероятное значение. Пусть N — их количество; оцениваем N.

139. Сначала я предположу, что $\alpha_1 = \alpha_2$. Если мы поменяем местами y_1 и y_2, получим то же самое слагаемое. Если бы мы учитывали порядок, то у нас было бы столько слагаемых, сколько есть размещений по 2 из μ букв, выбранных из n букв $y_1, y_2, \ldots, y_n$; тогда бы наши слагаемые повторялись.

Если бы все μ показателей были различны, то N было бы равно числу размещений из n букв по μ, т. е. $\frac{n!}{(n-\mu)!}$.

Я считаю μ_1 показателей равными α_1, μ_2 равными $\alpha_2, \ldots, \mu_k$ равными α_k; кроме того, я считаю, что $\alpha_1, \alpha_2, \ldots, \alpha_k$ различны, и

$$\mu_1 + \mu_2 + \ldots + \mu_k = \mu.$$

Я рассматриваю одно из размещений с такими показателями; я каким-то образом меняю местами μ_1 букв с показателем α_1, μ_2 букв с показателем α_2, ... , μ_k букв с показателем α_k. Соответствующее слагаемое не меняется; следовательно, одно и то же слагаемое повторяется в

$$\mu_1! \cdot \mu_2! \cdot \ldots \cdot \mu_k!$$

размещениях. Тогда

$$N = \frac{n!}{(n-\mu)!\,\mu_1! \cdot \mu_2! \cdot \ldots \cdot \mu_k!},$$

$$N = \frac{n(n-1)\ldots(n-\mu+1)}{\mu_1! \cdot \mu_2! \cdot \ldots \cdot \mu_k!}.$$

Значит, N — многочлен степени μ от n. Наибольшее значение, которое может принять μ, равно p. Действительно,

$$\alpha_1 + \alpha_2 + \ldots + \alpha_\mu = 2p.$$

Все α четны, и максимальное значение μ соответствует набору

$$\alpha_1 = \alpha_2 = \ldots = \alpha_\mu = 2.$$

Итак, среди N есть только одно слагаемое степени p относительно n; при нашем порядке аппроксимации мы должны сохранить только его. Для этого слагаемого N равно

$$N = \frac{n(n-1)\ldots(n-p+1)}{p!}.$$

Действительно, если все α совпадают, получается, что все $\mu_1, \mu_2, \ldots, \mu_k$ сводятся к

$$\mu_1 = \mu = p, \qquad \mu_2 = \ldots = \mu_k = 0.$$

В записи числа N слагаемое с n^p равно

$$\frac{n^p}{p!}.$$

140. С другой стороны, $\alpha_1, \alpha_2, \ldots, \alpha_\mu$ равны 2, и

$$M_{\alpha_1} = M_{\alpha_2} = \ldots = M_{\alpha_\mu} = M_2,$$

где M_2 — вероятное значение квадрата индивидуальной ошибки, $\bar{y}_1^2$.

Итак, вероятное значение $(y_1 + y_2 + \ldots + y_n)^{2p}$ равно

$$\overline{(y_1 + y_2 + \ldots + y_n)^{2p}} = \frac{(2p)!}{2^p}\frac{n^p}{p!}M_2^p.$$

Вероятное значение y^{2p} найдем, разделив все на n^{2p}:

$$\overline{y^{2p}} = \frac{(2p)!}{p!\,2^p}\left(\frac{M_2}{n}\right)^p.$$

Сравниваем с результатом, полученным для закона Гаусса; должно выполняться равенство

$$\left(\frac{1}{2h}\right)^p = \left(\frac{M_2}{n}\right)^p,$$

или $h = \dfrac{n}{2M_2}$.

Только закон Гаусса приводит к этому выражению вероятного значения ошибки.

Если заранее знать, что систематических ошибок нет и производится очень много наблюдений, то, взяв их среднее, получим ошибку, вероятность которой подчиняется закону Гаусса.

Ошибка, допущенная при использовании инструментов, является результатом очень большого числа малых ошибок, не зависящих друг от друга, таких, что вклад каждой из них в конечное значение очень мал; результирующая ошибка подчиняется закону Гаусса(14).

141. Поставим задачу по-другому.

В наблюдениях допущено несколько индивидуальных ошибок, y_1, $y_2, \ldots, y_n$, не зависящих друг от друга; полная ошибка равна

$$y = y_1 + y_2 + \ldots + y_n.$$

Предположим сначала, что все эти ошибки подчиняются одному и тому же закону, и систематических среди них нет. Получаем ту же самую задачу.

Вероятное значение y^{2p+1} будет нулем.

В параграфе 136 мы оценили вероятное значение

$$\left(\frac{y_1 + y_2 + \ldots + y_n}{n}\right)^{2p}.$$

Здесь мы будем искать вероятное значение

$$(y_1 + y_2 + \ldots + y_n)^{2p}.$$

Вероятность того, что y_i лежит между двумя данными числами α и β, то есть

$$\int\limits_{\alpha}^{\beta} \varphi(y_i)\, dy_i,$$

согласно гипотезе, одинакова для $y_1, y_2, \ldots, y_n$.

Нам достаточно умножить полученный для вероятного значения

$$(y_1 + y_2 + \ldots + y_n)^{2p}$$

результат, т. е.

$$\frac{(2p)!}{p!}\left(\frac{M_2}{2n}\right)^{p},$$

на n^{2p}, чтобы получить вероятное значение y^{2p}

$$\overline{y^{2p}} = \frac{(2p)!}{p!}\left(\frac{nM_2}{2}\right)^{p},$$

и, полагая $nM_2 = M$,

$$\overline{y^{2p}} = \frac{(2p)!}{p!}\left(\frac{M_2}{2}\right)^{p}.$$

Таким образом, вид этого выражения остается прежним, и, рассуждая как раньше, приходим к выводу, что вероятность, с которой полная ошибка y лежит в данных пределах, вновь подчиняется закону Гаусса.

142. Это рассуждение все еще не будет удовлетворительным, т. к. мало правдоподобно, что все индивидуальные ошибки подчинены одному закону. Предположим, что это не так, но все индивидуальные ошибки имеют практически одинаковый порядок малости, и вклад каждой из них в полную ошибку очень мал; пусть

$\int\limits_\alpha^\beta \varphi_1(y_1)\,dy_1$ — вероятность того, что y_1 лежит между α и β;

$\int\limits_\alpha^\beta \varphi_2(y_2)\,dy_2$ — вероятность того, что y_2 лежит между α и β;

. . . .

$\int\limits_\alpha^\beta \varphi_n(y_n)\,dy_n$ — вероятность того, что y_n лежит между α и β;

Я предполагаю, что все φ_1, φ_2, ... , φ_n — четные функции, иначе говоря, что систематических ошибок нет.

Я возьму M — сумму средних значений квадратов ошибок,

$$\bar{y}_1^2 + \bar{y}_2^2 + \ldots + \bar{y}_n^2 = M.$$

Если бы все частичные ошибки подчинялись одному вероятностному закону, величина M_2 для $\bar{y}_1^2$, $\bar{y}_2^2$, ... , $\bar{y}_n^2$ была бы одной и той же, а сумма M была бы равна nM_2.

Я ограничусь вычислениями, соответствующими нескольким первым степеням $\bar{y}$.

Предполагая, что все ошибки подчинены одному закону, мы нашли, что средняя величина для y^2 равна M; для y^4, $3M^2$; для y^6, $15M^3$ ...

Проведем те же вычисления, считая законы разными для разных индивидуальных ошибок. Я замечаю, что вероятное значение произведения $y_1^m y_2^n$ — это вероятное значение y_1^m, умноженное на вероятное значение y_2^n. Действительно, вероятное значение произведения $y_1^m y_2^n$ представляется как

$$\iint \varphi_1(y_1)\varphi_2(y_2)y_1^m y_2^n\,dy_1\,dy_2,$$

а произведение вероятных значений y_1^m и y_2^n представляется как

$$\int \varphi_1(y_1)y_1^m\,dy_1 \int \varphi_2(y_2)y_2^n\,dy_2.$$

Значит, эта теорема будет верна, как и в предыдущем случае; если m нечетно, вероятное значение равно нулю.

Надо отметить, что разница заключается в том, что вероятные значения y_1^m, y_2^m, ... , y_n^m больше не равны между собой.

Заметим также, что величины φ_1, φ_2, ... , φ_n имеют одинаковый порядок.

143. Ищем среднее значение y^2. Имеем

$$y^2 = \sum y_1^2 + 2\sum y_1 y_2.$$

Итак,

$$\bar{y}^2 = \sum \bar{y}_1^2 + 2\sum \bar{y}_1 \bar{y}_2;$$

последнее слагаемое исчезает; остается

$$\bar{y}^2 = \sum \bar{y}_1^2.$$

Для среднего значения y^4 находим

$$\bar{y}^4 = \sum \bar{y}_1^4 + 6\sum \bar{y}_1^2 \bar{y}_2^4;$$

мы отбросили члены с четными показателями.

Второе слагаемое будет гораздо больше первого: первая $\sum$ состоит из n членов, вторая $\sum$ — из $\frac{n(n-1)}{2}$; это числа порядка n и n^2 соответственно; первая сумма пренебрежимо мала по сравнению со второй. Различные члены этих двух $\sum$, между тем, по гипотезе, очень малы и практически одного и того же порядка. С другой стороны,

$$M^2 = \sum (\bar{y}_1^2)^2 + 2\sum \bar{y}_1^2 \bar{y}_2^2$$

или

$$3M^2 = 3\sum (\bar{y}_1^2)^2 + 6\sum \bar{y}_1^2 \bar{y}_2^2.$$

Первое слагаемое по-прежнему пренебрежимо мало по сравнению со вторым, а второе одинаково в обоих выражениях.

Итак, учитывая, что вычисления производятся приближенно, получим

$$\bar{y}^4 = 3M^2.$$

Для среднего значения y^6 имеем

$$\bar{y}^6 = \sum \bar{y}_1^6 + 15 \sum \bar{y}_1^4 \bar{y}_1^2 + 90 \sum \bar{y}_1^2 \bar{y}_2^2 \bar{y}_3^2.$$

С другой стороны, найдем $15M^3$:

$$15M^3 = 15 \sum (\bar{y}_1^2)^3 + 45 \sum (\bar{y}_1^2)^2 \bar{y}_2^2 + 90 \sum \bar{y}_1^2 \bar{y}_2^2 \bar{y}_3^2.$$

Сравниваем две правые части. В первой $\sum$ есть n членов, во второй $\sum$ — $n(n-1)$, в третьей $\sum$ их $\dfrac{n(n-1)(n-2)}{1 \cdot 2 \cdot 3}$; эти числа имеют порядки n, n^2, n^3, и первые две $\sum$ пренебрежимо малы по сравнению с третьей. Выражение, которым нельзя пренебречь, одинаково, поэтому

$$\bar{y}^4 = 15M^3.$$

144. Итак, мы предполагаем, что конечная ошибка является результатом очень большого числа частичных, не зависящих друг от друга, среди которых нет систематических; предполагаем также, что вклад каждой из этих ошибок, имеющих практически один и тот же порядок, в полную ошибку очень мал.

В этом случае *результирующая ошибка практически подчиняется закону Гаусса.*

Таково, как мне кажется, наилучшее основание для применения закона Гаусса.

Характеристические функции. — Я назову *характеристической функцией* $f(\alpha)$ вероятное значение $e^{\alpha x}$; таким образом, это

$$f(\alpha) = \sum p e^{\alpha x},$$

если величина x дискретна и принимает только конечное число значений, и

$$f(\alpha) = \int \varphi(x) e^{\alpha x}\, dx,$$

если x меняется непрерывно, а $\varphi(x)$ представляет ее вероятностный закон. Понятно, что

$$f(\alpha) = 1 + \frac{\alpha}{1!}(x) + \frac{\alpha^2}{1 \cdot 2}(x^2) + \frac{\alpha^3}{1 \cdot 2 \cdot 3}(x^3) + \ldots,$$

где (x^p) обозначено вероятное значение x^p. Видно, что $f(0) = 1$.

Характеристической функции достаточно для определения вероятностного закона. Действительно, по формуле Фурье имеем

$$f(i\alpha) = \int_{-\infty}^{+\infty} \varphi(x) e^{i\alpha x}\, dx,$$

$$2\pi\varphi(x) = \int_{-\infty}^{+\infty} f(i\alpha) e^{-i\alpha x}\, d\alpha.$$

Если две величины x и y ***независимы***, а $f(\alpha)$, $f_1(\alpha)$ — соответствующие характеристические функции, то характеристической функцией для $x + y$ будет произведение $f(\alpha) f_1(\alpha)$. Действительно, как мы видели в параграфе 130, вероятное значение произведения $e^{\alpha(x+y)}$ равно произведению вероятных значений $e^{\alpha x}$ и $e^{\alpha y}$.

Для закона Гаусса

$$\varphi(x) = \sqrt{\frac{h}{\pi}}\, e^{-hx^2}$$

легко найдем

$$f(\alpha) = e^{\frac{\alpha^2}{4h}}.$$

Предположим, что две независимые величины x и y подчиняются закону Гаусса, одна с константой h, другая — с h'; соответствующие характеристические функции будут

$$e^{\frac{\alpha^2}{4h}}, \qquad e^{\frac{\alpha^2}{4h'}}.$$

Характеристическая функция для $x + y$ будет их произведением

$$e^{\frac{\alpha^2}{4h''}}, \quad \text{где} \quad \frac{1}{h''} = \frac{1}{h} + \frac{1}{h'}.$$

Тогда сумма $x + y$ тоже подчиняется закону Гаусса с константой h''; это теорема г. Д'Оканя, доказанная в параграфе 60.

Предполагаем теперь, что результирующая ошибка равна сумме большого числа частичных, очень малых и независимых. Пусть $f_1(\alpha)$, $f_2(\alpha)$, ... , $f_n(\alpha)$ — соответствующие характеристические функции.

Каждая из них имеет вид

$$f_R(\alpha) = e^{\beta_1\alpha+\beta_2\alpha^2+\beta_3\alpha^3+\dots}.$$

При этом ошибка очень мала, и коэффициенты β очень быстро убывают.

Характеристическая функция, соответствующая результирующей ошибке, равна произведению

$$f_1(\alpha),\ f_2(\alpha),\ \dots,\ f_n(\alpha),$$

то есть

$$e^{\alpha\sum\beta_1+\alpha^2\sum\beta_2+\alpha^3\sum\beta_3+\dots}.$$

Ошибки не носят систематического характера, и

$$\sum\beta_1 = 0.$$

С другой стороны, коэффициенты β убывают очень быстро, поэтому $\sum\beta_3$, $\sum\beta_4$ и т.д. пренебрежимо малы по сравнению с $\sum\beta_2$, и остается

$$e^{\alpha^2}\sum\beta_2,$$

что соответствует характеристической функции гауссовского закона[15].

145. Закон Гаусса можно проверить a posteriori, используя теорему Бернулли.

Если какое-то испытание может привести к появлению новых событий, таких, что только одно из них происходит за один раз, и если это испытание повторяется очень много раз, то частота каждого из произошедших событий будет практически пропорциональна его вероятности.

Величина z измеряется много раз; результаты равны $x_1, x_2, \dots, x_n$, а ошибки — y_1, y_2, ..., y_n.

Если нам известна z, то нам известны y_1, y_2, ..., y_n; если мы считаем, сколько ошибок лежит в двух данных пределах a и b, то это число пропорционально

$$\int\limits_a^b \varphi(y_1)\,dy_1.$$

Можно построить кривую, соответствующую $\varphi(y_1)$. Разделим ось абсцисс на некоторое число частей, равных α: каждый из этих малых

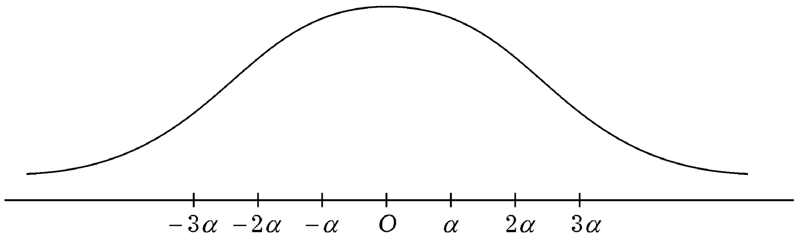


Рис. 16

интервалов достаточно велик, чтобы число ошибок, попавших в этот интервал, было велико; в середине интервала проведем ординату, равную числу ошибок.

Если закон Гаусса верен, то полученная кривая имеет уравнение

$$y = \sqrt{\frac{h}{\pi}} e^{-hx^2}.$$

Эта кривая симметрична относительно оси ординат; ось абсцисс является ее асимптотой.

Оказывается, этот результат подтверждается. Например, Бессель смог представить именно в таком виде результаты очень большого числа наблюдений за отклонением звезды, выполненных Брэдли.

146. Можно также не находить кривую, а проверять закон Гаусса с помощью вычислений.

Заметим в первую очередь, что не зная истинного значения измеряемой величины, можно принять за него среднее значение наблюдений.

Проверяем с помощью подсчетов: согласно закону Гаусса, вероятное значение y^{2p} должно равняться

$$\overline{y^{2p}} = \frac{(2p)!}{p!\,2^{2p}} \cdot \frac{1}{h^p};$$

для y^{2p} имеем

$$\overline{y^{2q}} = \frac{(2q)!}{q!\,2^{2q}} \cdot \frac{1}{h^q};$$

Исключаем h:

$$\frac{\sqrt[p]{\overline{y^{2p}}}}{\sqrt[q]{\overline{y^{2q}}}} = \frac{\sqrt[p]{\dfrac{(2p)!}{p!}}}{\sqrt[q]{\dfrac{(2q)!}{q!}}}.$$

Следует проверить, что это соотношение выполняется.

Мы знаем все допущенные ошибки $y_1, y_2, \ldots, y_n$. Выражение

$$\frac{\sqrt[p]{y_1^{2p} + y_1^{2p} + \ldots + y_n^{2p}}}{\sqrt[q]{y_1^{2q} + y_1^{2q} + \ldots + y_n^{2q}}}$$

должно быть равно

$$\frac{\sqrt[p]{\dfrac{(2p)!}{p!}}}{\sqrt[q]{\dfrac{(2q)!}{q!}}};$$

и проверка этого соотношения осуществляется точно так же.

Можно еще рассмотреть y^{2p+1}. Вероятное значение будет равно нулю, если брать y вместе с его знаком. Если рассматривать только абсолютную величину, то по закону Гаусса вероятным значением будет

$$2\int\limits_0^\infty \sqrt{\frac{h}{\pi}} y^{2p+1} e^{-hy^2}\, dy;$$

этот эйлеров интеграл не равен нулю.

Если брать y вместе со знаком, получится

$$\int\limits_{-\infty}^{+\infty} \sqrt{\frac{h}{\pi}} y^{2p+1} e^{-hy^2}\, dy,$$

равный нулю.

147. Исключения из закона Гаусса. До сих пор я приводил аргументы в пользу закона Гаусса, следствия из которого мы только что получили. Однако может случиться, что источник этих следствий не так уж хорош.

Не стоит испытывать какое-то суеверное чувство перед методом наименьших квадратов, к которому нас приводит закон Гаусса. Мы видели, что иногда бывают основания отказаться от этого закона.

Действительно, мы предполагали, что систематических ошибок нет, а *они есть всегда.*

С другой стороны, мы видели, что часто приходится отказываться от поиска среднего, и, например, отбрасывать одно наблюдение, чрезмерно отличающееся от других. Такого бы не было, если бы закон Гаусса всегда был верен.

Итак, в некоторых случаях могут найтись априорные основания отказаться от закона Гаусса. О чем это говорит? Почему мы отбрасываем наблюдение, непохожее на другие? Потому что предполагаем, что оно связано с грубой ошибкой, произошедшей случайно. То есть мы не считаем a priori грубую ошибку *совершенно невероятной*, какой она была бы по закону Гаусса.

148. Остановимся на некоторых деталях. В определенных случаях можно предполагать, что полная ошибка складывается из двух частичных; первая из них — результат накопления большого числа очень малых ошибок, она подчиняется закону Гаусса; вторая — грубая ошибка, чья вероятность, правда, очень мала, но чья величина может быть очень заметной.

Именно это могло бы произойти, например, если по недосмотру пренебречь какой-то существенной предосторожностью, какой-то необходимой регулировкой прибора, так и не узнав об этом. Вот одна из гипотез, в которой нет ничего неразумного. Пусть тогда

$$y = y_1 + y_2$$

полная ошибка, y_1 и y_2 — две частичные. Тогда y_1 подчиняется закону Гаусса, а функция $\varphi(y_1)$, определяющая вероятностный закон, равна

$$\varphi(y_1) = \sqrt{\frac{h}{2\pi}} e^{-hy_1^2},$$

соответствующая характеристическая функция равна $e^{\frac{\alpha^2}{4h}}$.

Что касается ошибки y_2, она может быть равна 0, $+k$ и $-k$, соответствующие вероятности этих трех ошибок равны $1-2\varepsilon$, ε и ε; соответствующая характеристическая функция —

$$1-2\varepsilon+\varepsilon(e^{\alpha k}+e^{-\alpha k}),$$

а характеристическая функция полной ошибки будет произведением

$$e^{\frac{\alpha^2}{4\pi}}[1-2\varepsilon+\varepsilon(e^{\alpha k}+e^{-\alpha k})].$$

Что касается функции $\varphi(y)$, она равна

$$\sqrt{\frac{h}{2\pi}}\left[(1-2\varepsilon)e^{-hy^2}+\varepsilon(e^{-h(y-k)^2}+e^{-h(y+k)^2})\right].$$

Предположим теперь, что мы бы хотели найти вероятное значение измеряемой величины z, n наблюдений над которой дали результаты $x_1, x_2, \ldots, x_n$. Мы знаем, что все зависит от произведения

$$\varphi(x_1-z)\varphi(x_2-z)\ldots\varphi(x_n-z);$$

когда у функции φ три слагаемых, у этого произведения их 3^n; для искомого вероятного значения найдем[1]

$$\frac{\sum z_k\mu_k e^{-\alpha_k}}{\sum \mu_k e^{-\alpha_k}},$$

в числителе и знаменателе по 3^n членов; кроме того,

$$z_k=\frac{x_1+x_2+\ldots+x_n}{n}+k\frac{\theta_1+\theta_2+\ldots+\theta_n}{n}=\frac{\sum x}{n}+k\frac{\sum\theta}{n},$$

числа θ могут принять значения 0, 1, и -1; поскольку всего чисел θ будет n, а каждое может иметь три значения, получим 3^n комбинаций.

Затем имеем

$$\mu_k=(1-2\varepsilon)^n\left(\frac{\varepsilon}{1-2\varepsilon}\right)^{\sum|\theta|}$$

[1]Здесь суммирование ведется по переменной θ. — *Прим. ред.*

и

$$\alpha_k = h\sum_i (x_i - z_k - \theta_i k)^2.$$

Для простоты предположим, что было только три наблюдения с результатами 0, 0 и k и будем пренебрегать квадратом $\sum$; тогда вероятное значение станет

$$\frac{k}{3}\left[1 - \varepsilon(2e^{-2\beta} + e^{\beta})\right] + \frac{2k}{3}\varepsilon(2 + \varepsilon^{-3\beta}),$$

где $\beta = \frac{2}{3}hk^2$; если β велико, т. е. разница между несогласованным наблюдением $x_3 = k$ и согласованными $x_1 = x_2 = 0$ велика по сравнению с точностью инструмента, то можно сделать так, что εe^{β} будет конечно[2], а затем, сохраняя только конечные величины, найти вероятное значение

$$\frac{k}{3}(1 - \varepsilon e^{\beta})$$

или, точнее, т. к. мы не можем больше пренебрегать квадратом εe^{β},

$$\frac{1}{3}\frac{k}{1 + \varepsilon e^{\beta}}.$$

Итак, это значение меньше среднего арифметического, т. е. мы должны приписать согласованным наблюдениям бóльший вес, чем несогласованному[(16)]. Вообще, считается более естественным предполагать, что несогласованное наблюдение произошло из-за грубой ошибки второго рода, а не из-за ошибки первого рода, подчиненной закону Гаусса, т. к. очень неправдоподобно, чтобы эта последняя могла достигать таких больших значений.

149. Можно попробовать произвести подсчеты для других законов; займемся сперва теми, где большие ошибки более вероятны, чем в законе Гаусса; например, можно взять

$$\varphi(y) = \frac{1}{\pi(1 + y^2)}, \qquad \varphi(y) = \frac{1}{2\lambda}e^{-\left|\frac{y}{\lambda}\right|}.$$

[2]Имеется в виду, что εe^{β} не будет ни бесконечно малым, ни бесконечно большим. — ***Прим. ред.***

Вычисления не представляют никакой трудности; пусть есть три наблюдения $x_1 = x_2$ и $x_3 > x_2$, находим вероятное значение z для последнего закона:

$$\frac{x_1 + \frac{2}{3}\lambda - e^{\frac{x_1 - x_3}{\lambda}}\left(\frac{x_3}{3} + \frac{2}{3}\lambda\right)}{1 - \frac{1}{2}e^{\frac{x_1 - x_3}{\lambda}}}.$$

Для очень большого числа x_3 оно вырождается в

$$x_1 + \frac{2}{3}\lambda,$$

что показывает, что при фиксированных x_1 и x_2 вероятное значение z не растет бесконечно вместе с x_3. Здесь снова вес несогласованного наблюдения меньше, чем у согласованных.

Рассмотрим теперь закон, в котором большие ошибки еще менее правдоподобны, чем в законе Гаусса; пусть, например,

$$\varphi(y) = Ke^{-hy^4};$$

тогда можно найти, что вес несогласованного значения больше, чем согласованных. Сначала это кажется довольно парадоксальным. Но этот результат легко объяснить. Предположим, что произведены три наблюдения

$$0, \quad 0, \quad k,$$

и среднее арифметическое равно $\frac{k}{3}$, а ошибки равны

$$\frac{k}{3}; \quad \frac{k}{3}; \quad \frac{2k}{3}.$$

Если бы мы взяли $\frac{k}{2}$, все три ошибки равнялись бы $\frac{k}{2}$. Вероятность ошибки очень быстро убывает вместе с ростом этой ошибки, и может случиться, что ошибка в $\frac{2k}{3}$ настолько неправдоподобна, что ее можно считать практически невозможной, тогда как ошибка $\frac{k}{2}$ остается приемлемой. Действительно,

$$\left(\frac{k}{3}\right)^4 + \left(\frac{k}{3}\right)^4 + \left(\frac{2k}{3}\right)^4 > \left(\frac{k}{2}\right)^4 + \left(\frac{k}{2}\right)^4 + \left(\frac{k}{2}\right)^4.$$

150. Какие изменения надо внести в закон Гаусса в этих последних случаях? Кривая

$$y = \sqrt{\frac{h}{\pi}}\, e^{-hx^2},$$

которую мы ранее начертили, должна пройти выше, чем раньше, в области, удаленной от оси ординат.

Другой пример; если бы закон Гаусса был верен, можно было бы, накопив достаточно наблюдений, получить сколь угодно большую точность. Но в большинстве случаев возникает чувство, что это заблуждение. С метром, разделенным на миллиметры, мы никогда не сможем, как бы часто мы ни повторяли измерения, определить длину с точностью до миллионной доли миллиметра.

Как пришлось бы изменить кривую Гаусса, чтобы учесть этот факт? Наши попытки достичь бесконечной точности являются заблуждением, потому что у нас не больше шансов не ошибиться вообще, чем ошибиться только на один или два микрона. Тогда кривая должна содержать маленькую ступеньку, маленький горизонтальный отрезок $y = \mathrm{const}$ вправо и влево от пересечения с осью ординат.

Но подобная гипотеза все еще будет недостаточной для учета всех факторов. В рассуждениях параграфа 135 мы видели, что каков бы ни был закон распределения ошибок, если он имеет вид $\varphi(x_i - z)$, вероятное значение квадрата ошибки, допущенной в среднем при n наблюдениях, стремится к нулю, когда n неограниченно растет. То же можно увидеть с помощью характеристических функций. Пусть $F(\alpha)$ — характеристическая функция, связанная с отдельным наблюдением; характеристической функцией среднего из n наблюдений будет

$$\left[F\left(\frac{\alpha}{n}\right)\right]^n.$$

Предположим, что $F(\alpha)$ — аналитическая и

$$F(\alpha) = 1 + A_1\alpha + A_2\alpha^2 + \ldots$$

Пусть также ошибки не имеют систематического характера, т. е. $A_1 = 0$; тогда

$$\log F(\alpha) = B_2\alpha^2 + B_3\alpha^3 + \ldots,$$

откуда

$$\log\left[F\left(\frac{\alpha}{n}\right)\right]^n = B_2\frac{\alpha^2}{n} + B_3\frac{\alpha^3}{n^2} + \ldots$$

и предел[1]

$$\log\left[F\left(\frac{\alpha}{n}\right)\right]^n = 0,$$

что показывает, что ошибка равна нулю.

На самом деле можно было бы предположить, что $F(\alpha)$ не аналитическая, и взять, например,

$$F(\alpha) = e^{-|\alpha|}.$$

Отсюда бы получилось

$$\left[F\left(\frac{\alpha}{n}\right)\right]^n = F(\alpha) = e^{-|\alpha|},$$

так что средняя ошибка не стремилась бы к нулю; этой гипотезе соответствует вероятностный закон со следующей функцией:

$$\varphi(y) = \frac{1}{\pi}\frac{1}{1+y^2}.^{(17)}$$

Какой недостаток в рассуждениях параграфа 135 возникает в этом случае?

Согласно этим рассуждениям, вероятное значение квадрата средней ошибки в n раз меньше, чем вероятное значение квадрата ошибки при отдельном наблюдении. Рассуждение остается верным, но здесь это последнее вероятное значение, представленное интегралом

$$\frac{1}{\pi}\int\limits_{-\infty}^{+\infty}\frac{y^2\,dy}{1+y^2},$$

бесконечно.

Данное решение — не то, которое мы ищем. Мы бы не увидели, что ошибка в среднем стремится к нулю, если бы функция $\varphi(y)$ не стремилась к нулю быстрее, чем $\frac{1}{y^2}$ при больших y, т. е. если бы очень большие ошибки не были очень маловероятными.

Но если мы можем рассчитывать только на инструмент, дающий бесконечное приближение при условии достаточного повторения наблюдений, то виноваты в этом не очень большие ошибки, а наоборот, очень малые; здравый смысл вполне доказывает это.

[1]При $n \to +\infty$. — *Прим. ред.*

151. Итак, надо допустить, что функция φ имеет вид не $\varphi(x_i - z)$, а $\varphi(x_i, z)$. Предположим, например, что прибор имеет деления, и мы можем оценить их десятые доли; вообще-то мы записываем целое число десятых: сотые или тысячные могут появиться в записях лишь случайно. Предположим, что ошибка может возникнуть только при чтении результата; возьмем за единицу десятую часть деления, чтобы x_i всегда было целым. Мы можем допустить, например, что если z лежит между $p-\varepsilon$ и $p+\varepsilon$ (p целое), то мы, конечно, запишем $x_i = p$; а если z лежит между $p+\varepsilon$ и $p+1-\varepsilon$, мы с вероятностью $\frac{1}{2}$ запишем $x_i = p$, и с вероятностью $\frac{1}{2}$ — $x_i = p+1$.

Если z лежит между $p+\varepsilon$ и $p+1-\varepsilon$, и мы производим n наблюдений, то мы n' раз напишем p и n'' раз — $p+1$; отношение n' и n'' к $\frac{n}{2}$ стремится к 1 при росте числа наблюдений; среднее

$$\frac{n'p + n''(p+1)}{n}$$

стремится к $p + \frac{1}{2}$, а никак не к z.

Можно прийти к аналогичным результатам с аналогичными и более сложными законами, если вероятность написать p меняется непрерывным образом вместе с z, но не является линейной функцией от $z-p$; правдоподобней всего то, что кривая, соответствующая этой вероятности, образует ступеньки, как это было в только что рассмотренном частном примере; пусть, наконец, ошибка отсчета имеет одну из этих особенностей, превосходит ошибки, проистекающие из других источников, и подчиняется закону Гаусса.

Если $\varphi(x_i, z)$ зависит не только от разности $y_i = x_i - z$, то вероятное значение ошибки, равное

$$\int y_i \varphi(x_i, z)\, dy_i$$

зависит только от z. Оно может не быть нулем, даже если сама ошибка не имеет систематического характера. Действительно, если это вероятное значение равно $\theta(z)$, достаточно, чтобы интеграл

$$\int\limits_{-\infty}^{+\infty} \theta(z)\, dz$$

(или лучше $\int\limits_{-\infty}^{+\infty} \psi(z)\theta(z)\,dz$, где $\psi(z)$ — априорная вероятность того, что z принимает данное значение), как в главе X, был равен нулю, и ошибку можно не называть систематической; но само $\theta(z)$ не должно быть тождественным нулем. Возьмем, например, гипотезу, изложенную выше, и предположим сверх того, что мы изначально знаем, что z лежит между p и $p+1$; то есть $\psi(z)$ равно нулю, когда z не находится в этих пределах, и может считаться константой иначе. Тогда имеем

$$\begin{aligned} &\theta(z) = p - \varepsilon && (p < z < p+\varepsilon),\\ &\theta(z) = p + \tfrac{1}{2} - \varepsilon && (p+\varepsilon < z < p+1-\varepsilon),\\ &\theta(z) = p + 1 - \varepsilon && (p+1-\varepsilon < z < p+1). \end{aligned}$$

Легко видеть, что

$$\int\limits_{-\infty}^{+\infty} \psi(z)\theta(z)\,dz = \int\limits_{p}^{p+1} \theta(z)\,dz = 0.$$

152. Вернемся теперь к формуле из главы XI, которая давала нам вероятное значение квадрата ошибки, допущенной в среднем:

$$\frac{\sum \bar{y}_1^2}{n^2} + \frac{2\sum \overline{y_1 y_2}}{n^2},$$

что можно еще записать как

$$\frac{\bar{y}^2}{n} + \frac{n(n-1)}{n^2}(\bar{y})^2.$$

Эта формула существует, только когда вероятные значения $\bar{y}$ и $\bar{y}^2$, ошибки в отдельном наблюдении и ее квадрата, будут зависеть от z. Согласно предыдущим рассуждениям,

$$\bar{y} = \theta(z)$$

может не быть нулем; первое слагаемое стремится к нулю, когда n неограниченно растет; но это неверно для второго слагаемого, которое стремится к $\theta^2(z)$, так что вероятное значение квадрата ошибки, допущенной в среднем, надо будет записать (не считая больше z равной какой-то конкретной величине) как

$$\int\limits_{-\infty}^{+\infty} \theta^2(z)\psi(z)\,dz$$

— интеграл, который не обращается в нуль, т. к. все элементы положительны.

Также можно было бы воспользоваться характеристической функцией, если записать ее как

$$F(\alpha,\, z) = e^{\beta_1\alpha + \beta_2\alpha^2 + \ldots};$$

все β зависят от z, и β_1, где β_1 — не что иное, как $\theta(z)$, может не совпадать с нулем. Для допущенной в среднем ошибки характеристическая функция равна

$$\left[F\left(\frac{\alpha}{n},\, z\right)\right]^n$$

и при очень больших n вырождается в

$$e^{\beta_1\alpha} = 1 + \beta_1\alpha + \frac{\beta_1^2\alpha^2}{2} + \ldots$$

Вероятные значения допущенной в среднем ошибки и ее квадрата равны β_1 и β_1^2 для какого-то частного значения z, а если не приписывать z частного значения, получим

$$\int \beta_1\psi(z)\,dz = 0, \qquad \int \beta_1^2\psi(z)\,dz > 0.$$

153. Кроме того, в законе Гаусса предполагается, что φ зависит только от y, где y равно $x - z$, тогда как φ может быть функцией от x и от z.

Например, чтобы учесть *десятичную ошибку*, мы могли взять

$$\varphi(y) = \sqrt{\frac{h}{\pi}}\, e^{-hy^2}\theta(x),$$

где $\theta(x)$ — периодическая функция от x с периодом 1, достигающая максимума на десятичных знаках, которые нравятся некоторым наблюдателям и все время попадают в их записи. С другой стороны, меньше шансов совершить ошибку, если мы попадем точно на деление, имеющееся в приборе, чем если мы вынуждены добавлять более мелкие деления на глазок. Итак, h больше не будет константой, но функцией от x с периодом 1; точность будет выше, когда x — целое число, нежели иначе; если тот же наблюдатель очень хорошо оценивает десятичные доли, но не может найти другие десятичные знаки, $\theta(x)$ будет точный нуль при любом значении x, в записи которого есть только десятые доли.

Глава 12
Ошибка при нахождении координат одной точки

154. Задача об ошибках, допущенных при нахождении координат одной точки. Вместо измерения одной единственной величины часто приходится измерять сразу несколько, две координаты звезды, например, и т. д.

Я предполагаю, что для оценки положения одной точки на плоскости сделано n измерений, и для двух прямоугольных координат x и y найдены величины

$$x_1, y_1, \quad x_2, y_2, \quad \dots, \quad x_n, y_n.$$

Допущенные ошибки равны

$$\begin{array}{ccc} x_1 - x & \text{и} & y_1 - y, \\ x_2 - x & \text{и} & y_2 - y, \\ \dots & \dots & \dots \\ x_n - x & \text{и} & y_n - y. \end{array}$$

Я обозначу

$$\begin{array}{cc} x_1 = x + \xi_1, & y_1 = y + \eta_1, \\ x_2 = x + \xi_2, & y_2 = y + \eta_2, \\ \dots\dots\dots & \dots\dots\dots \\ x_n = x + \xi_n, & y_n = y + \eta_n. \end{array}$$

Тогда допущенные ошибки равны

$$\begin{array}{ccc} \xi_1 & \text{и} & \eta_1, \\ \xi_2 & \text{и} & \eta_2, \\ & \dots & \\ \xi_n & \text{и} & \eta_n. \end{array}$$

155. Можно спросить, с какой вероятностью ошибки, допущенные при первом наблюдении, лежат между ξ_1 и $\xi_1 + d\xi_1$, η_1 и $\eta_1 + d\eta_1$. Пусть эта вероятность равна

$$\varphi(\xi_1,\, \eta_1)d\xi_1\, d\eta_1;$$

Может оказаться, что φ зависит от x и y, но я предполагаю, что это не так и φ зависит только от ошибок.

Эти ошибки могут быть независимы; тогда φ будет произведением двух других функций, а вероятность будет представляться как

$$\varphi(\xi_1) \cdot \varphi(\eta_1)\, d\xi_1\, d\eta_1;$$

но я все-таки предположу, что ошибка, допущенная в абсциссе, зависит от ошибки, допущенной в ординате.

Рассуждение аналогично тому, что приводит Гаусс в случае одной переменной. Будем основываться на следующем постулате: *если сделано столько-то измерений, и найденные точки отмечаются на плоскости, то наиболее вероятным положением искомой точки будет центр масс всех этих точек в предположении, что их массы равны.*

156. В примере Гаусса мы считали, что такая точка зрения обоснована. Каков будет закон распределения ошибок, оправдывающий эту гипотезу?

Ищем вероятность того, что координаты точки лежат между x и $x + dx$, y и $y + dy$. Это задача на условные вероятности, и нам надо искать

$$\frac{p_i \bar{\omega}_i}{\sum p_i \bar{\omega}_i}.$$

Величина $\bar{\omega}_i$ — априорная вероятность того, что координаты лежат между x и $x + dx$, y и $y + dy$; представим ее как

$$\bar{\omega}_i = \psi(x,\, y).$$

Величина p_i — вероятность того, что при воздействии данной причины происходит наблюдаемое явление: здесь наблюдаемым явлением будет то, что одни координаты лежат между

$$\begin{array}{ccc}
x_1 & \text{и} & x_1 + dx_1, \\
x_2 & \text{и} & x_2 + dx_2, \\
\ldots & \ldots & \ldots \\
x_n & \text{и} & x_n + dx_n,
\end{array}$$

другие — между

$$\begin{array}{ccc} y_1 & \text{и} & y_1 + dy_1, \\ y_2 & \text{и} & y_2 + dy_2, \\ \dots & \dots & \dots \\ y_n & \text{и} & y_n + dy_n. \end{array}$$

Вероятность p_i равна произведению вероятностей, соответствующих каждому из измерений

$$p_i = \varphi(\xi_1, \eta_1)\varphi(\xi_2, \eta_2)\dots\varphi(\xi_n, \eta_n)\, d\xi_1\, d\eta_1\, d\xi_2\, d\eta_2\, \dots\, d\xi_n\, d\eta_n.$$

Пусть

$$\Pi = \varphi(\xi_1, \eta_1)\varphi(\xi_2, \eta_2)\dots\varphi(\xi_n, \eta_n).$$

Тогда

$$\frac{p_i\bar{\omega}_i}{\sum p_i\bar{\omega}_i}$$

будет

$$\frac{\Pi\psi\, dx\, dy\, d\xi_1\, d\eta_1\, d\xi_2\, d\eta_2\, \dots\, d\xi_n\, d\eta_n}{\int \Pi\psi\, dx\, dy\, d\xi_1\, d\eta_1\, d\xi_2\, d\eta_2\, \dots\, d\xi_n\, d\eta_n};$$

интеграл берется по x и y; остается

$$\frac{\Pi\psi\, dx\, dy}{\int \Pi\psi\, dx\, dy}.$$

Знаменатель — константа, не зависящая от x и y; значит, вероятность пропорциональна числителю.

157. Каково наиболее вероятное значение? То, где числитель $\Pi\psi$ достигает максимума. Но

$$\Pi = \varphi(x_1 - x, y_1 - y)\varphi(x_2 - x, y_2 - y)\dots\varphi(x_n - x, y_n - y).$$

При каких значениях x и y это произведение максимально? Согласно постулату, это среднее арифметическое x_1, x_2, , … , x_n в качестве x и среднее арифметическое y_1, y_2, … , y_n в качестве y. Приравняем логарифмические производные к следующим функциям:

$$\frac{\frac{d\varphi}{d\xi_1}}{\varphi} = F(\xi_1, \eta_1), \quad \frac{\frac{d\varphi}{d\eta_1}}{\varphi} = F_0(\xi_1, \eta_1),$$

$$\frac{\frac{d\psi}{dx}}{\psi} = \theta, \qquad \frac{\frac{d\psi}{dy}}{\psi} = \theta_0.$$

Логарифмическая производная $\Pi\psi$ по x, которая должна обращаться в нуль в точке максимума, равна

$$-F(\xi_1, \eta_1) - F(\xi_2, \eta_2) - \ldots - F(\xi_n, \eta_n) + \theta = 0,$$

откуда

$$F(\xi_1, \eta_1) + F(\xi_2, \eta_2) + \ldots + F(\xi_n, \eta_n) = \theta.$$

Логарифмическое дифференцирование по y даст

$$F_0(\xi_1, \eta_1) + F_0(\xi_2, \eta_2) + \ldots + F_0(\xi_n, \eta_n) = \theta_0.$$

Эти два соотношения должны выполняться всякий раз, когда

$$x_1 + x_2 + \ldots + x_n = nx,$$
$$y_1 + y_2 + \ldots + y_n = ny,$$

то есть

$$\xi_1 + \xi_2 + \ldots + \xi_n = 0,$$
$$\eta_1 + \eta_2 + \ldots + \eta_n = 0.$$

158. Речь идет об определении функций φ и ψ так, чтобы эти две системы уравнений были совместны.

Дифференцируем эти две системы, считая x и y постоянными и меняя $\xi_1, \eta_1, \xi_2, \eta_2, \ldots, \xi_n, \eta_n$. Производные от θ и θ_0 в правой части равны нулю, т. к. θ и θ_0 не зависят от этих величин.

Я обозначу F_i и F_i^0 следующие функции

$$F_i = F(\xi_i, \eta_i), \qquad F_i^0 = F_0(\xi_i, \eta_i).$$

Первая система преобразуется в

$$dF_1 + dF_2 + \ldots + dF_n = 0, dF_1^0 + dF_2^0 + \ldots + dF_n^0 = 0,$$

а вторая система — в

$$d\xi_1 + d\xi_2 + \ldots + d\xi_n = 0, d\eta_1^0 + d\eta_2^0 + \ldots + d\eta_n^0 = 0.$$

Таковы две системы, которые должны быть эквивалентны.

Два первых уравнения должны быть комбинацией двух последних; выполняется тождество

$$dF_1 + dF_2 + \ldots + dF_n = \\ = A d\xi_1 + B d\eta_1 + A d\xi_2 + B d\eta_2 + \ldots + A d\xi_n + B d\eta_n.$$

Приравниваем соответствующие коэффициенты:

$$A = \frac{dF_1}{d\xi_1} = \frac{dF_2}{d\xi_2} = \ldots = \frac{dF_n}{d\xi_n},$$
$$B = \frac{dF_1}{d\eta_1} = \frac{dF_2}{d\eta_2} = \ldots = \frac{dF_n}{d\eta_n}.$$

Здесь A должно зависеть только от ξ_1 и η_1, т. к. оно равно $\frac{dF_1}{d\xi_1}$; так же оно зависит только от ξ_2 и η_2, т. к. оно равно $\frac{dF_2}{d\xi_2}$; ...

Итак, A и B — константы: F_1 — линейная функция ξ_1 и η_1.

Так же и F_1^0 — линейная функиця ξ_1 и η_1. Добавим, что в представлении F_1 как

$$F_1 = A\xi_1 + B\eta_1 + C$$

произвольная постоянная C равна нулю.

Действительно, если я заменю F_1, F_2, ... , F_n их значениями, я получу

$$A(\xi_1 + \xi_2 + \ldots + \xi_n) + B(\eta_1 + \eta_2 + \ldots + \eta_n) + nC = \theta.$$

Это уравнение должно быть верным всякий раз, когда

$$\xi_1 + \xi_2 + \ldots + \xi_n = 0,$$
$$\eta_1 + \eta_2 + \ldots + \eta_n = 0.$$

Итак, $nC = \theta$, и, т. к. θ не зависит от n,

$$\theta = 0, \qquad C = 0.$$

Отсюда следует, что ψ константа.

Чтобы теория Гаусса была применима, надо, чтобы a priori у нас не было никаких сведений о значении искомой величины.

Точно так же получим

$$F_1^0 = D\xi_1 + E\eta_1.$$

Кроме того, $B = D$, т. к.

$$B = \frac{dF_1}{d\eta_1}, \qquad D = \frac{dF_1^0}{d\xi_1},$$

а

$$\frac{dF_1}{d\eta_1} = \frac{dF_1^0}{d\xi_1}$$

в силу определения F и F_0.

Наконец, приходим к

$$\log\varphi = A\xi_1^2 + 2B\xi_1\eta_1 + E\eta_1^2 + G,$$

где G — константа, которую мы запишем как $\log H$.

Многочлен $A\xi_1^2 + 2B\xi_1\eta_1 + E\eta_1^2$ должен быть отрицательно определенным, т. к. бесконечно большая ошибка должна иметь нулевую вероятность; обозначим его — P.

$$\log\varphi = -P + \log H;$$

откуда

$$\varphi = He^{-P}.$$

Интеграл

$$\iint \varphi \, d\xi_1 \, d\eta_1$$

должен быть равен единице при всех возможных значениях x и y.

159. Это рассуждение может стать объектом тех же возражений, что и рассуждение Гаусса.

Если, несмотря на все возражения, мы принимаем закон Гаусса, то можно построить кривые

$$P = \text{const},$$

поместив начало в намеченную точку, т. е. очень близко к средней точке. Ситуация похожа на задачу о стрельбе по мишени.

Точки распределяются согласно обобщенному закону Гаусса.

Кривые $P = \text{const}$ будут концентрическими эллипсами с одинаковыми направлениями осей.

Рассмотрим один из этих эллипсов: я проведу из O несколько прямых, которые делят плоскость на сектора 1, 2, 3, 4.

Пусть s_1 — внутренняя часть первого сектора, S_1 — его внешняя часть.

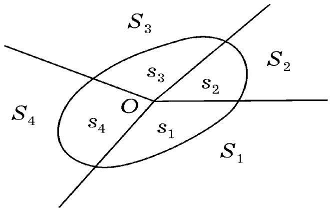


Рис. 17

Если на плоскости отмечено очень много точек, они распределяются почти пропорционально их вероятностям. В s_1 будет n_1 точек, в S_1 их будет N_1. Теорема, немедленно следующая из формулы предыдущего параграфа, такова:

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \frac{n_3}{N_3} = \frac{n_4}{N_4}.$$

В каждом секторе отношение количества внутренних и внешних точек эллипса одинаково.

160. Если наблюдения независимы, то φ равно произведению двух функций. Тогда P будет суммой двух функций, каждая из которых зависит от одной переменной; слагаемое с $\xi_1\eta_1$ исчезает.

Это означает, что оси эллипсов параллельны осям координат.

Для того чтобы разрыв между намеченной точкой и полученной в результате наблюдений не зависел от направления, эллипс должен выродиться в окружность.

Еще можно будет считать независимыми ошибку в абсциссе и ошибку в радиус-векторе.

Все сказанное выше явилось основанием для доказательства закона Гаусса. Эти вопросы уже упоминались в параграфе 16; они лишены смысла.

Глава 13

Метод наименьших квадратов

161. Метод наименьших квадратов служит для определения величин $u_1, u_2, \ldots, u_p$, если непосредственно их измерить нельзя, но можно найти какие-то функции $z_1, z_2, \ldots, z_n$ от них:

$$\begin{gathered} z_1 = F_1(u_1, u_2, \ldots, u_p), \\ z_2 = F_2(u_1, u_2, \ldots, u_p), \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ z_n = F_n(u_1, u_2, \ldots, u_p). \end{gathered}$$

Пусть при измерении $z_1, z_2, \ldots, z_n$ n наблюдений дали результаты $x_1, x_2, \ldots, x_n$; допущены какие-то ошибки. Для z_i ошибка y_i равна

$$y_i = x_i - z_i.$$

Задачу можно поставить только для $n > p$, т. к. для $n = p$ система уравнений дает единственное решение, а для $n < p$ уравнений будет недостаточно, и задача окажется неопределенной.

Для $n > p$ уравнений больше, чем нужно: каковы тогда будут самые приемлемые величины, которые можно взять за u?

Исключим p величин u; приходим к $n - p$ уравнениям

$$\theta_i(z_1, z_2, \ldots, z_n) = 0, \qquad i = 1, 2, \ldots, n - p.$$

Это уравнения связи.

Иногда выгоднее взять их в той форме, которая приведена выше.

Иногда предпочтительнее выразить z как функции от u.

Мы разберем примеры с первым и со вторым методом: первый будет связан с планетами, второй — с задачей о триангуляции.

162. Первая часть метода наименьших квадратов, которым мы займемся прямо сейчас, это определение *самых подходящих значений* u. Во второй части мы будем изучать допущенные ошибки.

По неизвестной причине величины u заключены в определенных пределах; наблюдаемые величины x_i тоже заключены в определенных пределах. Применим еще раз формулу

$$\frac{p_i\bar{\omega}_i}{\sum p_i\bar{\omega}_i},$$

где $\bar{\omega}_i$ — априорная вероятность того, что воздействует данная причина; здесь это означает, что величины u лежат между u_1 и $u_1 + du_1$, u_2 и $u_2 + du_2$, … , u_p и $u_p + du_p$. Эту вероятность можно представить как

$$\psi(u_1, u_2, \ldots, u_p)\, du_1\, du_2 \ldots du_p.$$

Функция ψ может иметь разный вид в зависимости от того, как мы a priori представляем себе величины u: здесь степень произвола очень велика.

Величина p_i равна вероятности того, что x_i лежит между x_i и $x_i + dx_i$, в предположении, что все u имеют те значения, которые мы им приписали; числа $z_1, z_2, \ldots, z_n$ даны как функции от u, и вероятность допущенной в каждом из них ошибки будет соответственно

$$\begin{gathered}
\varphi_1(y_1)\, dy_1 = \varphi(x_1 - z_1)\, dx_1,\\
\varphi_2(y_2)\, dy_2 = \varphi(x_2 - z_2)\, dx_2,\\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\varphi_n(y_n)\, dy_n = \varphi(x_n - z_n)\, dx_n.
\end{gathered}$$

Искомая вероятность p_i — та, с которой все эти обстоятельства происходят одновременно; это произведение вероятностей:

$$p_i = \varphi(y_1)\varphi(y_2)\ldots\varphi(y_n)\, dy_1\, dy_2 \ldots dy_n.$$

Я немного сокращу запись, полагая

$$du_1\, du_2 \ldots du_p = d\omega,$$

а с другой стороны,

$$\begin{gathered}
\varphi_1(y_1)\varphi_2(y_2)\ldots\varphi_n(y_n) = \Pi,\\
dy_1\, dy_2 \ldots dy_n = dx_1\, dx_2 \ldots dx_n = d\omega'.
\end{gathered}$$

Тогда

$$p_i\bar{\omega}_i = \Pi\psi\, d\omega\, d\omega'$$

и

$$\frac{p_i\bar{\omega}_i}{\sum p_i\bar{\omega}_i} = \frac{\Pi\psi\, d\omega\, d\omega'}{\sum \Pi\psi\, d\omega\, d\omega'}.$$

Интегрирование производится только по $d\omega$, поэтому $d\omega'$ сокращается.

Такова вероятность, которую надо знать, чтобы найти апостериорную вероятность того, что u_i лежит между u_i и $u_i + du_i$. Так как знаменатель равен константе, эта вероятность пропорциональна Π, т. е. функции от u, и $\psi\, d\omega$, представляющей априорную вероятность; $\psi\, d\omega$ — тоже функция от u.

163. Чтобы получить *наиболее вероятное значение величин* u, надо найти максимум $\Pi\psi\, d\omega$.

Самой простой гипотезой насчет ψ будет $\psi = 1$: остается определить максимум Π.

Самой простой гипотезой насчет φ будет предположение, что ошибки подчиняются закону Гаусса:

$$\varphi_k(y_k) = \sqrt{\frac{h_k}{\pi}}\, e^{-h_k y_k^2}.$$

Тогда

$$\Pi = Ce^{-(h_1 y_1^2 + h_2 y_2^2 + \ldots + h_n y_n^2)}.$$

Максимум достигается, когда выражение в скобках, т. е.

$$h_1 y_1^2 + h_2 y_2^2 + \ldots + h_n y_n^2,$$

минимально.

Это известная функция от u: величины $h_1, h_2, \ldots, h_n$ представляют собой веса наблюдений, и достигать минимума должна *сумма квадратов* допущенных ошибок, в которой каждый квадрат умножен на вес соответствующего наблюдения.

164. Можно было бы прийти к тому же результату без закона Гаусса, достаточно лишь предполагать случайные ошибки малыми, а систематические — нулевыми.

Действительно, предполагаем, что величины y очень малы; какой бы вид не имели функции φ, мы придем к какому-то аналогичному выражению. Речь пойдет о достижении максимума произведением всех φ_i.

Я считаю φ_i четными и раскладываю по формуле Тейлора $\log \varphi_i$, меняя знак:

$$\begin{gathered}
-\log \varphi_1(y_1) = a_1 + b_1 y_1^2 + c_1 y_1^4 + \ldots , \\
-\log \varphi_2(y_2) = a_2 + b_2 y_2^2 + c_2 y_2^4 + \ldots , \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
-\log \varphi_n(y_n) = a_n + b_n y_n^2 + c_n y_n^4 + \ldots
\end{gathered}$$

Сумма логарифмов с противоположным знаком должна быть минимальной, т. е. к минимуму стремится выражение

$$\sum a_i + \sum b_i y_i^2 + \sum c_i y_i^4 + \ldots$$

Я дифференцирую по u_k:

$$2 \sum b_i y_i \frac{dy_i}{du_k} + 4 \sum c_i y_i^3 \frac{dy_i}{du_k} + \ldots = 0.$$

Так как y_i считаются очень малыми, высшими степенями можно пренебречь, и все будет происходить так же, как если бы мы искали минимум $\sum b_i y_i^2$. Итак, допуская, что закон Гаусса неверен, получим, что истинный вероятностный закон не очень отличается от него *на нужном нам интервале.*

Одно из двух: либо наблюдения практически согласованы и, как мы только что видели, примени́м метод наименьших квадратов, либо они не согласованы, и в этом случае данные наблюдения ничего не стоят, из них ничего нельзя вывести.

165. Тем не менее мы не смогли бы удовлетвориться предыдущим рассуждением, т. к. то, что мы хотим получить, это *вероятное* значение u (а не *наиболее вероятное значение*).

Если вероятность того, что x лежит между x и $x+dx$, равна $\varphi(x)\,dx$, то наиболее вероятным значением x будет то, где достигается максимум $\varphi(x)\,dx$.

Вероятное значение x равно $\int x\varphi(x)\,dx$.

Если вероятное значение u равно u_0, то вероятное значение функции z,

$$z = u^2,$$

не совпадает с u_0^2.

Но если *все* z *— линейные функции от* u, их вероятные значения соответствуют вероятным значениям u;

$$z_i = A_{i_1} u_1 + A_{i_2} u_2 + \ldots + A_{i_p} u_p + B_i.$$

Дифференциал $C\Pi\, d\omega$ представляет собой вероятность того, что i-я величина u_i лежит между u_i и $u_i + du_i$.

Пусть u_k^0 — вероятное значение u_k,

$$u_k^0 = \int C\Pi u_k\, d\omega.$$

Вероятное значение z_i, обозначим его z_i^0, равно

$$z_i^0 = \int C\Pi(A_{i_1} u_1 + A_{i_2} u_2 + \ldots + A_{i_p} u_p + B_i)\, d\omega$$

или

$$z_i^0 = A_{i_1} \int C\Pi u_1\, d\omega + A_{i_2} \int C\Pi u_2\, d\omega + \ldots + A_{i_p} \int C\Pi u_p\, d\omega + B_i \int C\Pi\, d\omega,$$

и, т. к. $\int C\Pi\, d\omega$ равен вероятному значению единицы, т. е. 1,

$$z_i^0 = A_{i_1} u_1^0 + A_{i_2} u_2^0 + \ldots + A_{i_p} u_p^0 + B_i.$$

166. С одной стороны, я буду считать ψ константой; с другой стороны, я предположу, что закон распределения ошибок — гауссовский; наконец, что все z связаны с u линейными соотношениями.

Я хочу установить, что *вероятное значение* u совпадает с тем, которое дает метод наименьших квадратов.

Произведение $p_i \bar{\omega}_i$ достигает максимума, когда функция P

$$P = h_1 y_1^2 + h_2 y_2^2 + \ldots + h_n y_n^2$$

минимальна.

Пусть $u_1^0, u_2^0, \ldots, u_p^0$ — значения u, доставляющие минимум этому выражению. Они же будут, как мы сейчас увидим, вероятными значениями соответствующих u.

Имеем
$$\Pi = Ce^{-P}.$$

Вероятное значение u_k равно
$$\int \frac{\Pi \, d\omega}{\int \Pi \, d\omega} u_k = \frac{\int \Pi u_k \, d\omega}{\int \Pi \, d\omega}.$$

Надо показать, что
$$u_k^0 = \frac{\int \Pi u_k \, d\omega}{\int \Pi \, d\omega},$$
то есть
$$\int \Pi (u_k - u_k^0) \, d\omega = 0.$$

Имеем
$$y_i = x_i - z_i;$$
где x_i известно, z_i имеет первую степень по u: тогда P — многочлен второй степени по u. Многочлен P достигает минимума, когда
$$u_1 = u_1^0, \quad u_2 = u_2^0, \quad \ldots u_k = u_k^0, \quad \ldots$$

Это означает, что
$$p = p_0 + p_2,$$
где p_2 — однородный многочлен второй степени по $(u_1 - u_1^0)$, $(u_2 - u_2^0), \ldots, (u_k - u_k^0), \ldots$; а p_0 — значение минимума P.

Мы хотим показать, что
$$\int Ce^{-p_0} e^{-p_2} (u_k - u_k^0) \, d\omega = 0.$$

Здесь $Ce^{-p_0}e^{-p_2}$ — четная функция от величин $(u_i - u_i^0)$; $u_k - u_k^0$ — нечетная функция. Так как интеграл берется от $-\infty$ до $+\infty$, он равен нулю.

Итак, данные значения u будут не только наиболее вероятными, но и вероятными значениями.

167. Будет ли это верно в общем случае? Да, если измерения практически согласованы, ошибки малы и все происходит так же, как по закону Гаусса.

Если область изменения u очень ограничена, мы можем рассматривать z как линейные, какой бы вид не имели функции F.

Из тех же самых соображений, что область, на которой могут меняться u, очень ограничена, функция ψ, которая изначально была произвольной, может считаться константой.

Именно из-за всей совокупности этих обстоятельств метод наименьших квадратов может считаться применимым всякий раз, когда наблюдения практически согласованы и лишены систематических ошибок.

168. Установив это, посмотрим, *как должны проводиться вычисления.*

Мы знаем функции z от u: всего их p штук, а количество наблюдений равно n и больше p; вместо $z_1, z_2, \ldots, z_n$ нам даны значения $x_1, x_2, \ldots, x_n$.

Уравнений больше, чем неизвестных; ищем наилучший способ приблизительно удовлетворить им.

Первое приближение даст $u_1, u_2, \ldots, u_p$. Предположим, его достаточно, чтобы можно было пренебречь квадратом допущенной ошибки; пусть b_i равно этому первому приближению для u_i:

$$u_i = b_i + v_i.$$

Если мы раскладываем z_i по возрастанию степеней v_i, по формуле Тейлора, и пренебрегаем квадратами v_i,

$$z_i = A_{i_1} v_1 + A_{i_2} v_2 + \ldots + A_{i_p} v_p + B_i,$$

то мы получим таким образом n уравнений, которые станут линейными.

Надо найти минимум суммы

$$\sum h_i y_i^2 = \sum h_i (z_i - x_i)^2.$$

Пишем, что p производных по p величинам $v_1, v_2, \ldots, v_p$ равны нулю:

$$\sum h_i (z_i - x_i) \frac{dz_i}{dv_k} = 0.$$

Но

$$\frac{dz_i}{dv_k} = A_{ik};$$

остается

$$\sum h_i A_{ik}(z_i - x_i) = 0.$$

169. Итак, мы пришли к следующему правилу. Ниже я записываю приблизительные уравнения:

$$\begin{array}{l|l} x_1 - B_1 = A_{11}v_1 + A_{12}v_2 + \ldots + A_{1p}v_p & h_1 A_{11}, \\ x_2 - B_2 = A_{21}v_1 + A_{22}v_2 + \ldots + A_{2p}v_p & h_2 A_{21}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \ldots\ldots \\ x_n - B_n = A_{n1}v_1 + A_{n2}v_2 + \ldots + A_{np}v_p & h_n A_{n1}. \end{array}$$

Эти уравнения не только приблизительны, но и несовместны, т. к. $n > p$. Я умножаю обе части первого из них на $h_1 A_{11}$; обе части второго — на $h_2 A_{21}$; ... ; обе части n-го — на $h_n A_{n1}$ и почленно складываю результаты. Так я получаю первое уравнение, линейное по x и v; я пользуюсь коэффициентами

$$h_1 A_{1k},\ h_2 A_{2k},\ \ldots,\ h_n A_{nk},$$

где $k = 1$. Если я последовательно приравняю k к $2, \ldots, p$, я получу новые серии коэффициентов

$$h_1 A_{12},\ h_2 A_{22},\ \ldots,\ h_n A_{n2},$$

$$\ldots,\ \ldots,\ \ldots,\ \ldots,$$

$$h_1 A_{1p},\ h_2 A_{2p},\ \ldots,\ h_n A_{np},$$

и, соответственно, p линейных уравнений для v.

Решив эту систему, мы ответим на вопрос задачи.

170. Я проведу вычисления иначе.

У меня есть n функций z_i от $u_1, u_2, \ldots, u_p$:

$$z_i = F_i(u_1, u_2, \ldots, u_p).$$

Я могу исключить u; отсюда я получу $n - p$ уравнений, являющихся уравнениями связи:

$$\varphi_1(z_1, z_2, \ldots, z_n) = 0,$$

$$\varphi_2(z_1, z_2, \ldots, z_n) = 0,$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\varphi_q(z_1, z_2, \ldots, z_n) = 0,$$

где $q = n - p$.

Я могу разложить все уравнения связи, пренебрегая членами второго порядка по y, при условии, что наблюдения достаточно согласованы:

$$\begin{gathered}A_{11}y_1 + A_{12}y_2 + \ldots + A_{1n}y_n = B_1,\\ A_{21}y_1 + A_{22}y_2 + \ldots + A_{2n}y_n = B_2,\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ A_{q1}y_1 + A_{q2}y_2 + \ldots + A_{qn}y_n = B_q.\end{gathered}$$

Как лучше всего удовлетворить этим уравнениям связи? Устремив к минимуму

$$\sum h_i y_i^2.$$

Итак,

$$\sum h_i y_i\, dy_i = 0.$$

171. Надо заметить, что dy_1, dy_2, ... , dy_n не будут независимыми: они связаны соотношениями, которые можно получить, дифференцируя уравнения связи.

Общий вид этих уравнений —

$$\sum A_{ki} y_i = B_k,$$

откуда

$$\sum A_{ki}\, dy_i = 0.$$

Итак, соотношения, связывающие dy_1, dy_2, ... , dy_n, имеют вид

$$\begin{gathered}\sum A_{1i}\, dy_i = 0,\\ \sum A_{2i}\, dy_i = 0,\\ \ldots\ldots\ldots\ldots\\ \sum A_{qi}\, dy_i = 0;\end{gathered}$$

и

$$\sum h_i y_i\, dy_i = 0$$

должно быть следствием этих q уравнений.

Мы должны получить (при неопределенных коэффициентах ε)

$$h_i y_i = \varepsilon_1 A_{1i} + \varepsilon_2 A_{2i} + \ldots + \varepsilon_q A_{qi}.$$

Все ε можно определить, подставив в уравнения связи значения y_i как функции от ε; первое уравнение превратится в

$$\frac{A_{11}}{h_1}[\varepsilon_1 A_{11}+\varepsilon_2 A_{21}+\ldots+\varepsilon_q A_{q1}]+\frac{A_{12}}{h_2}[\varepsilon_1 A_{12}+\varepsilon_2 A_{22}+\ldots+\varepsilon_q A_{q2}]+\ldots = B_1,$$

Таким образом получается q уравнений для определения ε. Это второе решение задачи.

172. Возьмем две переменные и четыре наблюдения. Два уравнения связи равны

$$Ay_1 + By_2 + Cy_3 + Dy_4 = H,$$
$$A'y_1 + B'y_2 + C'y_3 + D'y_4 = H'.$$

Я считаю веса одинаковыми: $h_1 = h_2 = h_3 = h_4$.

$$y_1 = A\varepsilon_1 + A'\varepsilon_2,$$
$$y_2 = B\varepsilon_1 + B'\varepsilon_2,$$
$$y_3 = C\varepsilon_1 + C'\varepsilon_2,$$
$$y_4 = D\varepsilon_1 + D'\varepsilon_2,$$

После подстановки y в уравнения связи получим уравнения для ε:

$$\Big(\sum A^2\Big)\varepsilon + \Big(\sum AA'\Big)\varepsilon_2 = H$$
$$\Big(\sum AA'\Big)\varepsilon + \Big(\sum A'^2\Big)\varepsilon_2 = H'.$$

Из двух указанных методов один будет выгоднее другого в зависимости от обстоятельств.

Трудность заключается в необходимости решать большое число линейных уравнений; нашей целью будет по возможности уменьшить их количество.

В первом способе имеется p уравнений, во втором их $q = n - p$. Итак, надо применять первый способ, если n больше, чем $2p$, второй способ, если n меньше, чем $2p$.

173. Мы попытались найти минимум

$$\sum h_i(z_i - x)^2,$$

где h_i — вес наблюдения i. Можно свести задачу к случаю, когда все веса равны. Полагаем

$$z_i' = z_i\sqrt{h_i} = \sqrt{h_i}F_i(u_1, u_2, \ldots, u_p)$$

и

$$x_i' = x_i\sqrt{h_i};$$

тогда

$$h_i(z_i - x_i)^2 = (\sqrt{h_i}z_i - \sqrt{h_i}x_i)^2 = (z_i' - x_i')^2,$$

и надо искать минимум выражения такого же, как

$$\sum (z_i - x_i)^2.$$

174. Мы видели, что вычисления можно было бы провести двумя способами. Приведем пример для каждого из них.

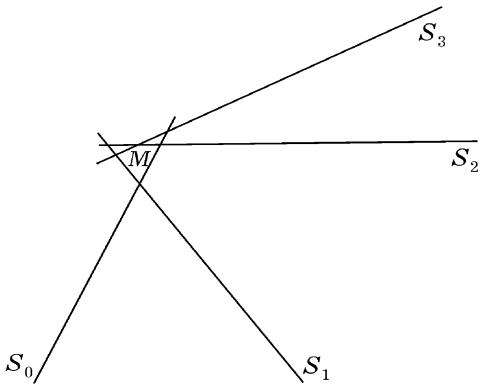


Рис. 18

Пример 1. Точку M наблюдают с нескольких позиций S_1, S_2, S_3, ... , их положение прекрасно известно; измеряют угол MS_1 с каким-то фиксированным направлением MS_0, иначе говоря, азимут MS_1; допустим, это φ_1; и так для всех MS_i.

Каково наиболее вероятное положение M после этих измерений?

Пусть x, y — координаты M; a_i, b_i — координаты S_i; φ_i — угол MS_i с MS_0, взятой в качестве оси абсцисс:

$$\varphi_i = \operatorname{arctg}\frac{y - b_i}{x - a_i}.$$

Уравнения для φ_i, вообще говоря, несовместны, намеченные линии не проходят точно через M и образуют вокруг этой точки маленький многоугольник.

Произвольная точка, взятая внутри этого многоугольника, будет первым приближением, $M_0(x_0, y_0)$. Я полагаю

$$x = x_0 + \xi,$$
$$y = y_0 + \eta,$$

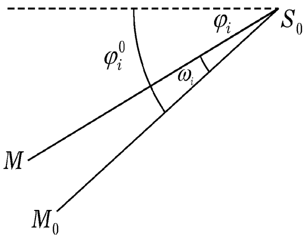


Рис. 19

где ξ и η — очень малые величины. Пусть

$$\varphi_i = \varphi_i^0 + \omega_i.$$

(В случае, изображенном на рис. 19, ω_i отрицательно.) Кроме того,

$$\varphi_i^0 = \operatorname{arctg}\frac{y_0 - b_i}{x_0 - a_i}.$$

Разложим φ_i по возрастающим степеням ξ и η, остановившись на членах второго порядка:

$$\varphi_i = \varphi_i^0 + \frac{\eta(x_0 - a_i) - \xi(y_0 - b_i)}{(x_0 - a_i)^2 + (y_0 - b_i)^2}.$$

Пусть

$$A_i = -\frac{y_0 - b_i}{(x_0 - a_i)^2 + (y_0 - b_i)^2},$$

$$B_i = -\frac{x_0 - a_i}{(x_0 - a_i)^2 + (y_0 - b_i)^2}.$$

Тогда

$$\omega_i = \varphi_i - \varphi_i^0 = A_i\xi + B_i\eta.$$

Наблюдаемое значение φ_i равно ψ_i,

$$\psi_i = \varphi_i^0 + \varepsilon_i.$$

Здесь ε_i будет наблюдаемым значением ω_i. Уравнения, полученные из наблюдений, будут следующими:

$$\varepsilon_i = A_i\xi + B_i\eta.$$

Эти уравнения, вообще говоря, несовместны, т. к. наблюдения неточны; надо выбрать ξ и η так, чтобы

$$\sum (A_i\xi + B_i\eta - \varepsilon_i)^2$$

была минимальна.

Я дифференцирую по ξ и η:

$$\sum A_i(A_i\xi + B_i\eta - \varepsilon_i) = 0,$$
$$\sum B_i(A_i\xi + B_i\eta - \varepsilon_i) = 0;$$

откуда имеем два линейных уравнения для определения ξ и η:

$$\xi \sum A^2 + \eta \sum AB = \sum A\varepsilon,$$
$$\xi \sum AB + \eta \sum B^2 = \sum A\varepsilon.$$

175. Пример 2. Предположим, что измерили девять углов: пусть z_i — истинные значения этих углов, x_i — их наблюдаемые значения, y_i — ошибки. Представим, что эти девять углов связаны четырьмя уравнениями:

$$z_1 + z_2 + z_3 = \pi,$$
$$z_4 + z_5 + z_6 = \pi,$$
$$z_7 + z_8 + z_9 = \pi,$$
$$z_3 + z_4 + z_7 = 2\pi.$$

Если h_1 — очень малая величина, полученная из наблюдений и равная превышению суммы найденных углов над развернутым:

$$x_1 + x_2 + x_3 = \pi + h_1,$$

то

$$y_1 + y_2 + y_3 = h_1.$$

Так же

$$y_4 + y_5 + y_6 = h_2,$$
$$y_7 + y_8 + y_9 = h_3.$$

Четвертое уравнение связи выражает тот факт, что

$$y_3 + y_4 + y_7 = h_4.$$

Требуется определить все y так, чтобы сумма

$$\sum y_i^2$$

была минимальна.

Я хочу ввести четыре вспомогательные величины ε_1, ε_2, ε_3, ε_4, соответствующие четырем величинам h_1, h_2, h_3, h_4. Так, для y_3 я воспользуюсь коэффициентом при y_3 в первом уравнении, где он равен 1, затем во втором и третьем, где он равен 0, затем в четвертом, где он равен 1:

$$y_2 = \varepsilon_1 + \varepsilon_4.$$

Так же находятся

$$y_1 = y_2 = \varepsilon_1,$$
$$y_5 = y_6 = \varepsilon_2,$$
$$y_8 = y_9 = \varepsilon_3,$$
$$y_7 = \varepsilon_2 + \varepsilon_4.$$

Чтобы определить ε, я заменяю все y их значениями:

$$\begin{aligned} 3\varepsilon_1 + \varepsilon_4 &= h_1, \\ 3\varepsilon_2 + \varepsilon_4 &= h_2, \\ 3\varepsilon_3 + \varepsilon_4 &= h_3, \\ \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + 3\varepsilon_4 &= h_4; \end{aligned}$$

откуда

$$3(\varepsilon_1 + \varepsilon_2 + \varepsilon_3) + 3\varepsilon_4 = h_1 + h_2 + h_3$$

и

$$6\varepsilon_4 = 3h_4 - h_1 - h_3 - h_3,$$

и так далее.

Здесь второй способ решения удобнее, чем первый. Есть девять углов и четыре уравнения связи, т. е. пять свободных переменных; мы свели задачу к четырем уравнениям с четырьмя неизвестными, а при первом способе нам бы пришлось решать пять уравнений с пятью неизвестными.

176. Еще один пример. Пример 3. Наблюдали некоторое число точек $M_1, M_2, \ldots, M_n$, не лежащих на одной прямой, хотя в действительности это должно выполняться. Какая прямая будет наиболее вероятной?

Следует определить n новых точек, лежащих на одной прямой так, чтобы сумма квадратов ошибок была наименьшей.

Ошибка будет двойной: в абсциссе и в ординате. Если я предположу, что вероятность ошибки в абсциссе будет совпадать с вероятностью ошибки в ординате, то сумма квадратов ошибок в абсциссе и в ординате будет суммой величин таких же, как

$$\overline{M_iP_i}^2.$$

В действительности нам нужны не точки P_i, а прямая D, проходящая через точки P_i: я заявляю, что M_iP_i должны быть перпендикулярны D.

Пусть это не так, и этот перпендикуляр равен M_iP_i'; заменяя M_iP_i на M_iP_i', я уменьшаю сумму, которая должна достигать минимума, следовательно, она не была минимальной.

Я больше не буду заниматься точками P: я буду искать прямую такую, чтобы сумма квадратов расстояний от точек M_i до этой прямой была минимальна.

Величина, стремящаяся к минимуму, — не что иное, как момент инерции точек относительно этой прямой, если считать, что каждая точка имеет массу, равную 1.

Первым свойством прямой будет то, что она проходит через центр масс. Если поворачивать эту прямую, момент инерции, как известно, меняется по очень простому закону, приводящему к определению эллипсоида инерции. Здесь этот эллипсоид был бы бесконечно сплющен, т. к. точки лежат на плоскости; таким образом, прямая совпадает с большой осью эллипса, в который вырождается эллипсоид. Кроме того, этот эллипс инерции будет очень вытянут, т. к. точки лежат почти на прямой линии.

177. В задаче об определении координат точки на плоскости вероятность ошибки в абсциссе может не совпадать с вероятностью ошибки в ординате. Кроме того, эти две ошибки могут не быть независимыми. Мы детально изучили этот вопрос в предыдущей главе.

В случае с точкой на плоскости мы пришли к рассмотрению ряда малых эллипсов; вероятность того, что координаты точки лежат между x и $x + dx$, y и $y + dy$, выражается функцией

$$e^p\,dx\,dy,$$

и многочлен второй степени P, приравненный к константе, дает нам уравнение одного из этих эллипсов.

Вернемся к точкам на прямой.

Из точки M_1 как центра я описываю эллипс, гомотетичный нормальному и касающийся D. Я составляю сумму квадратов больших осей эллипсов, описанных вокруг разных точек M, и пишу, что она минимальна.

Этот случай легко сводится к предыдущему. С помощью гомографического преобразования эллипсы могут превратиться в окружности. Если, например, малая ось равна половине большой, умножаем все абсциссы на 2, и надо только найти момент инерции, как было только что сделано.

Глава 14
Вычисление ошибок, которых следует избегать

178. Пусть выполнен закон Гаусса. Несколько наблюдений дали нам в качестве результатов измерений $x_1, x_2, \ldots, x_n$. Речь идет об оценке значения h и z; все, что мы знаем, это $x_1, x_2, \ldots, x_n$, а кроме того, мы допускаем, что закон распределения ошибок — гауссовский.

Положим

$$x_i - z = y_i.$$

Требуется знать вероятность, с которой z лежит между z и $z + dz$, а h в то же время лежит между h и $h + dh$.

Это задача на условные вероятности: причины неизвестны, они совпадают с описанным выше двойным событием; следствие известно, это тот факт, что n наблюдений дали результаты $x_1, x_2, \ldots, x_n$.

Если обозначить, как и раньше, $\bar{\omega}_i$ априорную вероятность рассмотренной причины, и p_i — вероятность события, которое происходит, если допустить, что действует эта причина, то апостериорная вероятность причины будет равна

$$\frac{\bar{\omega}_i p_i}{\sum \bar{\omega}_i p_i}.$$

В наиболее общем виде

$$\bar{\omega}_i = \psi(z, h)\, dz\, dh,$$

мы не считаем, что $\psi = 1$ и полагаем a priori, что искусство наблюдателя должно влиять на вероятность, которую мы приписываем константе h.

Величина p_i равна вероятности того, что наблюдения дали результаты, заключенные между

$$x_1 \text{ и } x_1 + dx_1, \quad x_2 \text{ и } x_2 + dx_2, \quad \text{и} \quad x_n \text{ и } x_n + dx_n.$$

Полагаем

$$\Pi = \varphi(x_1 - z)\varphi(x_2 - z)\ldots\varphi(x_n - z),$$

где

$$\varphi(x_i - z) = \sqrt{\frac{h}{\pi}}\, e^{-h(x_i - z)^2};$$

тогда

$$p_i = \Pi\, dx_1\, dx_2 \ldots dx_n,$$

и

$$\frac{\bar{\omega}_i p_i}{\sum \bar{\omega}_i p_i} = \frac{\Pi\psi\, dz\, dh\, dx_1\, dx_2 \ldots dx_n}{\int \Pi\psi\, dz\, dh\, dx_1\, dx_2 \ldots dx_n}.$$

По x интегрирование не производится, поэтому соответствующие дифференциалы исчезают вверху и внизу. По z интеграл берется от $-\infty$ до $+\infty$, по h — от 0 до $+\infty$.

179. Можно записать

$$\Pi = \left(\frac{h}{\pi}\right)^{\frac{n}{2}} e^{-hP},$$

где

$$P = (x_1 - z)^2 + (x_2 - z)^2 + \ldots + (x_n - z)^2.$$

Здесь P — многочлен второй степени по z, достигающий своего минимума при z, равном среднему арифметическому величин x; обозначим этот минимум, который будет положительным, $n\alpha^2$. Я полагаю

$$y = z - \frac{x_1 + x_2 + \ldots + x_n}{n},$$

откуда

$$dz = dy$$

и

$$P = n(y^2 + \alpha^2).$$

Тогда

$$\frac{\bar{\omega}_i p_i}{\sum \bar{\omega}_i p_i}$$

становится

$$\frac{h^{\frac{n}{2}}\, e^{-nh(y^2 + \alpha^2)}\psi\, dy\, dh}{\int h^{\frac{n}{2}}\, e^{-nh(y^2 + \alpha^2)}\psi\, dy\, dh}.$$

По y, как и по z, интегрирование проводится от $-\infty$ до $+\infty$.

180. Ищем вероятность того, что h лежит между h и $h+dh$, а y принимает все возможные значения:

$$\frac{dh \int\limits_{-\infty}^{+\infty} h^{\frac{n}{2}} e^{-nh(y^2+\alpha^2)} \psi \, dy}{\int\limits_{0}^{+\infty} dh \int\limits_{-\infty}^{+\infty} h^{\frac{n}{2}} e^{-nh(y^2+\alpha^2)} \psi \, dy}.$$

Функция ψ зависела от z и h; теперь она зависит от y и h. Сначала сузим задачу, считая все значения y равновероятными, т. е. ψ — не зависящей от y. Имеем

$$\int\limits_{-\infty}^{+\infty} e^{-nh(y^2+\alpha^2)} \, dy = e^{-nh\alpha^2} \int\limits_{-\infty}^{+\infty} e^{-nhy^2} \, dy = e^{-nh\alpha^2} \sqrt{\frac{\pi}{nh}}.$$

Тогда вероятность, связанная только с h, станет

$$\frac{dh \cdot h^{\frac{n-1}{2}} e^{-nh\alpha^2} \psi}{\int\limits_{0}^{\infty} dh \cdot h^{\frac{n-1}{2}} e^{-nh\alpha^2} \psi}.$$

181. Прежде всего, чему равно наиболее вероятное значение h? Ищем максимум числителя, т. к. вероятность пропорциональна ему. Этот числитель можно записать как

$$dh \frac{\psi}{\sqrt{h}} \left(\sqrt{h} e^{-h\alpha^2}\right)^n.$$

Между тем, максимум выражения $dh \cdot \Phi^n \times f$ в общем случае имеет место, когда

$$\frac{f'}{f} + \frac{n\Phi'}{\Phi} = 0,$$

и, если n очень велико, членом $\frac{f'}{f}$ можно пренебречь; итак, максимум достигается одновременно с максимумом функции Φ.

Здесь Φ равна $\sqrt{h}e^{-h\alpha^2}$; приравняем нулю логарифмическую производную:

$$\frac{1}{2h} - \alpha^2 = 0.$$

Так получаем наиболее вероятное значение h.

Можно заметить, что при очень большом количестве наблюдений произвольная функция ψ не играет никакой роли. Это неудивительно: гипотеза, которую мы a priori основывали на большем или меньшем искусстве наблюдателя, меняется из-за большого числа проверенных нами результатов.

182. Каково вероятное значение h^p? Это вероятное значение равно

$$\frac{\int\limits_0^\infty h^{p+\frac{n-1}{2}} e^{-nh\alpha^2}\psi\,dh}{\int\limits_0^\infty h^{\frac{n-1}{2}} e^{-nh\alpha^2}\psi\,dh}$$

или

$$\frac{\int\limits_0^\infty h^p f\Phi^n\,dh}{\int\limits_0^\infty f\Phi^n\,dh}.$$

Мы знаем, к какому пределу стремится отношение этих интегралов, если n очень велико; этот предел равен

$$\frac{h_0^p f(h_0)}{f(h_0)} = h_0^p,$$

где h_0 — наиболее вероятное значение. Итак, вероятное значение h^p равно h_0^p. Этот вывод верен только при условии, что n очень велико, а p конечно.

Полагаем $\psi = 1$: мы приходим к интегралам Эйлера

$$\bar{h}^p = \frac{\int\limits_0^\infty h^{p+\frac{n-1}{2}} e^{-nh\alpha^2}\,dh}{\int\limits_0^\infty h^{\frac{n-1}{2}} e^{-nh\alpha^2}\,dh} = \frac{\Gamma\left(p+\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{n+1}{2}\right)} \cdot \frac{(n\alpha^2)^{-p-\frac{n+1}{2}}}{(n\alpha^2)^{-\frac{n+1}{2}}}.$$

Отношение асимптотических значений дает

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

что совпадает с нашим результатом. Но если не считать p конечным, это рассуждение использовать нельзя. Пусть p очень велико, например, равно $\frac{n}{2}$:

$$\frac{\Gamma(2p)}{\Gamma(p)}(n\alpha^2)^{-p} = \frac{(2p)^{2p}e^{-2p}\sqrt{4\pi p}}{p^p e^{-p}\sqrt{2\pi p}}(2p\alpha^2)^{-p}.$$

Правая часть вырождается в $2^p e^{-p}\sqrt{2}(\alpha^2)^{-p}$. Найдено совсем другое выражение для среднего значения h^p:

$$\frac{2^p e^{-p}\sqrt{2}}{(\alpha^2)^{-p}} \quad \text{или} \quad \left(\frac{2}{e\alpha^2}\right)^p \sqrt{2}.$$

183. Вероятность допущенной ошибки. Эта задача разделяется на три.

1° Можно попытаться вычислить вероятность a priori.

Наблюдения еще не сделаны; известно только, что собираются провести n наблюдений и применить метод наименьших квадратов. Кроме того, нам известно искусство наблюдателя.

2° Совершенно другая задача появляется, если мы не знаем изначально, какое значение надо приписать константе, входящей в формулу Гаусса. Нам неизвестно искусство наблюдателя, но мы знаем результаты наблюдений.

3° Нам известно искусство наблюдателя и результаты наблюдений[(18)].

184. Первая задача. Результаты неизвестны, но известно искусство наблюдателя. Тогда все веса считаются одинаковыми.

Пусть будут допущены ошибки, равные $y_1, y_2, \ldots, y_n$; $x_1, x_2, \ldots, x_n$ — приближенные значения истинных величин $z_1, z_2, \ldots, z_n$.

Их соединяют $q = n - p$ уравнений связи:

$$\Phi_i(z_1, z_2, \ldots, z_n) = 0.$$

Если подставить вместо $z_1, z_2, \ldots, z_n$ величины $x_1, x_2, \ldots, x_n$, эти уравнения не будут выполняться, и мы получим

$$\Phi_i(x_1, x_2, \ldots, x_n) = \mu_i,$$

где μ_i очень мало.

Я заменяю x_k на y_k+z_k и раскладываю по возрастанию степеней y_k, оставляя только члены первого порядка:

$$A_1y_1 + \ldots + A_ny_n = \mu_i.$$

Для наглядности возьмем $n = 3$ и предположим, что есть два уравнения связи

$$A_1y_1 + A_2y_2 + A_3y_3 = \mu, \quad B_1y_1 + B_2y_2 + B_3y_3 = \mu'.$$

Величины μ и μ' очень малы, но я не знаю их значений, т. к. наблюдения еще не проведены. Речь идет о вычислении поправок y_i', которые надо сделать к наблюдаемым значениям x_i; очевидно, эти поправки зависят от μ и μ', например,

$$y_1' = \theta(\mu,\, \mu').$$

Я не знаю вид функции θ, но могу разложить ее по степеням μ и μ' и, т. к. оба μ очень малы, пренебречь их квадратами; я прихожу к системе

$$\begin{aligned} y_1' &= \lambda_1\mu + \lambda_1'\mu', \\ y_2' &= \lambda_2\mu + \lambda_2'\mu', \\ y_3' &= \lambda_3\mu + \lambda_3'\mu', \end{aligned}$$

в общем виде

$$y_i' = \lambda_i\mu + \lambda_i'\mu',$$

где константы λ еще надо найти. Гаусс определяет их так, чтобы вероятное значение для $(y_i - y_i')^2$, т. е. для

$$(y_i - \lambda_i\mu - \lambda_i'\mu')^2,$$

было минимально.

Если предположить, что изначально известно искусство наблюдателя, но не результаты наблюдений, надо оценивать вероятное значение. Если бы было известно и то, и другое, решение, вообще говоря, было бы иным.

185. Чтобы как следует объяснить эту разницу, я изучу более простой случай: пусть несколько раз наблюдают одну и ту же величину z. Из наблюдений берется среднее, что согласуется с методом наименьших квадратов. Ошибка, допущенная в среднем, будет равна

$$\frac{y_1 + y_2 + \ldots + y_n}{n}.$$

Априорная вероятность ошибки, допущенной, например, при первом наблюдении, будет равна $\varphi(y_1)\,dy_1$.

Априорная вероятность того, что z лежит между z и $z + dz$, будет $\psi(z)\,dz$.

Ищем вероятное значение $\left(\frac{y_1 + y_2 + \ldots + y_n}{n}\right)^2$, когда результаты наблюдений *неизвестны*, но известна ψ:

$$A = \int \ldots \iint \left(\frac{y_1 + y_2 + \ldots + y_n}{n}\right)^2 \varphi(y_1)\varphi(y_2)\ldots\varphi(y_n)\,dy_1\,dy_2\,\ldots\,dy_n.$$

Это первое вероятное значение.

Если одновременно известны и искусство наблюдателя, и результаты наблюдений, имеем второе, совсем другое вероятное значение:

$$B = \frac{\displaystyle\int \left(\frac{y_1 + y_2 + \ldots + y_n}{n}\right)^2 \varphi(y_1)\varphi(y_2)\ldots\varphi(y_n)\psi(z)\,dz}{\int \varphi(y_1)\varphi(y_2)\ldots\varphi(y_n)\psi(z)\,dz}.$$

В первом случае мы имели дело с n независимыми переменными $y_1, y_2, \ldots y_n$, почему и интеграл был порядка n; во втором случае имелась только одна переменная z.

Предположим, что вместо того чтобы применить правило среднего и, следовательно, придать z значение

$$\frac{x_1 + x_2 + \ldots + x_n}{n},$$

берут другое значение

$$\frac{x_1 + x_2 + \ldots + x_n}{n} + \varepsilon;$$

допущенная ошибка была бы равна

$$\frac{y_1 + y_2 + \ldots + y_n}{n} + \varepsilon.$$

Тогда *при неизвестных результатах наблюдений* мы бы нашли в качестве вероятного значения квадрата этой ошибки

$$A = \int \ldots \iint \left(\frac{y_1 + y_2 + \ldots + y_n}{n} + \varepsilon \right)^2 \varphi(y_1)\varphi(y_2)\ldots\varphi(y_n)\, dy_1\, dy_2\, \ldots\, dy_n,$$

а в качестве вероятного значения того же квадрата *при известных результатах* —

$$B = \frac{\displaystyle\int \left(\frac{y_1 + y_2 + \ldots + y_n}{n} + \varepsilon \right)^2 \varphi(y_1)\varphi(y_2)\ldots\varphi(y_n)\psi(z)\, dz}{\int \varphi(y_1)\varphi(y_2)\ldots\varphi(y_n)\psi(z)\, dz}.$$

Величина A достигает минимума при $\varepsilon = 0$ *всегда, когда функция φ четная.*

Напротив, для того чтобы B достигало своего минимума при $\varepsilon = 0$, *должен быть верен закон Гаусса.*

В рассуждениях, которые мы воспроизвели в главе XI, параграфах 129 и далее, Гаусс принял первую точку зрения и показал, что правило среднего всегда обосновано.

Он принял вторую точку зрения в рассуждениях, которые мы воспроизвели в главе X, параграфах 108 и далее, и показал, что это правило обосновано, только если верен закон Гаусса.

186. Вернемся к нашей задаче и постараемся определить λ. Надо устремить к минимуму выражение

$$\overline{(y_1 - \lambda_1 \mu - \lambda_1' \mu')^2}.$$

Здесь μ и μ' — линейные функции от y; итак, к минимуму стремится однородный многочлен второй степени по $y_1, y_2, \ldots, y_n$.

Согласно гипотезе, все веса равны. Пусть m^2 — вероятное значение y_1^2, m^2 также будет вероятным значением y_2^2 и y_3^2.

Если бы нам надо было сделать три наблюдения, мы бы не имели права *заранее* считать, что y_1 больше y_2 или y_3.

Вероятное значение $y_1 y_2$ совпадает с нулем: оно равно произведению вероятных значений y_1 и y_2, а вероятное значение y_1 равно нулю, т. к. систематических ошибок нет. Это рассуждение верно, т. к. результаты наблюдений неизвестны, и оно не будет верным, если узнать их.

Как найти вероятное значение многочлена? Заменяем все квадраты на m^2, все двойные произведения — на 0.

Если подставить вместо μ и μ' их выражения как функции от y, этот многочлен превратится в

$$(y_1 - \lambda_1 A_1 y_1 - \lambda_1' B_1 y_1)^2 + \dots ;$$

коэффициент при y_1^2 равен $(\lambda_1 A_1 + \lambda_1' B - 1)^2$; при y_2^2 — $(\lambda_1 A_2 + \lambda_1' B_2)^2$; при y_3^2 — $(\lambda_1 A_3 + \lambda_1' B_3)^2$.

Итак, $\overline{(y_1 - \lambda_1 \mu - \lambda_1' \mu')}^2$ равно

$$m^2[(A_1\lambda_2 + B_1\lambda_1' - 1)^2 + (A_2\lambda_2 + B_2\lambda_1')^2 + (A_3\lambda_2 + B_3\lambda_1')^2].$$

Это и есть то выражение, которое должно достигать минимума.

187. Назовем это выражение $m^2 P$: оно представляет собой вероятное значение квадрата ошибки $(y_1 - y_1')^2$, существующей после внесения поправки.

Приравниваем к 0 две производные:

$$\frac{dP}{d\lambda_1} = \frac{dP}{d\lambda_1'} = 0,$$

то есть

$$A_1(A_1\lambda_2 + B_1\lambda_1' - 1)^2 + A_2(A_2\lambda_2 + B_2\lambda_1')^2 + A_3(A_3\lambda_2 + B_3\lambda_1')^2 = 0,$$

и, симметрично,

$$B_1(A_1\lambda_2 + B_1\lambda_1' - 1)^2 + B_2(A_2\lambda_2 + B_2\lambda_1')^2 + B_3(A_3\lambda_2 + B_3\lambda_1')^2 = 0.$$

Эта система переписывается как

$$\lambda_1 \sum A^2 + \lambda_1' \sum AB = A_1,$$
$$\lambda_1 \sum AB + \lambda_1' \sum B^2 = B_1.$$

Она линейна по λ_1 и λ_1'.

Я добавляю сюда уравнение, задающее y'_1:

$$\lambda_1\mu + \lambda'_1\mu' = y'_1.$$

Исключая λ_1 и λ'_1, получаем

$$\begin{vmatrix} \sum A^2 & \sum AB & A_1 \\ \sum AB & \sum B^2 & B_1 \\ \mu & \mu' & y'_1 \end{vmatrix} = 0.$$

Отсюда находится y'_1 и, рассуждая точно также, можно было бы получить y'_2 и y'_3.

188. Таковы поправки, которые надо внести в результаты наблюдений, чтобы вероятное значение квадрата ошибки достигало минимума.

Эти поправки согласуются с методом наименьших квадратов.

Действительно, y'_1, y'_2, y'_3 удовлетворяют уравнениям:

$$A_1y'_1 + A_2y'_2 + A_3y'_3 = \mu, \quad B_1y'_1 + B_2y'_2 + B_3y'_3 = \mu'.$$

Чтобы вычислить все y', надо устремить к минимуму сумму $y_i^{2\prime}$. Итак,

$$\sum y'_i\,dy'_i = 0.$$

С другой стороны,

$$\sum A_i\,dy'_i = 0,$$
$$\sum B_i\,dy'_i = 0.$$

Первое из этих уравнений должно быть следствием двух других; поэтому

$$y'_i = \varepsilon A_i + \varepsilon' B_i,$$

откуда

$$\varepsilon \sum A^2 + \varepsilon' \sum AB = \mu,$$

и, симметрично,

$$\varepsilon \sum AB + \varepsilon' \sum B^2 = \mu'.$$

Исключаем ε и ε' из этих трех уравнений:

$$\begin{vmatrix} \sum A^2 & \sum AB & \mu \\ \sum AB & \sum B^2 & \mu' \\ \mu & \mu' & y_i' \end{vmatrix} = 0.$$

Определитель совпадает с предыдущим.

Таким образом, результат одинаков, применяется ли метод наименьших квадратов или поправка вносится так, чтобы после нее вероятное значение квадрата ошибки достигало минимума.

189. Теперь можно спросить, каково это минимальное значение: оно равно минимуму m^2P. Многочлен P может быть представлен в очень простом виде. Делаем его однородным:

$$P = (A_1\lambda_1 + B_1\lambda_1' - \lambda'')^2 + (A_2\lambda_1 + B_2\lambda_1')^2 + (A_3\lambda_1 + B_3\lambda_1' - \lambda'')^2$$

и применяем теорему об однородных функциях; имеем

$$2P = \frac{dP}{d\lambda_1}\lambda_1 + \frac{dP}{d\lambda_1'}\lambda_1' + \frac{dP}{d\lambda_1''}\lambda_1''.$$

Но $\dfrac{dP}{d\lambda_1}$ и $\dfrac{dP}{d\lambda_1'}$ — нули; $\lambda_1'' = 1$; остается

$$2P = \frac{dP}{d\lambda_1''}$$

$$P = \frac{1}{2}\frac{dP}{d\lambda_1''} = 1 - A_1\lambda_1 - B_1\lambda_1'.$$

Его произведение с m^2 равно вероятному значению квадрата ошибки, возникающей после поправки.

Когда количество наблюдений становится все больше и больше, вероятное значение квадрата ошибки убывает.

190. Введем третью величину и третье уравнение связи:

$$C_1y_1 + C_2y_2 + C_3y_3 + C_4y_4 = \mu''.$$

Только что мы устремляли к минимуму

$$\overline{(y_1 - \lambda_1\mu - \lambda_1'\mu')^2}; \tag{1}$$

теперь —

$$\overline{(y_1 - \lambda_1\mu - \lambda_1'\mu' - \lambda_1''\mu'')}^2. \tag{2}$$

Здесь больше неопределенности, т. к. добавляется λ_1''; минимум выражения (2), очевидно, меньше, чем у выражения (1), т. к. достаточно положить $\lambda_1'' = 0$ в выражении (2), чтобы вернуться к выражению (1).

191. Идем дальше. Пусть y — действительно допущенная ошибка. Если y' — поправка, то $y - y'$ равна ошибке, возникающей после коррекции.

Сумма $\sum y^2$ равна сумме квадратов допущенных ошибок; вероятное значение этой суммы равно nm^2.

Ищем вероятное значение суммы квадратов поправок, $\sum \bar{y}^{2\prime}$, и вероятное значение суммы квадратов ошибок после коррекции, $\sum \overline{(y-y')}^2$.

192. Я замечаю, что у нас

$$y_i' = \lambda_i\mu + \lambda_i'\mu';$$

то есть y' — линейная функция от обоих μ, которые в свою очередь являются линейными функциями от y. Итак, y' — линейная функция от y. Эти функции y' не будут линейно независимы, т. к. их можно линейно выразить через любые две из них — в нашем случае, а вообще — как функцию от стольких из них, сколько есть величин μ, т. е. от $n - p$ штук, т. к. различных μ будет столько же, сколько есть уравнений связи.

Рассмотрим $y_i - y_i'$: это тоже линейные функции от y, но тоже не линейно независимые; они связаны условиями

$$A_1(y_1 - y_1') + A_2(y_2 - y_2') + A_3(y_3 - y_3') = 0,$$
$$B_1(y_1 - y_1') + B_2(y_2 - y_2') + B_3(y_3 - y_3') = 0.$$

У нас есть два линейных соотношения, а вообще их p.

Итак, все y' выражаются линейно через $n-p$ из них, а все $y-y'$ — через p из них.

193. Я заявляю, что точно так же

$$\sum y_i y_i' = \sum y_i'^2.$$

Действительно,

$$y_i'^2 = \varepsilon A + \varepsilon' B_i, \sum y_i y_i' = \varepsilon \sum A_i y_i + \varepsilon' \sum B_i y_i = \varepsilon\mu + \varepsilon'\mu',$$
$$\sum y_i'^2 = \varepsilon \sum A_i y_i' + \varepsilon' \sum B_i y_i' = \varepsilon\mu + \varepsilon'\mu'.$$

Выполняется и другое тождество:

$$\sum y^2 = \sum y'^2 + \sum (y-y')^2.$$

Действительно, раскладывая $\sum (y-y')^2$, получим

$$\sum y^2 = \sum y'^2 + \sum y^2 - 2\sum yy' + \sum y'^2,$$

что будет верно, т. к. в силу предыдущего тождества

$$\sum y'^2 - 2\sum yy' + \sum y'^2 = 0.$$

194. Ищем вероятное значение $\sum \bar{y}'^2$. Это квадратичная форма по y.

Умножаем на m^2 сумму коэффициентов, стоящих при квадратах, или иначе говоря, рассматриваем *уравнение на S.*

Пусть F и F' — две квадратичные формы по n переменным; если S — константа, то

$$F - SF'$$

снова будет квадратичной по n переменным.

Написав, что дискриминант равен нулю, получают уравнение порядка n, т. е. уравнение для S.

Это уравнение имеет свойство не меняться при линейной замене переменных: оно инвариантно.

Предположим теперь, что мы намерены вычислить вероятное значение для квадратичной формы F; я беру $F' = \sum y^2$. Пишем, что дискриминант от

$$F - S\sum y^2$$

равен нулю.

Сумма корней этого уравнения равна сумме коэффициентов при квадратах.

Пусть

$$F = Ay_1^2 + A'y_2^2 + A''y_3^2 + 2By_2y_3 + 2B'y_3y_1 + 2B''y_1y_2.$$

Дискриминант равен

$$\begin{vmatrix} A-S & B'' & B' \\ B'' & A'-S & B \\ B' & B & A''-S \end{vmatrix} = 0$$

или

$$-S^3 + (A + A' + A'')S^2 + \ldots = 0.$$

Итак, сумма корней равна $A+A'+A''$; с другой стороны, если вероятное значение y_1^2 равно m^2, а для y_1y_2 оно равно 0, то вероятное значение F будет

$$m^2(A + A' + A'').$$

Иначе говоря, следует составить форму $F-S\sum y^2$, взять сумму корней уравнения на S и умножить ее на m^2.

195. Применим это правило к квадратичной форме $\sum y^2$.

Пусть форма

$$\Phi = \sum y'^2 - S\sum y^2$$

или, согласно параграфу 193,

$$\Phi = (1-S)\sum y'^2 - S\sum (y-y')^2.$$

Составляем уравнение на S и ищем сумму корней.

Величины y' линейно выражаются через $n-p$ из них; $\sum y'^2$ раскладываются таким образом на сумму $n-p$ квадратов, и если ξ — линейные функции от y, то

$$\sum y'^2 = \xi_1^2 + \xi_2^2 + \ldots + \xi_{n-p}^2.$$

Все $y-y'$ выражаются через p из них, и если η — линейные функции от y, то

$$\sum (y-y')^2 = \eta_1^2 + \eta_2^2 + \ldots + \eta_p^2.$$

Полученные таким образом n линейных функций будут линейно независимы. Действительно, заметим, что

$$\sum y^2 = \sum \xi^2 + \sum \eta^2.$$

Левая часть — сумма n квадратов

$$y_1^2,\ y_2^2,\ \dots\ y_n^2;$$

дискриминант левой части равен 1. Правая часть не может иметь дискриминантом 0. Итак,

$$\Phi = (1 - S)\sum \xi^2 - S\sum \eta^2.$$

Дискриминант равен

$$(S-1)^{n-p} \cdot S^p = 0;$$

$n-p$ корней равны 1, p корней равны 0, сумма корней равна $n-p$.

Вероятное значение $\sum y^2$ равно nm^2; по изложенному выше правилу вероятное значение $\sum \bar{y'}^2$ будет равно

$$\sum \bar{y'}^2 = (n-p)m^2.$$

Итак,

$$\sum \overline{(y-y')^2} = pm^2,$$

то есть эта сумма равна разности двух других.

В итоге:

1° вероятное значение суммы квадратов допущенных ошибок равно nm^2;

2° вероятное значение суммы квадратов внесенных поправок равно $(n-p)m^2$;

3° вероятное значение суммы квадратов ошибок после коррекции равно pm^2.

196. Вероятное значение $\sum \bar{y'}^2$ меньше, чем вероятное значение $\sum \bar{y}^2$.

Это легко было предвидеть. Мы пытаемся определить поправки так, чтобы сумма их квадратов была минимальна: принцип тот же, что и в методе наименьших квадратов.

Сейчас мы покажем, что если рассматривать ошибку, допущенную при наблюдении, то по мере увеличения числа наблюдений она стремится к нулю.

Предположим, что число наблюдений постоянно растет; число p остается константой, как и pm^2; число слагаемых увеличивается; возможно, каждое из них постоянно убывает.

Если мы рассмотрим наименьшую из величин $\overline{(y_k - y'_k)^2}$, то она обязательно будет меньше, чем $\frac{pm^2}{n}$.

Наблюдаем одну и ту же величину n раз; есть единственная независимая переменная: $p = 1$,

$$\sum \overline{(y - y')^2} = m^2.$$

У нас n слагаемых, n наблюдений сделаны в одинаковых условиях; итак,

$$\overline{(y - y')^2} = \frac{m^2}{n}.$$

197. До сих пор мы предполагали, что известна точность наблюдений, но сами они еще не проведены.

Поставим задачу по-другому: о точности ничего не известно, но наблюдения уже проведены.

Мы хотим вывести значение h или m^2. Вот решение этой задачи. Все y известны; все y' найдены по методу наименьших квадратов. Сумма $\sum {y'}^2$ известна.

Я приравниваю значение этой суммы к вероятному значению, вычисленному a priori:

$$\sum {y'}^2 = (n - p)m^2,$$

откуда нахожу m^2.

198. Ж. Бертран критиковал этот метод. Действительно, если бы мы использовали другое выражение, например, $\sum {y'}^4$, то найденное значение m не было бы тем же самым. Метод может показаться сомнительным.

Эта задача — задача на условные вероятности; приме́ним соответствующие правила вычислений.

Требуется найти апостериорную вероятность того, что h лежит в определенных пределах. Эта вероятность равна

$$\frac{p_i \bar{\omega}_i}{\sum p_i \bar{\omega}_i}.$$

Здесь $\bar{\omega}_i$ — априорная вероятность причины, т. е. того, что h лежит между h и $h + dh$; p_i — вероятность, с которой, если причина действовала, наблюдения дали результаты, заключенные между x_1 и $x_1 + dx_1$, x_2 и $x_2 + dx_2$, … , x_n и $x_n + dx_n$ соответственно.

Ищем вероятное значение $f(h)$ — какой-то функции от h; это вероятное значение равно

$$\overline{f(h)} = \frac{\sum f(h) p_i \bar{\omega}_i}{\sum p_i \bar{\omega}_i}.$$

Замечаем, что результат будет зависеть от априорной вероятности; тогда результат Гаусса уже не может быть совершенно точным.

Если я определяю h из соотношения

$$f(h) = \overline{f(h)},$$

то вероятное значение h будет зависеть от функции f.

Если я ищу наиболее вероятное значение, происходит то же самое.

Эту сложность можно обойти при одном условии: n должно быть очень велико. Множитель $\bar{\omega}_i$ больше не будет оказывать большого влияния; так, для

$$f(h) = h^{\nu}$$

все методы приводят к одному результату, если число наблюдений n всегда очень велико.

199. Когда наблюдают какую-то величину z, и наблюдения дают результаты $x_1, x_2, \ldots, x_n$, можно представить $\bar{\omega}_i$ и p_i как

$$\bar{\omega}_i = \psi(h,\, z)\, dh\, dz,$$
$$p_i = \Pi\, dx_1\, dx_2 \ldots dx_n.$$

Апостериорная вероятность того, что h лежит между h и $h + dh$, а z — между z и $z + dz$, равна

$$\frac{\Pi\psi(h,\, z)\, dh\, dz\, dx_1\, dx_2 \ldots dx_n}{\int \Pi\psi(h,\, z)\, dh\, dz\, dx_1\, dx_2 \ldots dx_n}.$$

Дифференциалы $dx_1\, dx_2 \ldots dx_n$ сокращаются, а интеграл надо взять по h от 0 до $+\infty$ и по z от $-\infty$ до $+\infty$.

Представим, что вместо одной величины z их n: $z_1, z_2, \ldots, z_n$, — и они зависят от p переменных $u_1, u_2, \ldots, u_p$, а наблюдаемые значения этих z равны $x_1, x_2, \ldots, x_n$; допущенные ошибки равны $y_1, y_2, \ldots, y_n$.

Величина $\bar{\omega}_i$ равна априорной вероятности причины; здесь это означает, что h лежит между h и $h + dh$, u_1 лежит между u_1 и $u_1 + du_1$, u_2 — между u_2 и $u_2 + du_2$; ... ,

$$\bar{\omega}_i = \psi(h,\, u_1,\, u_2,\, \ldots\, ,\, u_p)\, dh\, du_1\, du_2 \ldots du_p.$$

Величина p_i — вероятность следствия в предположении, что действовала данная причина,

$$p_i = \Pi\, dx_1\, dx_2 \ldots dx_n,$$

где

$$\Pi = \sqrt{\frac{h}{\pi}}\, e^{-h \sum y^2}.$$

Апостериорная вероятность равна

$$\frac{\Pi\psi(h,\, u_1,\, u_2,\, \ldots\, ,\, u_p)\, dh\, du_1\, du_2 \ldots du_p\, dx_1\, dx_2 \ldots dx_n}{\int \Pi\psi(h,\, u_1,\, u_2,\, \ldots\, ,\, u_p)\, dh\, du_1\, du_2 \ldots du_p\, dx_1\, dx_2 \ldots dx_n}.$$

Интеграл берется по h от 0 до $+\infty$, по всем u от $-\infty$ до $+\infty$; что касается дифференциалов от x, они сокращаются, как и раньше.

Итак, найденная вероятность записывается в виде

$$\frac{\Pi\psi\, dh\, du_1\, du_2 \ldots du_p}{\int \Pi\psi\, dh\, du_1\, du_2 \ldots du_p}.$$

200. Она зависит от функции ψ. Эта функция совершенно произвольна и соответствует представлению, которое мы составляем a priori о значениях u, и точности, которую мы a priori приписываем наблюдениям; но ψ не играет большой роли, если наблюдений много.

Я хочу придать функции ψ особый вид, предположив, что она не зависит от h.

Эту точку зрения можно оправдать, говоря, что измерения дали всем u значения, очень близкие друг к другу; что они лежат на малом интервале, где z изменяется очень мало, если наблюдения согласованы.

Вероятное значение h^ν будет равно

$$\overline{h_\nu} = \frac{\int \Pi\psi h^\nu\, dh\, du_1\, du_2 \ldots du_p}{\int \Pi\psi\, dh\, du_1\, du_2 \ldots du_p}.$$

Эти два интеграла должны вычисляться так же: h меняется от 0 до $+\infty$, а все u от $-\infty$ до $+\infty$.

201. Что такое $\sum y^2$? Это функция от x_i, которые известны, и от z_i, которые зависят от u.

По методу наименьших квадратов получают значение u_i^0, которое надо взять вместо u_i; определенные таким образом величины u не точны, но наиболее приемлемы,

$$u_i = u_i^0 + v_i,$$

где v_i очень мало.

Все y_i — функции от v_i, и их можно считать линейными, если пренебречь членами второго порядка.

Многочлен

$$\sum y^2 = P$$

имеет вторую степень по v_i, но он не однороден; он достигает минимума, когда все v_i — нули. Уравнения

$$\frac{dP}{dv_i} = 0$$

должны выполняться при нулевых v_i.

Итак, P включает члены только второго и нулевого порядка; членов первого порядка нет.

$$P = P_0 + P_2.$$

Тогда P_0 — минимум $\sum y^2$, то, что мы назвали $\sum {y'}^2$, чье вероятное значение равно $(n-p)m^2$. Таким образом, P_0 очень велико, и вероятность того, что P_0 велико, тем больше, чем больше произведено наблюдений.

$$P_0 = (n-p)A,$$

где A — некоторая константа.

Многочлен P_2 получается при сложении друг с другом членов второго порядка; их количество очень велико, оно равно n, и коэффициенты многочлена P_2 имеют тот же порядок, что и n.

$$P_2 = (n-p)Q,$$

где Q имеет один порядок с A.

Вероятное значение h^ν будет равно

$$\overline{h^\nu}\frac{\displaystyle\int \psi h^\nu \left(\frac{h}{\pi}\right)^{\frac{n}{2}} e^{-h(n-p)(A+Q)} dh\, dv_1\, dv_2 \dots dv_p}{\displaystyle\int \psi \left(\frac{h}{\pi}\right)^{\frac{n}{2}} e^{-h(n-p)(A+Q)} dh\, dv_1\, dv_2 \dots dv_p};$$

интегрирование проводится по h от 0 до $+\infty$ и по всем v от $-\infty$ до $+\infty$.

202. Первый интеграл связан с

$$e^{-h(n-p)A},$$

которое зависит от h, и с

$$e^{-h(n-p)Q},$$

которое зависит от v. Внизу и вверху надо сосчитать

$$\int e^{-h(n-p)Q} dv_1\, dv_2 \dots dv_p.$$

Здесь Q — однородный многочлен второй степени от v; я полагаю

$$\omega_i = v_i\sqrt{h}.$$

Выражение hQ становится однородным многочленом второго порядка по ω, который мы обозначим Q'. Интеграл превращается в

$$\int e^{-h(n-p)Q'} d\omega_1\, d\omega_2 \dots d\omega_p h^{-\frac{p}{2}}.$$

Пределы интегрирования остаются прежними, и значение интеграла равно

$$Bh^{-\frac{p}{2}},$$

где B не зависит от h. Тогда

$$\overline{h^\nu} = \frac{\int\limits_0^\infty dh\, \psi h^\nu \pi^{-\frac{n}{2}} h^{\frac{n-p}{2}} e^{-hA(n-p)} B}{\int\limits_0^\infty dh\, \psi \pi^{-\frac{n}{2}} h^{\frac{n-p}{2}} e^{-hA(n-p)} B}$$

или

$$\overline{h^\nu} = \frac{\int\limits_0^\infty \psi h^\nu h^{\frac{n-p}{2}} e^{-hA(n-p)} dh}{\int\limits_0^\infty \psi h^{\frac{n-p}{2}} e^{-hA(n-p)} dh}.$$

Я обозначу

$$\Phi = \sqrt{h} e^{-hA},$$

$$\overline{h^\nu} = \frac{\int\limits_0^\infty \psi h^\nu \Phi^{n-p} dh}{\int\limits_0^\infty \psi \Phi^{n-p} dh},$$

и формула будет зависеть от ψ.

203. Если бы мы хотели решать задачу дальше, следовало бы ввести какую-нибудь гипотезу относительно ψ. Однако, если предположить $n-p$ очень большим, то вид функции ψ не имеет значения. Когда есть

$$\frac{\int F \Phi^n dh}{\int F_1 \Phi^n dh},$$

и n бесконечно растет, предел этого отношения равен

$$\frac{F(h_0)}{F_1(h_0)},$$

где h_0 — значение, при котором Φ достигает максимума. Итак, в нашем случае имеем

$$\overline{h^\nu} = \frac{\psi(h_0) h_0^\nu}{\psi(h_0)},$$

то есть

$$\overline{h^\nu} = h_0^\nu,$$

где h_0 — значение, на котором Φ достигает максимума.

При условии очень большого n вероятное значение h^ν больше не зависит от ψ; вероятное значение h всегда равно h_0, какой бы ни была ψ.

Это не будет верно, если не предполагать n очень большим.

Более того, n должно быть очень велико не только по абсолютной величине, но и по отношению к p, с одной стороны, и к ν, с другой стороны.

Так, если бы

$$\nu = \frac{n-p}{2},$$

то надо было бы искать максимум не Φ, а $\Phi h^{\frac{1}{2}}$.

204. Если бы n было не очень велико, нам бы пришлось искать максимум

$$\psi\Phi^{n-p},$$

откуда

$$\frac{\psi'(h)}{\psi(h)} + (n-p)\frac{\Phi'(h)}{\Phi(h)} = 0;$$

если n велико, значение h очень близко к тому, на котором достигается максимум $\Phi(h)$. Имеем

$$h_0 = \frac{1}{2A} = \frac{n-p}{2\sum y'^2},$$

так как

$$\sum y'^2 = (n-p)A;$$

этот результат сочетается с законом Гаусса,

$$\sum y'^2 = (n-p)m^2.$$

Вероятное значение квадрата ошибки

$$m^2 = \frac{1}{2h_0},$$

откуда

$$\sum y'^2 = \frac{n-p}{2h_0},$$

следовательно,

$$h_0 = \frac{n-p}{2\sum y'^2}$$

и совпадает с предыдущим.

Не следует придавать большое значение тому, что вместо n рассматривалось $n-p$, т. к. n очень велико и $\dfrac{n-p}{n}$ близко к 1.

Итак, это правило обосновано, если число наблюдений очень велико.

Глава 15
Теория интерполяции

205. Я хочу применить метод наименьших квадратов к новой задаче: *поиску неизвестной функции $f(x)$.*

Мы измеряем некоторые значения этой функции.

$$\begin{gathered} f(a_1) = A_1, \\ f(a_2) = A_2, \\ \ldots\ldots \\ f(a_n) = A_n. \end{gathered}$$

Строим кривую

$$y = f(x);$$

нам известны несколько ее точек.

Через эти точки M_1, M_2, … , M_n всегда можно провести некоторую кривую, но такое решение не было бы наилучшим: нам надо провести кривую *близко* к этим точкам, но при этом как можно более *гладко.*

Другой способ, так же как и геометрический, имеет определенную степень произвола: я хочу, чтобы моя кривая была как можно меньшей степени q,

$$f(x) = C_0 + C_1 x + \ldots + C_q x^q;$$

где q меньше, чем $n-1$, т. к. если бы q было в точности равно $n-1$, то функция бы полностью удовлетворяла всем условиям задачи.

Какое значение приписать q? Это значение произвольно. Сначала его выбирают довольно малым, потом, если этого недостаточно, вводят еще одно слагаемое в правую часть и так далее.

206. Оставим в стороне этот метод проб и ошибок и считаем q выбранным.

Мы определим коэффициенты многочлена так, чтобы

$$\sum [f(a_i) - A_i]^2$$

была минимальна. Функция $f(x)$ линейна по коэффициентам C. Я положу

$$F(x) = (x - a_1)(x - a_2)\dots(x - a_n).$$

Наша задача связана с разложением в непрерывную дробь отношения

$$\frac{F'(x)}{F(x)}.$$

Это разложение производится так, словно мы ищем наибольший общий делитель F и F'. Получаем последовательно

$$\begin{aligned}
F &= Q_1 F' + R_1,\\
F' &= Q_2 R_1 + R_2,\\
R_1 &= Q_3 R_2 + R_3,\\
&\dots\dots\dots\\
R_{n-3} &= Q_{n-1} R_{n-2} + R_{n-1},\\
R_{n-2} &= Q_n R_{n-1}.
\end{aligned}$$

В последнем уравнении нет члена R_n, т. к. $R_n = 0$.

В общем случае F имеет степень n, F' — степень $n-1$, R_1 — степень $n-2$, R_p — степень $n-p-1$, R_{n-1} — степень 0, и все Q — степень 1. Имеем

$$\begin{aligned}
\frac{F'}{F} &= \frac{1}{Q_1 + \dfrac{R_1}{F'}},\\
\frac{R_1}{F'} &= \frac{1}{Q_2 + \dfrac{R_2}{R_1}},\\
&\dots\dots\dots\dots\dots\\
\frac{R_{n-2}}{R_{n-3}} &= \frac{1}{Q_{n-1} + \dfrac{R_{n-1}}{R_{n-2}}},\\
\frac{R_{n-1}}{R_{n-2}} &= \frac{1}{Q_n}.
\end{aligned}$$

Итак,

$$\frac{F'}{F}=\cfrac{1}{Q_1+\cfrac{1}{Q_2+\cfrac{1}{Q_3+\ddots+\cfrac{1}{Q_{n-1}+\cfrac{1}{Q_n}}}}}.$$

207. Мы должны рассмотреть последовательные остатки, возникающие в этом разложении. Так как

$$F-Q_1F'=R_1,$$

вместо

$$F'-Q_2R_1=R_2$$

я могу написать

$$-Q_2F+F'(1+Q_1Q_2)=R_2.$$

Если я полагаю

$$N_1=1,\qquad D_1=Q_1,$$

я получу

$$N_1F-D_1F'=R_1.$$

Если я полагаю

$$N_2=-Q_2,\qquad D_2=-(1+Q_1Q_2),$$

я получу

$$N_2F-D_2F'=R_2.$$

Теперь выразим точно так же произвольный остаток:

$$N_iF-D_iF'R_2,$$
$$N_{i+1}F-D_{i+1}F'=R_{i+1}.$$

Так как

$$R_i=Q_{i+2}R_{i+1}+R_{i+2},$$
$$R_{i+2}=(N_iF-D_iF')-Q_{i+2}(N_{i+1}F-D_{i+1}F'),$$

если я положу

$$N_{i+2} = N_i - Q_{i+2}N_{i+1},$$
$$D_{i+2} = D_i - Q_{i+2}D_{i+1},$$

я получаю

$$R_{i+2} = N_{i+2}F - D_{i+2}F'.$$

Из этих рекуррентных соотношений устанавливаем, что

$$R_1,\ R_2,\ \ldots \text{ и } R_{n-2}$$

имеют степени $n-2$, $n-3$, ... и 1 соответственно, а R_{n-1} равен константе. Многочлен N_1 имеет степень 0, N_2 — степень 1, ... , N_i — степень $i-1$. Легко видеть, что если это предположение верно для N_i и N_{i+1}, то оно верно и для N_{i+2}. Многочлен D_i имеет степень i.

208. Теперь я заявляю, что N_i и D_i — это числитель и знаменатель дроби порядка i. Рекуррентные соотношения делают этот факт очевидным; но заметить его можно и иначе.

Я пишу последовательно:

$$F = Q_1F' + R_1,$$
$$F' = Q_2R_1 + R_2,$$
$$\ldots\ldots\ldots\ldots\ldots$$
$$R_{i-2} = Q_iR_{i-1} + R_i,$$

откуда я вывожу

$$\frac{F'}{F} = \cfrac{1}{Q_1 + \cfrac{1}{Q_2 + \ddots + \cfrac{1}{Q_i + \cfrac{R_i}{R_{i-1}}}}}.$$

Еще я могу записать следующее уравнение:

$$N_iF - D_iF' = R_i.$$

Предположим, нам хочется вычислить дробь

$$\frac{\alpha_i}{\beta_i} = \cfrac{1}{Q_1 + \cfrac{1}{Q_2 + \ddots + \cfrac{1}{Q_i}}}.$$

Мне надо только сделать $R_i = 0$ в предыдущем равенстве, а также положить

$$F = \beta_i, \qquad F' = \alpha_i,$$
$$\beta_i = Q_1\alpha_i + R_1',$$
$$\alpha_i = Q_2 R_1' + R_2',$$

все R превратились в R'. Так же вводим R' в рекуррентные соотношения

$$R_1' = Q_3 R_2' + R_3',$$
$$\dots\dots\dots\dots\dots$$
$$R_{1-2}' = Q_i R_{i-1}';$$

здесь R_i равно нулю, N_i и D_i те же, что и выше,

$$N_i\beta_i - D_i\alpha_i = 0,$$
$$\frac{\alpha_i}{\beta_i} = \frac{N_i}{D_i}.$$

Итак, $\frac{N_i}{D_i}$ — это i-я дробь. Какое отношение имеет эта дробь к предложенной задаче?

209. Фундаментальное уравнение имеет вид:

$$N_i F - D_i F' = R_i.$$

Обратим внимание на степень всех многочленов. Вспоминаем, что N_i имеет степень $i-1$; F — степень n; D_i — степень i; F' — степень $n-1$; R_i — степень $n-1-i$.

Из фундаментального уравнения я вывожу, что

$$\frac{D_i F'}{F} = N_i - \frac{R_i}{F}.$$

Я умножаю обе части на x^μ:

$$\frac{x^\mu D_i F'}{F} = N_i x^\mu - \frac{x^\mu R_i}{F}.$$

Я собираюсь оценивать сумму вычетов справа и слева; замечаем сначала, что для этих вычислений нам надо принимать в расчет только две дроби, где F входит в знаменатель.

Затем предположим, что в некоторой дробно-рациональной функции степень числителя на единицу меньше, чем у знаменателя, например,

$$\frac{A''x^{n-1} + B''x^{n-2} + \ldots}{A'x^n + B'x^{n-1} + \ldots};$$

тогда эта дробь раскладывается в

$$\sum \frac{A}{x-a} + \sum \frac{B}{(x-a)^\lambda} + \ldots$$

Если я умножу все на x, то при неограниченно растущем x сумма

$$\sum \frac{A}{x-a}$$

будет стремиться к $\sum A$, или сумме вычетов, а другие $\sum$ (то есть те, где $x - a$ входит в знаменатель в степени большей, чем 1) обнуляются. Но

$$\lim \frac{xP}{Q} = 0 \quad \text{при} \quad x = \infty,$$

если степень xP будет меньше, чем степень Q. Итак, сумма вычетов будет нулевой, если степень числителя меньше степени знаменателя больше, чем на единицу. Рассмотрим

$$\frac{x^\mu R_i}{F};$$

знаменатель имеет степень n, числитель — степень $n - 1 - i + \mu$. Если

$$n - 1 - i + \mu < n - 1,$$

то есть

$$\mu < i,$$

то сумма вычетов равна нулю. Таким образом, при μ, равном $0, 1, \ldots, i-1$ сумма вычетов равна нулю.

210. Когда есть рациональная дробь $\frac{P}{Q}$, у которой Q не имеет кратных корней и обращается в нуль при $x = a$, вычет при $x = a$ равен

$$\frac{P(a)}{Q'(a)}.$$

Возьмем в качестве a одно из значений, по которым мы вычисляем $f(x)$; вычет выражения

$$\frac{x^\mu D_i F'}{F}$$

в точке a будет равен

$$\frac{a^\mu D_i(a) F'(a)}{F'(a)}$$

или

$$a^\mu D_i(a).$$

Так, если только μ меньше i, то

$$\sum a^\mu D_i(a) = 0,$$

суммирование ведется по всем значениям a,

$$a_1,\ a_2,\ \ldots,\ a_n.$$

Отсюда следует, что если P — какой-то многочлен порядка $i - 1$, то

$$\sum P_{i-1}(a) D_i(a) = 0.$$

Возьмем

$$P_{i-1} = D_k;$$

тогда

$$\sum D_k(a) D_i(a) = 0.$$

Это уравнение выполняется всегда, как только k меньше i.

Но так как эти два индекса ничем не отличаются, данное уравнение будет также выполняться всякий раз, когда k *отлично* от i.

Аналогичное свойство имеют многочлены Лежандра[19]; для любых двух из них, ν_m, ν_n, имеем

$$\int\limits_{-1}^{+1} \nu_m \nu_n \, d\tau = 0.$$

211. Мы хотим получить многочлен порядка q, меньшего, чем $n-1$, чьи коэффициенты были бы выбраны так, чтобы

$$\sum [A_i - f(a_i)]^2$$

была минимальна.

Если $f(x)$ — многочлен порядка q, его всегда можно записать в виде

$$f(x) = C_0 + C_1 D_1(x) + C_2 D_2(x) + \ldots + C_q D_q(x).$$

Надо определить коэффициенты C так, чтобы достигала минимума сумма квадратов

$$\sum (A - C_0 - C_1 D_1 - C_2 D_2 - \ldots - C_q D_q)^2.$$

212. Раскладываем эту сумму.

В первой строке мы запишем квадраты слагаемых:

$$\sum A^2 + nC_0^2 + C_1^2 \sum D_1^2 + C_2^2 \sum D_2^2 + \ldots + C_q^2 \sum D_q^2.$$

Во второй строке запишем сумму удвоенных произведений:

$$-2C_0 \sum A - 2C_1 \sum AD_1 - \ldots - 2C_n \sum AD_q,$$

где

$$\sum AD_1 = A_1 D_q(a_1) + \ldots + A_n D_q(a_n).$$

В третьей строке запишем сумму следующих членов

$$2C_0 C_i \sum D_i + 2C_i C_k \sum D_i D_k.$$

Но

$$\sum D_i(a) = 0,$$
$$\sum D_i D_k = 0.$$

Остаются первая и вторая строки. Кроме того, я могу сократить запись, полагая

$$1 = D_0(x).$$

Сумма, стремящаяся к минимуму, превращается в

$$\sum A^2 + \sum C_i^2 \sum D_i^2 - 2 \sum C_i \sum AD_i.$$

Я дифференцирую по C_i и делю на 2; приравняв к нулю производную по C_i, получаю

$$C_i \sum D_i^2 = \sum AD_i,$$

откуда находим следующее выражение для C_i:

$$C_i = \frac{\sum AD_i}{\sum D_i^2},$$

то есть

$$C_i = \frac{A_1 D_i(a_1) + A_2 D_i(a_2) + \ldots + A_n D_i(a_n)}{D_i^2(a_1) + D_i^2(a_2) + \ldots + D_i^2(a_n)}.$$

Аналогия с другой задачей из анализа очевидна.

Когда функция $f(x)$ хотят разложить в ряд по многочленам Лежандра, приходят к сумме

$$f(x) = \sum C_i X_i,$$

где

$$C_i = \frac{\int\limits_{-1}^{+1} f(x) X_i \, dx}{\int\limits_{-1}^{+1} X_i^2 \, dx}.$$

У нас соотношение того же рода, только вместо интегралов появляются суммы.

213. В чем преимущество многочленов D?

Допустим, кто-то попытался сначала представить результаты наблюдений с помощью многочлена степени q и нашел константы $C_0, C_1, \ldots, C_q$. Затем оказывается, что сумма квадратов допущенных ошибок недопустимая: тогда приходится продолжать решение с многочленом степени $q+1$. При обычных способах решения все надо начинать сначала; здесь, наоборот, надо только добавить член $C_{q+1} D_{q+1}(x)$: предыдущие коэффициенты $C_0, C_1, \ldots, C_q$ не меняются, как можно видеть из выражения C_i.

214. Подобная задача возникает и тогда, когда просто хотят *интерполировать* функцию: почему в качестве решения обычно берется многочлен $(n-1)$-го порядка?

Пусть функция $f(x)$ голоморфна внутренности какого-то контура, например, кругу. Значения, которые мы придаем переменным, очень малы по сравнению с радиусом круга. Когда переменная равна a_1, a_2, ... , a_n, значение функции остается известным. Надо узнать значение функции для какой-то величины x внутри круга. Интеграл

$$\int \frac{f(z)\,dz}{(z-x)(z-a_1)(z-a_2)\dots(z-a_n)},$$

взятый вдоль окружности, обращается в нуль. С другой стороны, интеграл

$$J = \frac{1}{2i\pi}\int \frac{f(z)\,dz}{(z-x)(z-a_1)(z-a_2)\dots(z-a_n)}$$

равен сумме вычетов функции $f(z)$ в полюсах x, a_1, a_2, ... , a_n, т. е.

$$\frac{f(x)}{(x-a_1)(x-a_2)\dots(x-a_n)} + \frac{f(a_1)}{(a_1-x)(a_1-a_2)\dots(a_1-a_n)} + \dots$$

Вообще, я назову p_i многочлен, который получается, если убрать из

$$(z-x)(z-a_1)(z-a_2)\dots(z-a_n)$$

множитель $z-a_i$ и заменить z на a_i. Я полагаю также

$$F(x) = (x-a_1)(x-a_2)\dots(x-a_n).$$

Тогда

$$J = \frac{f(x)}{F(x)} + \frac{f(a_1)}{p_1} + \frac{f(a_2)}{p_2} + \dots + \frac{f(a_n)}{p_n},$$

откуда

$$f(x) = -\sum \frac{f(a_i)F(x)}{p_i} + JF(x).$$

215. Такова общая формула при интерполяции; справа участвуют
1° многочлен;
2° допущенная ошибка.
Эта ошибка равна

$$JF = \frac{1}{2i\pi} \int \frac{f(x)dz}{z-x} \frac{x-a_1}{z-a_1} \frac{x-a_2}{z-a_2} \cdots \frac{x-a_n}{z-a_n}.$$

Если R обозначает радиус сходимости, а x настолько близко к a_1, что

$$|x-a_i| < \frac{R}{2},$$

то каждый из множителей $\frac{x-a_i}{z-a_i}$ меньше, чем $\frac{1}{2}$, и под знаком $\int$ стоят n таких множителей.

Итак, при очень больших n значение JF очень мало.

Если бы радиус сходимости принимал какое-то определенное значение только с некоторой вероятностью, мы бы имели вероятностную задачу.

216. Я предполагаю, что a priori известно, что функция $f(x)$ раскладывается в ряд в какой-то определенной области по возрастанию степеней x:

$$f(x) = A_0 + A_1 x + \ldots$$

Мы ничего не знаем о коэффициентах A, кроме того, что вероятность, с которой один из них, A_i, находится в определенных пределах, y и $y+dy$, равна

$$\sqrt{\frac{h_i}{\pi}} e^{-h_i y^2} dy.$$

После n наблюдений мы знаем, что

$$f(a_1) = B_1,$$
$$f(a_2) = B_2,$$
$$\ldots\ldots$$
$$f(a_n) = B_n.$$

Мы ищем вероятное значение $f(x)$ при другом x.

Это тоже интерполяционная задача, с той разницей, что здесь ищется предельный многочлен.

Я спешу добавить, что рассматриваю этот вопрос как простое вычислительное упражнение, т. к. я *произвольно* ввел закон Гаусса; иначе задача осталась бы неопределенной.

217. Мы должны найти бесконечное число коэффициентов A нашей функции с помощью n наблюдений; здесь неизвестных больше, чем наблюдений, и мы можем руководствоваться лишь мыслью, что вероятностный закон задан a priori.

Так мы подходим к еще большим обобщениям, чем те, что мы делали до сих пор, потому что теперь нам надо определить неизвестную функцию.

Сначала я возьму только конечное число коэффициентов.

218. Вообще, пусть есть конечное число неизвестных,

$$u_1,\, u_2,\, \dots\, ,\, u_p;$$

где p известно.

Я предполагаю, что вероятность, с которой u_i лежит между u и $u + du$, представляется законом Гаусса:

$$\sqrt{\frac{h_i}{\pi}}e^{-h_i u^2}du.$$

Вероятность того, что одно из u уклоняется от нуля, будет тем меньше, чем больше h.

Мы знаем значения определенных функций от u,

$$x_1,\, x_2,\, \dots\, ,\, x_n,$$

причем наблюдения считаются совершенно точными.

Величина n меньше p, неизвестных больше, чем наблюдений.

Я предполагаю, что x — линейные функции от u; так же как в предыдущем примере все B были линейными функциями от A, и мы имели

$$B_k = A_0 + A_1 a_k + \dots$$

Тогда я полагаю

$$y{=}C_k^1 u_1 + C_k^2 u_2 + \dots + C_k^p u_p.$$

Наблюдения сообщили нам, что y_k лежат между x_k и $x_k + dx_k$. Найти u означает решить задачу на условные вероятности. Причиной будет тот факт, что все u лежат в определенных пределах; наблюдаемые следствия — это то, что все x лежат в определенных пределах.

Здесь формула

$$\frac{p_i \bar{\omega}_i}{\sum p_i \bar{\omega}_i}$$

будет упрощена.

219. Если значения всех u определены, то значения линейных функций x_k тоже будут определены, и вероятность этих значений будет вероятностью достоверного события; в зависимости от того, попадают эти функции в пределы, заданные наблюдением, или нет, вероятность равна 1 или 0. Тогда формула апостериорной вероятности упрощается до

$$\frac{\bar{\omega}_i}{\sum \bar{\omega}_i},$$

если заменить p_i на вероятность следствия при условии, что действует данная причина.

Сумма $\sum \bar{\omega}_i$ состоит из всех вероятностей, относящихся к значениям u, согласующимся с наблюдениями; $\bar{\omega}_i$ — априорная вероятность того, что различные величины u лежат в определенных пределах. При этом k-е неизвестное лежит между u_k и $u_k + du_k$ с вероятностью

$$\sqrt{\frac{h_k}{\pi}} e^{-h_k u_k^2} du_k;$$

а $\bar{\omega}_i$ состоит из p аналогичных множителей:

$$\bar{\omega}_i = \sqrt{\frac{h_1 h_2 \ldots h_p}{\pi^p}} e^{-(h_1 u_1^2 + h_2 u_2^2 + \ldots + h_p u_p^2)} du_1\, du_2 \ldots du_p,$$

для краткости я буду писать

$$\bar{\omega}_i = \Pi\, du_1\, du_2 \ldots du_p,$$

откуда

$$\sum \bar{\omega}_i = \int \Pi\, du_1\, du_2 \ldots du_p.$$

Интегрировать надо по всем значениям u, согласующимся с наблюдениями, т. е. удовлетворяющим неравенствам

$$x_h < C'_k u_1 + C_k^2 u_2 + \ldots + C_k^p u_p < x_h + dx_h.$$

Искомой вероятностью будет

$$\frac{\bar{\omega}_i}{\sum \bar{\omega}_i} = \frac{\Pi\, du_1\, du_2 \ldots du_p}{\int \Pi\, du_1\, du_2 \ldots du_p},$$

когда все u удовлетворяют неравенствам, написанным выше. Иначе вероятность будет равна 0.

220. Если я ищу вероятное значение какой-то функции F от u, то

$$\overline{F} = \frac{\int F\Pi\, du_1\, du_2 \ldots du_p}{\int \Pi\, du_1\, du_2 \ldots du_p}.$$

Преобразуем эти два интеграла. Имеем

$$x_h < y_h < x_h + dx_h.$$

Я хочу взять в качестве неизвестных $y_1, y_2, \ldots, y_n$ и добавить $p-n$ совершенно произвольных линейных функций от u, $z_1, z_2, \ldots, z_{p-n}$. Выражение

$$\Pi\, du_1\, du_2 \ldots du_p$$

превратится в

$$\Pi\Delta\, dy_1\, dy_2 \ldots dy_p\, dz_1\, dz_2 \ldots dz_{p-n}.$$

Здесь Δ — функциональный определитель u по y и z; так как это линейные функции, Δ — константа. Получается

$$\overline{F} = \frac{\int F\Pi\, dy_1\, dy_2 \ldots dy_p\, dz_1\, dz_2 \ldots dz_{p-n}}{\int \Pi\, dy_1\, dy_2 \ldots dy_p\, dz_1\, dz_2 \ldots dz_{p-n}}.$$

Сначала мы должны интегрировать по y; например, y_1 будет меняться от x_1 до $x_1 + dx_1$, т. е. очень мало; функция под знаком $\int$ останется практически константой, и можно будет записать, что

$$\overline{F} = \frac{dx_1\, dx_2 \ldots dx_n \int F\Pi\, dz_1\, dz_2 \ldots dz_{p-n}}{dx_1\, dx_2 \ldots dx_n \int \Pi\, dz_1\, dz_2 \ldots dz_{p-n}}.$$

Дифференциалы от x исчезнут. И $F\Pi$, и Π — функции от z, и мы интегрируем по z от $-\infty$ до $+\infty$.

Это произведение постоянного множителя на экспоненту

$$e^{-(h_1u_1^2 + h_2u_2^2 + \ldots + h_pu_p^2)};$$

показатель здесь — многочлен P второго порядка, неоднородный по z.

221. Считаем F линейной функцией от z. Ищем вероятное значение F. Вероятные значения различных величин z получают, найдя те z_i, на которых достигается минимум показателя P; пусть z_i^0 — значение z_i, при котором P минимально. Тогда

$$P = P_2 + P_0,$$

где P_0 — константа, а P_2 — однородный многочлен второй степени от разностей $z_i - z_i^0$. Интеграл

$$\int (z_i - z_i^0) e^{-P} dz_1\, dz_2 \ldots dz_{p-n}$$

имеет подынтегральную функцию, нечетную по $z_i - z_i^0$; интегрируя от $-\infty$ до $+\infty$, получаем, что значение этого интеграла равно нулю.

Отсюда выводим, что если при замене z_i на z_i^0 функция F равна F_0, то

$$\int (F - F_0) e^{-P} dz_1\, dz_2 \ldots dz_{p-n},$$

взятый от $-\infty$ до $+\infty$, равен нулю. Так же

$$\int (F - F_0) \Pi\, dz_1\, dz_2 \ldots dz_{p-n} = 0;$$

а, следовательно,

$$\int F \Pi\, dz_1\, dz_2 \ldots dz_{p-n} = F_0 \int \Pi\, dz_1\, dz_2 \ldots dz_{p-n},$$

то есть

$$\overline{F} = F_0.$$

Таким образом, вероятное значение F получается при замене z_i на значения, доставляющие минимум многочлену P.

222. Применим эти принципы к задаче, поставленной в параграфе 216.

В случае с $f(x)$ наши неизвестные u равны A, и мы имеем

$$P = h_0 A_0^2 + h_1 A_1^2 + \ldots + h_i A_i^0 + \ldots$$

Надо найти минимум этого многочлена.

Надо сказать, что значения A произвольны, исключение составляют линейные соотношения, взятые из наблюдений

$$f(ah) = B_h.$$

Пишем, что P минимально. Приращение dP должно равняться нулю, когда A_i возрастают на dA_i:

$$dP = h_0 A_0\, dA_0 + \ldots + h_i A_i\, dA_i + \ldots = 0.$$

Приращения $dA_0, \ldots, dA_i, \ldots$ связаны соотношением

$$df(a_k) = 0.$$

Между тем $f(a_1)$, например, равно

$$f(a_1) = A_0 + A_1 a_1 + \ldots + A_i a_1^i + \ldots$$

Итак,

$$\begin{gathered} dA_0 + a_1\, dA_1 + \ldots + a_1^i\, dA_i + \ldots = 0, \\ dA_0 + a_2\, dA_1 + \ldots + a_2^i\, dA_i + \ldots = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ dA_0 + a_n\, dA_1 + \ldots + a_n^i\, dA_i + \ldots = 0. \end{gathered}$$

223. Чтобы P было минимально, первое уравнение,

$$dP = 0,$$

должно выполняться, какими бы ни были dA; это первое уравнение должно быть следствием других n соотношений между dA.

Пусть выбраны подходящие коэффициенты $\varepsilon_1, \varepsilon_2, \ldots \varepsilon_n$, на которые мы умножаем соответственно обе части каждого из этих n соотношений,

$$\begin{gathered} h_i A_i = \varepsilon_1 a_1^i + \varepsilon_2 a_2^i + \ldots + \varepsilon_n a_n^i, \\ A_i x^i = \varepsilon_1 \frac{(x a_1)^i}{h_i} + \varepsilon_2 \frac{(x a_2)^i}{h_i} + \ldots + \varepsilon_n \frac{(x a_n)^i}{h_i}. \end{gathered}$$

Если я полагаю

$$\varphi(x) = \frac{1}{h_0} + \frac{x}{h_1} + \frac{x^2}{h_2} + \ldots + \frac{x^i}{h_i} + \ldots,$$

то коэффициент при ε_1 в

$$f(x) = \sum A_i x^i$$

будет равен $\varphi(xa_1)$, и т. д. Окончательно получается

$$f(x) = \varepsilon_1 \varphi(xa_1) + \varepsilon_2 \varphi(xa_2) + \ldots + \varepsilon_n \varphi(xa_n).$$

Мы располагаем коэффициенты ε так, чтобы удовлетворять наблюдениям, откуда и возникали n уравнений с n неизвестными.

224. Вид $f(x)$ зависит от h. Чтобы ряд, представляющий $\varphi(x)$, был сходящимся, надо, чтобы коэффициенты h возрастали с достаточной скоростью.

Если имеем

$$|x| < \rho \quad \text{и} \quad |ah| < \rho,$$

то есть

$$|xah| < \rho^2,$$

то ряд будет сходящимся, когда

$$h_i > \rho^{2i}.$$

В общем можно сказать, что вероятность, с которой последние коэффициенты при h_i уклоняются от нуля, становится все меньше. Достаточно предположить, что начиная с n-го слагаемого все h бесконечны. Тогда в $\varphi(x)$ крайние члены с h_n, $h_{n+1} \ldots$ обнуляются, и $\varphi(x)$ будет многочленом порядка n.

Глава 16
Различные вопросы

225. Тасование карт. Во введении я занимался задачами, связанными с игроком, тасующим колоду карт. Почему, когда колода тасуется достаточно долго, мы допускаем, что все перестановки карт, т.е. все порядки, в которых могут располагаться карты, должны быть равновероятны? Этим вопросом мы займемся прямо сейчас.

Пусть q — количество карт; пусть S_i — какая-то перестановка, т.е. операция, переводящая на место α карту, которая до перестановки занимала место β; тогда α — определенная функция от β. Полное число всех возможных перестановок равно $q!$ Найдется какой-то порядок карт, который мы будем считать нормальным и обозначим S_0; мы будем называть S_i порядок, в котором располагаются карты, если, имев изначально нормальный порядок, они подвергались перестановке S_i. Таким образом, S_0 представляет собой нормальный порядок и, одновременно, *тождественную* перестановку, которая не изменяет порядок карт. Если это установлено, то две последовательные перестановки S_i и S_j будут эквивалентны единственной перестановке S_k, что выражается соотношением

$$S_i S_j = S_k. \tag{1}$$

Множество перестановок образует группу, когда произведение двух произвольных перестановок из этого множества вновь принадлежит этому множеству.

Итак, пусть

$$S_0, S_1, \ldots, S_r \text{ —}$$

различные перестановки из группы G и соответствующие порядки карт. Допустим, мы знаем, что порядок карт в нашей колоде принадлежит группе G, и различные перестановки из группы G имеют вероятности

$$p_0, p_1, \ldots, p_r$$

соответственно, такие, что

$$p_0 + p_1 + \ldots + p_r = 1.$$

Символически мы можем представить этот вероятностный закон с помощью гиперкомплексного числа. Известно, что можно задать гиперкомплексные числа в виде

$$X = x_0e_0 + x_1e_1 + \ldots + x_re_r,$$

где все x — обычные величины, а все e — базисные единицы. Операции над этими гиперкомплексными числами проводятся по обычным правилам с той разницей, что умножение, которое остается дистрибутивным и ассоциативным, может не быть коммутативным.

Вводя правила умножения, т.е. определяя произведение e_ie_j двух произвольных базисных единиц, получают систему гиперкомплексных чисел.

Мы определили гиперкомплексную систему, соответствующую нашей G; каждой перестановке S_i из этой группы мы сопоставим базисную единицу e_i; и если верно уравнение (1): $S_1S_j = S_k$, то мы договариваемся, что произведение e_ie_j равно e_k. Это правило допустимо, т.к. операция ассоциативна.

Итак, мы сможем символически представить рассмотренный вероятностный закон гиперкомплексным числом

$$P = p_0e_0 + p_1e_1 + \ldots + p_re_r.$$

226. У игрока, тасующего карты, есть какие-то привычки, так что при каждом движении он подвергает карты перестановке S_i с вероятностью p_i. Этот вероятностный закон, который нам, вообще говоря, неизвестен, символически представляется числом $P = \sum p_ie_i$. Если тасование начинается с нормального порядка, то вероятность того, что после перестановки получится порядок S_i, равна p_i, так что вероятностный закон для различных порядков по-прежнему представляется символически с помощью P. Если вместо нормального порядка S_0 мы начнем с какого-то порядка S_j, то вероятностный закон будет представляться как e_jP. Если до перестановки вероятностный закон представлялся комплексным числом Q, то после перестановки он станет QP. Итак, если мы начинаем с нормального порядка и делаем n перестановок, то в конце вероятностный закон будет представляться гиперкомплексным числом P^n.

Мы хотим показать, что если n очень велико, то практически будет выполняться равенство

$$P^n = \frac{1}{r+1}(e_0 + e_1 + \ldots + e_r),$$

то есть все возможные порядки будут равновероятны. И этот результат не зависит от P, т.е. от неизвестного закона, неизвестных привычек игрока.

227. Г. Картан ввел в теорию гиперкомплексных чисел понятие характеристического уравнения. Пусть A — данное гиперкомплексное число, X — неизвестное гиперкомплексное число, ω — неизвестное *обычное* число; рассмотрим уравнение

$$AX = \omega X. \tag{2}$$

Обе его части — гиперкомплексные числа, и, приравнивая коэффициенты при $e_0, e_1, \ldots, e_r$, получаем $r+1$ уравнение для $r+1$ коэффициента x_i неизвестного гиперкомплексного числа X; с одной стороны, эти уравнения линейны по x_i, с другой стороны, по ω и $r+1$ коэффициенту a_i числа A. Итак, у нас $r+1$ линейное уравнение с $r+1$ неизвестным x_i. Пишем, что определитель Δ этих уравнений равен нулю; получим алгебраическое уравнение $(r+1)$-го порядка, которое определяет ω. По теореме о линейных подстановках каждому простому корню этого уравнения $\Delta = 0$ соответствует одно гиперкомплексное число X, удовлетворяющее (2). Двойному корню соответствуют два гиперкомплексных числа X и X_1 такие, что

$$AX = \omega X, \qquad AX_1 = \omega X_1 + \varepsilon_1 X. \tag{2'}$$

Тройному корню — три числа X_1, X_2, X_3 такие, что

$$AX = \omega X, \quad AX_1 = \omega X_1 + \varepsilon_1 X, \quad AX_2 = \omega X_2 + \varepsilon_2 X_1 \tag{2''}$$

и так далее; все ε — обычные постоянные числа, которые можно считать равными 0 или 1. Замечаем, что если $\varepsilon_1 = \varepsilon_2 = 0$, то

$$A(\lambda X + \lambda_1 X_1 + \lambda_2 X_2) = \omega(\lambda X + \lambda_1 X_1 + \lambda_2 X_2),$$

каковы бы ни были константы λ, λ_1, λ_2 (можно условиться, что это обычные числа).

Если гиперкомплексные числа принадлежат некоторой группе, задача упрощается. Пусть X — какое-то гиперкомплексное число и

$$e_i X = Y = \sum y_j e_j;$$

очевидно, что все y — не что иное как x, расположенные в другом порядке. То же самое происходит, если взять $Xe_i = Y$; кроме того

$$e_0 X = X e_0 = X,$$

так что уравнение (2) можно записать как

$$(A - \omega e_0)X = 0.$$

228. Я составляю характеристическое уравнение для числа P:

$$PX = \omega X$$

и собираюсь показать сначала, что у него есть один корень, равный 1, а все остальные меньше 1 по абсолютной величине.

Действительно, пусть

$$PX = \sum y_i e_i;$$

тогда

$$y_i = \sum p_k \cdot x_h,$$

индексы i, k и h связаны соотношением $e_k e_h = e_i$, так что если я пишу

$$y_i = \sum p_{hi} x_h,$$

то все p_{hi} — не что иное, как p_h, взятые в другом порядке. Тогда уравнение (2) даст нам

$$\sum p_{hi} x_h = \omega x_i. \tag{3}$$

Эти уравнения будут выполняться, если взять все x_i равными между собой, откуда мы выводим, что

$$\sum p_{hi} = \omega.$$

Но

$$\sum p_{hi} = \sum p_h = 1,$$

откуда $\omega = 1$; и это показывает, что есть один корень, равный 1.

В том, что касается других корней, уравнения (3) дадут нам

$$\sum |p_{hi} x_h| \geqslant |\omega x_i| \tag{4}$$

или, т. к. все p действительны и положительны,

$$\sum p_{hi} |x_h| \geqslant |\omega| |x_i|.$$

Сложим все эти неравенства; получится

$$\sum_i \sum_h p_{hi} |x_h| \geqslant |\omega| \sum |x_i|.$$

Но

$$\sum_i p_{hi} = 1.$$

Значит,

$$\sum_i \sum_h p_{hi} |x_h| = \sum |x_h|,$$

откуда

$$\sum |x_h| \geqslant |\omega| \sum |x_i|,$$

откуда

$$|\omega| \leqslant 1.$$

Итак, ни один корень не может быть больше 1 по абсолютной величине. Что и требовалось доказать.

229. Пусть ω — корень, отличный от 1; рассмотрим гиперкомплексное число, *принадлежащее* этому корню ω, т. е., по терминологии, принятой г. Картаном, одно из чисел X, X_1, …, таких, что

$$PX = \omega X, \qquad PX_1 = \omega X_1 + \varepsilon_1 X_1, \ \ldots \ ;$$

я заявляю, что для всех этих гиперкомплексных чисел

$$\sum x_i = 0.$$

Действительно, заменяем во всех наших гиперкомплексных числах все базисные единицы e_i на обыкновенную единицу; равенства, возможные между этими гиперкомплексными числами, останутся верны. Если мы имеем

$$P = \sum p_i e_i, \qquad X = \sum x_i e_i, \qquad X_1 = \sum x_i^1 e_i,$$

то после подстановки эти гиперкомплексные числа превратятся в

$$\sum p_i = 1, \qquad \sum x_i, \qquad \sum x_i^1$$

соответственно, и наши равенства примут вид

$$\sum x_i = \omega \sum x_i, \qquad \sum x_i^1 = \omega \sum x_i^1 + \varepsilon_1 \sum x_i,$$

следовательно, если ω не равно 1, то

$$\sum x_i = 0, \qquad \sum x_i^1 = 0.$$

230. Мы сказали, что имеется один корень, равный 1; остается узнать, может ли здесь быть несколько корней с модулем 1, и, кроме того, будет ли 1 кратным корнем. Для того чтобы неравенство (4) превратилось в равенство, надо, чтобы все x_i имели одинаковый модуль, и, т. к. эти x_i определены только отношениями друг к другу, мы можем предполагать, что все они действительны и положительны. Поскольку все p действительны и положительны, то же будет верно для ω, т. е. мы получим

$$\omega = 1.$$

Пусть тогда x_j — наибольшее из всех x_i; имеем

$$\sum p_{hj} x_h \leqslant \sum p_{hj} x_j = x_j,$$

это неравенство превращается в равенство, только когда все p_{hj}, кроме тех, что умножаются на x_h, равные x_j, совпадают с нулем. Но уравнение (3) дает нам

$$\sum p_{hj} x_h = \omega x_j = x_j.$$

Отсюда мы должны сделать вывод, что

$$p_{hj} = 0, \quad \text{если} \quad x_h < x_j,$$

и это показывает нам, что если ни одно из p_i (и, следовательно, p_{hj}) не равно нулю, то все x_i равны между собой.

Будем говорить, что перестановка S_i принадлежит классу C, если x_i равно x_j, т.е. наибольшему из всех x.

С другой стороны, я буду говорить, что подстановка S_k принадлежит множеству E, если всегда, когда S_j принадлежит классу C, то же будет верно для

$$S_h = S_k^{-1} S_j.$$

Необходимым условием того, что $p_k = p_{hj}$ может не равняться нулю, будет принадлежность S_k множеству E.

Теперь я заявляю, что множество E образует подгруппу в G. Действительно, если S_k и S_c принадлежат этому множеству, S_j принадлежит C, только когда то же верно для $S_k^{-1} \cdot S_j$ и $S_c^{-1} \cdot S_j$. Но тогда то же самое должно выполняться для

$$S_k^{-1}(S_c^{-1} S_j) = (S_c S_k)^{-1} S_j,$$

что означает, что $S_c S_k$ также принадлежат C.

Итак, интересующее нас событие может произойти, только когда все p_k — нули, кроме тех, которые соответствуют перестановкам из некоторой подгруппы; т.е. если бы заранее было известно, что игрок, тасующий карты, производит только перестановки, принадлежащие к этой подгруппе.

Если оставить в стороне этот единственный исключительный случай, уравнение

$$PX = X$$

может выполняться, только когда все x равны между собой.

231. Может существовать еще одна возможность; можно было бы предположить, что для некоторого гиперкомплексного числа X такого, что

$$\begin{gathered} PX_1 = X_1 + \varepsilon_1 X_1, \qquad \varepsilon_1 \neq 0, \\ X = e_0 + e_1 + \ldots + e_r \end{gathered} \tag{5}$$

и корень 1 — кратный. Но из уравнения (1) следует, что

$$P^n X_1 = X_1 + n\varepsilon_1 X.$$

Коэффициенты в записи X_1 и $\varepsilon_1 X$ не зависят от n; коэффициенты P^n зависят от n, но они остаются действительными и положительными, и их сумма все время равна 1, потому что P^n символически представляет вероятностный закон после n перестановок.

Итак, коэффициенты $P^n X$ остаются ограниченными. Наоборот, коэффициенты $X_1+n\varepsilon_1 X$ будут многочленами первой степени от n; таким образом, они не могут быть ограниченными и уравнение (5) невозможно. Если мы возвращаемся к уравнению в виде

$$PX = \omega X,$$

то мы уже знаем, что ω задается уравнением $(n+1)$-го порядка, и корню кратности μ *принадлежат* μ различных гиперкомплексных чисел. Сумма всех кратностей равна $r+1$, и найдется $r+1$ гиперкомплексных чисел, принадлежащих разным корням, причем они будут линейно независимы. Итак, произвольное гиперкомплексное число можно рассматривать как линейную комбинацию чисел, принадлежащих различным корням.

В нашем случае только одно число принадлежит корню 1, это

$$e_0 + e_1 + \ldots + e_r;$$

все остальные принадлежат корням меньшим 1 по абсолютной величине, и сумма коэффициентов каждого из этих гиперкомплексных чисел равна нулю. Отсюда следует, что *любое гиперкомплексное число, у которого сумма коэффициентов равна нулю, можно рассматривать как линейную комбинацию гиперкомплексных чисел, принадлежащих корням, меньшим* 1 *по абсолютной величине.*

232. Теперь рассмотрим корень ω такой, что $|\omega| < 1$; если этот корень кратный, мы имеем

$$PX = \omega X, \qquad PX_1 = \omega X_1 + \varepsilon_1 X, \qquad PX_2 = \omega X_2 + \varepsilon_2 X_1, \ \ldots,$$

откуда легко выводим, что

$$\begin{aligned}
P^n X &= \omega^n X, \\
P^n X_1 &= \omega^n X_1 + n\omega^{n-1}\varepsilon_1 X, \\
P^n X_2 &= \omega^n X_2 + n\omega^{n-1}\varepsilon_2 X_1 + \frac{n(n-1)}{2}\omega^{n-2}\varepsilon_1\varepsilon_2 X, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

Так как ω^n, $n\omega^{n-1}$, $\dfrac{n(n-1)}{2}\omega^{n-2}$, ... стремятся к нулю, когда n бесконечно растет, можно видеть, что для произвольного комплексного числа X, принадлежащего корню, меньшему 1 по абсолютной величине, получаем

$$\lim P^n X = 0 \qquad (n = \infty).$$

Но любое гиперкомплексное число, сумма коэффициентов которого равна нулю, является комбинацией чисел, принадлежащих корням, меньшим 1. Значит, кроме того, получаем

$$\lim P^n X = 0$$

всякий раз, когда сумма коэффициентов X равна нулю.

Если же, напротив, X принадлежит корню 1, т.е. если все его коэффициенты равны между собой, получаем

$$P^n X = X.$$

Если X — произвольное гиперкомплексное число, мы сможем положить

$$X = SX_0 + X',$$

где S — сумма коэффициентов X, а

$$X_0 = \frac{1}{r+1}(e_0 + e_1 + \ldots + e_r)$$

и сумма коэффициентов X' равна нулю. Тогда получим

$$\lim P^n X = SX_0.$$

Теперь заметим, что

$$X_0 X = X X_0 = SX_0.$$

Действительно, имеем

$$X = \sum x_i e_i, \qquad X_0 = \sum \frac{e_j}{r+1}, \qquad X_0 X = \sum \frac{e_k}{r+1} x_{kj},$$

где $x_{kj} = x_i$ если $e_j e_i = e_k$, или же

$$X_0 X = \sum \frac{e_k}{r+1} (\sum_j x_{kj}).$$

Но все x_{kj}, фигурирующие под знаком $\sum_j$, равны самим x_i, взятым в другом порядке, т. е.

$$\sum_j x_{kj} = \sum x_i = S,$$

откуда окончательно

$$X_0X = SX_0$$

и

$$\lim(P^n - X_0)X = 0.$$

Но X — произвольное гиперкомплексное число, тогда мы можем считать $X = e_0$, откуда

$$(P^n - X_0)X = (P^n - X_0)e_0 = P^n - X_0.$$

Итак, остается

$$\lim P^n = X_0,$$

что означает, что в пределе все вероятности, т. е. все коэффициенты гиперкомплексного числа P^n, символически представляющего вероятностный закон, будут равны. Именно это мы хотели доказать.

Я сошлюсь на некоторые работы, связанные с гиперкомплексными числами и их отношением к группам. В первую очередь я называю работы г. Фробениуса, (Frobenius, Sitzungsberichte Берлинской Академии с 1896 по 1901 годы), а затем — статью г. Картана (Cartan, Sur les groupes bilinéaires et les systémes de nombres complexes, Annales de la Faculte de Toulouse, t. XII). Я сам занимался этим вопросом и, в частности, постарался приблизиться к настоящим результатам этих двух выдающихся ученых в совершенно другом виде в работе «Sur l'intégration algébrique des équations linéaires» (Journal de Liouville, 5^e série, t. IX).

233. Распределение десятичных знаков в числовой таблице. Предположим, что в числовой таблице берется большое число последовательных логарифмов и рассматривается, например, третий по счету десятичный знак. Видно, что десять цифр 0, 1, 2, 3, ... , 9 будут в этом списке распределены одинаково, а следовательно, вероятность того, что третий десятичный знак четен, равна $\frac{1}{2}$. Так заставляет думать непреодолимый инстинкт, а кроме того, это проверяется a posteriori. Можно ли объяснить этот факт?

Рассмотрим числа

$$\log\left(1+\frac{x}{100\,000}\right),$$

где x принимает все целые значения от 1 до 100 000. Рассмотрим функцию

$$F\left[\log\left(1+\frac{x}{100\,000}\right)\right],$$

где $F(y)$ — функция, равная $+1$, если третий десятичный знак y четен, и -1, если он нечетен. Я собираюсь доказать, что среднее значение $F(y)$ равно нулю или очень мало.

Определив $F(y)$, имеем

$$F\left(y+\frac{1}{500}\right)=F(y),$$

что показывает, что $F(y)$ — функция периодическая с периодом $\frac{1}{500}$, как и функция

$$\sin(1000\pi y),$$

которую мы сначала рассмотрим. Итак, мы должны оценить сумму

$$S=\frac{1}{10\,000}\sum_{x=1}^{10\,000}\sin\left[1000\pi\log\left(1+\frac{x}{100\,000}\right)\right].$$

Заменим ее интегралом

$$J=\frac{1}{10\,000}\int\limits_{1/2}^{10\,000+1/2}\sin\left[1000\pi\log\left(1+\frac{x}{100\,000}\right)\right]dx.$$

Сначала оценим разность $S-J$; воспользуемся формулой Тейлора

$$\varphi(n+h)=\varphi(n)+h\varphi'(n)+\frac{h^2}{1\cdot 2}\varphi''(n+\theta h)\quad(0<\theta<1),$$

откуда

$$\int\limits_{n-\frac{1}{2}}^{n+\frac{1}{2}}\varphi(x)\,dx=\varphi(n)+\frac{1}{12}\theta' M,$$

где θ' лежит между -1 и $+1$, а M — максимум $|\varphi''(x)|$ на рассмотренном интервале. Если мы берем

$$\varphi(x) = \sin\left[1000\pi \log\left(1 + \frac{x}{100\,000}\right)\right],$$

считаем $n = 1, 2, \ldots, 10\,000$ и складываем, то получается

$$J - S = \frac{1}{120\,000}\sum \theta' M.$$

Что такое M? Обозначим для краткости

$$1000\pi \log\left(1 + \frac{x}{100\,000}\right) = y; \qquad \frac{x}{100\,000} = z;$$

тогда

$$\varphi'(x) = \frac{1000\pi}{100\,000}\frac{1}{1+z}\cos y,$$

$$\varphi''(x) = -\frac{(1000\pi)^2}{(100\,000)^2}\frac{1}{(1+z)^2}\sin y - \frac{1000\pi}{(100\,000)^2}\frac{1}{(1+z)^2}\cos y,$$

откуда

$$M < \frac{\pi^2}{(100)^2} + \frac{\pi}{(100)^2\cdot 1000} < \frac{1}{1000}$$

и

$$\sum \theta' M < \frac{10\,000}{1000} = 10,$$

$$|J - S| < \frac{1}{12\,000}.$$

Теперь оценим верхний предел J. Пусть

$$z = \frac{x}{100\,000}, \qquad \alpha = 1000\pi, \qquad u = \log(1+z),$$

тогда

$$J = 10\int \sin\alpha u\, dz = 10\int \sin\alpha u e^u\, du.$$

Интегрирование по частям дает мне

$$\frac{J}{10} = -\frac{\cos\alpha u e^u}{\alpha} + \int\limits_{u_0}^{u_1}\frac{\cos\alpha u}{\alpha}e^u\, du,$$

$$\left|\frac{J}{10}\right| = \frac{e^{u_0} + e^{u_1}}{\alpha} + \int\limits_{u_0}^{u_1}\frac{e^u\, du}{\alpha} < \frac{2e^{u_1}}{\alpha}.$$

Между тем

$$e^{u_0} = 1 + \frac{1}{200\,000}, \qquad e^{u_1} = 1 + \frac{10\,000}{100\,000} = 1{,}1.$$

Итак,

$$|J| < \frac{2{,}2}{100\pi} < \frac{1}{100},$$

или, сравнивая с пределом $|J - S|$,

$$|S| < \frac{1}{100}.$$

234. Доказательство было приведено для очень частного примера, но легко видеть его суть и, следовательно, применимость в общем случае.

В действительности все было основано на трех фактах:

1° последовательные производные логарифма остаются конечными на рассмотренном интервале;

2° число $\alpha = 1000\pi$ очень велико;

3° очень большое в абсолютном смысле, оно очень мало по сравнению с числом 100 000, которое стоит в знаменателе выражения

$$\log\left(1 + \frac{x}{100\,000}\right).$$

Видно, что те же самые обстоятельства повторяются в большом числе аналогичных случаев, и те же самые рассуждения применимы ко всем непрерывным функциям. Пусть в еще более общей задаче надо оценить сумму

$$S = \beta \sum F[\alpha \cdot \varphi(\beta x)],$$

где α — очень большое число, F — ограниченная периодическая функция, β — очень маленькое число, такое, что даже $\alpha\beta$ очень мало, а x принимает все целые значения от 1 до $\frac{1}{\beta}$. Сравним эту сумму с интегралом

$$J = \int F[\alpha \cdot \varphi(z)]\, dz.$$

Как и выше, получается

$$J - S = \frac{\beta}{12} \sum \theta' M,$$

где M — максимум второй производной функции $F[\alpha \cdot \varphi(z)]$ по x; но эта вторая производная равна

$$(\alpha\beta)^2 F''\varphi'^2 + \alpha\beta^2 F'\varphi'',$$

буквы со штрихами обозначают производные от F и φ по их аргументам $\alpha\varphi$ и βx соответственно. Видно, что если эти производные ограничены, $J - S$ имеет порядок $(\alpha\beta)^2$.

Пусть теперь $\Phi(u)$ — первообразная функция для $F(u)$, так что

$$F(u) = \Phi'(u);$$

интегрирование по частям даст нам

$$J = \int \Phi'(\alpha\varphi) \frac{dz}{d\varphi}\, d\varphi = \frac{\Phi}{\alpha} \frac{dz}{d\varphi} - \frac{1}{\alpha} \int \Phi \frac{d^2 z}{d\varphi^2}\, d\varphi,$$

что показывает, что J имеет порядок $\frac{1}{\alpha}$. Итак, для применимости нашего рассуждения достаточно, чтобы $\frac{1}{\alpha}$ и $(\alpha\beta)^2$ были очень малы.

235. Однако существуют еще две трудности. Не будет ли результат верен даже тогда, когда $\alpha\beta$ не очень мало, если малы β и $\frac{1}{\alpha}$? Пусть, вместо того чтобы рассуждать о третьем десятичном знаке в таблице с 5 знаками, мы рассуждали о пятом, и вместо того чтобы взять

$$\alpha = 1000\pi, \quad \beta = \frac{1}{100\,000},$$

нам пришлось взять

$$\alpha = 100\,000\pi, \quad \beta = \frac{1}{100\,000},$$

и $\alpha\beta$ уже не было очень мало. А между тем инстинкт, который побуждал верить, что третий десятичный знак должен быть распределен равномерно, так же силен и в том, что касается пятого знака.

В рассуждениях, связанных с оценкой J, не может быть сложностей даже для не очень малого $\alpha\beta$; надо заняться только разностью $J-S$. С другой стороны, периодическая функция F, которую мы определили в начале этого исследования и которая возникает, когда хочется узнать распределение десятичных знаков в таблице логарифмов, эта функция, которую мы заменили синусом, чтобы упростить вычисления, была равна ± 1 в зависимости от того, оказывалась ли определенная цифра четной или нет. Итак, это была *разрывная* функция, и при вычислении $J - S$ мы не можем опираться на тот факт, что ее производные ограничены.

Правда, можно было бы разложить эту периодическую функцию F в ряд Фурье и вернуться к синусам, к которым можно применить всё предшествующее; но в этом ряду для слагаемых высокого порядка α оказалось бы настолько велико, что произведение $\alpha\beta$ уже бы не было слишком мало; итак, мы бы вернулись к первой трудности.

236. Таким образом, приходится прибегнуть к другим соображениям. Пусть $F(x)$ — функция с ограниченными производными. Составим таблицу, где x примет все значения, кратные $\frac{1}{10\,000}$. Возможно ли, что в этой таблице пятый десятичный знак всегда будет равен 0? Тогда окажется, что

$$F(\frac{\nu}{10\,000}) = \frac{n}{10\,000} + \varepsilon,$$

где ν и n целые, а $\varepsilon < \frac{1}{100\,000}$. Воспользуемся известной формулой

$$F(x + 2h) + F(x) - 2F(x + h) = h^2 F''(x + 2\theta h) \qquad (0 < \theta < 1).$$

Придаем x значение, кратное $\frac{1}{10\,000}$, и делаем

$$h = \frac{1}{10\,000}.$$

Пусть n_1 и ε_1, n_2 и ε_2, n_3 и ε_3 — значения n и ε, соответствующие x, $x + h$, $x + 2h$; получаем

$$h^2 F'' = \frac{n_1 + n_3 - 2n_2}{10\,000} + \varepsilon_1 + \varepsilon_3 - 2\varepsilon_2.$$

Так как ε меньше, чем $\frac{1}{100\,000}$, мы сможем положить

$$\varepsilon_1 + \varepsilon_3 - 2\varepsilon_2 = \frac{4\theta}{100\,000} \qquad (\theta < 1),$$

$$n_1 + n_3 - 2n_2 = N,$$

откуда

$$F'' = \left(N + \frac{4\theta}{10}\right) 10\,000.$$

Если ограниченная функция F'' не имеет порядок $10\,000$, то должно выполняться $N = 0$, или же

$$n_1 + n_3 = 2n_2,$$

то есть целые n, или числа, фигурирующие в нашей таблице, представляются многочленом 1-й степени, или же все разности второго порядка в нашей таблице — нули. Это, очевидно, имеет место только в очень частном случае, и для большинства функций можно утверждать, что это не так.

Таким образом, хотя я больше не настаиваю на этой точке зрения, видно, на чем можно основать теорию вероятностей распределения десятичных знаков в числовой таблице.

237. Смесь жидкостей. Я скажу только несколько слов о другой проблеме очень большой важности; я не в состоянии разрешить ее. Рассмотрим жидкость, целиком заполняющую некоторый сосуд. Молекулы этой жидкости находятся в постоянном движении; законы этого движения известны, они выражаются дифференциальными уравнениями, которые я считаю заданными. Пусть x, y, z — координаты какой-то молекулы, компоненты ее скорости имеют вид

$$\frac{dx}{dt} = X, \qquad \frac{dy}{dt} = Y, \qquad \frac{dz}{dt} = Z.$$

Если движение постоянно, то X, Y, Z — функции от координат x, y, z, не зависящие от времени t, и я предполагаю, что эти функции известны. Так как жидкость несжимаема, имеем соотношение

$$\frac{dX}{dx} + \frac{dY}{dy} + \frac{dZ}{dz} = 0.$$

Если уравнение стенки сосуда —

$$\varphi(x, y, z) = 0,$$

то для всех точек этой стенки получаем соотношение

$$X\frac{d\varphi}{dx} + Y\frac{d\varphi}{dy} + Z\frac{d\varphi}{dz} = 0,$$

которое означает, что нормальная составляющая скорости равна нулю.

Установив это, я предполагаю, что некоторые молекулы жидкости отличаются от других каким-то явным свойством, например, имеют розовый цвет, тогда как остальные бесцветны; но при этом они подчиняются одному и тому же закону движения.

Пусть в момент времени $t = 0$ молекулы розового цвета каким-то образом распределены в сосуде; опыт показывает нам, что через некоторое время они полностью смешаются с остальными и будут распределены по сосуду равномерно.

Рассмотрим в данном сосуде внутренний объем v: какое количество розовой жидкости заключено в этом объеме в произвольный момент времени, или, если предпочесть другую формулировку, чему равна P — вероятность того, что одна случайно взятая в нем молекула будет розовой? Если распределение равномерно, то эта вероятность будет константой, каков бы ни был объем v, выбранный внутри сосуда.

Если рассмотреть два объема v_1 и v_2, для которых эти вероятности равны P_1 и P_2, и обозначить P вероятность, относящуюся к полному объему $v_1 + v_2$, мы, очевидно, получим

$$P(v_1 + v_2) = P_1 v_2 + P_2 v_2$$

и сможем написать, что для произвольного объема

$$P_v = \int p\, d\tau,$$

где интеграл берется по всем элементам $d\tau$, принадлежащим объему v, а p — вероятность, относящаяся к объему $d\tau$.

Эта вероятность p будет функцией от x, y, z, t, определенной с помощью уравнения

$$\frac{dp}{dt} + X\frac{dp}{dx} + Y\frac{dp}{dy} + Z\frac{dp}{dz} = 0.$$

Тогда рассмотрим два произвольных объема v и v' и соответствующие вероятности p и p'; можно ли считать, что каково бы ни было начальное распределение жидкости, т. е. значение p при $t = 0$, отношение $\frac{P'}{P}$ стремится к 1, когда t неограниченно возрастает, и это происходит тем быстрее, чем больше объемы v и v' и чем проще их вид?

Если мы не можем утверждать этого, то можем ли мы по меньшей мере считать, что отношение

$$\frac{\int\limits_0^T P\,dt}{\int\limits_0^T P'\,dt}$$

стремится к 1 при бесконечно большом T?

238. Таков возникающий здесь вопрос, и этот вопрос еще не решен; я хотел бы объяснить в нескольких словах, что делает его таким важным. Представим механическую систему в состоянии, определенном с помощью n координат $q_1, q_2, \ldots, q_n$; ее скорость удовлетворяет уравнениям Гамильтона,

$$\frac{dq_i}{dt} = \frac{dF}{dp_i}, \qquad \frac{dp_i}{dt} = -\frac{dF}{dq_i}.$$

Если T — кинетическая энергия, U — потенциальная, $F = T + U$ — полная, то имеем $p_i = \dfrac{dT}{dq'_i}$. Рассмотрим $q_1, q_2, \ldots, q_n, p_1, p_2, \ldots, p_n$ как координаты точки в $2n$-мерном пространстве. Запишем уравнения Гамильтона в виде

$$\frac{dq_i}{dt} = Q_i = \frac{dF}{dp_i}, \qquad \frac{dp_i}{dt} = P_i = -\frac{dF}{dq_i}.$$

Эти уравнения имеют тот же вид, что и уравнения движения нашей жидкости, т. к. они удовлетворяют соотношению

$$\sum \frac{dQ_i}{dq_i} + \sum \frac{dP_i}{dp_i} = 0,$$

аналогичному условию несжимаемости. Тогда мы должны изучать движение жидкости в $2n$-мерном сосуде. Какой бы ни была вероятность того или иного состояния системы в нулевой момент времени, не получим ли мы равномерной вероятности для ее состояния в момент t, если t достаточно велико?

Именно это утверждается в кинетической теории газа, в частности, когда хотят установить теорему Больцмана–Максвелла. И было бы очень полезно обосновать этот постулат.

239. В общем случае этот постулат должен быть верен, но он, разумеется, допускает и исключения. Если дифференциальные уравнения

$$\frac{dx}{x} = \frac{dy}{y} = \frac{dz}{z} = dt$$

имеют интеграл

$$F(x,\, y,\, z) = \mathrm{const},$$

где F симметричная однозначная функция, то интеграл

$$Pv = \int p\, d\tau,$$

взятый по объему, ограниченному двумя поверхностями

$$F(x,\, y,\, z) = a, \qquad F(x,\, y,\, z) = b,$$

будет константой; тогда, если мы рассмотрим два одинаковых объема v и v', определенных неравенствами

$$a < F < b, \qquad a' < F < b',$$

соответственно, и назовем P и P' соответствующие вероятности, отношение $\frac{P'}{P}$ будет константой, не зависящей от времени, и, т. к. начальное значение этой константы произвольно, она не может стремиться к 1.

Вот как в этом случае должен измениться наш постулат. Рассмотрим объем, ограниченный двумя бесконечно близкими поверхностями

$$a < F < a + da.$$

Пусть $V(a)\, da$ — этот объем, $P(a)\, da$ — соответствующий интеграл $\int p\, d\tau$.

Рассмотрим теперь два произвольных объема v' и v''. Пусть P' и P'' две соответствующие вероятности. Пусть объем $V'(a)\, da$ — пересечение $V(a)\, da$ и v', аналогично $V''(a)\, da$ — пересечение $V(a)\, da$ и v''; при $t = \infty$ получаем

$$\lim \frac{P'}{P''} = \frac{\int V'(a)\, da}{\int V''(a)\, da}.$$

Другими словами, в каждом из бесконечно тонких слоев, определенных неравенствами $a < F < a + da$, вероятность ***в конце концов*** будет распределена равномерно, но «плотность» этой конечной вероятности будет меняться от слоя к слою.

Именно это происходит, когда уравнения Гамильтона допускают первый интеграл $F = \text{const}$. Но это не единственное возможное исключение. Допустим, что у уравнений нет общего интеграла $F = 0$, т. е. из уравнения $F = 0$ следует, что

$$X\frac{dF}{dx} + Y\frac{dF}{dy} + Z\frac{dF}{dz} = 0.$$

Тогда замкнутая поверхность $F = 0$ разделит сосуд на две области; в каждой из них предельная вероятность распределяется равномерно, но ее плотность не будет одинаковой в этих двух областях.

240. Другое исключение. Чтобы лучше объяснить его, я начну с частного примера. Сначала я предположу, что есть интеграл $F = \text{const}$, и поверхность $F = 0$ — это тор. Если бы постулат был верен, даже с дополнением, рассмотренным в предыдущий раз, то предельная вероятность должна была бы распределиться равномерно в бесконечно тонком слое, лежащем между двумя поверхностями $F = 0$ и $F = \varepsilon$, где ε очень мало. Чтобы представить положение точки на торе $F = 0$, мы воспользуемся двумя углами; один из них, φ, будет долготой, другой, ω, будет отсчитываться от меридионального сечения, и его можно было бы назвать широтой, если бы он не менялся от $0°$ до $360°$, вместо того чтобы меняться от $-90°$ до $+90°$. Пусть тогда

$$\frac{d\varphi}{dt} = \Phi, \qquad \frac{d\omega}{dt} = \Omega$$

— дифференциальные уравнения движения, и предположим сначала, что Φ и Ω — константы. Мы можем считать, что поверхность $F\varepsilon$ выбрана так, чтобы эти уравнения были совместимы с условием несжимаемости (это сводится к предположению, что расстояние по нормали между двумя поверхностями $F = 0$ и $F = \varepsilon$ должно быть обратно расстоянию до оси вращения). Тогда мы получим

$$\varphi = \varphi_0 + \Phi t, \qquad \omega = \omega_0 + \Omega t,$$

где φ_0 и ω_0 — начальные значения φ и ω; при этих условиях вероятное значение выражения

$$\sin(n\varphi + n\omega + h),$$

где m и n — целые, а h — константа, будет равно

$$\int p_0 \sin(m\varphi + n\omega + h)\, d\varphi_0\, d\omega_0,$$

где p_0 — известная функция, не зависящая от выбора φ_0 и ω_0. Это дает

$$A \sin(m\Phi + n\Omega)t + B\cos(m\Phi + n\Omega)t,$$

где

$$\begin{matrix} A \\ B \end{matrix} = \int p_0 \begin{matrix} \cos \\ \sin \end{matrix} (m\varphi_0 + n\omega_0 + h)\, d\varphi_0\, d\omega_0.$$

Очевидно, что это выражение колеблется, не стремясь ни к какому определенному пределу. Итак, в этом случае первый постулат неверен. Не так обстоит дело со вторым постулатом, относящимся к интегралу $\int\limits_0^T P\, dt$. Действительно, мы должны рассмотреть выражение

$$\frac{1}{T}\int\limits_0^T dt \left[\int p_0 \sin(m\varphi + n\omega + h)\, d\varphi_0\, d\omega_0\right],$$

которое равно

$$\frac{A}{T}\int\limits_0^T \sin(m\Phi + n\Omega)t\, dt + \frac{B}{T}\int\limits_0^T \cos(m\Phi + n\Omega)t\, dt.$$

Это выражение стремится к нулю, когда T бесконечно растет, за исключением случая нулевого $m\Phi + n\Omega$.

Но $m\Phi + n\Omega$ может изначально равняться нулю, если Φ и Ω рационально соизмеримы друг с другом, и отношение $\frac{m}{n}$ равно $-\frac{\Omega}{\Phi}$. В этом случае траектория, описываемая молекулой жидкости, будет замкнутой кривой, имеющей уравнение

$$F = 0, \qquad F_1 = \text{const};$$

итак, здесь больше нет одного общего интеграла, их два, и мы возвращаемся к исключительному случаю в предыдущем примере.

Еще коэффициент $m\Phi + n\Omega$ может обратиться в нуль, если Φ и Ω рационально несоизмеримы, но m и n — нули. Пусть тогда θ — некоторая периодическая функция от ω и φ; рассмотрим ее вероятное значение

$$\Pi(t) = \int p_0 \theta \, d\omega_0 \, d\varphi_0$$

и интеграл

$$J = \frac{1}{T} \int\limits_0^T \Pi(t) \, dt.$$

Мы можем разложить θ в ряд Фурье. Каждому из членов этого ряда соответствует слагаемое из J. Согласно предыдущим рассуждениям, все члены J будут стремиться к нулю при очень большом T, исключая член с $m = n = 0$, т. е. тот, который соответствует среднему значению θ, постоянному члену ряда Фурье.

Рассуждение, совершенно аналогичное приведенному в параграфе 239, может показать нам, что в этом случае вероятность, представленная интегралом $\int \Phi \, dt$, распределена равномерно.

241. Очень существенно отдавать себе отчет в истинных причинах этого нового исключения, которое мы только что обозначили. Рассмотрим одну из молекул нашей жидкости, занимающую в момент 0 точку x_0, y_0, z_0, а в момент t — точку x, y, z; затем рассмотрим молекулы, которые в момент 0 заполняют сферу радиуса ε с центром в точке x_0, y_0, z_0; в момент времени t они заполняют очень малый объем; при бесконечно малом ε этот объем будет подобен эллипсоиду с центром в точке x, y, z.

Как будет вести себя этот эллипсоид при изменении t? В общем случае его оси станут все более и более разными, так что отношение этих осей стремится к бесконечности. Это существенно, для того чтобы постулат был верен; в приведенном исключительном случае это не так; если мы обозначим x_0, y_0, z_0 начальные координаты молекулы и из точки x_0, y_0, z_0, как из центра, опишем сферу радиуса ε, то эта сфера вырежет на поверхности тора область, которую можно считать кругом очень малого радиуса. При возрастании t эта область будет перемещаться по поверхности тора, оставаясь похожей на малый эллипс; но легко понять, что сжатие этого эллипса, вместо того чтобы бесконечно расти, будет колебаться в каких-то пределах. Если бы мы на мгновение рассмотрели ω и φ как координаты точки на плоскости, мы бы получили развертку тора; наш маленький эллипс был бы тогда представлен на плоскости другим маленьким эллипсом, *который всегда оставался бы равен себе самому*; если же мы возвращаемся к бесконечно маленькому эллипсу на торе, то отношение его осей будет зависеть только от параллельного к тору радиуса, на котором находится его центр, иначе говоря, будет линейной функцией от $\cos\omega$; это тоже периодическая по времени функция.

242. Рассмотрим теперь пример, в котором нет этого исключения, но ограничимся только частным случаем. Вернемся к нашему тору $F = 0$ и нашим уравнениям

$$\frac{d\varphi}{dt} = \Phi, \qquad \frac{d\omega}{dt} = \Omega,$$

но пусть Φ и Ω не будут константами; мы всегда сможем расположить поверхность $F = \varepsilon$ так, чтобы удовлетворить условию несжимаемости. Считаем, что

$$\Phi = \alpha M, \qquad \Omega = \beta M,$$

где α и β — две константы с иррациональным отношением, а M — периодическая функция от φ и ω, которая не обращается ни в нуль, ни в бесконечность.

Мы сможем ввести вспомогательную переменную τ такую, что

$$\frac{d\varphi}{d\tau} = \alpha, \qquad \frac{d\omega}{d\tau} = \beta, \qquad \frac{d\tau}{dt} = M,$$

откуда

$$\varphi = \alpha\tau + \varphi_0 m, \qquad \omega = \beta\tau + \omega_0, \qquad t = \int \frac{d\tau}{M};$$

$\dfrac{1}{M}$ — функция от τ, более того, это периодическая функция от ω_0 и φ_0. Рассмотренная как функция от τ, она раскладывается в тригонометрический ряд вида

$$\sum A\cos(\gamma t + h)$$

с нецелыми коэффициентами; именно эти функции г. Эсклангон называет *квазипериодическими*; отсюда получаем

$$t = A_0\tau + f(\tau, \omega_0, \varphi_0),$$

где f — квазипериодическая функция, а следовательно (при определенных условиях),

$$\tau = \frac{t}{A_0} + f_1(t, \omega_0, \varphi_0),$$

где f_1 — квазипериодическая функция от t. Если бы A_0 зависело от ω_0 и φ_0, анализ можно было бы осуществить без особых сложностей; но это не так. A_0 — член, не зависящий от t, т. е. от ω и φ в разложении $\frac{1}{M}$; он не зависит ни от ω_0, ни от φ_0.

Рассмотрим разность

$$f_1(t, \omega_0 + \varepsilon, \varphi_0 + \eta) - f_1(t, \omega_0, \varphi_0),$$

где ε и η очень малы; это еще одна квазипериодическая функция; если эта функция остается ограниченной, когда t меняется от $-\infty$ до $+$ $+\infty$, мы получаем результаты, аналогичные результатам предыдущего примера, и постулат не будет верен. Если, наоборот, эта квазипериодическая функция может стать больше любого предела (и я показал в Le Bulletin astronomique, Tome I, что существуют квазипериодические функции, для которых это выполнено), то постулат, вероятно, верен, но чтобы установить это, нам придется встретиться со всеми трудностями, связанными с теорией квазипериодических функций.

Случилось бы тоже самое, если бы мы предположили, в более общем виде,

$$\Phi = \alpha + \varepsilon\Phi_1, \qquad \Omega = \beta + \varepsilon\Omega_1,$$

где α и β — константы, ε — очень малая константа, Φ и Ω — периодические функции от φ и ω. Тогда можно было бы интегрировать с помощью последовательных приближений, раскладывая в ряд по степеням ε; получилось бы

$$\begin{aligned} \varphi &= at + f(t, \omega_0, \varphi_0), \\ \omega &= bt + f_1(t, \omega_0, \varphi_0), \end{aligned}$$

где a и b — константы, f и f_1 — функции, периодические по ω_0 и φ_0 и квазипериодические по t. Только в этом случае a и b снова не будут зависеть от ω_0 и φ_0, так что мы столкнемся с теми же трудностями.

243. Трудности, которые мы встретили в этом, таком простом, примере, показывают, что нас будет ожидать в общем случае. Скажем только несколько слов о методе рассуждений, которого можно было бы придерживаться. Разделим объем сосуда на очень большое число равных объемов; пусть v_1, v_2, ... , v_n — эти объемы; пусть p_i — вероятность того, что молекула находится в объеме v_i; пусть

$$x_1,\, x_2,\, \ldots\, ,\, x_n$$

— n вспомогательных переменных; рассмотрим выражение

$$P = p_1x_1 + p_2x_2 + \ldots + p_nx_n.$$

Пусть q_{ik} — вероятность, с которой молекула в момент времени $t+\tau$ находится в объеме v_i, если известно, что в момент времени t она находилась в объеме v_k. Тогда, если вероятностный закон в момент t представлен выражением P, то в момент $t+\tau$ он будет представляться выражением PS, в которое превратится P после линейной замены S, т. е. когда x_k заменяется на

$$q_{1k}x_1 + q_{2k}x_2 + \ldots + q_{nk}x_n.$$

В момент времени $t+2\tau$ он будет представляться выражением PS^2, в момент времени $t+h\tau$ — PS^h. Можно легко показать, что при бесконечно большом h вероятностный закон, представленный выражением PS^h стремится к равномерному вероятностному закону.

Но на это рассуждение имеется серьезное возражение. Мы не доказали, что вероятность, с которой молекула попадает в момент времени $t+2\tau$ в объем v_i, если в момент $t+\tau$ она была в объеме v_k, остается той же, было ли совершенно неизвестно, где она находилась в момент t, или же было известно, например, что она в этот момент находилась в объеме v_l.

Тем не менее, я счел необходимым привести это рассуждение, т. к. оно, несомненно, того же типа, к которому относятся многие рассуждения в кинетической теории газа, и в некоторых случаях они могут быть правдоподобны; так, когда рассматривают вероятность, с которой молекула газа отклоняется после соударения с другой молекулой, эта вероятность почти не зависит от предыдущих соударений, испытанных той же молекулой.

Значительная часть трудностей исчезнет, если считать, что функции X, Y, Z не полностью заданы, но зависят от какой-то функции от t

(или даже от многих функций), чье значение неизвестно, и мы не знаем ничего, кроме вероятности, с которой эта функция принимает значение, лежащее в данных пределах, например, между a и $a + da$. Тогда можно было бы рассуждать почти так же, как мы делали это в случае с тасованием карт.

Приведем пример из кинетической теории газа. Предположим, что молекулы газа заключены в сосуде, имеющем форму прямоугольного параллелепипеда, и могут ударяться о его стенки, но *не могут испытывать соударений друг с другом.* Если все они имеют *одну и ту же скорость*, то они не будут равномерно распределены в сосуде по истечении какого-то времени, если они не были равномерно распределены в момент времени $t = 0$. Напротив, это будет верно, если их скорость изменяется по какому-то вероятностному закону, например, по закону Максвелла, и этот вероятностный закон может быть любым.

Примечания редактора

[1](к стр. 26) Равновероятность всех исходов является главным условием классического определения вероятности, когда вероятность события равна отношению числа исходов, благоприятствующих этому событию, к числу всевозможных исходов.

[2](к стр. 28) Здесь А. Пуанкаре подчеркивает, что решение любой вероятностной задачи состоит из двух частей: первая — построение математической модели случайного эксперимента, вторая — вычисление вероятностей случайных событий с использованием математических формул. Следует заметить, что под построением математической модели сейчас понимается построение вероятностного пространства $(\Omega, \mathscr{F}, P)$, состоящего из пространства Ω всевозможных исходов случайного эксперимента, множества $\mathscr{F}$ случайных событий и определенной на $\mathscr{F}$ числовой функции P, называемой *вероятностью события.* При этом функция P выбирается с условием, что вероятность случайного события тем больше, чем более вероятно, что это событие произойдет. Само вероятностное пространство $(\Omega, \mathscr{F}, P)$ и есть математическая интерпретация изучаемого случайного явления.

[3](к стр. 55) Фактически А. Пуанкаре поделил все вероятностные пространства на четыре группы по числу исходов эксперимента.

К первой группе он относит пространства с конечным числом исходов, ко второй — пространства с бесконечным, но счетным множеством исходов, к третьей — с несчетным числом исходов, но с конечным числом размерностей («произвольных констант»). В том случае, когда исходом эксперимента является целая функция (бесконечномерное множество исходов), вероятностные задачи по Пуанкаре образуют четвертую группу.

[4](к стр. 39) Такое простое и изящное рассуждение А. Пуанкаре заключает неверным выводом. Выписав плотность распределения абсциссы попадания при стрельбе из пистолета в виде

$$\varphi(x) = Ce^{\frac{hx^2}{2}},$$

он должен был найти, что

$$f(x,\,y)=\varphi(x)\varphi(y)=C^2e^{\frac{h(x^2+y^2)}{2}}=C^2e^{\frac{h\rho^2}{2}}=f(\rho).$$

Это и есть плотность двумерного гауссовского распределения. В главе 6 А. Пуанкаре доказывает, что координаты вектора с таким распределением являются независимыми. Более того, нетрудно убедиться, что само распределение инвариантно относительно любого поворота плоскости с центром в начале координат. Таким образом, предположение А. Пуанкаре о равновероятности попадания во всех направлениях выполнено и $f(x,\,y)$ и есть решение задачи. Известно, что гауссовское распределение является наиболее адекватным распределению попаданий при стрельбе (как, впрочем, и распределению скоростей молекул газа — смотрите следующий параграф). Последнее подтверждает правильность красивых рассуждений А. Пуанкаре.

(5) (к стр. 49) Решая задачу об игре с выходящим, А. Пуанкаре фактически строит дерево игры. Сложность этой задачи состоит в том, что игра может продолжаться бесконечно, значит и количество исходов этого эксперимента бесконечно. Однако, после каждой партии мы оказываемся в одном из конечного числа состояний игры. Таким образом, А. Пуанкаре строит конечный граф игры и определяет вероятность перехода из каждого состояния в каждое.

Выписав соответствующую систему уравнений, он находит вероятности достижения окончательных результатов. Такого рода случайные процессы известны сейчас как конечные цепи Маркова с поглощающими состояниями. Вероятности поглощения в такой модели принято находить методом, предложенным А. Пуанкаре для решения этой задачи.

(6) (к стр. 55) А. Пуанкаре не вводит понятие случайной величины, как это принято в современных учебниках по теории вероятностей. Однако он понимает важность рассмотрения функции от случая и ее числовых характеристик. Здесь и далее под понятием «функция» имеется в виду функции от исходов случайного эксперимента. А. Пуанкаре не дает точного определения независимости функций от случая. Под независимостью функций следует понимать независимость любых случайных событий, относящихся к этим функциям.

(7) (к стр. 80) Фактически А. Пуанкаре доказал более сильный ре-

зультат, чем сформулированный закон больших чисел, а именно, центральную предельную теорему.

(8)(к стр. 95) А. Пуанкаре доказал устойчивость гауссовского (нормального) распределения, состоящую в том, что сумма независимых гауссовских случайных величин имеет также гауссовское распределение, причем дисперсия суммы случайных величин равна сумме дисперсий.

(9)(к стр. 108) Это верно только в том случае, если длина отрезка MP не превосходит расстояния между параллельными прямыми. Это можно понять из следующего простого рассуждения.

Допустим, что утверждение А. Пуанкаре верно: вероятность пересечения удваивается вместе с удвоением отрезка. Возьмем какой-нибудь отрезок длины l. Пусть вероятность пересечения с параллелями равна $p > 0$. Тогда вероятность пересечения для отрезка длины $2l$ равна $2p$. Нетрудно по индукции показать, что для любого натурального n вероятность пересечения для отрезка длины $2^n l$ равна $2^n p$. Но для любого $p > 0$ всегда можно найти такое n, что $2^n > p$. Это противоречит условию о том, что вероятность любого события не превосходит 1.

Эта ошибка А. Пуанкаре при решении задачи о бросании иглы (в современной литературе она известна как задача Бюффона) основана на том, что вероятность пересечения иглой параллельных линий не является линейной функцией от длины иглы (в отличие от математического ожидания числа пересечений). Покажем это.

Пусть расстояние между параллельными прямыми равно $2a$, а длина иглы равна $2l$. Найдем вероятность p того, что игла пересечет какую-нибудь прямую.

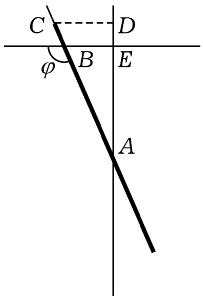


Рис. 1

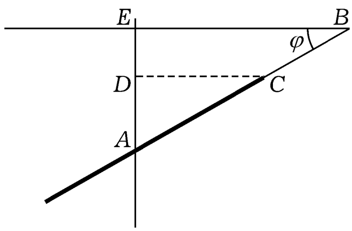


Рис. 2

Предположим, что мы бросили иглу. Рассмотрим ближайшую от центра иглы A прямую. Проведем перпендикуляр из A на эту прямую. Точку пересечения прямой и перпендикуляра обозначим через E. Пусть C — ближний конец иглы к прямой, а D — проекция точки C на перпендикуляр AE. Обозначим через φ угол между параллельной прямой и направлением иглы AC (см. рис. 1 и 2). Тогда положение иглы относительно ближайшей прямой полностью определяется длиной x отрезка AE и углом φ. При этом возможные значения x от 0 до a, а возможные значения φ от 0 до π. Другими словами, исходом нашего эксперимента (бросание иглы) можно считать пару чисел (x, φ) — точку прямоугольника со сторонами a и π (см. рис. 3 и 4):

$$\Omega = \{(x, \varphi),\ 0 \leqslant x \leqslant a \quad 0 \leqslant \varphi < \pi\}.$$

Площадь S_Ω равна πa.

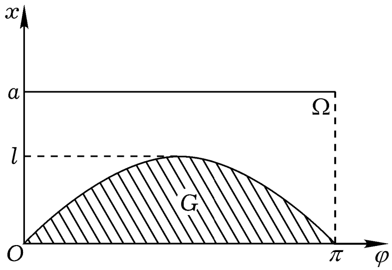


Рис. 3

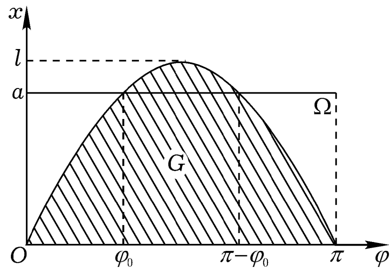


Рис. 4

Найдем теперь фигуру G, соответствующую событию, состоящему в том, что игла пересечет ближайшую параллель. Как видно из рисунков 1 и 2, игла пересечет прямую, если $AD \geqslant AE$. Выразим AD чере φ. Так как $AC = l$ и $\sin\widehat{ACD} = \sin\varphi$, то из прямоугольного треугольника ACD $Ad = l\sin\varphi$. Поэтому $G = \{(x, \varphi) \in \Omega\colon x \leqslant l\sin\varphi\}$ — это фигура, ограниченная сверху синусоидой.

Следует рассмотреть два случая. В первом случае $l \leqslant a$ (см. рис. 3). Тогда синусоида $x = l\sin\varphi$ не пересекает прямую $x = a$. Нетрудно найти, что площадь фигуры G равна

$$S_G = \int\limits_0^\pi l\sin\varphi\, d\varphi = -l\cos\varphi|_0^\pi = 2l.$$

Тогда вероятность p находится по формуле

$$p = \frac{S_G}{S_\Omega} = \frac{2l}{\pi a}.$$

Как видно, вероятность будет линейной функцией от l, но только в том случае, когда $l \leqslant a$ и игла может пересечь не более одной линии.

Если же $l > a$ и игла может пересечь сразу две параллельные линии, синусоида $x = l \sin\varphi$ выходит за пределы прямоугольника Ω (см. рис. 4). Пусть $\varphi_0 = \arcsin\frac{a}{l}$. Тогда площадь фигуры G представляется в виде суммы

$$S_G = \int\limits_0^{\varphi_0} l \sin\varphi \, d\varphi + \int\limits_{\varphi_0}^{\pi-\varphi_0} a \, d\varphi + \int\limits_{\pi-\varphi_0}^{\pi} l \sin\varphi \, d\varphi.$$

Первое и третье слагаемое равны

$$\int\limits_0^{\varphi_0} l \sin\varphi \, d\varphi = l(1 - \cos\varphi_0) = l\left(1 - \sqrt{1 - \frac{a^2}{l^2}}\right).$$

Второе слагаемое есть площадь прямоугольника со сторонами a и $(\pi - 2\varphi_0)$. Поэтому

$$S_G = 2l\left(1 - \sqrt{1 - \frac{a^2}{l^2}}\right) + a\bigl(\pi - 2\arcsin\frac{a}{l}\bigr).$$

Значит вероятность p равна

$$p = \frac{S_G}{S_\Omega} = \frac{2l}{\pi a}\left(1 - \sqrt{1 - \frac{a^2}{l^2}}\right) + \bigl(1 - \frac{2}{\pi}\arcsin\frac{a}{l}\bigr).$$

Полученная функция не является линейной по l, что опровергает утверждение А. Пуанкаре.

(10)(к стр. 108) Следует заметить, что в отличие от вероятности пересечения математическое ожидание числа пересечений есть линейная функция от длины отрезка. А. Пуанкаре очень изящно переходит от прямых к кривым. Затем в качестве эталона берется окружность диаметра, равного расстоянию между прямыми. Как бы мы ни бросали такую окружность на плоскость, мы всегда будем иметь две точки пересечения с прямыми. Так как длина такой окружности равна πd, математическое ожидание m числа пересечений кривой длины s удовлетворяет пропорции

$$\frac{m}{2} = \frac{s}{\pi d}, \quad \text{откуда} \quad m = \frac{2s}{\pi d}.$$

В частности, если мы бросаем иглу длины $2l$ и $d = 2a$, то $m = \frac{2l}{\pi a}$. В том случае, когда $l \leqslant a$, число пересечений иглой прямых не может быть более одного. Поэтому $p = m = \frac{2l}{\pi a}$, а это и есть решение задачи Бюффона. Более простого и изящного решения этой задачи, пожалуй, придумать невозможно!

(11)(к стр. 119) Здесь пропорциональность вероятности высоте, предложенная А. Пуанкаре, является спорной. Аналогично парадоксу Бертрана (см. параграф 63) можно предложить, что вероятность пропорциональна площади той области на сфере, куда должен попасть подвижный полюс P', или даже объему части шара, находящейся между соответствующими плоскостями. Однако, предположение А. Пуанкаре существенно упрощает задачу и делает ее более естественной.

(12)(к стр. 145) Другими словами, А. Пуанкаре пытается проверить, при каких условиях на φ оценка максимального правдоподобия будет совпадать со средним арифметическим полученных измерений. Как видно, метод максимального правдоподобия обсуждается в книге достаточно тщательно. Выводы, сделанные в этой главе, являются существенными для понимания этого метода при оценке математического ожидания случайной величины по заданной выборке.

(13)(к стр. 162) В частном случае для нормального распределения А. Пуанкаре решена задача, известная сейчас как *проблема моментов.* Она состоит в том, что можно ли по известным значениям моментов однозначно определить распределение случайной величины. Известно, что в некоторых случаях ответ на этот вопрос отрицательный. В общем случае эта проблема остается открытой.

(14)(к стр. 168) Таким образом А. Пуанкаре доказал центральную предельную теорему (сходимость к гауссовскому закону) для сумм независимых одинаково распределенных случайных величин, имеющих моменты всех порядков.

[(15)](к стр. 174) Приведенные рассуждения можно считать наброском доказательства центральной предельной теоремы для независимых случайных величин, имеющих разное распределение. Метод, который здесь применяется, известен как *метод характеристических функций.*

[(16)](к стр. 179) А. Пуанкаре здесь предлагает производить так называемое взвешенное сглаживание, тем самым разделяя ошибки первого и второго рода.

[(17)](к стр. 182) Такое распределение известно как *распределение Коши.* А. Пуанкаре привел его в качестве примера случайных величин, для которых центральная предельная теорема не выполнима. Действительно, в силу формулы

$$\left[F\left(\frac{\alpha}{n}\right)\right]^n = F(\alpha)$$

имеем, что среднее арифметическое случайных величин, имеющих распределение Коши, есть снова случайная величина, распределенная по Коши. И никакой сходимости к гауссовскому распределению здесь быть не может.

[(18)](к стр. 214) Под вероятностью допущенной ошибки здесь следует понимать ее вероятное значение. Вычисление вероятного значения ошибки А. Пуанкаре разделяет на три случая. Первый — это теоретическое вычисление. Второй — оценка по результатам наблюдений, когда дисперсия ошибки («искусство наблюдателя») неизвестна. Третий — это оценка по результатам наблюдений с заданной дисперсией.

[(19)](к стр. 239) Напомним, что многочленами Лежандра называются многочлены, определяемые по формуле:

$$\nu_n(\tau) = \frac{1}{n!\,2^n} \cdot \frac{d}{d\tau^n}(\tau^2 - 1)^{2n}.$$

Анри Пуанкаре

Теория вероятностей

Дизайнер В. В. Вельш
Корректор Г. В. Гребнева
Компьютерная подготовка И. В. Рылова
А. В. Широбоков
Компьютерная графика А. В. Широбоков

Лицензия ЛУ № 056 от 06.01.98. Подписано к печати 30.11.99.
Формат $60 \times 84^1/_{16}$. Печать офсетная. Усл. печ. л. 16,28. Уч. изд. л. 16,34.
Гарнитура Computer Modern Roman. Бумага офсетная № 1.
Заказ № К178. Тираж 1130 экз.
Ижевская республиканская типография,
426057, г. Ижевск, ул. Пастухова, 13.