Теория вероятностей и математическая статистика

История и методология прикладной математики и информатики

Ю.Б.Буркатовская, доцент ОИТ

Предыстория теории вероятностей

- У теории вероятностей по существу не было античных или средневековых предшественников, она целиком — создание Нового времени.
- Долгое время теория вероятностей считалась чисто опытной наукой и «не совсем математикой», её строгое обоснование было разработано только в 1929 году, то есть даже позже, чем аксиоматика теории множеств (1922).
- В наши дни теория вероятностей занимает одно из первых мест в прикладных науках по широте своей области применения; «нет почти ни одной естественной науки, в которой так или иначе не применялись бы вероятностные методы»[

Предыстория теории вероятностей

Предыстория, до XVI века включительно.

- В античные времена и в Средневековье натурфилософы ограничивались метафизическими рассуждениями о происхождении случайности и её роли в природе.· Математики в этот период рассматривали и иногда решали задачи, связанные с теорией вероятностей, но никаких общих методов и тематических понятий ещё не появилось.
- Главным достижением данного периода можно считать развитие комбинаторных методов, которые позже пригодились создателям теории вероятностей.
- Правило умножения и правило сложения.
- Перестановки, размещения и сочетания.
- Биномиальные коэффициенты (бином Ньютона, треугольник Паскаля).

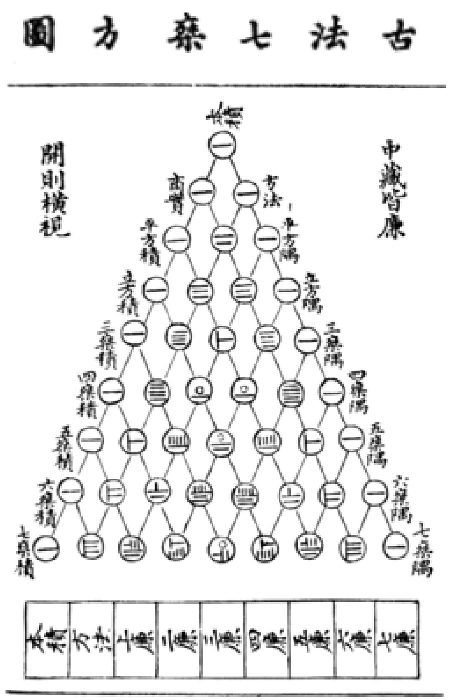
Предыстория теории вероятностей

Треугольник Паскаля.

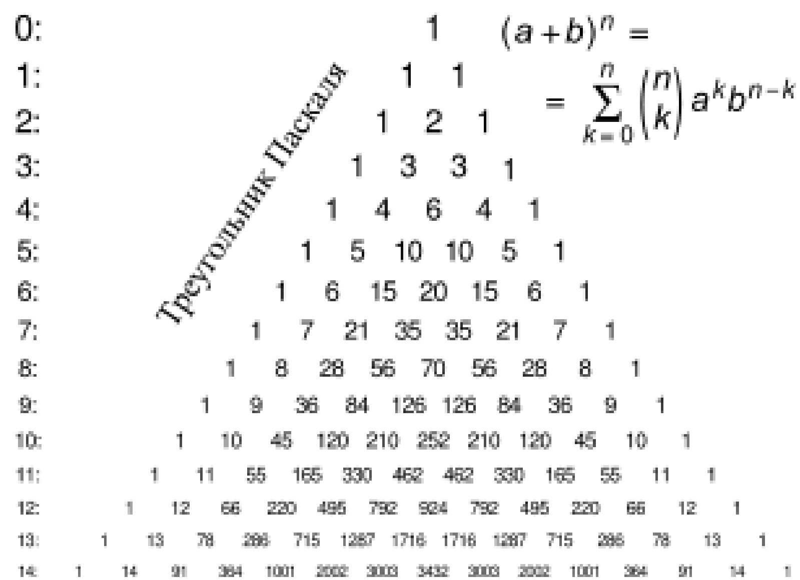
- Первое упоминание треугольной последовательности биномиальных коэффициентов под названием *meru-prastaara* встречается в комментарии индийского математика X века Халаюдхи к трудам другого математика, Пингалы.
- Треугольник исследуется также Омаром Хайямом около 1100 года, поэтому в Иране эту схему называют треугольником Хайяма.
- В 1303 году была выпущена книга «Яшмовое зеркало четырёх элементов» китайского математика Чжу Шицзе, в которой был изображен треугольник Паскаля на одной из иллюстраций; считается, что изобрёл его другой китайский математик, Ян Хуэй (поэтому китайцы называют его треугольником Яна Хуэя).
- В Италии треугольник Паскаля иногда называют «треугольником Тартальи», поскольку Никколо Тарталья описал эту таблицу на сто лет раньше Паскаля.
- На титульном листе учебника арифметики, написанном в 1529 году Петром Апианом, астрономом из Ингольштадтского университета, также изображён треугольник Паскаля.
- В 1665 году[1] вышла книга Блеза Паскаля «Трактат об арифметическом треугольнике», которая была специально посвящена данной таблице и по содержательности далеко опережала своих предшественников.

Предыстория теории вероятностей

Треугольник Яна Хуэя в китайском средневековом манускрипте, 1303 год



Первые 15 строк треугольника Паскаля



Предыстория теории вероятностей

Треугольник Паскаля.

- Числа треугольника симметричны (равны) относительно вертикальной оси.
- В строке с номером n:
  - первое и последнее числа равны 1.
  - второе и предпоследнее числа равны n.
  - m-е число (при нумерации с 0) равно биномиальному коэффициенту C(n,m).
- Сумма чисел восходящей диагонали, начинающейся с первого элемента (n-1)-й строки, есть n-е число Фибоначчи.
- Сумма чисел n-й строки треугольника Паскаля равна 2^n.
- Все числа в n-й строке, кроме единиц, делятся на число n, тогда и только тогда, когда n является простым числом.
- Если в строке с нечётным номером сложить все числа с порядковыми номерами вида 3n, 3n+1, 3n+2, то первые две суммы будут равны, а третья на 1 меньше.
- Каждое число в треугольнике равно количеству способов добраться до него из вершины, перемещаясь либо вправо-вниз, либо влево-вниз.

Зарождение теории вероятностей

Игра в кости.

- Французский каноник XIII века Ришар де Фурниваль правильно подсчитал все возможные суммы очков после броска трёх костей и указал число способов, которыми может получиться каждая из этих сумм. Это число способов можно рассматривать как первую числовую меру ожидаемости события, аналогичную вероятности.
- До Фурниваля, а иногда и после него, эту меру часто подсчитывали неверно, считая, например, что суммы 3 и 4 очка равновероятны, так как оба могут получиться «только одним способом»: по результатам броска «три единицы» и «двойка с двумя единицами» соответственно. При этом не учитывалось, что три единицы в самом деле получаются только одним способом: 1 + 1 + 1, а двойка с двумя единицами — тремя: 1 + 1 + 2 ; 1 + 2 + 1 ; 2 + 1 + 1, так что эти события не равновероятны

Зарождение теории вероятностей

Игра в кости.

- В обширной математической энциклопедии «Сумма арифметики, геометрии, отношений и пропорций» итальянца **Луки Пачоли** (1494) содержатся оригинальные задачи на тему: как разделить ставку между двумя игроками, если серия игр прервана досрочно.
- **Пример подобной задачи**: игра идёт до 60 очков, победитель получает всю ставку в 22 дуката, в ходе игры первый игрок набрал 50 очков, второй — 30, и тут игру пришлось прекратить; требуется справедливо разделить исходную ставку. Решение зависит от того, что понимать под «справедливым» разделом; сам Пачоли предложил делить пропорционально набранным очкам (55/4 и 33/4 дуката); позднее его решение было признано ошибочным.

Зарождение теории вероятностей

Игра в кости.

- **Джероламо Кардано** посвятил анализу игры содержательную монографию «Книга об игре в кости» (1526 год, опубликована посмертно). Кардано провёл полный и безошибочный комбинаторный анализ для значений суммы очков и указал для разных событий ожидаемое значение доли «благоприятных» событий: например, при бросании трёх костей доля случаев, когда значения всех 3 костей совпадают, равна 6/216 или 1/36.
- Кардано сделал проницательное замечание: *реальное количество исследуемых событий может при небольшом числе игр сильно отличаться от теоретического, но чем больше игр в серии, тем доля этого различия меньше*. По существу, Кардано близко подошёл к понятию вероятности.
- *Итак, имеется одно общее правило для расчёта: необходимо учесть общее число возможных выпадений и число способов, которыми могут появиться данные выпадения, а затем найти отношение последнего числа к числу оставшихся возможных выпадений.*

Зарождение теории вероятностей

Игра в кости.

- **Джероламо Кардано** посвятил анализу игры содержательную монографию «Книга об игре в кости» (1526 год, опубликована посмертно). Кардано провёл полный и безошибочный комбинаторный анализ для значений суммы очков и указал для разных событий ожидаемое значение доли «благоприятных» событий: например, при бросании трёх костей доля случаев, когда значения всех 3 костей совпадают, равна 6/216 или 1/36.
- Кардано сделал проницательное замечание: *реальное количество исследуемых событий может при небольшом числе игр сильно отличаться от теоретического, но чем больше игр в серии, тем доля этого различия меньше*. По существу, Кардано близко подошёл к понятию вероятности.
- *Итак, имеется одно общее правило для расчёта: необходимо учесть общее число возможных выпадений и число способов, которыми могут появиться данные выпадения, а затем найти отношение последнего числа к числу оставшихся возможных выпадений.*

Зарождение теории вероятностей

Игра в кости.

- **Никколо Тарталья**, раскритиковал подход Пачоли к решению задачи о разделе ставки: ведь если один из игроков ещё не успел набрать ни одного очка, то алгоритм Пачоли отдаёт всю ставку его сопернику, но это трудно назвать справедливым, поскольку некоторые шансы на выигрыш у отстающего всё же имеются. Кардано и Тарталья предложили свои (различные) способы раздела, но впоследствии и эти способы были признаны неудачными.
- Исследованием данной темы занимался и **Галилео Галилей**, написавший трактат «О выходе очков при игре в кости» (1718 год, опубликован посмертно). Изложение теории игры у Галилея отличается исчерпывающей полнотой и ясностью.
- В своей главной книге «*Диалог о двух главнейших системах мира, птолемеевой и коперниковой*» Галилей также указал на возможность оценки погрешности астрономических и иных измерений, причём заявил, что малые ошибки измерения вероятнее, чем большие, отклонения в обе стороны равновероятны, а средний результат должен быть близок к истинному значению измеряемой величины. Эти качественные рассуждения стали первым в истории предсказанием нормального распределения ошибок.

Зарождение теории вероятностей

Пьер Ферма и Блез Паскаль.

- Задачи **шевалье де Мере**.
- 1) Сколько раз нужно подбросить две кости, чтобы число случаев, благоприятствующих выпадению хотя бы раз двух шестерок, было больше, чем число случаев, когда ни при одном бросании не появляются две шестерки одновременно?
- 2) Два игрока играют и они договорились, что то, кто первым выиграет 6 партий, получит весь приз. Предположим, то на самом деле игра остановилась, до того, как один из них выиграл приз (например, первый игрок выиграл 5 партий, второй − 3). Как справедливо следует разделить приз? Большинство математиков XVI − XVII вв. считали, что в отношении 5:3, один из них − Тарталья считал, что 2:1. Паскаль и Ферма установили, что 7:1.
- Паскаль и Ферма вступили в переписку друг с другом по поводу данной задачи и родственных вопросов (1654). В рамках этой переписки учёные обсудили ряд проблем, связанных с вероятностными расчётами; в частности, рассматривалась старая задача о разделе ставки, и оба учёных пришли к решению, что надо разделить ставку соответственно остающимся шансам на выигрыш.

Зарождение теории вероятностей

Христиан Гюйгенс (14 апреля 1629, Гаага – 8 июля 1695, там же) — нидерландский механик, физик, математик, астроном и изобретатель.

- "О расчётах в азартной игре" (1657 г.), первый трактат по теории вероятностей.
- *Я полагаю, что при внимательном изучении предмета читатель заметит, что имеет дело не только с игрой, но что здесь закладываются основы очень интересной и глубокой теории.*
- Главным достижением стало введение понятия математического ожидания, то есть теоретического среднего значения случайной величины. Гюйгенс также указал классический способ его подсчёта
- *Если число случаев, в которых получается сумма a, равно p, а число случаев, в которых получается сумма b, равно q, то стоимость моего ожидания равна (ap+bq)/(p+q).*

Зарождение теории вероятностей

Христиан Гюйгенс.

- В книге большое число задач, некоторые с решениями, другие «для самостоятельного решения». Из последних особый интерес и оживлённое обсуждение вызвала **«задача о разорении игрока»**: у игроков А и В есть a и b монет соответственно, в каждой игре выигрывается одна монета, вероятность выигрыша А в каждой игре равна p, требуется найти вероятность полного его разорения. Полное общее решение «задачи о разорении» дал **Абрахам де Муавр** полвека спустя (1711). В наши дни вероятностная схема «задачи о разорении» используется при решении многих задач типа «случайное блуждание».
- Гюйгенс проанализировал и задачу о разделе ставки, дав её окончательное решение: ставку надо разделить пропорционально вероятностям выигрыша при продолжении игры.
- Он также впервые применил вероятностные методы к демографической статистике и показал, как рассчитать среднюю продолжительность жизни.

Зарождение теории вероятностей

Первые результаты статистики.

- К этому же периоду относятся публикации английских статистиков **Джона Граунта** (1662) и **Уильяма Петти** (1676, 1683). Обработав данные более чем за столетие, они показали, что многие демографические характеристики лондонского населения, несмотря на случайные колебания, имеют достаточно устойчивый характер — например, соотношение числа новорождённых мальчиков и девочек редко отклоняется от пропорции 14 к 13, невелики колебания и процента смертности от конкретных случайных причин.
- Граунт также впервые составил таблицы смертности — таблицы вероятности смерти как функции возраста.
- Вопросами теории вероятностей и её применения к демографической статистике занялись также **Иоганн Худде** и **Ян де Витт** в Нидерландах, которые в 1671 году также составили таблицы смертности и использовали их для вычисления размеров пожизненной ренты. Более подробно данный круг вопросов был изложен в 1693 году **Эдмундом Галлеем**.

Развитие теории вероятностей

Якоб Бернулли.

- Изучил теорию вероятностей по книге Гюйгенса «О расчётах в азартной игре», в которой ещё не было определения и понятия вероятности (её заменяет количество благоприятных случаев).
- Ввел **классическое определение вероятности события** как *отношения числа исходов, связанных с этим событием, к общему числу исходов* (у достоверного события вероятность равна единице, у невозможного — нулю).
- Единственное важное уточнение — о том, что все «элементарные исходы» обязаны быть равновероятны, — сделал Пьер-Симон Лаплас в 1812 году.
- Если для события невозможно подсчитать классическую вероятность (например, из-за отсутствия возможности выделить равновероятные исходы), то Бернулли предложил использовать **статистический подход**, то есть оценить вероятность по результатам наблюдений этого события или связанных с ним. Это **статистическое определение вероятности**.
- Сформулировал первый вариант закона больших чисел.
- *Среднее значение конечной выборки из фиксированного распределения близко к математическому ожиданию этого распределения.*

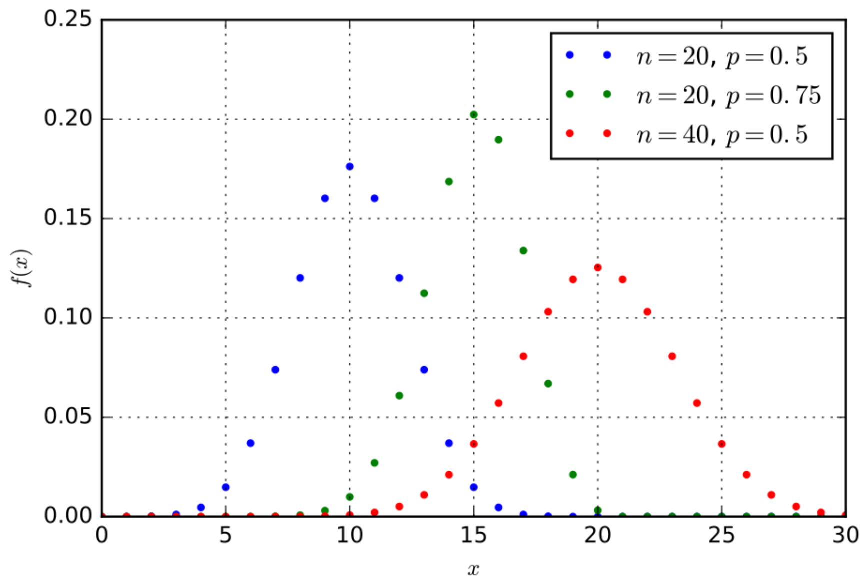
Развитие теории вероятностей

Якоб Бернулли.

- Имя Якоба носит важное в комбинаторике распределение Бернулли (случайная величина принимает значение 1 с вероятностью *p* и значение 0 с вероятностью 1−*p*).
- **Формула Бернулли:** если вероятность события равна *p*, то вероятность того, что в *n* испытаниях событие случится *m* раз, равна

$$C_n^m p^m (1-p)^{n-m}$$

- Распределение числа «благоприятных» исходов появления события с вероятностью *p* в *n* независимых испытаниях сейчас называется **биномиальным распределением**.



Развитие теории вероятностей

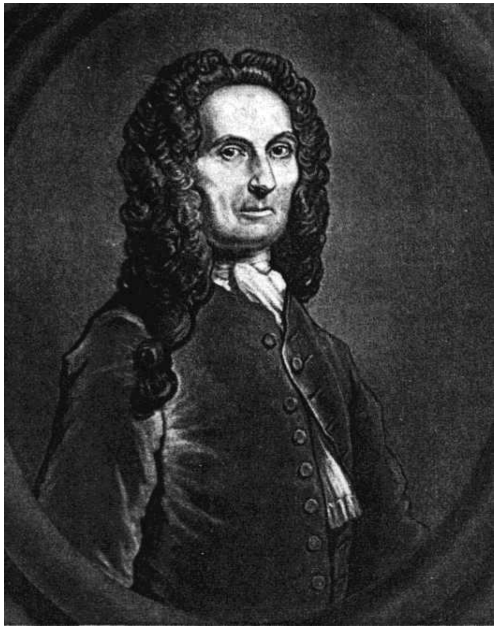
Якоб Бернулли.

- Якоб Бернулли подготовил монографию в этой области, однако издать её не успел. Она была напечатана посмертно, в 1713 году, его братом Николаем, под названием **«Искусство предположений».** Это содержательный трактат по теории вероятностей, статистике и их практическому применении, итог комбинаторики и теории вероятностей XVII века.
- В последней части книги, оставшейся недописанной, Бернулли собирался рассмотреть экономические и другие практические приложения теории.

JACOBI BERNOULLI,
Profeſſ. Baſil. & utriuſque Societ. Reg. Scientiar.
Gall. & Pruſſ. Sodal.
MATHEMATICI CELEBERRIMI,
ARS CONJECTANDI,
OPUS POSTHUMUM.
Accedit
TRACTATUS
DE SERIEBUS INFINITIS,
Et EPISTOLA Gallicè ſcripta
DE LUDO PILÆ
RETICULARIS.

BASILEÆ,
Impenſis THURNISIORUM, Fratrum.
cIↃ IↃcc XIII.

Развитие теории вероятностей



Абрахам де Муавр (26 мая 1667, Витри-ле-Франсуа — 27 ноября 1754, Лондон) — английский математик французского происхождения.

- Член Лондонского королевского общества, Парижской и Берлинской академий наук.
- Родился во Франции, в недворянской семье врача-гугенота; частицу *де* перед своей фамилией он добавил сам. В 11 лет поступил в Протестантскую академию в Седане, где успел проучиться 4 года, после чего академия была запрещена властями (1682). Муавр продолжил образование в Сомюре (2 года). Вероятно, в это время он познакомился с теорией вероятностей по трудам Гюйгенса.
- Далее около года Муавр слушал лекции по физике и математике в Париже но в 1685 году Людовик XIV официально отменил Нантский эдикт, возобновились притеснения протестантов, а сам Муавр попал в тюрьму. Подробности его заключения неизвестны, но так или иначе, он вынужден был покинуть родину.

Развитие теории вероятностей

- Далее около года Муавр слушал лекции по физике и математике в Париже, но в 1685 году Людовик XIV официально отменил Нантский эдикт, возобновились притеснения протестантов, а сам Муавр попал в тюрьму. Подробности его заключения неизвестны, но так или иначе, он вынужден был покинуть родину.
- В 1688 году он осел в Лондоне, где и прожил всю оставшуюся жизнь. На жизнь зарабатывал частным преподаванием и шахматной игрой. Вскоре Муавр стал известен как талантливый математик, однако как иностранец не имел никаких шансов на кафедру в английском учебном заведении. Религиозная дискриминация сменилась национальной.
- Незадолго до его приезда вышла книга Ньютона «Математические начала натуральной философии» в трёх томах. Она так увлекла Муавра, что он разобрал её по листам и постоянно носил с собой очередную порцию для чтения, чтобы не терять времени при переездах от одного ученика к другому.

Развитие теории вероятностей

- В 1692 году де Муавр познакомился с Галлеем, а через него — с Ньютоном. Если верить сплетням той поры, Ньютон выпроваживал посетителей, досаждавших ему мелкими делами математического характера, с помощью следующей фразы: «Идите к де Муавру, он разбирается в этом лучше меня». Муавр также постоянно помогал Ньютону в издании и редактировании трудов (особенно «Оптики»).
- В 1695 году был опубликован первый труд Муавра по анализу, *«Метод флюксий»*. В 1710 году он участвовал в комиссии, разбиравшей приоритетный спор Ньютона и Лейбница.
- В 1718 году опубликован главный труд по теории вероятностей: «The Doctrine of Chances: A method of calculating the probabilities of events in play». Книга вызвала большой интерес и выдержала 3 издания.
- В 1724 году вышло в свет вероятностно-статистическое исследование «Annuities on lives» (переиздано четырежды).
- В 1730 году Муавр вернулся к анализу и опубликовал труд «Miscellanea Analytica», где впервые появилась формула Стирлинга.
- 27 ноября 1754 года де Муавр умер в Лондоне. Несколько важных работ де Муавра были опубликованы посмертно.

Развитие теории вероятностей

Абрахам де Муавр.

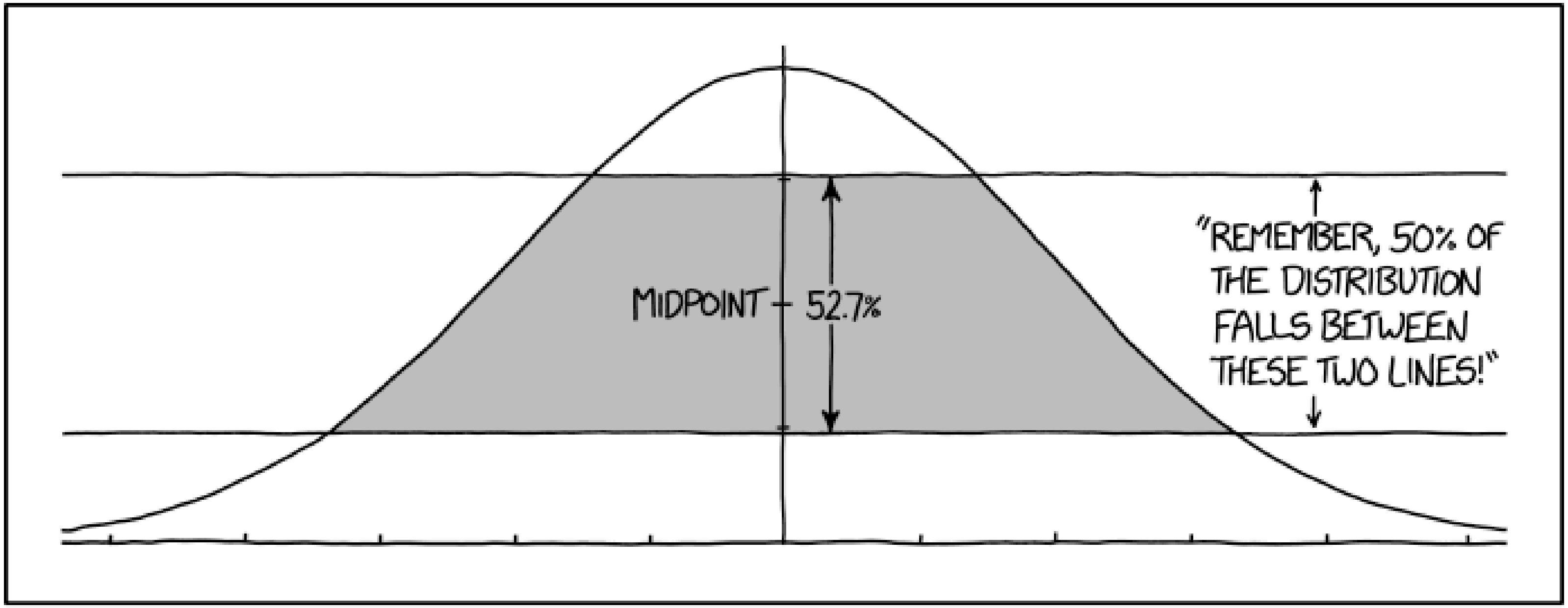
- Муавр не только полностью решил упоминавшуюся выше «задачу о разорении игрока», но и оценил для неё среднюю продолжительность игры и вероятности выигрыша за заданное число игр для каждого игрока.
- Теорема умножения вероятностей (1718 год) «*вероятность появления двух зависимых событий равна произведению вероятности появления одного из них на вероятность того, что другое должно появиться, если первое из них уже появилось*»[1].
- Первый вариант теоремы Муавра—Лапласа, исследующей распределение отклонений статистической частоты (m/n) от вероятности. Муавр рассмотрел только случай, когда вероятность равна 1/2, общий же случай доказал Лаплас. Здесь впервые появилось **нормальное распределение**.
- **Теорема Муавра-Лапласа.** Если в схеме Бернулли n стремится к бесконечности, величина $p\in(0,1)$ постоянна, то вероятность m успехов в n испытаниях равна

$$P_n(m) = \frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{(m-np)^2}{2npq}\right)\left(1+O\left(\frac{1}{\sqrt{n}}\right)\right)$$

Развитие теории вероятностей

Рэндалл Манро. Комикс о нормальном распределении.

https://www.explainxkcd.com/wiki/index.php/2118:_Normal_Distribution



HOW TO ANNOY A STATISTICIAN

Развитие теории вероятностей

Даниил Бернулли (29 января (8 февраля) 1700 — 17 марта 1782) — швейцарский физик, механик и математик, один из создателей кинетической теории газов, гидродинамики и математической физики. Сын Иоганна Бернулли.

- Независимо от Муавра, исследовал нормальное распределение для ошибок наблюдений.
- Первым применил к вероятностным задачам методы математического анализа (до этого использовался комбинаторный подход)
- Опубликовал первый из вероятностных парадоксов.
- Продвинул также математическую статистику, рассмотрев с применением вероятностных методов ряд практически важных задач. Например, оценивал эффективность вариоляции (прививки путем вдыхания перетертых в пыль струпьев переболевших оспой).

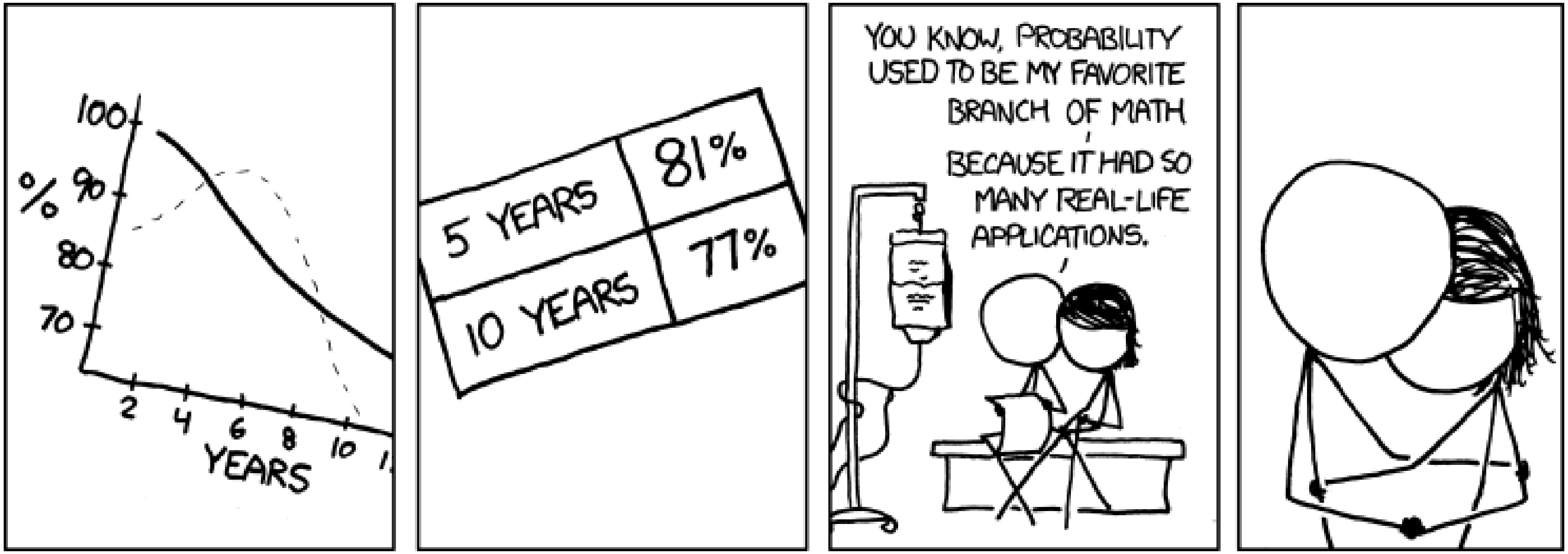
Развитие теории вероятностей

Даниил Бернулли

- Для определения общей эффективности вариоляции он предложил уравнение, которое описывало долю людей в каждой возрастной группе, никогда не болевших оспой и, следовательно, рискующих ею заразиться. Свое уравнение он выверял по таблице смертности, составленной Эдмундом Галлеем (прославившегося наблюдением за кометами), которая описывала, какая часть из всех родившихся доживала до определённого возраста. На этой основе Бернулли смог вычислить соотношение выздоровевших и умерших к общему числу тех, кто болел оспой.
- С помощью второго уравнения он подсчитал число жизней, которые могли бы быть спасены, если бы вариоляции регулярно подвергали всё население. Бернулли пришёл к выводу, что при всеобщей вариоляции почти 50% новорожденных доживали бы до 25 лет. По сегодняшним меркам число удручающее, но по сравнению с показателем в 43% при свободном распространении оспы это существенный прогресс. Он также показал, что прививка способна увеличить среднюю ожидаемую продолжительность жизни более чем на три года.

Развитие теории вероятностей

Рэндалл Манро. Комикс «Вероятность».
https://www.explainxkcd.com/wiki/index.php/881:_Probability



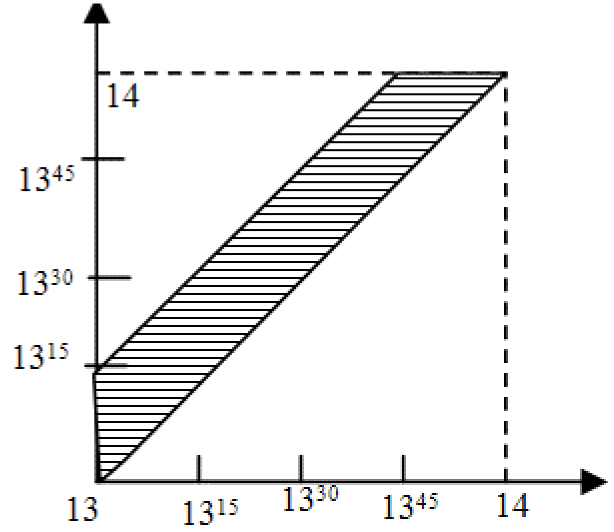
Развитие теории вероятностей

- **Санкт-петербургский парадокс** (или **санкт-петербургская лотерея**) в экономической науке — парадокс, иллюстрирующий расхождение между теоретически оптимальным поведением игрока и «здравым смыслом».
- Рассматривается следующая задача. Вступая в игру, игрок платит некоторую сумму, а затем подбрасывает монету (вероятность каждого исхода — 50 %), пока не выпадет орёл. При выпадении орла игра заканчивается, а игрок получает выигрыш, рассчитанный по следующим правилам. Если орёл выпал при первом броске, игрок получает 2^0 дукатов, при втором броске — 2^1 дукатов и так далее (при n-ном броске — 2^{n-1} дукатов). Другими словами, выигрыш, возрастая от броска к броску вдвое, последовательно пробегает степени двойки — 1, 2, 4, 8, 16, 32 и так далее.
- **Вопрос:** при каком вступительном взносе игра становится справедливой?
- Парадокс заключается в том, что хотя вычисленное значение (математическое ожидание выигрыша) этого справедливого взноса и равно бесконечности, то есть выше любого возможного выигрыша, реальные игроки ощущают, что даже 25 дукатов — слишком высокая цена для входа в игру.

Развитие теории вероятностей

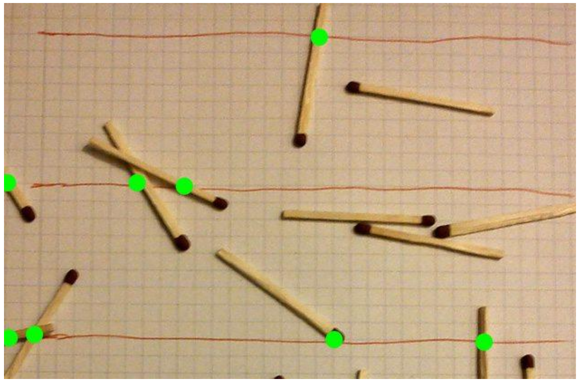
Томас Симпсон (20 августа 1710 — 14 мая 1761) — английский математик.

- В ходе занятий численным анализом в книге «Природа и законы случая» (1740) фактически использовал третье (наряду с классическим и статистическим) определение вероятности — **геометрическое**, пригодное для исследования непрерывных случайных величин с бесконечным числом значений
- На рисунке – иллюстрация к **задаче о встрече**. Какова вероятность встречи, если двое приходят от 13-00 до 14-00 и ждут друг друга 5 минут.
- **Вероятность** – отношение площадей фигур, соответствующих благоприятным событиям и всем событиям).



Развитие теории вероятностей

Жорж-Луи Леклерк, граф де Бюффон (или просто **Бюффон**; 7 сентября 1707, Монбар, Бургундия — 16 апреля 1788, Париж) — французский натуралист, биолог, математик, естествоиспытатель и писатель XVIII века.



- **Задача Бюффона о бросании иглы** — один из первых примеров применения **метода Монте-Карло** и рассмотрения понятия геометрической вероятности. Эта задача сделала возможным определение числа Пи вероятностными методами.
- Суть метода была в бросании иглы длиной L на плоскость, расчерченную параллельными прямыми, расположенными на расстоянии r друг от друга. Вероятность того, что игла пересечет прямую, связана с числом Пи.

$$p = \frac{2L}{r\pi}$$

Развитие теории вероятностей

Томас Байес (1702, Лондон — 17 апреля 1761, Танбридж-Уэллс) — британский математик, пресвитерианский священник.

- Отец — Джошуа Байес — пресвитерианский священник. Томас обучался дома и в 1719 году поступил в Эдинбургский университет изучать логику и богословие. По возвращении домой в 1722 году Байес помогал отцу в часовне проводить службу, а вскоре, в 30-х годах, сам стал священником в пресвитерианской церкви. В 1734 году переехал в Танбридж Уэллс, графство Кент. В 1752 году он вышел в отставку. Умер в 1761 году.
- в 1742 году был избран в члены лондонского Королевского общества, хотя не было опубликовано ни одной работы Байеса по математике.

Развитие теории вероятностей

Томас Байес.

- Фундаментальное исследования Байеса в области теории вероятностей было изложено им в "Эссе о решении проблем в теории случайных событий". Эту работу математика лишь после его смерти обнаружил друг Ричард Прайс, который и переслал статью в академию. В 1764 году это "Эссе" было опубликовано в "Трудах Лондонского Королевского общества".
- **Теорема Байеса** (или **формула Байеса**) позволяет определить вероятность какого-либо события при условии, что произошло другое статистически взаимозависимое с ним событие. Вводится понятие **условной вероятности**.
- Формула Байеса позволяет «переставить причину и следствие»: по известному факту события вычислить вероятность того, что оно было вызвано данной причиной.

$$P(A|B) = \frac{P(AB)}{P(A)} = \frac{P(B|A)\,P(A)}{P(B)}$$

Развитие теории вероятностей

Применения теоремы Байеса.

- Для её практического применения требуется большое количество расчётов, вычислений, поэтому байесовские оценки стали активно использовать только после революции в компьютерных и сетевых технологиях.
- Пионером здесь стала британская интернет-компания Autonomy, для интеллектуального поиска информации созданная математиком Майком Линчем. Программное обеспечение Autonomy, построенное на базе байесовых оценок, позволяет компьютерам "понимать" содержание неструктурированной информации, такой как текстовые участки веб-страниц.
- В Microsoft этот же статистический аппарат заложен в программы выявления неполадок в новой ОС WinXP, а еще ранее - был использован при создании для MS Office столь доставшего всех своими советами "мистера Скрепки".
- Фильтры для почтового сервера — такие, как DSPAM, SpamAssassin, SpamBayes, SpamProbe, Bogofilter, CRM114 — используют методы байесовского фильтрования спама.

Развитие теории вероятностей

Парадокс теоремы Байеса.

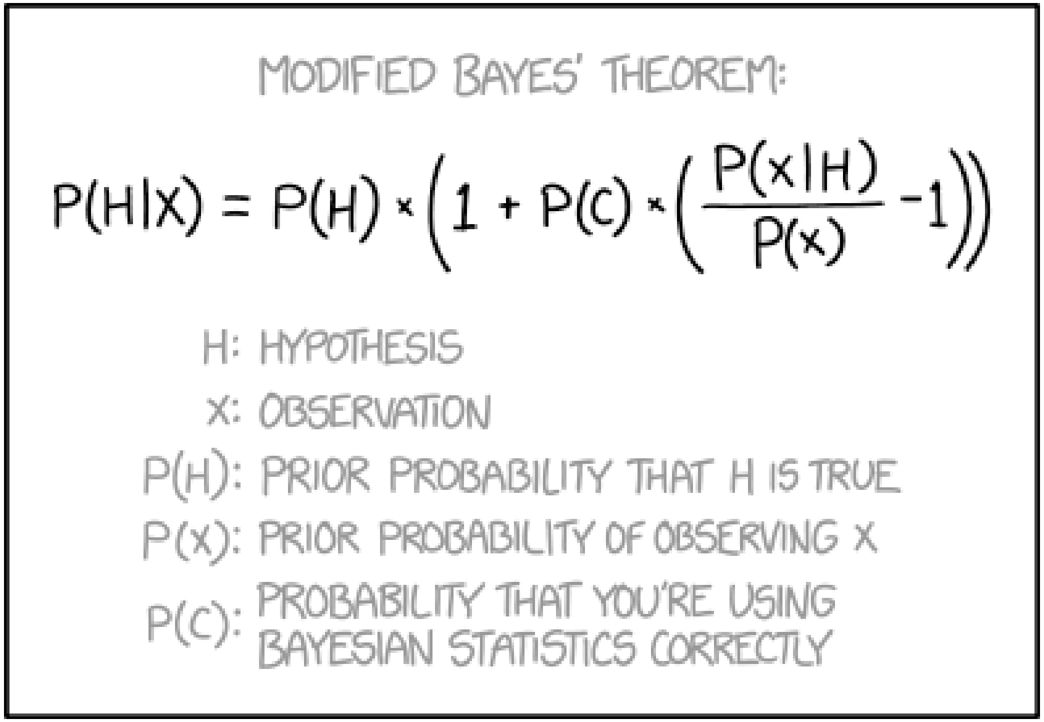
- *«При рентгеновском обследовании вероятность обнаружить заболевание туберкулёзом у больного туберкулёзом равна 0.9, вероятность принять здорового человека за больного равна 0.01. Доля больных туберкулёзом по отношению ко всему населению равна 0.001»*.
- Если для таких условий посчитать вероятность ложноположительного результата, получится, что **91.7% людей, у которых обследование показало результат «болен», на самом деле здоровы.**
- **Объяснение:** доля больных значительно меньше, чем вероятность принять здорового за больного.
- A – тест положительный, B – человек здоров.

$$P(B|A)=\frac{P(A|B)\,P(B)}{P(A|B)\,P(B)+P(A|\bar{B})\,P(\bar{B})}=\frac{0.01\times 0.999}{0.01\times 0.999+0.9\times 0.001}=0{,}917.$$

Развитие теории вероятностей

Рэндал Манро. Модификация теоремы Байеса.

https://www.explainxkcd.com/wiki/index.php/2059:_Modified_Bayes%27_Theorem



Развитие теории вероятностей

- **Леонард Эйлер** дал подробный анализ разных типов лотерей, в Пруссии ему было поручено проведение лотерей.
- Статистике и страхованию Эйлер посвятил немало работ; он, в частности, решал задачу: оценить по статистическим таблицам, какова вероятность того, что человек в возрасте m лет проживёт ещё n лет.
- В XIX веке число работ по теории вероятностей продолжало расти, были даже компрометирующие науку попытки распространить её методы далеко за разумные пределы — например, на область морали, психологии, правоприменения и даже богословия.
- Основной сферой её применения в тот период была математическая обработка результатов наблюдений, содержащих случайные погрешности, а также расчёты рисков в страховом деле и других статистических параметров.
- Уже к середине XIX века формируется вероятностная теория артиллерийской стрельбы.

Вероятность и статистика в XIX веке

- Среди главных прикладных задач теории вероятностей и математической статистики XIX века можно назвать следующие:
  - найти вероятность того, что сумма независимых случайных величин с одинаковым (известным) законом распределения находится в заданных пределах. Особую важность эта проблема представляла для теории ошибок измерения, в первую очередь для оценки погрешности наблюдений;
  - установление статистической значимости различия случайных значений или серий таких значений. Пример: сравнение результатов применения нового и старого видов лекарств для принятия решения о том, действительно ли новое лекарство лучше;
  - исследование влияния заданного фактора на случайную величину (факторный анализ).
- В большинстве крупных стран Европы были созданы национальные статистические организации. В конце века область применения вероятностных методов начала успешно распространяться на физику, биологию, экономику, социологию.

Вероятность и статистика в XIX веке

Иоганн Карл Фридрих Гаусс (30 апреля 1777, Брауншвейг — 23 февраля 1855, Гёттинген) — немецкий математик, механик, физик, астроном и геодезист. Считается одним из величайших математиков всех времён, «королём математиков».

- Дед Гаусса был бедным крестьянином; отец, Гебхард Дитрих Гаусс, – садовником, каменщиком, смотрителем каналов; мать, Доротея Бенц, – дочерью каменщика.
- С именем Гаусса связаны фундаментальные исследования почти во всех основных областях математики: в алгебре, теории чисел, дифференциальной и неевклидовой геометрии, математическом анализе, теории функций комплексного переменного, теории вероятностей, а также в аналитической и небесной механике, астрономии, физике и геодезии.

Вероятность и статистика в XIX веке

Иоганн Карл Фридрих Гаусс

- Гаусс, постоянно занимавшийся астрономическими вычислениями, разработал вероятностную методику работы с измерениями, содержащими погрешности (1809).
- Он глубоко изучил **нормальное распределение**, показал, что оно во многих практических ситуациях является предельным для случайных значений, обосновал применение **метода наименьших квадратов** для оценки измеряемого значения и параметров его возможного диапазона разброса.
- Хотя нормальный закон был известен задолго до Гаусса, его вклад в теорию этого важнейшего распределения настолько велик, что долгое время нормальный закон называли «законом Гаусса»; современный термин закрепился благодаря работам Карла Пирсона в конце XIX века

Вероятность и статистика в XIX веке

Симеон Дени Пуассон (21 июня 1781, Питивье, Франция — 25 апреля 1840, Со, Франция) — французский математик, механик и физик.

- Его имя часто встречается в учебниках по математическому анализу и электромагнетизму, теории вероятностей и акустики, квантовой механики и теории упругости.
- *Жизнь украшается двумя вещами: занятием математикой и её преподаванием.*

Вероятность и статистика в XIX веке

Симеон Дени Пуассон

- В 1837 году обобщил закон больших чисел Бернулли, сняв условие о том, что вероятность события в каждой игре одна и та же; при этих новых условиях статистическая частота будет сходиться к среднему арифметическому для вероятностей отдельных игр.
- Опубликовал **формулу Пуассона** (**«закон редких событий»**) , удобную для описания схемы Бернулли в том случае, когда вероятность события близка к нулю или к единице.

$$P(X=m)\approx\frac{(np)^m}{m!}e^{-np}$$

- Применил теорию вероятностей к задачам стрельбы.
- **Распределение Пуассона** является одним из основных в прикладных задачах.

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$$

Вероятность и статистика в XIX веке

Пьер –Симон, маркиз де Лаплас

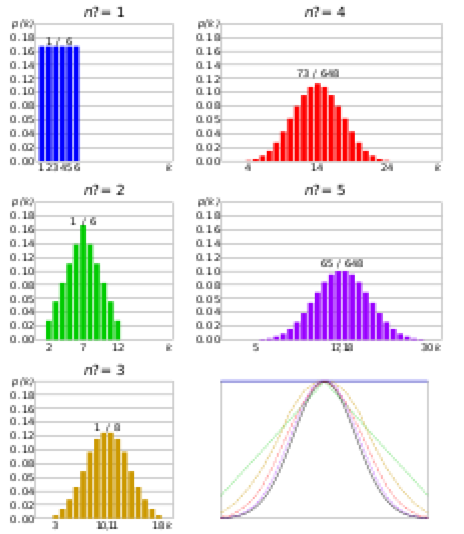
- **«Аналитическая теория вероятностей»** (1812 год) выдержала во Франции три переиздания и была переведена на многие языки.
- Лаплас исследовал как дискретные, так и непрерывные случайные величины (ещё не вводя термина **«случайная величина»**), причём для непрерывных дал ключевое понятие **плотности распределения вероятности**, ранее неявно и ограниченно использованное Даниилом Бернулли.
- Интегральное понятие **функции распределения** возникло гораздо позже (его в 1912 году ввёл А. М. Ляпунов); общий термин «случайная величина» также, по-видимому, впервые появился в работах русской вероятностной школы.
- Введение плотности вероятности и **характеристических функций** позволило Лапласу применить для решения вероятностных задач мощные аналитические средства, включая дифференциальные уравнения в частных производных.

Вероятность и статистика в XIX веке

Пьер–Симон, маркиз де Лаплас

- Лаплас привёл **формулу полной вероятности** для нескольких несовместных «причин» (в современной терминологии, «гипотез»), доказал ряд предельных теорем, в том числе **теорему Муавра—Лапласа** и сходимость **биномиального распределения** к нормальному при увеличении числа испытаний (на рисунке – плотности биномиального и нормального распределения).
- Указал, что наблюдаемые погрешности измерения являются результатом суммирования множества случайных ошибок, и их распределение должно быть близко к нормальному. Для оценки возможного диапазона значений измеряемой величины Лаплас, как и Гаусс, рекомендовал метод наименьших квадратов.

$$P(A) = \sum_{i=1}^{n} P(A|H_i)\,P(H_i)$$



Вероятность и статистика в XIX веке

Пьер–Симон, маркиз де Лаплас

- По его мнению, ход реальных процессов полностью предопределён («детерминирован»), случайность появляется лишь в человеческом восприятии и только там, где человек не владеет полным знанием происходящего.
- *Ум, которому были бы известны для какого-либо данного момента все силы, одушевляющие природу, и относительное положение всех её составных частей, если бы вдобавок он оказался достаточно обширным, чтобы подчинить эти данные анализу, обнял бы в одной формуле движение величайших тел вселенной наравне с движениями легчайших атомов; не осталось бы ничего, что было бы для него недостоверно, и будущее, так же, как и прошедшее, предстало бы пред его взором.*
- Ошибочность подобной предопределённости была отмечена задолго до появления вероятностной квантовой механики — ещё в начале XX века Анри Пуанкаре обнаружил принципиально непредсказуемые процессы, в которых ничтожное изменение исходного состояния вызывает со временем сколь угодно большие отклонения в конечном состоянии.

Вероятность и статистика в XIX веке

- В России в первой половине XIX века начали возникать собственные серьёзные исследования по теории вероятностей. Первый учебный курс начал читать **С. Ревковский** в Вильнюсском университете (1829 год), там же в 1830 году была создана первая в Российской империи кафедра теории вероятностей.
- В Петербургском университете лекции с 1837 года читал **В. А. Анкудович**, а с 1850 года — **В. Я. Буняковский**. Фундаментальный учебник «*Основания математической теории вероятностей*» Буняковский опубликовал в 1846 году, и придуманная им терминология стала общепринятой.
- В Московском университете курс появился в 1850 году, лекции читал **А. Ю. Давидов**, будущий президент Московского математического общества.
- Статьи по вероятностным темам публиковали многие крупные математики России, в том числе **М. В. Остроградский**, **Н. Д. Брашман**, **Н. И. Лобачевский**, **Н. Е. Зернов**. В значительной части этих работ ощущается сильное влияние трудов и взглядов Лапласа.

Вероятность и статистика в XIX веке

Виктор Яковлевич Буняковский (3 (15) декабря 1804 — 30 ноября (12 декабря) 1889) — русский математик, педагог, историк математики, вице-президент академии наук в 1864—1889 годах.

- В 1820 году Буняковский отправился за границу, где изучал преимущественно математические науки. Он имел возможность заниматься у Лапласа, Пуассона, Фурье, Коши, Ампера, Лежандра и других знаменитых учёных. Больше всего Буняковский работал у Коши.
- В 1824 году Буняковский получил степени бакалавра и лиценциата; 19 мая 1825 года защитил диссертацию, и получил от Парижского университета степень доктора математических наук.
- Пробыв за границей в общей сложности семь лет, Буняковский в 1826 году приехал в Петербург, где занялся педагогической деятельностью

Вероятность и статистика в XIX веке

Виктор Яковлевич Буняковский

- В 1846 году появился труд Буняковского, послуживший началом его всемирной известности, — **«Основания математической теории вероятностей»**. Кроме теории, он включал и историю возникновения и развития теория вероятностей; в нём впервые сведено вместе всё то, что было выработано по этой теории трудами известных математиков, начиная с Паскаля и Ферма, даны объяснения относительно новых решений самых трудных и запутанных вопросов.
- Много практических приложений теории вероятностей, например, **к вопросу о средней продолжительности жизни людей различных возрастов**, к определению достоверности свидетельств и преданий, к вспомогательным кассам и страховым учреждениям, к определению погрешностей при наблюдениях, к вопросам судебного дела, к вычислению вероятностных потерь в войске и т. д.
- Гаусс выучился русскому языку по этому сочинению.

Вероятность и статистика в XIX веке

Виктор Яковлевич Буняковский

- В 1848 году Буняковский поместил в «Современнике» обратившую на себя внимание статью: «О возможности введения определённых мер доверия к результатам некоторых наук и преимущественно статистики».
- Особенно большую практическую пользу оказали труды Буняковского по вопросу об эмеритальных кассах; он разработал основания эмеритальной пенсионной кассы морского ведомства, и его труды послужили к учреждению целого ряда подобных касс на выработанных им началах. Сделав в 1869 году выводы эмпирического закона о смертности, Буняковский упростил этим решение вопросов относительно страхования капиталов и пожизненных доходов.
- При богатстве и глубине содержания, лекции Буняковского всегда отличались поразительной ясностью, увлекательностью и в то же время литературной красотой изложения, делали легко доступными самые сложные математические положения и увлекали даже безучастных слушателей. По отношениям к лекциям Буняковский проявлял замечательную аккуратность и в течение всего времени своей службы в университете не пропустил ни одной лекции и не опоздал ни разу.

Вероятность и статистика в XIX веке

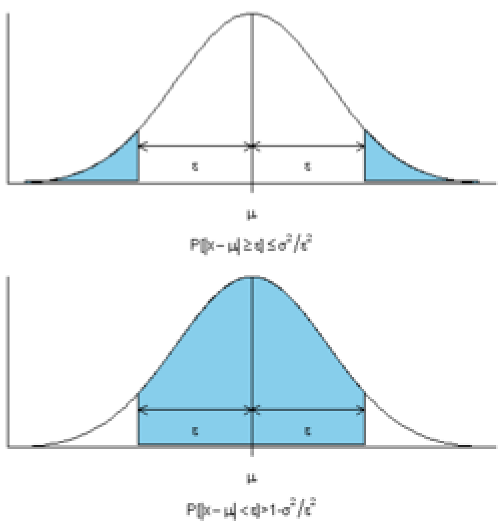
Пафнутий Львович Чебышёв (4 [16] мая 1821, Окатово, Калужская губерния, — 26 ноября [8 декабря] 1894, Санкт-Петербург) — русский математик и механик, академик Петербургской академии наук и ещё 24 академий мира.

- Сын богатого землевладельца, представителя старинного русского дворянского рода Чебышёвых Льва Павловича Чебышёва — участника Отечественной войны 1812 года и взятия Парижа в 1814 году
- Получил фундаментальные результаты в теории чисел и теории вероятностей, построил общую теорию ортогональных многочленов, теорию равномерных приближений и многие другие. Основал математическую теорию синтеза механизмов и разработал ряд практически важных концепций механизмов.

Вероятность и статистика в XIX веке

Пафнутий Львович Чебышёв

- Чебышёв стал первым русским математиком мирового уровня и в теории вероятностей. С 1860 года он сменил В. Я. Буняковского на кафедре теории вероятностей Петербургского университета и начал свой цикл лекций.
- Он опубликовал по данной теме всего четыре работы, но фундаментального характера.
- В статье **«О средних величинах»** (1866 год) было впервые доказано «неравенство Чебышёва», позднее усиленное Марковым. Эта формула означает, что вероятность отклонения любой случайной величины x от её среднего значения (математического ожидания) M[x] более чем на k стандартных отклонений σ не превышает $1/k^2$. Например, отклонение более чем на 5σ имеет вероятность не более 1/25, то есть 4%.



$$P\left(\left|x - M(x)\right| \geq k\sigma\right) \leq \frac{1}{k^2}$$

Вероятность и статистика в XIX веке

Пафнутий Львович Чебышёв

- Хотя неравенство впервые было опубликовано (без доказательства) **И.-Ж. Бьенэме** в 1853 году, за ним закрепилось название «неравенство Чебышёва» — в значительной мере потому, что П. Л. Чебышёв не только дал вывод этого неравенства, но и успешно применил его для решения важной проблемы — обоснования закона больших чисел.
- Именно, в качестве следствия данного неравенства Чебышёв получил чрезвычайно общую формулировку закона больших чисел: если математические ожидания серии n случайных величин и квадраты этих математических ожиданий ограничены в совокупности, то среднее арифметическое этих величин с ростом n сходится к среднему арифметическому для их математических ожиданий.
- Из этой теоремы получаются как следствия теоремы Бернулли и Пуассона; Чебышёв впервые строго оценил точность этих теорем и других приближений.
- В этой же статье П. Л. Чебышёв впервые чётко обосновал общепринятую сегодня точку зрения на понятие **случайной величины** как на одно из основных понятий теории вероятностей — величина, принимающая определенные значения с определенной вероятностью.

Вероятность и статистика в XIX веке

Пафнутий Львович Чебышёв

- В 1887 году появилась статья **«О двух теоремах относительно вероятностей».** В этой работе он установил, что при некоторых (достаточно общих) условиях выполняется центральная предельная теорема: сумма большого числа независимых случайных величин с нулевыми математическими ожиданиями (например, погрешностей измерения) распределена приближённо по нормальному закону, и тем точнее, чем больше слагаемых в сумме. Этот результат по своей общности далеко перекрывает теорему Муавра — Лапласа и все её аналоги.
- Доказывая свой вариант центральной предельной теоремы, Чебышёв допустил логический пробел: оказалось, что — в дополнение к указанным Чебышёвым условиям применимости теоремы — следует ещё потребовать, чтобы среднее арифметическое дисперсий при стремлении n к бесконечности имело предел. Данный недостаток был вскоре исправлен **А. А. Марковым**.

Вероятность и статистика в XIX веке

Пафнутий Львович Чебышёв

- Обе упомянутые теоремы Чебышёва занимают центральное место в теории вероятностей. Особенно важно то обстоятельство, что Чебышёв не только указал предельное распределение, но в обоих случаях детально проанализировал границы возможных отклонений от этого предела.
- Исследования П. Л. Чебышёва продолжили его ученики, в первую очередь А. А. Марков и А. М. Ляпунов.
- Характеристика его учёных заслуг очень хорошо выражена в записке академиков А. А. Маркова и И. Я. Сонина, зачитанной на первом после смерти Чебышёва заседании Академии.
- *Труды Чебышёва носят отпечаток гениальности. Он изобрёл новые методы для решения многих трудных вопросов, которые были поставлены давно и оставались нерешёнными. Вместе с тем он поставил ряд новых вопросов, над разработкой которых трудился до конца своих дней.*

Вероятность и статистика в XIX веке

Пафнутий Львович Чебышёв (прикладная математика)

- В течение сорока лет Чебышёв принимал активное участие в работе военного артиллерийского ведомства и работал над усовершенствованием дальнобойности и точности артиллерийской стрельбы, применяя для обработки результатов опытных стрельб методы теории вероятностей. В курсах баллистики до наших дней сохранилась **формула Чебышёва** для вычисления дальности полёта снаряда в зависимости от его угла бросания, начальной скорости и сопротивления воздуха при заданной начальной скорости.
- В ходе данных исследований он в 1873 году предложил новый тип квадратурных формул (*квадратурные формулы Чебышёва*). Эти формулы позволяют упростить вычисления и сократить их объём, обладая следующим важным свойством: они доставляют минимум дисперсии вычисленного по ним приближённого значения интеграла (при условии, что погрешности в узлах независимы и имеют одинаковую дисперсию и равное нулю математическое ожидание).

Вероятность и статистика в XIX веке

Андрей Андреевич Марков (2 (14) июня 1856, Рязань — 20 июля 1922, Петроград) — русский математик, академик, внёсший большой вклад в теорию вероятностей, математический анализ и теорию чисел. Член Санкт-Петербургского математического общества.

- В 1874 году А. А. Марков окончил гимназию и поступил в Санкт-Петербургский университет. Там он слушал лекции профессоров А. Н. Коркина и Е. И. Золотарёва, а также Пафнутия Львовича Чебышёва, оказавшего определяющее влияние на выбор научной деятельности Андрея Маркова.
- Преподавал в Петербургском университете, в частности, после Чебышёва читал курс теории вероятностей.

Вероятность и статистика в XIX веке

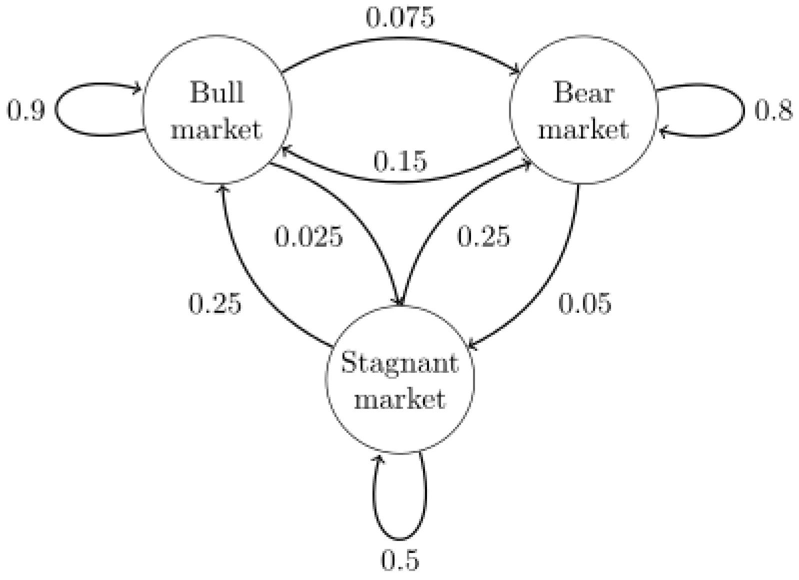
Андрей Андреевич Марков. Факты биографии.

- В 1874 году А. А. Марков окончил гимназию и поступил в Андрей Марков страдал туберкулёзом коленного сустава и до 10 лет ходил на костылях. После операции, проведённой известным хирургом Кадэ, он получил возможность ходить нормально.
- В 1866 году его отдали в 5-ю Петербургскую гимназию. Это классическое учебное заведение с преподаванием древних языков (латинского и греческого) пришлось ему не по вкусу; по большинству предметов он учился плохо, исключение составлял только один предмет — математика.
- В 1883 году А. А. Марков женился на Марии Ивановне Вальватьевой. Через 20 лет у них родился сын Андрей (полный тёзка отца), основоположник советской школы конструктивной математики.
- Ещё в 1901 году академик Марков протестовал против решения Синода об отлучении от церкви Льва Толстого. 12 февраля 1912 года А. А. Марков направил в Святейший синод письмо, в котором просил отлучить его от церкви. Академик писал: «*Я не усматриваю существенной разницы между иконами и мощами, с одной стороны, и идолами, которые, конечно, не боги, а их изображения, с другой, и не сочувствую всем религиям, которые, подобно православию, поддерживаются огнём и мечом и сами служат им*».

Вероятность и статистика в XIX веке

Андрей Андреевич Марков

- Является первооткрывателем обширного класса стохастических процессов с дискретной и непрерывной временной компонентой, названных его именем.
- **Цепь Маркова** — последовательность случайных событий с конечным или счётным числом исходов, где вероятность наступления каждого события зависит от состояния, достигнутого в предыдущем событии.
- В то время, когда эта теория была построена, она считалась абстрактной, однако в настоящее время практические применения данной теории чрезвычайно многочисленны.



Вероятность и статистика в XIX веке

Андрей Андреевич Марков

- А. А. Марков существенно продвинул классические исследования предшественников, касающиеся закона больших чисел и центральной предельной теоремы теории вероятностей, а также распространил их и на цепи Маркова.
- **Неравенство Маркова** даёт оценку вероятности что неотрицательная случайная величина превзойдёт по модулю фиксированную положительную константу, в терминах её математического ожидания.

$$P(X \geq a) \leq \frac{M(X)}{a} \qquad P(|x - M(x)| \geq a) \leq \frac{M|x - M(x)|}{a}$$

- А. А. Марков своим открытием цепей Маркова сделал крупнейший вклад в теорию случайных процессов и теорию вероятностей в целом.

Вероятность и статистика в XIX веке

Приложения цепей Маркова

- Моделирование систем массового обслуживания и социально-экономических процессов.
- Поисковые системы. Метод PageRank.
- Анализ текста и определение авторства. Работа начата еще А.А.Марковым, который провел анализ романа «Евгений Онегин» на частоту появления букв – гласных и согласных – в зависимости от предыдущей буквы.
- Автоматическая генерация текста и спама.
- Расчет сумм для обязательного страхования автогражданской ответственности.
- Моделирование химических и экологических процессов.

Вероятность и статистика в XIX веке

Александр Михайлович Ляпунов (25 мая [6 июня] 1857, Ярославль — 3 ноября 1918, Одесса, Херсонская губерния) — русский математик и механик, академик Петербургской Академии наук с 1901 года, член-корреспондент Парижской академии наук, член Национальной академии деи Линчеи (Италия) и ряда других академий наук и научных обществ.

- Ему принадлежит введение метода характеристических функций в учение о предельных теоремах теории вероятностей.
- Доказал центральную предельную терему при более общих условиях, чем его предшественники.

$$\varphi_X \quad t \quad = E\left[e^{itX} \right]$$

Аксиоматика теории вероятностей

- Проблема аксиоматизации теории вероятностей включена Д. Гильбертом в формулировку его 6-й проблемы «Математическое изложение основ физики»:
- *С исследованиями по основаниям геометрии близко связана задача об аксиоматическом построении по этому же образцу тех физических дисциплин, в которых уже теперь математика играет выдающуюся роль: это в первую очередь теория вероятностей и механика. Что касается аксиом теории вероятностей, то мне казалось бы желательным, чтобы параллельно с логическим обоснованием этой теории шло рука об руку строгое и удовлетворительное развитие метода средних значений в математической физике, в частности, в кинетической теории газов.*
- **Андрей Николаевич Колмогоров (1903−1987)** в 1933 году на основе теории множеств и теории меры дал наиболее совершенное аксиоматическое построение теории вероятностей.

Аксиоматика теории вероятностей

Аксиоматика Колмогорова

- ***Аксиома I (алгебра событий)***. *F* является алгеброй событий.
- ***Аксиома II (существование вероятности событий)***. Каждому событию *x* из *F* поставлено в соответствие неотрицательное вещественное число *P*(*x*), которое называется *вероятностью события x*.
- ***Аксиома III (нормировка вероятности)***. P (Ω) = 1.
- ***Аксиома IV (аддитивность вероятности)***. Если события *x* и y {\displaystyle *y* не пересекаются, то *P*(*x*+*y*) = *P*(*x*)+*P*(*y*).
- Совокупность объектов (Ω,*F*,*P*), удовлетворяющая *аксиомам I—IV*, называется вероятностным пространством (у Колмогорова: *поле вероятностей*).

Аксиоматика теории вероятностей

- Норберт Винер, «отец» кибернетики, свидетельствовал: *«…Хинчин и Колмогоров, два наиболее видных русских специалиста по теории вероятностей, долгое время работали в той же области, что и я. Более двадцати лет мы наступали друг другу на пятки: то они доказывали теорему, которую я вот-вот готовился доказать, то мне удавалось прийти к финишу чуть-чуть раньше их»*.
- Колмогоров А.Н. **Теория вероятностей и математическая статистика.**
- В ней помещены исследования по теории вероятностей (основания, предельные теоремы, случайные процессы, разнообразные приложения), математической статистике и некоторым другим вопросам.

Математическая статистика

- **Математическая статистика** — раздел математики, посвященный математическим методам систематизации, обработки и использования статистических данных для научных и практических выводов.
- **Разделы математической статистики:**
  - Оценка параметров
  - Проверка гипотез
  - Планирование эксперимента
  - Последовательный анализ
  - Дисперсионный анализ
  - Корреляционный анализ
  - Многомерный статистический анализ
  - Факторный анализ
  - Статистика случайных процессов
  - Статистика объектов нечисловой природы

Математическая статистика

Зарождение математической статистики

- Работы Гаусса и Лапласа. Астрономы хотели определить реальное положение звезд после проведения нескольких измерений Лаплас догадался, что реальное положение было причиной наблюдаемых положений, в то время как ошибки являлись случайными. Он переоткрыл теорему Байеса, и пришел к выводу, что существует кривая, описывающая распределение ошибок (экспоненциальное распределение).
- Лаплас пытался минимизировать сумму абсолютных ошибок (модулей ошибок), а другие астрономы сосредоточились на минимизации суммы квадратов. Метод описан Адриеном Мари Лежандром в книге «Новые метода определения орбит комет» (1805)
- Гаусс применил метод наименьших квадратов для оценки орбиты астероида Цереры (открыт 1 января 1801 года). Использовал нормальное распределение для описания ошибок. Метод описан в книге «Теория движения небесных тел» (1809).

Математическая статистика

Зарождение математической статистики

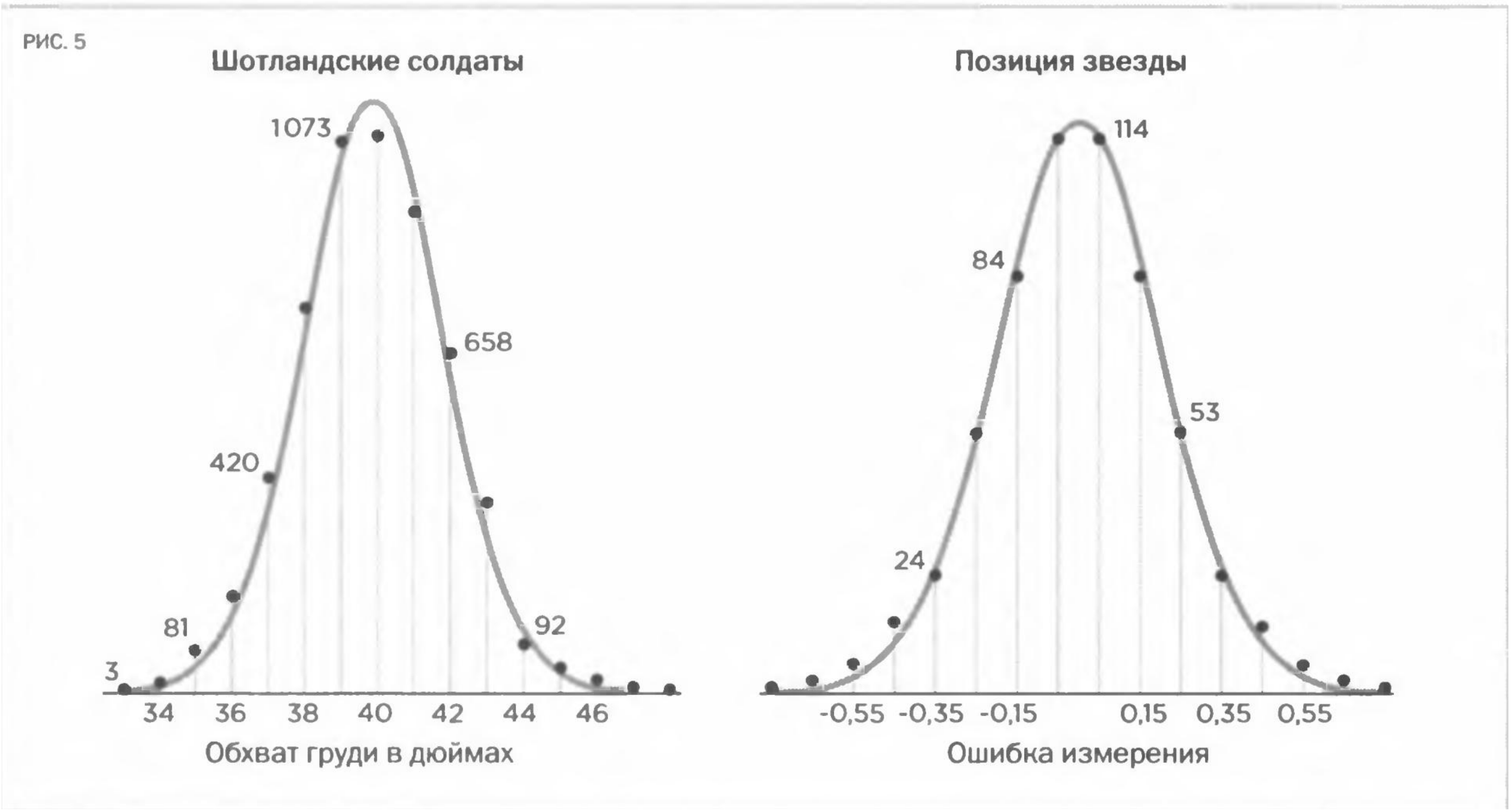
- **Адольф Кетле (1796-1874),** бельгийский астроном, получил образование в Париже, где благодаря своему учителю **Жозефу Фурье** познакомился с работами Гаусса и Лапласа.
- Кетле приводила в замешательство стабильность статистических данных: — криминальная статистика во Франции ежегодно давала почти одни и те же цифры. С 1825 по 1830 год число обвиняемых насчитывало около 7100, а число осужденных — 4400.
- Кетле применил вероятностный подход к пропорциям человеческого тела (именно ему мы обязаны термином «индекс массы тела») и затем с легкостью использовал его в астрономии и геодезии.

Математическая статистика

Зарождение математической статистики

- В 1845 году, изучив и графически представив данные по обхвату груди 5738 шотландских солдат, взятые из медицинского журнала того времени, Кетле заметил, что полученная кривая частот похожа на кривую, которую дает измерение положения звезд, сделанное Фридрихом Бесселем.
- Кетле утверждал, что измерение обхвата груди множества солдат эквивалентно многократному измерению одного и того же «среднего солдата», и предположил, что природа стремится к среднему человеку, а индивиды, находящиеся по бокам колокола, суть случайные отклонения от идеального канона.
- **Впервые использовал кривую Гаусса для описания не ошибок, а распределения случайной величины.**
- В Великобритании колоколообразная кривая получила название **нормального закона**. Нормальными считались люди, соответствовавшие центральной тенденции социальных законов, описывающих рост, вес и интеллект. Социология пошла дальше, назвав людей, чьи показатели удалялись от средних, патологическими, «анормальными».

Математическая статистика



Математическая статистика

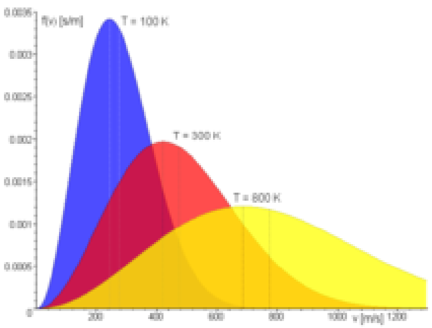
Зарождение математической статистики

- **Джеймс Клерк Максвелл** (13 июня 1831, Эдинбург, Шотландия — 5 ноября 1879, Кембридж, Англия) — британский (шотландский) физик, математик и механик. Член Лондонского королевского общества (1861).
- Работая над кинетической теорией газов, Максвелл вводит в физику понятия «вероятно», «это событие может произойти с большей степенью вероятности».
- Максвелл выступил в качестве продолжателя идей Рудольфа Клаузиуса, который ввёл понятия средней длины свободного пробега и средней скорости молекул (предполагалось, что в состоянии равновесия все молекулы имеют одну и ту же скорость).

Математическая статистика

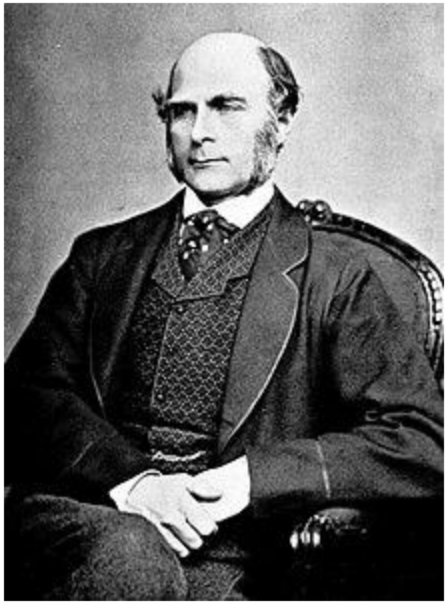
Зарождение математической статистики

- Максвелл исходил из представления о газе как об ансамбле множества идеально упругих шариков, хаотически движущихся в замкнутом пространстве и сталкивающихся друг с другом.
- Шарики-молекулы можно разделить на группы по скоростям, при этом в стационарном состоянии число молекул в каждой группе остаётся постоянным, хотя они могут менять скорость после столкновений. Из такого рассмотрения следовало, что в равновесии частицы имеют не одинаковую скорость, а распределяются по скоростям в соответствии с кривой Гаусса (распределение Максвелла).
- С помощью полученной функции распределения Максвелл рассчитал ряд величин, играющих важную роль в явлениях переноса: число частиц в определённом диапазоне скоростей, среднюю скорость и средний квадрат скорости.



Математическая статистика

Зарождение математической статистики



- Сэр **Фрэнсис Гальтон** (16 февраля 1822 — 17 января 1911) — английский исследователь, географ, антрополог и психолог; основатель дифференциальной психологии и психометрики, статистик.
- Гальтон был двоюродным братом Чарльза Дарвина по их деду — Эразму Дарвину. Его отцом был Самюэль Тертиус Гальтон, сын Самюэля «Джона» Гальтона.
- Среди прочего, открыл явление антициклона, показал уникальность отпечатков пальцев и способствовал введению дактилоскопии в полицейскую практику.

Математическая статистика

Зарождение математической статистики

- **Фрэнсис Гальтон (1822-1911)**
- Гальтон считал, что союз двух умных людей приведет к рождению еще более умных детей, точно так же как у двух высоких людей рождаются еще более высокие дети. Однако он открыл другую статистическую закономерность.
- В своей книге *«Естественное наследование»* (1889) он назвал ее «возвращением к посредственности», а позднее — **«регрессией к среднему».**
- Он заметил, что у высоких родителей обычно рождаются высокие дети, а у низких родителей дети, как правило, невысокого роста. Но когда родители особенно высокие или особенно низкие, наблюдается возврат к среднему росту в популяции (дети регрессируют).
- *Уравнение регрессии*: рост ребенка, см = 85 см + 0,5×рост родителя, см.
- Ввел понятие **корреляции**: линейной связи между величинами.

Математическая статистика

Зарождение математической статистики. Фрэнсис Гальтон

- Занимался не средними величинами, а отклонениями от среднего (*дисперсия*, *вариация*).
- Гальтон обращал внимание на значительно отличающихся от среднего индивидуумов: средний человек Кетле был для него не прообразом совершенства, а посредственностью, которая должна была развиваться.
- «*Точно так же, как методами тщательной и умелой селекции в рамках естественных ограничений удается получить стабильную породу собак или лошадей, обладающих особенными способностями к бегу или к чему-либо еще, представляется вполне возможным произвести высокоталантливую расу людей путем рассчитанных браков в течение нескольких последовательных поколений*».

«Наследственность таланта, ее законы и последствия» (1869)

Математическая статистика

Зарождение математической статистики

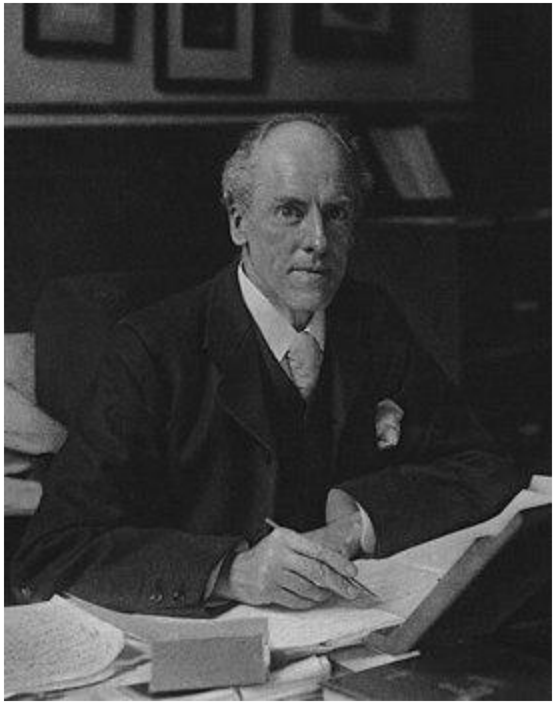
- **Фрэнсис Гальтон (1822-1911)**
- В 1883 году Гальтон ввел термин **евгеника** (наука об улучшении человеческой расы). В конце века это учение получило признание в британском обществе, погруженном в проблемы империи — как внешние (по отношению к другим империям), так и внутренние (усиление люмпен-пролетариата — низших социальных слоев, в которых уровень рождаемости был значительно выше, чем в привилегированных классах). Оно закрепилось в США и нацистской Германии вместе с провозглашением законов о принудительной стерилизации душевнобольных и неимущих.
- Евгеническое движение развивалось практически без остановки до тех пор, пока не были уничтожены концентрационные лагеря Центральной Европы и разделение человечества по расовому признаку не объявили аморальным мифом довоенной физической антропологии.

Математическая статистика

Итак, предпосылки математической статистики

- **Астрономия (Гаусс, Лаплас):** вероятность, распределение ошибок, закон больших чисел, центральная предельная теорема, метод наименьших квадратов.
- **Кинетическая теория газов (Максвелл):** моделирование поведения большого числа частиц через нормальное распределение.
- **Биология и антропология (Гальтона):** регрессия и корреляция.

Математическая статистика



Карл Пирсон (27 марта 1857, Лондон — 27 апреля 1936, там же) — английский математик, статистик, биолог и философ; основатель математической статистики.

- Карл Пирсон опубликовал основополагающие труды по математической статистике (более 400 работ по этой теме). Разработал теорию корреляции, критерии согласия, алгоритмы принятия решений и оценки параметров.
- Методы Пирсона имеют предельно общий характер и применяются практически во всех естественных науках.
- Свою невероятную работоспособность Пирсон в шутку объяснял привычкой не отвечать во время работы на телефонные звонки и никогда не участвовать в оргкомитетах.

Математическая статистика

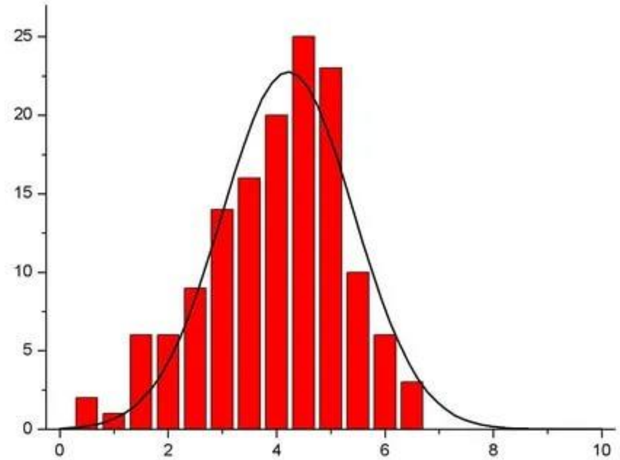
Карл Пирсон

- Благодаря дружбе с **Уолтером Фрэнком Рафаэлем Уэлдоном** (1860-1906), профессором зоологии, Пирсон в 1892 году заинтересовался разработкой статистических методов. Это должно было придать импульс исследованиям наследственности и эволюции.
- В 1890 году, основываясь на измерениях десятиногих ракообразных, Уэлдон продемонстрировал, что распределение вариабельности у этих животных примерно соответствует нормальному закону. Это стало первым в биологии применением статистических техник, разработанных Гальтоном в сфере антропологии. Впервые был вычислен коэффициент корреляции между размерами двух органов. Уэлдону потребовалась помощь математика.
- Уэлдон и Пирсон в 1893 году учредили Школу биометрики при прямом участии Гальтона. Термин **биометрия** означает «наука об измерениях живого».
- Школа заложила основы математической статистики между 1895 и 1915 годами.

Математическая статистика

Карл Пирсон

- Пирсон ввёл наглядное представление распределения случайной величины с помощью гистограммы, ввёл и исследовал понятия **стандартного отклонения**, **коэффициентов асимметрии и эксцесса** (островершинности) распределения.
- Предложил **метод моментов** для оценивания параметров распределения, отличного от нормального: метод состоит в сопоставлении теоретических и выборочных моментов.
- Ввел **коэффициент вариации**.



$$c_v = \frac{\sigma}{\mu}$$

$$\sigma = \sqrt{E\left[\; x-\mu \;^{2} \right]}$$

$$\gamma_1 = \frac{E\left[\; x-\mu \;^{3} \right]}{\sigma^3}$$

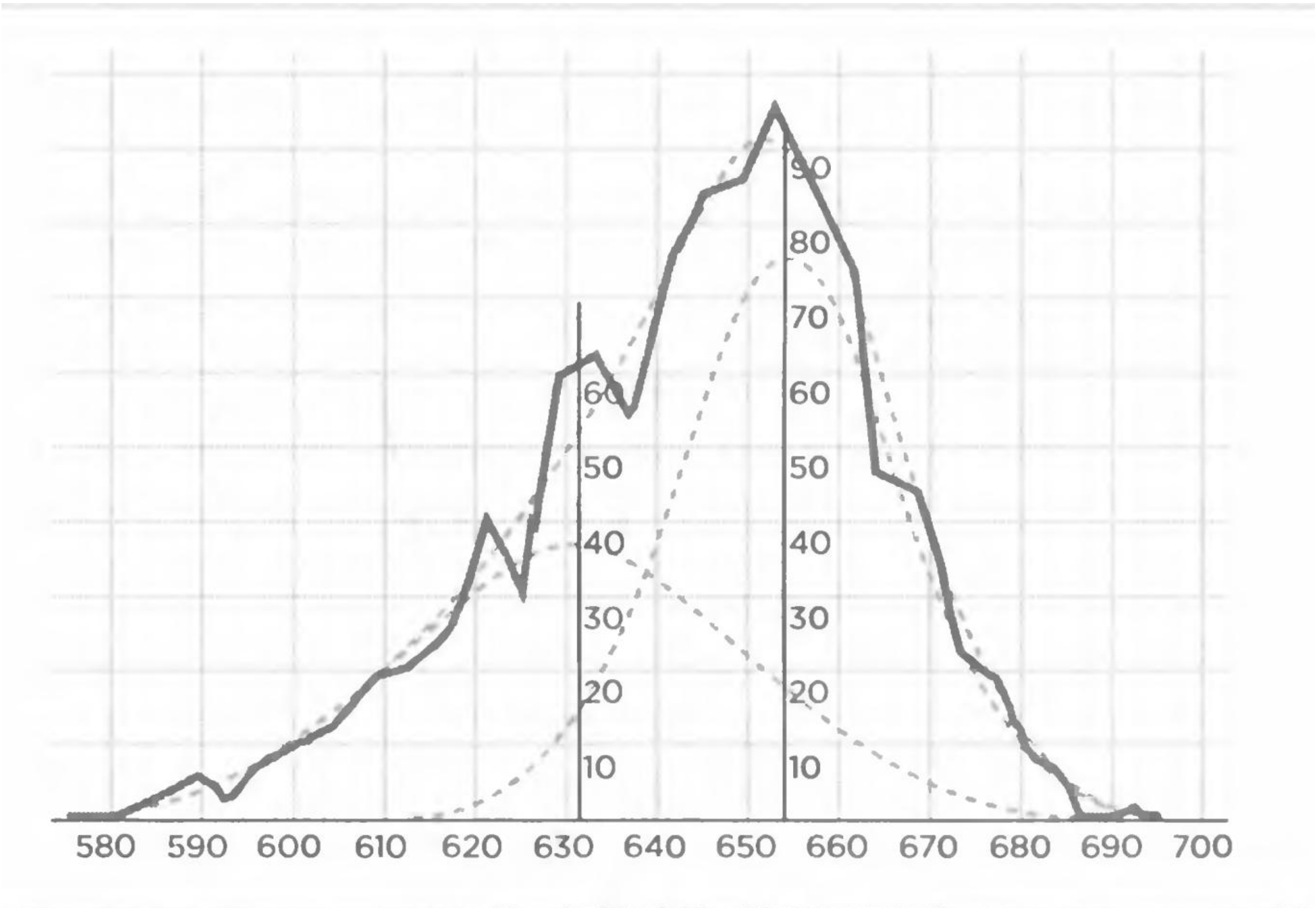
$$\gamma_2 = \frac{E\left[\; x-\mu \;^{4} \right]}{\sigma^4} - 3$$

Математическая статистика

Карл Пирсон

- Уэлдон попросил у Пирсона совета при анализе данных по размерам крабов (диаметр панциря, длина клешней и так далее), которые получил, проводя отпуск в бухте Неаполитанского залива. Наблюдения не подчинялись нормальному закону. Их распределение не было симметричным: вместо одного пика, как у нормального распределения, выделялись два горба.
- Пирсон помог Уэлдону разделить распределение на два нормальных компонента, согласно логике Гальтона (то есть все распределения были нормальными или суммой нормальных), и тот пришел к выводу, что перед ним два разных вида крабов, которых он по незнанию относил к одному, или же один вид, но находящийся в процессе разделения на два разных вида.
- Но Пирсон хотел найти способ интерпретировать данные без принудительной подгонки к нормальному распределению, без нарушения формы кривой частот. Он не желал отбрасывать возможность асимметрии в исходных данных.

Математическая статистика

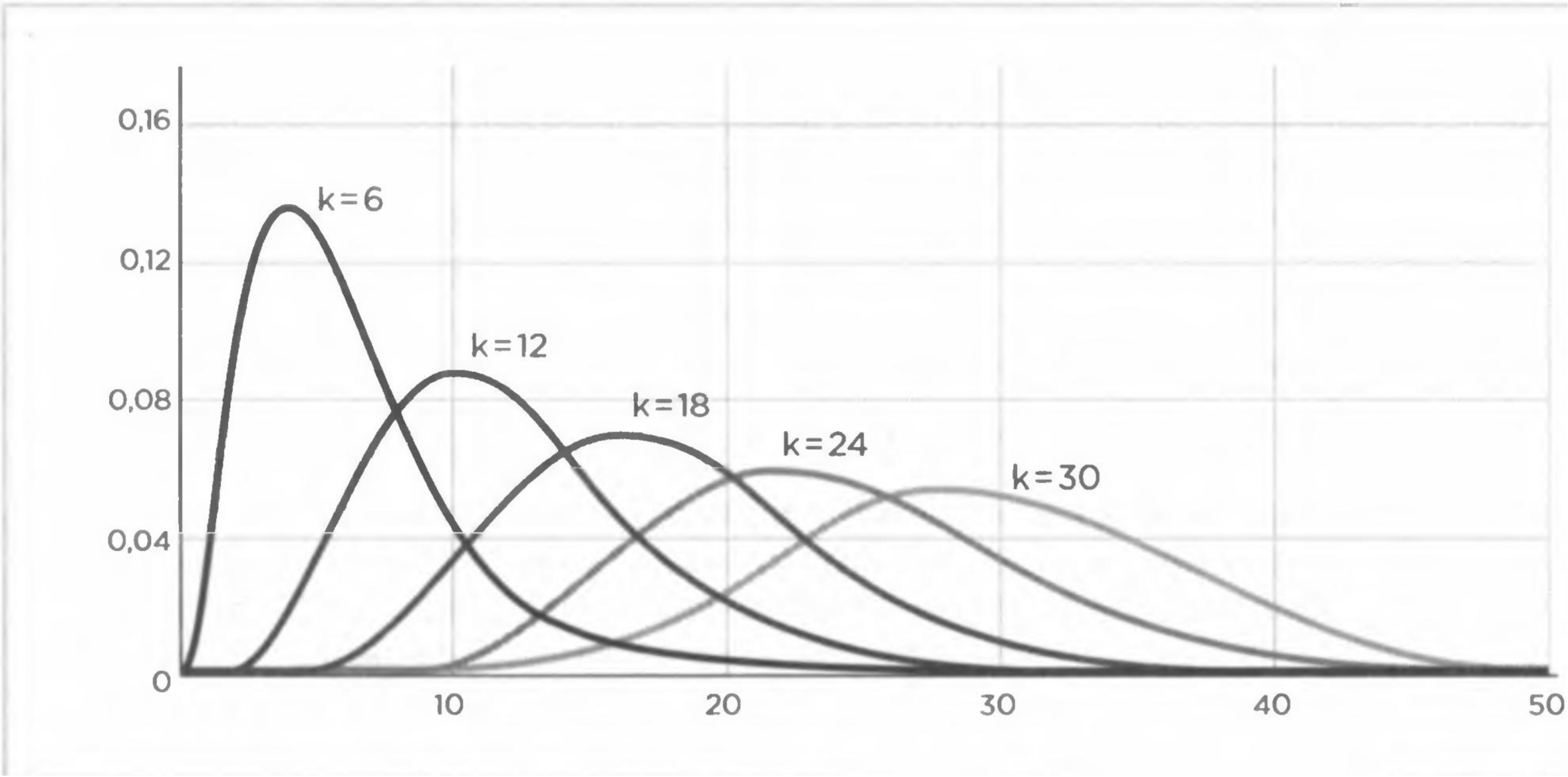


Математическая статистика

Карл Пирсон

- В 1894 году в своей первой публикации о статистике Пирсон обозначил целую систему частотных кривых, которые можно было использовать в биологических исследованиях, как альтернативу нормальному распределению (бета- и гамма-распределения, распределение $\chi 2$).
- Пирсон разработал **метод моментов**, позволявших оценивать параметры, определяющие каждую кривую, на основе имеющихся данных.
- В частности, оценка начиналась с вычисления четырех моментов, связанных соответственно со средним, стандартным отклонением, асимметрией и эксцессом.
- На следующем слайде показано **распределение хи-квадрат**, при увеличении числа степеней свободы оно стремится к нормальному.

Математическая статистика



Математическая статистика

Карл Пирсон

- Пирсон окончательно ввёл в науку понятие **корреляции** как вероятностный аналог причинно-следственной связи, но он же первым предупредил, что корреляционная связь шире, чем причинно-следственная, и, вообще говоря, доказанная корреляция двух факторов не означает, что один из факторов является причиной другого (например, они оба могут быть следствием третьего фактора). Эта ошибка была типична для статистики второй половины XX века.
- **Критерий хи-квадрат Пирсона** стал незаменимым средством для решения нескольких задач — проверка согласия реального и предполагаемого распределения случайной величины, проверка однородности разных выборок или независимости факторов.
- До изобретения компьютеров неоценимую помощь специалистам оказывали составленные Пирсоном таблицы типовых распределений.

Математическая статистика

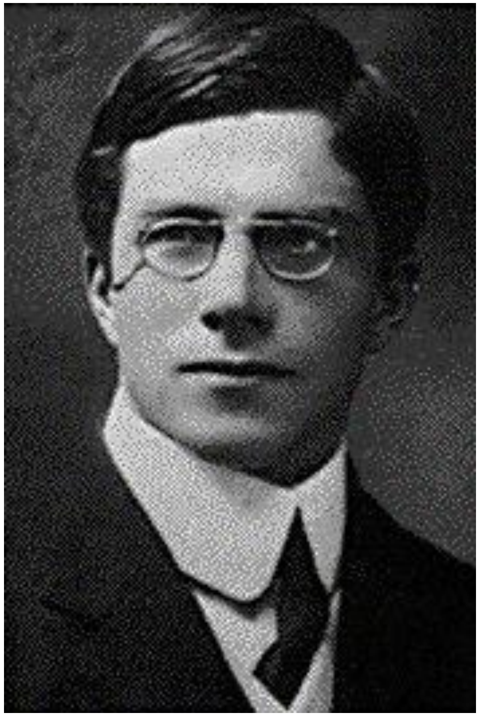
Карл Пирсон

- Чтобы содействовать внедрению математических методов в биологию, в 1900 году Пирсон и Уэлдон основали журнал **«Biometrika»**. Пирсон оставался постоянным редактором этого журнала до конца жизни.
- В 1925 году Пирсон создал ещё один журнал — **«Annals of Human Genetics»,** посвящённый генетике человека.
- В 1892 году опубликован основной философский труд Пирсона **«Грамматика науки»**; книга вызвала большой интерес, была многократно переиздана и переведена на многие языки. В этой книге Пирсон призвал сформировать систему моральных и культурных ценностей общества с научных позиций, отбросив исторические предрассудки. В частности, он отстаивал социализм, дарвинизм, евгенику, защищал принудительную отбраковку умственно отсталых и душевнобольных.

Математическая статистика

Сэр **Роналд Эйлмер Фишер** (17 февраля 1890 — 29 июля 1962) — английский статистик, биолог-эволюционист и генетик. Член Лондонского королевского общества (1929) и Королевского статистического общества, почётный член многих академий и научных обществ; почётный доктор наук и доктор права многих университетов (Лондона, Гарварда, Чикаго, Калькутты, Глазго и других).

- В математической статистике Фишер является виднейшим продолжателем классических работ и методов Карла Пирсона. Большинство методов Фишера имеют общий характер и применяются в естественных науках, в экономике и в других областях деятельности.



Математическая статистика

Роналд Эйлмер Фишер

- *Знакомые генетики спрашивали меня: правда ли, что великий генетик Р. Э. Фишер был также известным статистиком?*

  Леонард Джимми Сэвидж (1976)
- В 1952 году королева Великобритании Елизавета II удостоила Фишера дворянского титула за научные заслуги.
- Биографы учёного отмечают его сложный, противоречивый характер, временами проявлявшийся в грубых высказываниях или письмах, адресованным коллегам, не раз приводивший к ссорам с последними (Карл Пирсон, Эгон Пирсон, Ежи Нейман).

Математическая статистика

Роналд Эйлмер Фишер

- Он верил в наследственность интеллекта и таланта, и поэтому считал, что браки должны заключаться только между индивидуумами одного социально-физиологического уровня. Как следствие, он призывал финансово-политическими мерами повышать рождаемость в высших слоях общества и контролировать в низших, вплоть до принудительной стерилизации.
- В Великобритании эти призывы не имели успеха, однако в США, Германии, Дании законы о стерилизации были приняты и даже просуществовали некоторое время.
- Фишер также выступил против Декларации ЮНЕСКО о расовых предрассудках (*UNESCO Declaration on Race and Racial Prejudice*, 1978).
- Недоумение общественности вызвало и мнение Фишера о том, что между курением и раком лёгких нет связи. Оно было основано на том, что корреляция не означает зависимости одного от другого, что в общем верно.

Математическая статистика

Роналд Эйлмер Фишер. Основные результаты.

- В своей программной статье **«О математических основах теоретической статистики»** представленной в Лондонском королевском обществе в 1921 году и опубликованной в 1922-м, Фишер предложил номенклатуру, сегодня встречающуюся в каждом учебнике по статистическому выводу. Так, здесь впервые появился и был употреблен 57 раз термин «параметр» в его современном статистическом понимании.
- Фишер описывает понятие «статистической модели», позволяющее ясно различать генеральную совокупность (реальную или гипотетическую) и выборку из нее — две связанные сущности, граница между которыми до тех пор была нечеткой.

Математическая статистика

Роналд Эйлмер Фишер

- Фишер указал на три класса математических проблем, с которыми сталкивается статистический вывод.
- В первую очередь это проблемы **«спецификации»,** состоящие в определении модели генеральной совокупности, то есть семейства распределений, зависящих от одного или нескольких параметров Θ, из которого (предположительно) делаются выборки.
- Во вторую очередь это проблемы **«оценки»** — основного направления статистического вывода.
- И на третьем и последнем месте — проблемы **«распределений»**, то есть точное выведение распределения выборочной статистики на основе распределения в генеральной совокупности, считающегося известным. Выборочные распределения задают вероятность, с которой какая-либо статистика принимает значение между двумя заданными границами.

Математическая статистика

Роналд Эйлмер Фишер

- Сформулировал требования к оценкам:
    1. **Состоятельность** (оценка стремится к истинному значению параметра с увеличением объема выборки).
    2. **Эффективность** (стандартная ошибка оценки должна быть как можно более малой).
    3. **Достаточность** (статистика не должна терять никакую информацию, содержащуюся в выборке, она должна объединять всю информацию, нужную для оценки соответствующего параметра).
- Фишер показал, что оценки Пирсона по методу моментов состоятельные, но неэффективные.
- Предложил и развил **метод максимального правдоподобия.** В качестве оценки θ выбирается то значение, которое соответствует максимальной **функции правдоподобия** для наблюдаемых значений выборки. Это был принципиально новый метод оценивания параметров.
- Появление этого знаменитого труда Фишера открыло новую эру в статистике.

Математическая статистика

Роналд Эйлмер Фишер

- **«Статистические методы для исследователей»** (1925), на сегодняшний день выдержала 14 переизданий.
- Рассматривает **задачу проверки гипотез** и вводит понятие **уровня значимости критерия**.
- **Статистическая гипотеза** — предположение о виде распределения и свойствах случайной величины, которое можно подтвердить или опровергнуть применением статистических методов к данным выборки.
- H_0 – *нулевая гипотеза*, подвергающаяся проверке.
- H_1 – *альтернативная гипотеза*, противоположная нулевой.
- *Уровень значимости* α – вероятность *ошибки первого рода* (отвергнуть верную нулевую гипотезу).
- *Мощность критерия* $1 - \beta$, β – вероятность *ошибки второго рода* (принять неверную нулевую гипотезу)

Математическая статистика

Роналд Эйлмер Фишер

- Использовал статистики с различными распределениями ($\chi 2$ Пирсона, *t* Стьюдента, F-распределение Фишера-Снедекора).
- *Распределение $\chi 2$* – распределение суммы квадратов независимых нормальных случайных величин.
- *Распределение Стьюдента* – распределение нормальной случайной величины, деленной на корень из усредненной $\chi 2$.
- *Распределение Фишера-Снедекора* – распределение отношения двух независимых усредненных $\chi 2$.
- Американский математик **Джордж Снедекор (1881-1974)** уточнил логарифмическую (лог-нормальную) аппроксимацию, использованную Фишером.

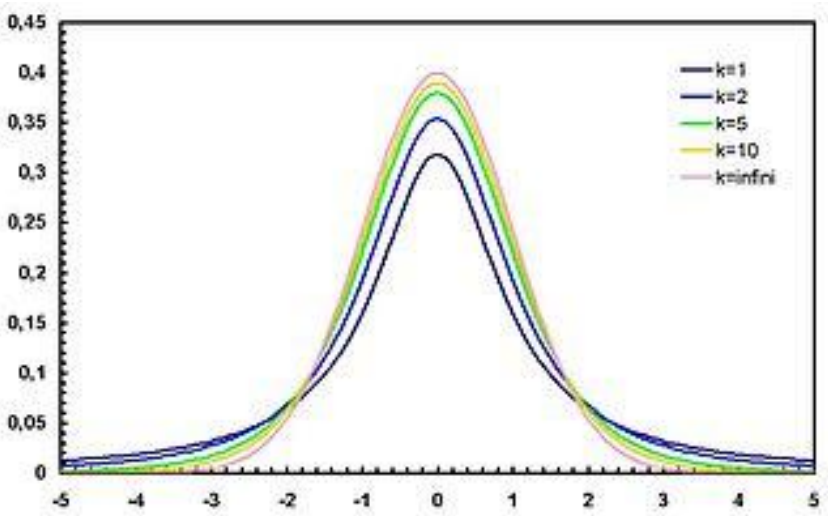
Математическая статистика

Роналд Эйлмер Фишер

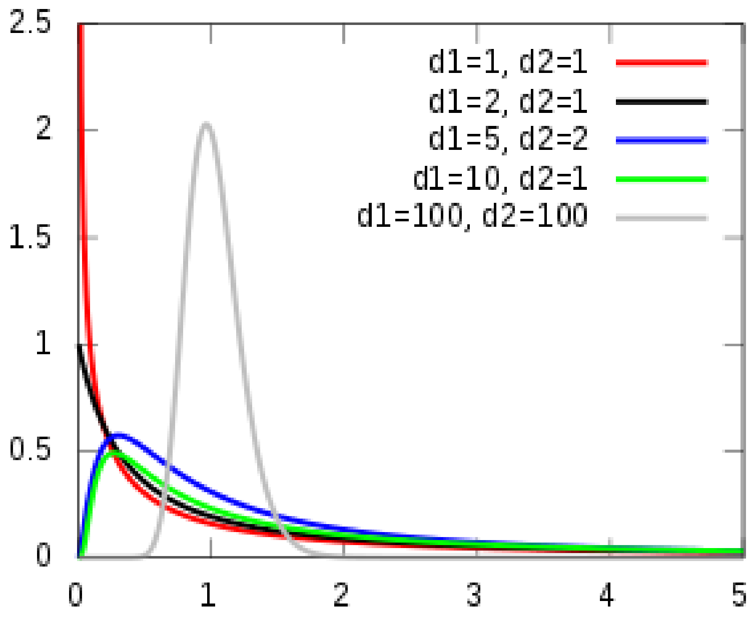
- Распределение Стьюдента берёт название из статьи **Уильяма Госсета** в журнале Пирсона «Биометрика», опубликованной под псевдонимом «Стьюдент». В широкое употребление было введено Фишером.
- Госсет работал в пивоварне Гиннесс в Дублине, Ирландия, и применял свои знания в области статистики как при варке пива, так и на полях — для выведения самого урожайного сорта ячменя. Исследования были обращены к нуждам пивоваренной компании и проводились на малом количестве наблюдений, что послужило толчком для развития методов, работающих на малых выборках.
- Госсету пришлось скрывать свою личность при публикации из-за того, что ранее другой исследователь, работавший на Гиннесс, опубликовал в своих материалах сведения, составлявшие коммерческую тайну компании, после чего Гиннесс запретил своим работникам публикацию любых материалов, независимо от содержавшейся в них информации.

Математическая статистика

Распределение Стьюдента



Распределение Фишера



Математическая статистика

Роналд Эйлмер Фишер

- **Дисперсионный анализ** —ANOVA (англ. ANalysis Of VAriance).
- Цель дисперсионного анализа – исследование наличия или отсутствия существенного влияния какого-либо качественного или количественного фактора на изменения исследуемого результативного признака. Этот метод позволяет сравнивать несколько (более двух) выборок по признаку, измеренному в метрической шкале.
- Изучается отличие между средними значениями признаков в группах на основе вычисления внутригрупповых и межгрупповых дисперсий.
- Успех «Статистических методов для исследователей» обозначил конец эпохи корреляций и подгонки кривых. До Фишера статистики посвящали большую часть своих усилий вычислению коэффициентов, следуя примеру Карла Пирсона, но не проверялось качество этих оценок. Кроме того, были разработаны методы для малых выборок.

Математическая статистика

Роналд Эйлмер Фишер

- **«Планирование экспериментов» (1935).**
- Статистика, по Фишеру, необходима для получения ответа на вопросы вида «Какое удобрение лучше?», «Какое лекарство эффективнее?» и так далее.
- **Выборочная техника** состоит в исследовании репрезентативной выборки из генеральной совокупности и измерении изучаемых показателей. **Планирование экспериментов**, наоборот, заключается в фиксации одних параметров и наблюдении за другими, с измерением возникающих изменений.
- *Спросить мнение у статистика после того, как сделаны выводы по эксперименту, — зачастую все равно что просить его сделать посмертную экспертизу. Может, у него получится установить причину смерти эксперимента.*

  Выступление Фишера на Первом индийском статистическом конгрессе (1938).

Математическая статистика

Роналд Эйлмер Фишер

- Статистический анализ стал ежедневной реальностью в эконометрике, метеорологии, эпидемиологии (биостатистике), промышленной инженерии (контроль качества) и других областях.
- Рост количества областей исследования, кафедр, книг и специальных журналов стимулировало также распространение пакетов статистический прикладных программ.
- Обнаружение знаменитого бозона Хиггса в июле 2012 года не обошлось без проверки гипотез: физики сообщили, что вероятность получения наблюдаемого эффекта в ускорителе частиц исходя из предположения, что это фоновый шум (нулевая гипотеза), составляет менее 0,0000003, и это значение интерпретируется как серьезное доказательство существования данной частицы.

Математическая статистика

- **Ежи Нейман** (при рождении **Юрий Чеславович Нейман**; 16 апреля 1894, Бендеры, Бессарабская губерния, Российская империя — 5 августа 1981, Окленд, Калифорния, США) — польский и американский математик и статистик.
- Член Национальной академии наук США (1963), иностранный член Лондонского королевского общества (1979).
- Юрий Нейман родился в уездном бессарабском городе Бендеры в польской католической семье. В 1909 году окончил Каменец-Подольскую губернскую мужскую гимназию. Учился в Харьковском университете. Окончил Варшавский университет (1923). В 1923—1934 годах преподавал в Варшавском и Краковском университетах колледжа Лондонского университета. В 1927 году организовал в Варшаве лабораторию биометрики. С 1938 года — профессор Калифорнийского университета.

Математическая статистика

Ежи Нейман. Основные результаты.

- Ввел понятие **доверительного интервала** (интервал, в котором неизвестный параметр находится с заранее заданной вероятностью.
- Развивал **бихевиористскую статистику** (методы принятия решений в условиях неопределённости), которая с успехом применялась исследователями в астрономии, физике, биологии, медицине — везде, где необходимо было снижать частоту ошибок.
- **Критерий Неймана-Пирсона (Эгон Пирсон, сын Карла Пирсона).** Критерий проверки гипотез. при заданном уровне значимости даёт наибольшую мощность решения по сравнению со всеми другими правилами.
- Во время Второй мировой войны Ежи Нейман принимал участие в разработке американской атомной бомбы. Многие аспекты его научных изысканий в тот период до сих пор засекречены. Также он работал над прицелами бомбардировщиков и системами наведения.

Математическая статистика

Абрахам Вальд (31 октября 1902, Клуж-Напока, Австро-Венгрия — 13 декабря 1950, Траванкор, Индия) — венгерский математик и статистик. В сферу его научных интересов входили теория принятия решений, эконометрика, геометрия, математическая статистика и теория вероятностей.

- Родился в религиозной еврейской семье. Получил домашнее образование под руководством родителей.
- Продолжил образование в Венском университете, в 1931 году стал доктором философии по математике.
- Был вынужден эмигрировать в США.
- В 1950 г. по приглашению индийского правительства читал лекции, погиб в результате авиационной катастрофы в горном массиве Нилгири.

Математическая статистика

Абрахам Вальд – основатель статистического последовательного анализа.

- **Статистический последовательный анализ** — раздел математической статистики, изучающий статистические методы, основанные на последовательной выборке, формируемой в ходе статистического эксперимента.
- Наблюдения производятся по одному (или, более общим образом, группами) и анализируются в ходе самого эксперимента с тем, чтобы на каждом этапе решить, требуются ли ещё наблюдения (*решение о продолжении эксперимента*) или наблюдений уже достаточно (*решение об остановке эксперимента*). Когда эксперимент остановлен, заключительное статистическое решение принимается на основе всех наблюденных в эксперименте данных.
- **Объём последовательной выборки** (общее число наблюдений, используемое для принятия статистического решения) является случайной величиной.

Математическая статистика

Абрахам Вальд

- **Последовательная проверка статистических гипотез**
- Существенным в последовательной проверке гипотез является то, что количество наблюдений, необходимых для принятия решения, зависит от исхода самих наблюдений и является случайной величиной.
- Пусть установлено некоторое правило, в соответствии с которым на каждом этапе эксперимента принимается одно из трех решений:
  - принять гипотезу;
  - отклонить гипотезу;
  - продолжить эксперимент и провести дополнительное наблюдение.
- **Критерий отношения вероятностей Вальда.** Решающая статистика – отношение вероятностей наблюдений при условии нулевой гипотезы к альтернативной.

Математическая статистика

Абрахам Вальд

- **Систематическая ошибка выжившего**
- Во время Второй мировой войны командование американских и британских ВВС поручило ученым выяснить, какие части фюзеляжа самолета нужно защитить дополнительной броней. У самолетов, возвращавшихся с боевых вылетов, отмечались места попаданий.
- Вальд рекомендовал установить дополнительную защиту на те участки (центральную и заднюю части фюзеляжа), где количество пробоин было **минимальным**. Рекомендация была основана на выводе, что защищать нужно от тех попаданий, которых Вальд не видел, — самолеты, которые их получили, не возвращались.

Математическая статистика

Непараметрическая статистика — раздел статистики, который не основан исключительно на параметризованных семействах вероятностных распределений (математическое ожидание, дисперсия и т.д.).

- Отрицание, содержащееся в названии этого направления, имеет исторические корни: в прошлом (30-е годы 20 века) оно возникло как альтернатива господствовавшей тогда системе обработки данных, основанной на гауссовском (нормальном) распределении.
- **Проверка статистических гипотез** (критерии Колмогорова-Смирнова, ранговые критерии Уилкоксона и Манна-Уитни, коэффициенты ранговой корреляции Спирмена и Кендалла).
- **Робастная статистика** (robust — крепкий, грубый, дюжий) объединяет разработки в области изучения устойчивости методов по отношению к отступлениям от предположений статистической модели.

Математическая статистика

Андрей Николаевич Колмогоров и **Николай Васильевич Смирнов**.

- *Критерий Колмогорова-Смирнова* – непараметрический критерий согласия, в классическом понимании предназначен для проверки простых гипотез о принадлежности анализируемой выборки некоторому известному закону распределения. Наиболее известно применение данного критерия для проверки исследуемых совокупностей на нормальность распределения.
- Основа – статистика Колмогорова, эмпирическая функция распределения.

Математическая статистика

Ранговые критерии

- Ранги – это порядковые номера единиц совокупности в упорядоченном ряду.
- Ранговый коэффициент корреляции Спирмена.
- n – число наблюдений
- d_i – разница рангов по двум переменным.
- https://100task.ru/sample/38.aspx

$$r_s = \frac{1 - 6\sum_{i=1}^{n} d_i^2}{n^3 - n}$$

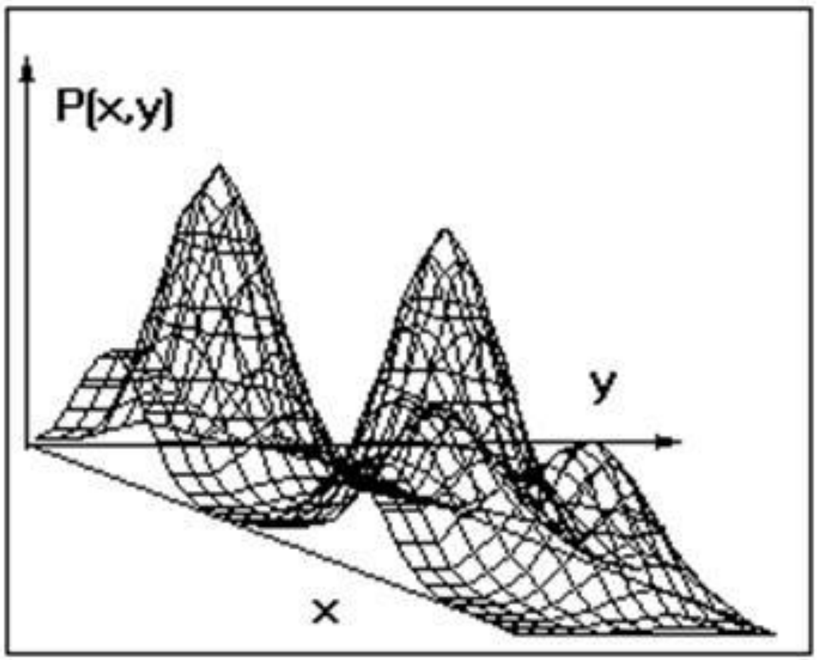
x	y	rx	ry	d
26	4.7	8	10	–2
22	4.4	7	9	–2
8	3.8	1	4	–3
12	3.7	3	3	0
15	4.2	4	7	–3
30	4.3	9	8	1
20	3.6	3	2	2
31	4	10	6	4
10	3.1	2	1	1
17	3.9	5	5	0

Математическая статистика

Ядерная оценка плотности (оценка Розенблата-Парзена)

- $K()$ – функция ядра, обладающая определенными свойствами;
- n – число наблюдений;
- x_i – наблюдения;
- h – параметр.

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)$$



Математическая статистика

Непараметрическая статистика

- В настоящее время непараметрические методы (в первую очередь, ранговые) образуют систему обработки статистических данных, по своим возможностям не уступающую классическому методу наименьших квадратов.
- **Достоинством** непараметрических методов является **широта** их применимости, **устойчивость** статистических выводов относительно ошибок, неточностей модели и т.д., математическая **простота** большей части статистических правил.
- **Недостаток** — как правило, требуют большого числа вычислений (упорядочивание, хранение всей выборки).
- **Приложения** непараметрических методов чрезвычайно распространены в экспериментальной и социальной психологии, а через них — в маркетинге, социологии, теории надежности, в политических исследованиях, в планировании, изучении рисков, анализе категоризированных (классифицированных) данных.

Литература

- *Вилейтнер Г.* . История математики от Декарта до середины XIX столетия. — М.: ГИФМЛ, 1960. — 468 с.
- *Гнеденко Б. В.* К истории основных понятий теории вероятностей // История и методология естественных наук. — М.: Изд. МГУ, 1986. — Вып. XXXII. Математика, механика. — С. 81—88.
- Математика XIX века. Математическая логика, алгебра, теория чисел, теория вероятностей. Том I / Под ред. А. Н. Колмогорова, А. П. Юшкевича. — М.: Наука, 1978. — 255 с.
- *Майстров Л. Е.* . Теория вероятностей. Исторический очерк. — М.: Наука, 1967. — 321 с.
- История математики. Т. II. Математика XVII столетия / Под ред. А. П. Юшкевича. — М.: Наука, 1970. — 301 с.
- История математики. Т. III. Математика XVIII столетия / Под ред. А. П. Юшкевича — М.: Наука, 1972. — 496 с.
- *Никифоровский В. А.* . Вероятностный мир. — М.: Наука, 1992. — С. 48. — (История науки и техники).
- *Стройк Д. Я.* . Краткий очерк истории математики. — Изд. 3-е. — М.: Наука, 1984. — 285 с.
- *Шейнин О. Б.* Теория вероятностей до П. Л. Чебышёва // Историко-математические исследования. — М.: Наука, 1978. — № 23. — С. 284—306.
- Возможно да, возможно нет. Фишер. Статистический вывод. / Пер. с итал. — М: Де Агостини, 2015. — 176 с.