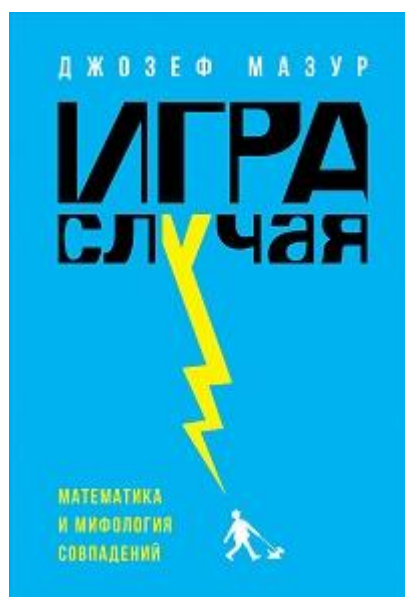


## Джозеф Мазур. Игра случая

Что есть случайность? Этим вопросом мы задаемся, сталкиваясь с неожиданными и, казалось бы, невозможными совпадениями. Однако с математической точки зрения шансы многих событий гораздо выше, чем любой из нас мог бы подумать. В книге «Игра случая» математик Джозеф Мазур открывает необыкновенный мир теории вероятностей, описывая сложные математические понятия простым, веселым языком. Могут ли присяжные быть абсолютно уверенными в результатах анализа ДНК, найденного на месте преступления? На многих примерах реальных событий Мазур показывает нам неотвратимость случайных событий. Эта книга понравится всем, кто когда-либо задавался вопросом, каким образом маленькие решения, которые мы принимаем в течение жизни, складываются в невероятное целое.

Джозеф Мазур. Игра случая. Математика и мифология совпадений – М.: [Альпина нон-фикшн](#), 2017. – 292 с.



Купить цифровую книгу в [ЛитРес](#), бумажную книгу в [Ozon](#)

### Глава 5. Дар Бернулли

Математическое понятие случайного было впервые описано в «Книге об азартных играх» (Liber de Ludo Aleae) Джироламо Кардано — сборнике работ, содержащих основы понимания природы случайности и того, что мы сейчас называем вероятностью; книга была написана около 1563 г. Но «Книга об азартных играх» оставалась неизданной еще сто лет (подробнее см. [Джироламо Кардано. О моей жизни](#)).

«Книга об азартных играх» — это 15 страниц бессвязных математических и философских заметок. Кардано не собирался ее публиковать. Но в книге мы находим полезные инструменты для изучения частотности совпадений. Она считается краеугольным камнем теории вероятности, расчетных величин, средних величин, таблиц распределения, свойств сложения вероятностей и различных способов вычисления  $k$  успешных испытаний из  $N$  — общего числа испытаний. В ней даже содержалось предположение о существовании математического закона, который позже станет известен как *слабый закон больших чисел*.

Галилей написал небольшой трактат об игре в кости, хотя маловероятно, что Галилей знал о «Книге об азартных играх» Кардано. Он перечислил все комбинации и обнаружил, что для трех игровых костей существует 27 различных способов получить в сумме 10 или 11, но только 25 способов получить 9 или 12.

В 1654 г. философ-математик Блез Паскаль прочел экземпляр «Книги об азартных играх» и вступил в переписку со своим другом, юристом и математиком Пьером Фермаб. Вместе они пришли к выводу, что шансы выбросить две шестерки немного меньше, чем 1 к 1, при 24 попытках и немного больше при 25 (см. [Альфред Реньи. Письма о вероятности: письма Паскаля к Ферма](#)).

Швейцарский математик Якоб Бернулли умер в 1705 г. Он оставил кипы неоконченных и неопубликованных рукописей своему племяннику Николаю Бернулли. В течение следующих восьми лет Николай разбирался в бумагах своего дяди и наконец опубликовал «Искусство предположений» (Ars Conjectandi).

«Искусство предположений» Бернулли дает нам слабый закон больших чисел. Если подбросить правильную монету  $N$  раз в надежде, что орел выпадет  $k$  раз, теорема говорит о вероятности того, насколько близко будет отношение  $k/N$  к  $1/2$ .

Математическое ожидание и стандартное отклонение — колеса и винтики статистики частотного распределения. показывающей насколько приближаются данные к некоему центральному значению. В 1657 г. голландский математик и астроном Христиан Гюйгенс опубликовал работу «О расчетах в азартных играх» (De Ratiociniis in Aleae Ludo), которая еще полвека оставалась главным учебником по теории вероятностей. Это первая из напечатанных работ, указывающая на отличие между числом успешных исходов и возможным числом успешных исходов.

## Глава 6. Длинная серия орлов

Согласно данным Всемирной организации здравоохранения, доля рождения мальчиков к общей рождаемости по всему миру составляет 0,515. Если рассмотреть данные по конкретным регионам или странам, то шансы далеки от равных. В Мексике доля новорожденных мальчиков очень низкая, тогда как в США и Канаде их доля выше 0,52. Однако для всего населения Земли — а оно уже больше 7 млрд — шансы рождения мальчиков по отношению к девочкам почти равны. Причина проста: у человеческого сперматозоида равное число  $X$  и  $Y$  хромосом, и у каждой из них равные шансы в момент зачатия. Это бросок правильной монеты.

Вероятность падения правильной монеты орлом вверх —  $1/2$ . Благодаря математике мы знаем, что по мере увеличения числа бросков монеты отношение орлов к решкам постепенно приближается к 1. Однако с увеличением числа бросков  $N$ , размер серий (число подряд выпавших орлов) возрастает. Например, при  $N = 1\,000\,000$  вероятность выпадения 20 орлов подряд более 50%.

## Глава 8. Задача о дне рождения

Есть по крайней мере две математические модели, которые дают нам надлежащие способы оценки совпадений. Одна из них — задача о дне рождения. Она также является моделью для осмысления совпадений — возможно, лучшей из имеющихся. В стандартном виде задача формулируется таким образом: в группе из  $N$  случайно выбранных людей насколько велико должно быть  $N$ , чтобы шансы на то, что у 2 людей в этой группе совпадают дни рождения (число и месяц), были выше, чем  $1$  к  $1$ ? Ответ:  $N = 23$ , удивительно малое число.

Пусть  $p(N)$  обозначает вероятность того, что у  $N$  человек дни рождения не совпадают. Сначала предположим, что  $N = 2$ . Тогда  $p(2) = 365/365 \times 364/365$ , потому что любой из двоих людей может быть рожден в любой из 365 дней, исключая при этом один день для другого человека. Эта  $p(2)$  очень-очень близка к 1. Что неудивительно. Далее: предположим, что  $N = 3$ . По той же причине, что и в случае с  $N = 2$ , день рождения третьего человека не может совпадать с днями рождения двух других, т.е.  $p(3) = 365/365 \times 364/365 \times 363/365$ . Продолжая таким образом, мы видим, что  $p(N)$  сокращается по мере того, как  $N$  увеличивается. В определенный момент мы дойдем до  $N = 23$  и произведем следующие расчеты:

$$p(23) = 365/365 \times 364/365 \times 363/365 \times \dots \times 343/365 = (1/365)^{23} \times (365 \times 364 \times 363 \times \dots \times 343) = 0,4927$$

$p(23)$  (вероятность того, что у двух людей в группе из 23 человек совпадают дни рождения) равняется 0,4927. Переведем отрицание в утверждение и найдем вероятность того, что у 2 людей в группе из 23 человек совпадают дни рождения, равной 0,5073 — шанс выше, чем  $1$  к  $1$ .

## Глава 9. Громадный мир

Скрытые переменные внушают нам ложную мысль о том, что причины либо нет, либо ее слишком сложно найти. Иногда кажется, что у двух абсолютно независимых переменных появляется статистическая связь через третью переменную. Когда такое происходит, мы обнаруживаем иллюзорную корреляцию, вызванную тем, как мы видим данные или как эти данные организованы. Например, мы можем обнаружить корреляцию между доходами во взрослой жизни и отметками в колледже. Мы можем сделать ошибочный вывод, что доход во взрослой жизни зависит от оценок,

которые человек получал в школе, тогда как в действительности скрытой переменной был объем работы, который ученик был готов усердно выполнять.

До последнего времени на Новых Гебридах — группе островов на юге Тихого океана — было распространено убеждение, что вши полезны для здоровья. В течение веков старейшины случайно замечали, что у здоровых местных жителей были вши, а у больных их не было. Они пришли к выводу, что вши положительно влияли на здоровье. В условиях более тщательного и контролируемого исследования было замечено, что вши были почти у всех туземцев большую часть времени. Вши также могли вызывать лихорадку, что, в свою очередь, вызывало гибель вшей. Путаница возникла из-за того, что у нездоровых людей случалась лихорадка и поэтому у них не было вшей. «Вот где причина и следствие искажены до неузнаваемости, перевернуты и перемешаны», — написал Даррелл Хафф в своей книге [Как лгать при помощи статистики](#), которой сейчас уже больше 60 лет, но она все еще остается бестселлером.

Наука любит прямые связи между причинами и следствиями, но не обязывает нас знать о существовании таких связей. Ученые могут допускать корреляцию между двумя сложными феноменами. Настоящая проблема в том, что люди склонны видеть связь там, где ее нет, а также игнорировать имеющиеся связи, которые слишком сложны для того, чтобы можно было их прогнозировать. Мы видим в совпадениях события, таинственным образом предусмотренные неким глубокомысленным замыслом. Может, и так, а может, и нет. В этом сложном мире взаимосвязанных феноменов некоторые связи сцеплены столь искусными и длинными цепочками опосредованных взаимоотношений, что мы даже представить не можем влияние одного на другое.

## Глава 10. К вопросу об историях

Совпадения — это выдающиеся события, которые возбуждают у нас интерес к вероятности. Никто не сомневается, что они чрезвычайно редки, но насколько редкой должно быть событие, чтобы сжать мир во времени и пространстве? Следующие истории в самом деле редкие, однако вполне могут происходить.

**История 2. Джек Фрост и другие истории.** Энн Парриш, находясь в Париже в 1929 г., солнечным воскресным утром после посещения мессы в Нотр-Дам и птичьего рынка вместе с мужем, промышленником Чарльзом Альбертом Корлиссом, отправились пообедать в Les Deux Magots. Оставив Чарльза допивать вино, она прошла по букинистическим лавкам вдоль Сены. Рыться в развалах книг на длинных столах для нее было обычным делом. В тот день она нашла книгу Хелен Вуд «Джек Фрост и другие истории». Немного поторговавшись с продавцом и уплатив один франк, она поспешила к мужу, который все еще сидел над бокалом вина, и вручила книгу ему со словами, что в детстве очень ее любила. Он медленно полистал страницы. Некоторое время молчал, потом вернул ей книгу, открыв форзац, где «нескладными детскими каракулями было написано: «Энн Парриш, Норд-Вебер стрит, 209 Колорадо Спрингс, Колорадо». Это была именно та книга — из детства.

Если ничего не знать о жизни Энн Парриш, ее история кажется поразительной. Но давайте соберем мозаику. Ее мать, имя которой также Энн, но называли ее Ани, изучала живопись в Пенсильванской академии изящных искусств в 1860 г.; тогда же там обучалась Мэри Кэссэт. Во время учебы в Пенсильванской академии Ани и Мэри стали близкими подругами. Мэри стала известной импрессионисткой и переехала в Париж. Тогда возможно ли, что Ани передала книгу своей хорошей подруге Мэри, а та забрала ее с собой в Париж? Мэри умерла в 1926 г. Ее имущество, вероятно, было распродано, как и библиотека, и американская книга Энн Парриш, «вероятно», оказалась на прилавке одного из парижских книжных киосков где-то между 1926 и 1929 гг., до того как Энн Парриш ее нашла.

Давайте выдвинем некоторые разумные предположения. Сначала предположим, насколько вероятно была поездка Энн в Париж летом 1929 г. Я бы дал этой вероятности умеренное значение, близкое к 0,1. Энн была замужем за богатым промышленником. Париж в 1929 г. был одним из самых популярных европейских туристических направлений среди богатых американцев. Какова вероятность того, что она посетила бы книжные киоски, пока была в Париже? Я бы сказал, что вероятность этого 0,3. Сложнее всего установить вероятность того, что книга была бы в нужном месте. А вот здесь поможет сопутствующая история — связь между матерью Энн и Мэри Кэссэт, смерть Мэри и несколько мест в Париже, имеющих дело с бывшими в употреблении книгами. Я

предположу, что вероятность будет примерно 0,01. Тогда вероятность подобной истории составит  $p = 0,1 \times 0,3 \times 0,01 = 0,0003$ . Да, есть много неучтенных скрытых переменных, усложняющих наши расчеты, но они не изменили бы значение вероятности больше чем на 1/10 000, а следовательно, шансы Энн Парриш все же немного выше, чем получить каре при игре в покер.

## Глава 11. Доказательство

Люди любят истории о совпадениях и думают, что они редки. Когда некоторые из них становятся присяжными в делах, которые могут закончиться смертным приговором, они полагают, что совпадение, способное привести к судебной ошибке, — вещь крайне маловероятная. Аргументы в пользу проб ДНК сильны, особенно для простого люда с весьма поверхностным пониманием того, как работают эти пробы.

До тестов ДНК стандартными инструментами были определение группы крови, серология и дактилоскопия. Эти традиционные методы судебной экспертизы дают очень неточные данные в сравнении с данными проб ДНК. Примерно у 40% американцев I положительная группа крови, а совпадающие отпечатки пальцев во многих уголовных делах не являются убедительным доказательством.

Генотипоскопия сейчас играет важную роль в оправдании неправомерно осужденных. И все же и защита, и обвинение могут использовать результаты тестов ДНК в свою пользу, пытаясь либо убедить присяжных в неопровержимой научной достоверности проведенного анализа, либо подвергнуть критике процессы сбора и хранения проб.

Давайте вкратце вспомним кое-что о геноме человека — генетической информации, закодированной в парах хромосом, находящихся в ядре каждой клетки человеческого тела. Хромосома — это набор молекул ДНК в ядре клетки. У человека 23 пары хромосом (22 пары плюс две половые хромосомы), т. е. 23 хромосомы из набора матери и 23 — из набора отца. Как только мы поймем, что наследование генетической информации — вопрос куда более сложный, чем текст, приведенный далее на нескольких страницах, то сможем составить для себя достаточно точную картину того, как можно установить личность человека по его или ее ДНК. ДНК — это сокращение от названия химического соединения дезоксирибонуклеиновая кислота, которое находится в живых клетках. Структура ДНК представлена в виде винтовой лестницы, двойной винтовой спирали.



### Структура двойной спирали

Ступени составлены из основанных на азоте соединений, называемых нуклеотидами, или основаниями: аденин, гуанин, тимин и цитозин, для простоты обозначаемые буквами А, Г, Т и С. Две спиралевидные нити, состоящие из соединенных молекул сахара и фосфата, формируют боковые

поверхности лестницы. Каждая ступень — соединение нуклеотидов от каждой из этих двух нитей. Сочетание букв определяет генотип человека или его генетический профиль.

Чтобы описать последовательность ДНК, мы сначала рассматриваем короткие tandemные повторы (КТП), являющиеся повторами комбинаций четырех нуклеотидов А, Т, Г и С. Существует  $4 \times 4 \times 4 \times 4 = 256$  возможных комбинаций такой последовательности. Рассмотрим организацию любых четырех последовательностей букв А, Т, Г и С, учитывая, что буквы могут повторяться. Тогда мы получим АААА, или АГТС, или любую другую из 254 комбинаций. У одного человека может быть хромосома с короткими tandemными повторами, которые выглядели бы как АГТТ, АГТТ, АГТТ, а у другого человека могла бы быть хромосома с повторами, которые выглядели бы как АГТТ, АГТТ, АГТТ, АГТТ. И все же у этого другого человека могло бы быть 6 повторов или 12.

Обратите внимание, что у первого человека было всего 3 повтора, тогда как у второго — 4. Это создает намного большую изменчивость в генетической матрице индивида. И если мы добавим сюда тот факт, что человек наследует одну последовательность каждой хромосомы от своей матери, а другую — от отца, вероятность того, что у двух человек среди всего мирового населения (исключая однояйцевых близнецов) будет одинаковая ДНК, близка к нулю, но все же нулю не равна. Чтобы представить, насколько мала и насколько длинна молекула двойной спирали ДНК в одной-единственной клетке, представьте: она заключена в ядре клетки, диаметр которой меньше, чем одна пятидесятитысячная сантиметра, а если ее полностью развернуть, ее длина составит 2 м. Это невообразимо плотная упаковка.

Чтобы понять всю сложность модели, задумайтесь: в каждой из 23 пар хромосом приблизительно 3 млрд последовательностей из 4 нуклеотидов, каждый из которых получен от матери и отца<sup>15</sup>. Без сомнения, это большое число. Проблема в том, что мы не знаем, какие из 3 млрд позиций в последовательности могут отличаться.

Чтобы сличить профили ДНК двух человек со 100%ным соответствием, мы должны были бы сравнить приблизительно 3 млрд пар нуклеотидов, а это нерациональный и очень дорогой процесс. Мы этого и не делаем. Вместо этого мы сравниваем очень небольшую часть, чтобы найти сходство. Если в этой малой доле есть соответствие, мы оцениваем, насколько вероятно то, что соответствие является результатом совпадения. Вопрос, который стоит перед нами, заключается в следующем: насколько мала должна быть эта «небольшая часть», чтобы мы могли уверенно сказать, что соответствие возникло не в результате совпадения?

У судмедэкспертов принято говорить о вероятности случайного соответствия, основанного всего на 13 различных КТП. Таким образом, они утверждают, что могут опознать человека по 13 различным КТП, распределенным по всему геному человека. Иными словами, они считают, что в случайной выборке из 13 КТП на 23 человеческих хромосомах обнаружатся несоответствия. Почему всего 13? Такое решение принято на основании двух соображений: практичности и цены, его объясняют тем, что число КТП на каждой из 13 позиций должно очень сильно отличаться в любой группе людей. Например, на хромосоме 3 один человек мог унаследовать 5 повторов от матери, а другой мог унаследовать 3 повтора от матери и 6 от отца. Во всей популяции некоторые повторы будут очень редкими, а другие — довольно распространенными. Требуется всего одно отличие, чтобы исключить, что ДНК подозреваемого совпадает с образцом, найденным на месте преступления. В одной хромосоме КТП могут и не быть настолько уж редкими. Довольно низкой частотой в популяции может быть, скажем, 0,1.

Но умножьте это на частоты КТП в 13 выбранных хромосомах, и вы обнаружите вероятность соответствия порядка 1 миллиона миллиардов. Но список подозреваемых в совершении преступления — это значительно меньшая группа, чем население всего мира. Поэтому судмедэксперты уверены: нет фактически никаких шансов, что у двух человек будет одинаковый набор копий. Шанс того, что у двух человек имеются одинаковые пары по всем 13 КТП, не равен нулю, но ограниченный группой подозреваемых в преступлении шанс этот так необычайно близок к нулю, что можно допустить, что это фактически ноль.

Другими словами, если профиль ДНК с места преступления и профиль подозреваемого совпадают, то доказательства указывают на вину подозреваемого. С другой стороны, если профили не соответствуют, то доказательства указывают на невиновность подозреваемого. Так работают генотипоскопия и судебная экспертиза. На что бы ни указывали доказательства, следствие должно

учитывать, что случайности, совпадения, человеческий фактор и неизвестные скрытые переменные обычно усложняют простую картину, особенно такую, которая составлена на основе единственного измерения.

### Глава 15. «Сэр Гавейн и Зеленый Рыцарь»

«Некогда один человек потерял бриллиантовую запонку в бескрайних просторах синего моря, — написал Владимир Набоков в своем романе «Смех в темноте». — И вот проходит 20 лет, и в тот же самый день — предположим, в пятницу — он ест большую рыбу, но, увы, никакого бриллианта в ней не обнаруживает. Вот какие совпадения мне нравятся».

Этот отрывок говорит нам о том, что представляет собой совпадение на самом деле. Совпадение — это в первую очередь неожиданность. Только в данном случае неожиданность заключается в отсутствии неожиданности. Неожиданность — фундаментальный структурный элемент повествования, а совпадения, по определению, всегда характеризуются именно неожиданностью. Антропологи утверждают, что с тех пор, как люди овладели языком в достаточной степени, чтобы можно было рассказывать истории, они рассказывали истории. Представители каждой культуры на земле рассказывали истории своим детям. В этих историях могло присутствовать зерно истины, порожденное реальностью, однако оживляет их именно глубина фантазии. Рассказы о легендарных героях особенно часто используют совпадения для описания встреч персонажей.

Совпадение в художественной литературе — не то же самое, что совпадение в реальной жизни. Умышленные или нет, они вызывают когнитивные эффекты, которые могут в ином случае приводить к различным вариантам толкования истории.