Закон арбузной корки
и нормальность ненормальности

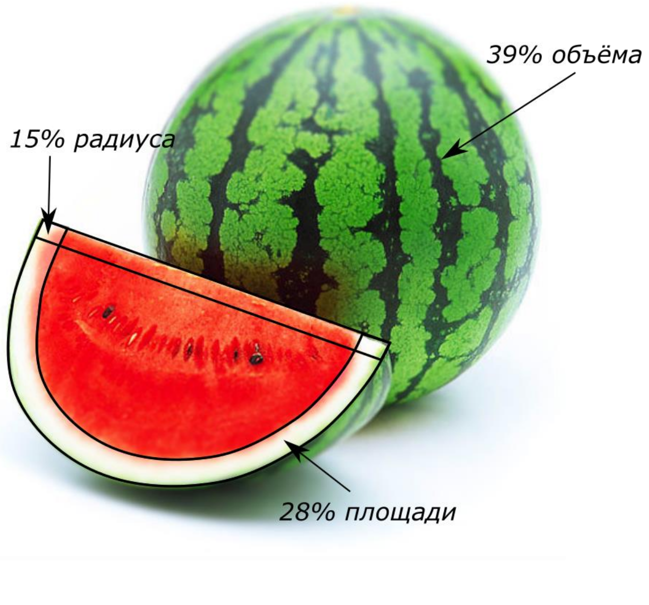
В этой главе мы начнём с анализа арбузов и их корок, выясним их связь со знаменитым законом Мерфи и убедимся со всей математической строгостью в том, что о вкусах не спорят.

Мне одному кажется, что я нормальный?

Глядя новости, или читая комментарии к ним, мы порой недоумеваем: "Есть в этом мире нормальные люди?!" Вроде как, должны быть, ведь нас много, и в среднем, мы и должны быть нормальны. Но при этом мудрецы говорят, что каждый из нас уникален. А подростки уверены, что они-то уж точно отличаются от серой массы "нормальных людей" и ни на кого не похожи.

Читатели, знакомые со статистикой, конечно же, много раз видели как для различных несимметричных распределений мода (максимум на графике плотности вероятности) не совпадает со средним значением или математическим ожиданием. То есть, среднее значение не соответствует самой большой плотности вероятности, но всё равно, на то оно и ожидаемое, чтобы быть если уже и не самым часто встречающимся, то, по крайней мере, доминирующим. Однако, не всё так просто. До сих пор мы рассматривали одновариантные распределения – распределения в одномерном пространстве исходов. Но жизнь многогранна, и уж точно не одномерна! А при добавлении дополнительных размерностей могут случаться весьма неожиданные вещи.

Одна из особенностей многомерной геометрии — увеличение доли пограничных значений в ограниченном объёме. Вот что имеется в виду. Рассмотрим классическую задачу об арбузе в пространствах с различной размерностью и зададимся целью выяснить, сколько же чудесной сахарной мякоти нам достанется от этого огромного, крепкого и аппетитного арбуза, если надрезав его, мы выяснили, что толщина его корки не превышает 15% от его радиуса? Кажется, что это уж больно много, но посмотрите на рисунок, пожалуй, арбуз с такими пропорциями мы сочтём вполне приемлемым. Рассмотрим сначала одномерный арбуз, в виде розового столбика. Его корка представляет собой два маленьких белых отрезочка по краям, суммарная длина корки будет мерой (обобщённым объёмом) в одномерном мире и составит 15% от общей меры арбуза. У двумерного, блинообразного арбуза, мера корка в ви-



Доли, которые занимает корка в арбузе различной размерности.

де площади белого кольца, будет меньше, чем его внутренняя часть, уже всего в три раза. В привычном нам трёхмерном мире, такая корка составит почти 40% общего объёма. Уже чувствуется подвох.

Такую возрастающую роль границ мы уже встречали, когда рассматривали туристический закон подлости. Но тогда мы ограничились двумерным случаем, вполне естественным для топографических карт. Сейчас мы пойдём в нашем исследовании дальше.

Для шара, как, впрочем, и для тела произвольной формы, можно точно вычислить зависимость отношения объёма корки к общему объёму тела. Её легко получить, вновь воспользовавшись анализом размерности, и общим понятием меры. Для сплошного тела в пространстве размерности m, его мера, или обобщённый объём, пропорциональна степенной функции от характерного размера тела d:

$$V \propto d^m.$$

Под *сплошным* я понимаю тело, не являющееся *фрактальным*. Такие объекты отличаются от сплошных именно тем, что их обобщённый объём пропорционален их размеру в некоторой дробной степени, отличной от размерности вмещающего пространства. С примерами фрактальных объектов: множеством Жулиа и губкой Менгера, мы уже встречались в одной из первых глав, когда рассматривали подмножества нулевой меры. Может показаться, что это экзотика, но природа находит фрактальные решения для очень многих задач: от роста кристаллов до разряда молнии, от корневой системы растений до устройства наших с вами лёгких.

С объёмом, как с мерой, мы разобрались в первой главе, а что такое характерный размер? Интуитивно ясно, что человек имеет характерный размер порядка метра, муравей — миллиметра, а галактика в сотню тысяч световых лет. Мы определим это понятие позже через среднее геометрическое, а пока остановимся на интуитивном понимании. Под знаком пропорциональности скрывается константа, которая называется *формфактором*. Она зависит от формы тела, но не зависит от размеров: для куба она равна 1, для сферы того же размера примерно в двое меньше — $\pi/6$ и т. д.

Объём корочки равен следующей разнице:

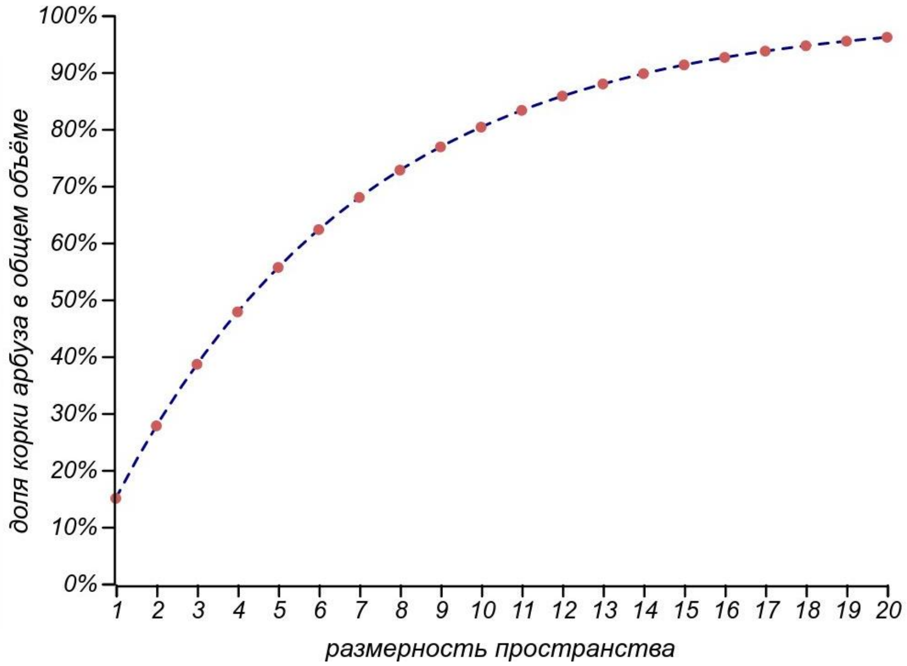
$$V_{\text{корки}} = V_{\text{общ}} - V_{\text{внутр}},$$

а отношение объёма корки, составляющей долю δ от размеров тела, к общему объёму выражается следующим образом:

$$\frac{V_{\text{корки}}}{V_{\text{общ}}} = 1 - \frac{V_{\text{внутр}}}{V_{\text{общ}}} = 1 - (1-\delta)^m.$$

Как хорошо получилось — мы перешли от пропорциональности к точному равенству. Это произошло благодаря отношениям, в которых сократились неизвестные нам формфактор и размеры тела. Таким образом, полученное соотношение объёма корки и объёма тела универсально и годится для арбузов сколь угодно сложной формы.

Вот как выглядит график роста доли пятнадцатипроцентной по радиусу корочки арбуза в его объёме, при дальнейшем увеличении размерности пространства.



В четырёхмерном пространстве наш условно тонкокорый арбуз оставит нам уже лишь половину мякоти, а в одиннадцатимерном мире мы сможем полакомится лишь 15% от всего арбуза, выбросив корочку, составляющую 15% его радиуса!

Итак, сейчас мы готовы сформулировать глубокомысленный закон ***арбузной корки***:

Покупая многомерный арбуз, ты приобретаешь, в основном, его корку.

Обидно, конечно, но какое это имеет отношение к нормальности нашего мира и к законам подлости? Увы, именно он препятствует отысканию так называемой "золотой середины", обесценивает результаты социологических опросов и повышает роль маловероятных неприятностей.

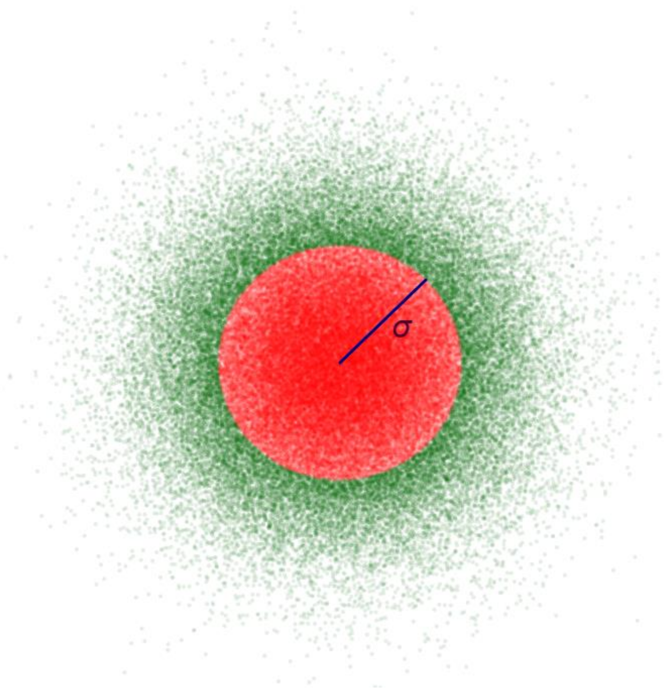
Дело в том, что пространство людей со всеми их параметрами существенно многомерно. Вполне независимыми размерностями можно счесть и очевидные рост, вес, возраст и достаток, а также, уровни интеллектуального (IQ) и эмоционального (EQ) развития, наконец, наблюдаемые, хоть и плохо формализуемые черты лица, либо черты характера, такие как уровень болтливости, упрямства или влюбчивости. Мы без труда насчитаем с десяток-полтора параметров, характеризующих человека. И для каждого из этих параметров существует некая статистически определяемая "норма" — наиболее ожидаемое и, более того, часто наблюдаемое значение. Сколько же в таком богатом пространстве параметров окажется людей, типичных во всех отношениях? Выражение, которое мы использовали для вычисления отношения объёмов корки и арбуза, можно использовать и для вычисления вероятности попасть в число хоть в чём-то но "ненормальных" людей. Действительно, вероятность удовлетворить всем критериям типичности одновременно равна произведению вероятностей оказаться типичным по каждому критерию в отдельности.

Колмогоровское определение вероятности, которое мы ввели в самом начале книжки, сильно упростит задачу, избавив нас от пугающих формул, по которым нельзя ничего толком вычислить. Полученная нами формула арбуза работает для любых, сколь угодно сложных форм, даже не имеющих границы, подобно атмосфере Земли, уходящей далеко в космическое пространство, становясь всё тоньше и тоньше. Так что нам не нужно знать каким именно распределениям подчиняются обсуждаемые качества людей, остаётся лишь предположить, что у этих распределений есть среднее значение (что, как мы увидим, бывает не всегда). Если обозначить за P_{out} вероятность оказаться за пределами области, которую мы сочли бы нормой, то вероятность оказаться в чём-то ненормальным, при рассмотрении m критериев, будет вычисляться по "арбузной" формуле:

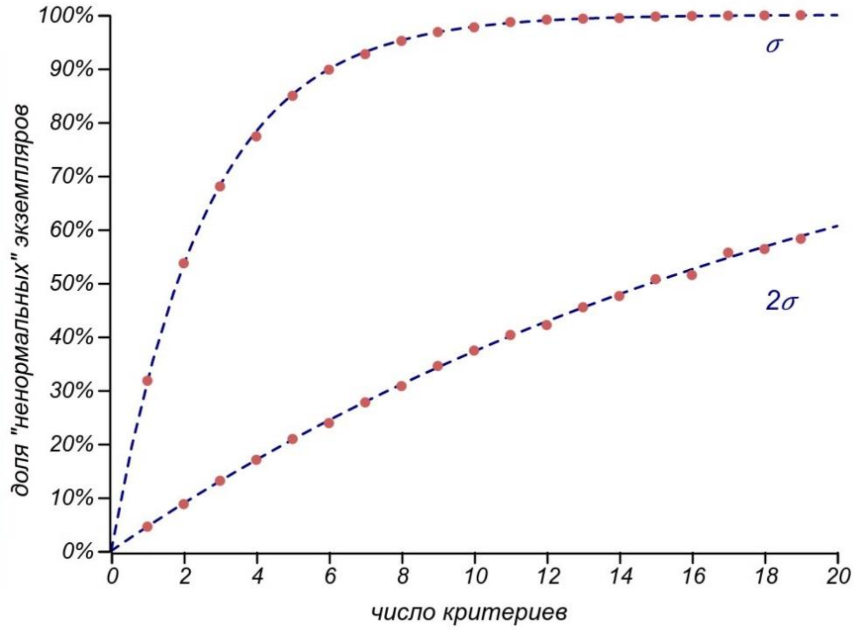
$$P = 1 - (1 - P_{out})^m.$$

Здесь мы вновь пользуемся колмогоровским пониманием вероятности, как меры. Толщину корки арбуза мы измеряли линейкой, попадание случайной величины в какой-нибудь диапазон мы измеряем вероятностью. Какой бы малой ни была вероятность P_{out}, при $m \geq P_{out}/2$, значение P превысит $1/2$.

Для хоть какой-то конкретики можно предположить, что параметры, о которых мы говорим, имеют нормальное распределение. Это вполне разумно для наших целей, ведь мы говорим не о каком-то конкретном наборе характеристик, а, прямо скажем, фантазируем, стараясь сформулировать хоть что-то определённое в столь зыбкой теме. Выбор нормального распределения адекватно отражает степень нашего неведения, и загружаться подробностями, пока не видна самая общая картина, рановато. Итак, наш арбуз превратился в размытое туманное пятно, что не мешает нам вычислить долю его бесконечной по протяжённости «корки». Для хорошего , в известном смысле, распределения за норму можно принять значения, не отклоняющиеся от среднего больше, чем на величину стандартного отклонения. Для нормального распределения доля значений выходящих за пределы нормы имеют $P_{out} = 16\%$, примерно, как в рассмотренном нами реальном арбузе. Применительно к нашему нечёткому арбузу здесь имеется в виду вероятность оказаться на удалении в одно стандартное отклонение от среднего, как показано на рисунке. При более толерантном понимании нормы можно ограничиться двумя стандартными отклонениями, получая $P_{out} = 2.3\%$.



Бесконечный абстрактный арбуз
в двумерном параметрическом пространстве.



Вероятности оказаться "ненормальным" для различного числа критериев сравнения и для различной "строгости" определения нормы. Верхний и нижний графики отличаются тем, что при определении "нормальности" используют радиус в одно и два стандартных отклонения, соответственно.

Что же, выходит, это нормально — быть хоть в чём-то ненормальным. Оценивая людей по десятку параметров, будьте готовы к тому, что полностью заурядными окажутся лишь 2% общей популяции. Причём, как только мы их разыщем, они тут же станут знаменитостями, потеряв свою заурядность!

Нетипичность нормы и ментальные ошибки, к которым может привести попытка усреднения многопараметрических систем, подробно рассматривается в книге Тодда Роуза «The End of Average». В частности, в ней приводится история, случившаяся в начале Второй Мировой войны. В попытке разобраться в причинах ошибок пилотов боевых самолётов командование ВВС США предприняло исследование, основной целью которого было уточнение средних характеристик воздушных бойцов. От этих параметров зависели конкретные инженерные решения по проектированию эргономики кабины, и считалось, что чем точнее будут известны эти характеристики, тем более эргономичным будет разработанная на их основе техника. Каково же было удивление молодого антрополога Гилберта С. Дэниэлса, которому поручили эту работу, когда выяснилось, что из четырёх тысяч обмеренных им пилотов не обнаружилось ни одного «среднего», для которого кабина самолёта оказалась бы удобной по всем измеряемым параметрам. Всего использовалось 10 физических характеристик, и Дэниэлс придерживался очень строгого критерия «нормальности»: выходящим за пределы нормы считалось отклонение параметра от среднего, превышающее 30% от всей выборки. Для десяти параметров вероятность попасть в нормальные значения по таким критериям составит 0.0006% — 1 человек на 170 тысяч! В конце концов, Дэниэлс пришёл к заключению, опубликованному уже после войны: в реальности среднего пилота не существует. Если вы проектируете кабину для среднего пилота, то она не будет подходить ни для кого. Чтобы повысить эффективность солдат, в том числе лётчиков, рекомендуются радикальные изменения: окружение должно соответствовать индивидуальным параметрам, а не средним.

Кроме того, Тодд Роуз приводит историю из мирной жизни. Газета Plain Dealer объявила конкурс среди женщин и девушек. Им предлагалось прислать параметры своего тела, и победить должны те представительницы прекрасного пола, которые окажутся ближе всего к параметрам «типичной женщины», Нормы, увековеченной в статуе из медицинского музея Кливленда. Норма родилась вследствие усреднения 15000 женщин разного возраста и должна была олицетворять идеал, «определённый самой Природой». Всего рассматривалось 9 параметров, и из 3864 конкурсанток ни одна не попала в средние параметры. По пяти критериям «нормальными» оказались лишь 10% участниц, что даёт нам возможность оценить используемую жюри «толщину корки» в 75%. С таким суровым подходом надеяться найти хотя бы один «идеал» в пространстве девяти измерений можно лишь рассмотрев 260 тысяч красавиц. На всё человечество их наберётся от силы пара тысяч человек.

Норма. Идеал, недостижимый в силу своей исключительной типичности. Авторы: Роберт Л. Дикинсон и Абрам Белски

Далее Роуз пишет: «Но хотя Дэниэлс и организаторы конкурса получили одинаковый результат, они сделали совершенно разные выводы из этого. Большинство врачей и учёных того времени вовсе не посчитали, что Норма представляет неправильный идеал. Совсем наоборот: они пришли к заключению, что большинство американских женщин нездоровы и не поддерживают нормальную форму. Одним из таких был доктор Бруно Гебхард (Bruno Gebhard), директор медицинского музея Кливленда: он сокрушался, что послевоенные женщины в значительной степени непригодны к службе в армии, и упрекал их, упоминая плохую физическую форму, что делает их «плохими производителями и плохими потребителями». Интерпретация Дэниэлса была в точности противоположной. «Склонность думать в терминах "среднего человека" — это ловушка, которая многих приводит к просчётам, — писал он. — Практически невозможно найти среднего лётчика не из-за каких-то индивидуальных черт его группы, а из-за большого разброса параметров в размерах тела у всех людей».

Тот самый закон подлости

Один из классических законов подлости, сформулированный в сердцах инженером Эдвардом Мёрфи, гласит:

Всё, что может пойти не так, пойдет не так.

Сейчас мы можем взглянуть на него не только с иронической позиции. Он несколько глубже, чем тривиальное утверждение о том, что в полной выборке наблюдаются все исходы, даже самые маловероятные.

Пусть для выполнения некоторой работы требуется совершить ряд действий, и для каждого из них существует маленькая, но отличная от нуля вероятность неудачи. Какова вероятность того, что всё задуманное пройдёт без сучка без задоринки? Мы имеем дело с пересечением множества событий, каждое из которых соответствует *успешному завершению* того или иного этапа работы. Как посчитать вероятность для пересечения двух событий мы уже знаем: для этого нам требуется перемножить вероятность для первого события при условии что второе событие случилось на вероятность для второго события:

$$P(A_1 \cap A_2) = P(A_1|A_2) \cdot P(A_2)$$

Операция пересечения *ассоциативна*, это значит, что мы можем в произвольном порядке расставлять скобки для трёх и более пересекающихся событий:

$$A_1 \cap A_2 \cap A_3 = A_1 \cap (A_2 \cap A_3) = (A_1 \cap A_2) \cap A_3.$$

Отсюда легко получить общую формулу для пересечения произвольного числа событий:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = \\ = P(A_1|A_2 \cap A_3 \cap \ldots \cap A_n) \cdot P(A_2|A_3 \cap \ldots \cap A_n) \cdot \ldots \cdot P(A_{n-1}|A_n) \cdot P(A_n)$$

Если события независимы, то мы получаем произведение вероятностей наступления каждого из них:

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2) \cdot \ldots \cdot P(A_n)$$

Но для нас важно, что в любом случае, вероятности, условные или нет, по определению должны быть меньше единицы, а значит, мы вправе использовать закон арбузной корки: чем больше число шагов, тем существеннее роль границ. В нашем случае, границами являются какие-либо нештатные ситуаций. Достаточно дюжины шагов, для того чтобы средняя вероятность ошибки для множества этапов в 5%, выросла до 50% вероятности провала всего дела!

Эти наши рассуждения чрезвычайно просты, а закон Мерфи скорее эмоционален, чем объективен и кажется трюизмом, но всё же, именно с этого наблюдения в сороковые-пятидесятые годы двадцатого века началась новая большая наука: *теория надёжности*. Она добавила в рассмотрение время, взаимосвязь элементов систем, экономику, а также человеческий фактор и нашла применение за пределами инженерных наук: в экономике, теории управления и, наконец, в программировании. Мы ещё вернёмся к этой теме, когда будем изучать *проклятие режиссёра*, заставляющее принтер барахлить именно в день сдачи проекта. Закон Мерфи с учётом времени — поистине страшная сила!

В связи с рассуждениями о пересечении множества событий может возникнуть интересный и непростой вопрос. Если вероятность определена как мера, то она должна обладать свойством аддитивности. То есть, мера целого должна быть суммой мер его частей. Но мы рассмотрели вероятность успеха для некоего дела со множеством этапов, и увидели другую картину: вероятность целого оказалась равна произведению вероятностей для его частей, а не суммой. Это соответствует свойству *мультипликативности*. Так аддитивна вероятность, или мультипликативна? Тут следует различать *вероятностное пространство*, на котором вероятность играет роль аддитивной меры, и в котором сложение целого из частей выполняется с помощью операции объединения событий, и *фазовое пространство* некоторой системы, содержащее все возможные её состояния. Фазовое пространство измеримо, но вероятность мерой в нём не является. Чтобы произошло событие, соответствующее попаданию системы в заданное состояние, нужно чтобы все составные части системы *одновременно* попали в свои конкретные состояния, что является пересечением соответствующих событий, таким образом, вероятности этих событий перемножаются. Однако, превратить вероятность в «нормальную» аддитивную меру на фазовом пространстве можно и нужно. Мы совершим это превращение, когда будем говорить об *энтропии* систем и распределений случайных величин в главе «Термодинамика классового неравенства».

Счастье — это найти друзей с тем же диагнозом, что и у тебя

А можно ли вообще ставить вопрос о соответствии какой-то норме, не пытаемся ли мы при этом оценивать и сравнивать? Вы спросите, что же в этом плохого? Мы всё время кого-нибудь с кем-нибудь сравниваем, чаще всего, себя с другими, но иногда позволяем оценить и кого-нибудь ещё. Однако, с точки зрения математики, всё не так просто.

Для того, чтобы сравнивать что-либо с чем-либо, нужно правильно определить *отношение порядка* или ввести *метрику*. Определить отношение порядка, значит обозначить, что один элемент некоего множества, в каком-то смысле, предшествует другому. Этому мы научились ещё в школе: 2 меньше чем 20, слон слабее кита, договор дороже денег и т. п. Но вот вам ряд вопросов. Что идёт раньше понедельник или вторник? А воскресенье или понедельник? А какое воскресенье — то, что перед понедельником, или то, что после субботы? А какое число больше: $2 + 3i$ или $3 + 2i$? Мы можем назвать по порядку цвета радуги и даже ассоциировать все промежуточные цвета с вещественным числом — частотой света, но кроме этих цветов существует множество неспектральных цветов, они образуют хорошо знакомый типографам и дизайнерам цветовой круг, можно ли все видимые глазом цвета выстроить по порядку? Эти примеры показывают, что с отношением порядка бывают трудности. Например, на множестве дней недели не работает транзитивность (из того, что за A следует B, а за B следует C нельзя сделать вывод, что C всегда следует за A), так же как не транзитивна игра «камень-ножницы-бумага». На множестве рациональных чисел отношение порядка определено, но невозможно указать наименьшее или наибольшее число на каком-либо открытом отрезке. Попытка ввести понятие больше/меньше на поле комплексных чисел не согласуется с арифметикой этих чисел, а цвета обладают обоими этими недостатками.

Итак, мы видим, что отношение порядка вовсе не так просто, как мы привыкли думать, а главное, оно не универсально. Но мы всё-таки можем сравнивать людей, книги, блюда, языки программирования и прочие объекты, имеющие множество параметров, пусть даже условно формализуемых? Можем, используя вместо сравнения другую концепцию — степень подобия объектов между собой или *метрику*. Фильмы про Индиану Джонса ближе к «Пиратам Карибского моря», чем к комедиям Вуди Аллена или документалистике. Русский язык ближе к польскому чем к немцкому и совсем не похож на суахили. Числа $2 + 3i$ или $3 + 2i$ ближе друг к другу, чем к числу 100. Если мера обобщает размеры (длину, объём и т. д.), то метрика — это обобщение понятия «расстояние», которое было введено в математику Морисом Фреше в 1906 году.

> На некотором пространстве **метрикой** называется функция, ставящая любым двум элементам этого пространства неотрицательное вещественное число, такая что
> 1) равенство нулю расстояния между элементами эквивалентно равенству этих элементов,
> 2) порядок аргументов не изменяет результата и
> 3) должно выполняться *правило треугольников*, знакомое нам их курса геометрии: для любых трёх объектов сумма метрик для любых двух пар объектов них не должна превышать метрики для третьей пары. Иными словами: «окружной путь» не должен быть короче «прямого пути».

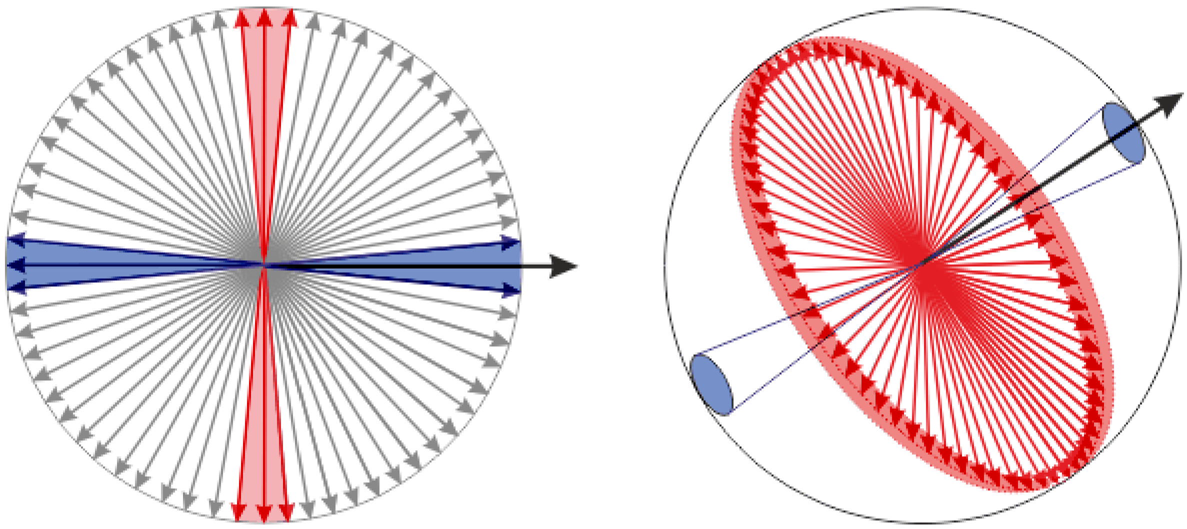
Понятие метрики позволяет вводить аналог расстояния (или степени близости) в совсем неочевидных случаях, например, на бесконечномерном пространстве функций, или между строками текста, либо изображениями, наконец, между распределениями случайных величин.

Понятие метрики не решает всех проблем, но в отсутствии внятной и корректной метрики легко увязнуть в бесконечном, бурном и бессмысленном споре, который в околокомпьютерной среде известен как «холивар» (от английского «holy war» — священная война). Увы, жаркие споры возникают чаще всего уже на этапе выбора метрик, поскольку они сами образуют некое множество, на котором тоже нужно определять отношение порядка: «лучше/хуже». Впрочем, можно предложить вполне осмысленный и однозначный способ рассуждений о сравнимости многомерных объектов, например, людей.

В многомерном пространстве параметров каждый объект может быть представлен вектором — набором чисел, определяющих значения критериев, которые его характеризуют. Рассматривая ансамбль векторов (например, человеческое общество), мы увидим, что какие-то из них окажутся сонаправлены, или, по крайней мере, близки по направлениям, вот их-то уже вполне можно сравнивать по длине. В тоже время, какие-то векторы будут *ортогональны* (в геометрическом смысле — перпендикулярны, в более широком — независимы), и соответствующие им люди будут попросту друг другу непонятны: они по ряду параметров окажутся в *сопряжённых пространствах*, как пресловутые физики и лирики. Нет смысла рассуждать о том, что хороший поэт в чём-либо лучше или хуже талантливого инженера или одарённого природой спортсмена. Единственное, о чём можно судить, это о длине вектора — о степени одарённости, о расстоянии от среднего.

В этой связи может возникнуть любопытный вопрос: а какая доля случайных векторов в пространстве заданной размерности будет сонаправленной, а какая ортогональной? Как много удастся найти единомышленников или, хотя бы, тех с кем можно себя сравнить?

В двухмерном мире каждому вектору соответствует одномерное пространство коллинеарных (сонаправленных) и одномерное пространство ортогональных векторов. Если мы рассмотрим "почти" сонаправленные и "почти" ортогональные вектора, то они образуют секторы одинаковой меры (неважно, площади или угла) при одинаковом выборе допустимого отклонения. То есть похожих и непохожих объектов, при рассмотрении двух критериев, будет одинаковое количество (под «количеством» мы понимаем меру).

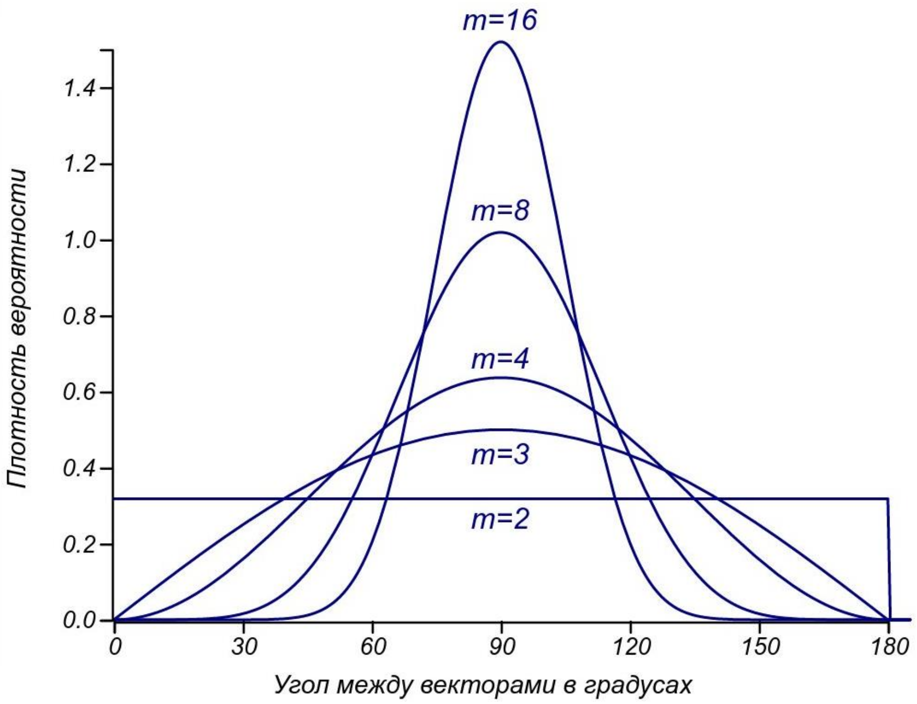


Почти коллинеарные и почти ортогональные векторы в двухмерном и трёхмерном пространстве.

В трёхмерном мире картина поменяется. Сонаправленные векторы всё также образуют одномерное пространство, а вот ортогональные уже заполняют плоскость, то есть, двухмерное пространство. С точки зрения ортогональных векторов, мера сонаправленных уже равна нулю, но давайте всё же позволим векторам немного отклониться от курса. Фиксируя длину векторов R и допуская небольшое отклонение от идеальных направлений на угол $\Delta\varphi$, можно число почти сонаправленных векторов сопоставить с площадью круговых областей вокруг полюсов $2\pi(R\,\Delta\varphi)^2$, а число почти ортогональных векторов — с площадью полосы вокруг экватора: $4\pi\,R^2\Delta\varphi$. Их отношение $2/\Delta\varphi$ растёт неограниченно при уменьшении отклонения $\Delta\varphi$. В четырёхмерном мире ортогональные векторы образуют уже трёхмерное пространство, тогда как сонаправленные векторы всё ещё лежат в одномерном, и разница в их количестве растёт уже пропорционально квадрату отклонения от идеала. Но на этом этапе лучше обратиться к теории вероятностей и выяснить, каковы шансы получить ортогональные или сонаправленные векторы, взяв наугад два вектора из пространства, размерности m? Об этом нам расскажет распределение углов между случайными векторами. К счастью, рассуждая о площадях многомерных сфер, распределение можно вычислить аналитически и даже представить в конечной форме:

$$p(\varphi) = \frac{\Gamma\left(\frac{m}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{m-1}{2}\right)} \sin^{m-2}\varphi\,, \qquad \text{для } 0 \le \varphi \le \pi,$$

здесь $\Gamma(x)$ — это гамма-функция, обобщение факториала на вещественные (и даже комплексные) числа. Её основное свойство: $\Gamma(x+1) = x\Gamma(x)$.



Распределения углов случайных векторов в пространствах различных размерностей.

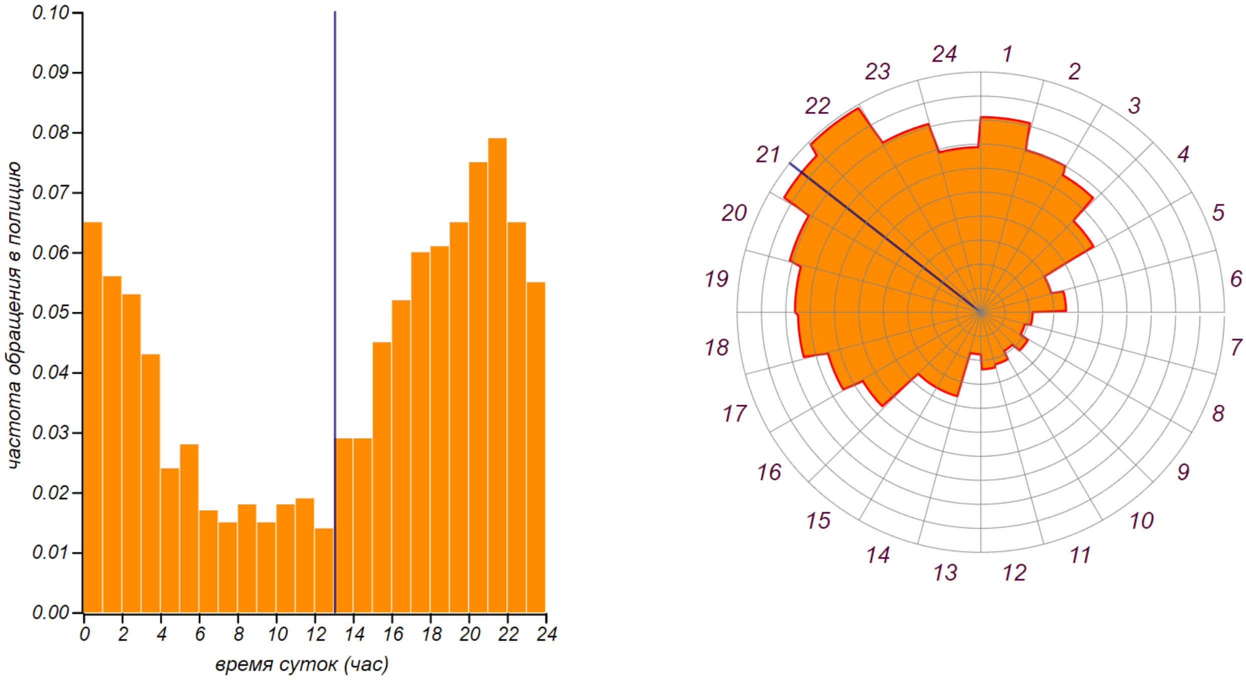
Для двумерного пространства углы распределяются равномерно, для трёхмерного — пропорционально синусоидальной функции. Свойства синуса приводят к тому, что плотность вероятности в нуле для $m > 2$ в точности равна нулю. Это согласуется с нашими рассуждениями о том, что сонаправленные вектора образуют множества нулевой меры. Для всех размерностей выше двух, мода распределения приходится на $90°$ и доля взаимно ортогональных векторов увеличивается, по мере увеличения числа параметров. Самое же главное наблюдение — сонаправленных векторов (имеющих угол около $0°$ или $180°$ практически не остаётся при достаточно высокой размерности пространства. Если считать более или менее похожими (сонаправленными, сравнимыми) векторы, имеющие угол мене $30°$, а это вполне малый угол: $30° = \pi/6 \approx 1/2 = \sin 30°$, то при увеличении размерности пространства на единицу, доля сравнимых векторов будет уменьшаться практически вдвое. При этом, при сравнении по двум критериям, похожей на какой-то выделенный вектор, окажется только треть всех случайных векторов. Таким образом, мы приходим к векторной формулировке закона арбузной корки:

В пространствах высокой размерности почти все вектора ортогональны друг другу.

или эквивалентно: на вкус и цвет товарищей нет.

Этот странный закольцованный мир

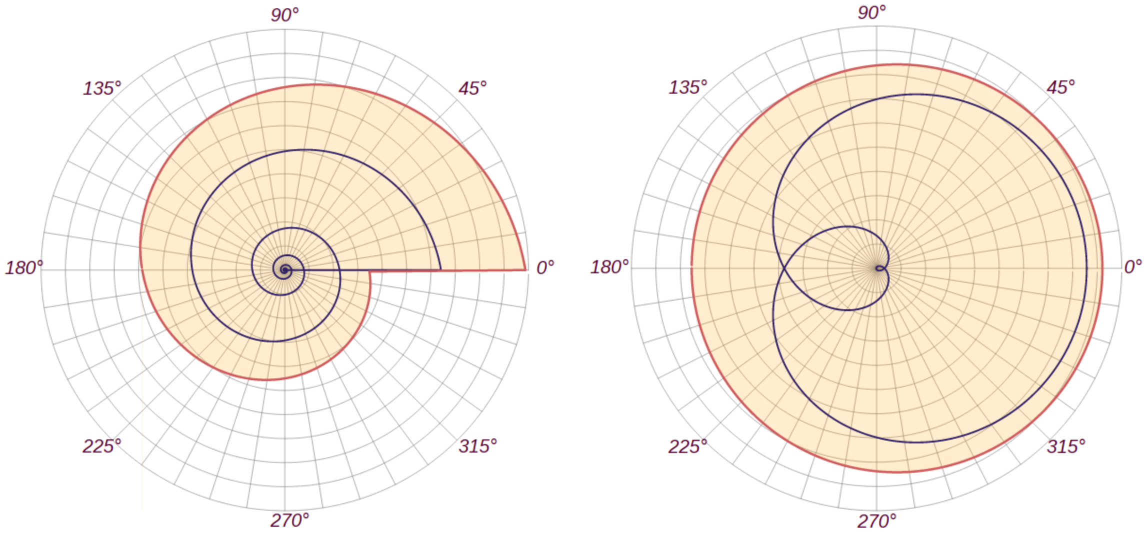
По мере повышении размерности распределение углов становится похожим на нормальное. Но это не оно, несмотря на характерную колоколообразную форму. Нормальное распределение определено для всей вещественной числовой оси, в нашем же случае значение угла зациклено в переделах от 0° до 180°. Мы попали из поля вещественных чисел на кольцо вычетов, математическую структуру, подобную циферблату на часах, дням недели или остаткам от деления. Применяя привычные нам операции в этом кольцевом мире, нужно быть аккуратным, даже выполняя простые расчёты. Скажем, чему равно среднее значение для двух углов: 30° и 350°? Простое сложение даст ответ 190°, тогда как чертёж покажет, что правильным ответом будет 10°. А чему равно среднее значение равномерного распределения на всей окружности? Оно не определено, хотя площадь под кривой распределения конечна. Даже просто вычислить среднее для набора измеренных углов уже становится нетривиальной задачей, требующей перехода на плоскость (декартову или комплексную). Представьте себе, что вы исследуете зависимость числа обращений граждан в полицию от времени суток, и получили гистограмму, показанную на рисунке.



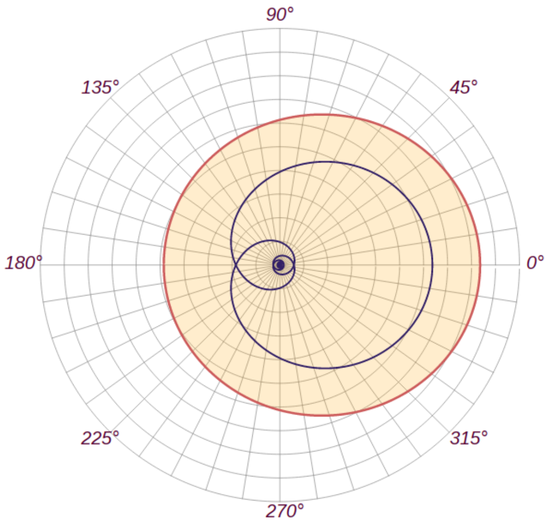
Гистограмма, показывающая распределение числа событий по времени суток не отражает цикличности времени и не даёт возможность правильно найти среднее значение.

Попытка вычислить математическое ожидание для самого неспокойного времени с помощью среднего арифметического даст невнятный результат. Он показан на рисунке вертикальной линией. Правильным будет изобразить нашу гистограмму в полярных координатах и там уже найти математическое ожидание, вычислив положение *центра масс* получившейся фигуры.

Привычные распределения вероятностей с хорошо известными свойствами на кольцах вычетов «зацикливаются» и становятся своеобразными. На рисунке показано, как можно построить аналоги некоторых распределении на окружности. Числовая ось как бы наматывается на окружность, при этом каждый слой получающейся спирали суммируется и в результате мы получаем циклический аналог распределения, который имеет единичную площадь.



Построение циклических экспоненциального и нормального распределений. Графики функций плотности для обыкновенных (линейных) распределений показаны синим.



Циклический аналог распределения Коши.

Например, циклическое экспоненциальное распределение описывает *случайное положительное отклонение от заданного угла с заданным средним значением*. С его помощью можно описать *время суток*, в которое ожидается появление пуассоновского события. Циклическое нормальное распределение можно использовать для описания погрешностей в измерении углов, хотя если быть точным, они будут подчиняться другому распределению, но об этом чуть позже. Циклические распределения, хоть и выглядят несколько однообразно, становятся важны при анализе данных на земном шаре, если их дисперсии сравнимы с длинной экватора, а это широкий класс задач геофизики, климатологии и других наук о Земле.

Любопытно, что при зацикливании свойства распределения могут поменяться радикально. Например, относительная погрешность при измерении нулевой величины описывается распределением Коши. Оно примечательно тем, что её функция плотности вероятности имеет бесконечную площадь под кривой, так что невозможно посчитать значение среднего и дисперсии для этого распределения, они, в отличие от моды и медианы, для распределения Коши попросту не определены. Однако круговой аналог этого распределения ведёт себя хорошо, интегрируется и имеет вычислимые значения среднего и дисперсии. Это распределение встречается, например, в физике, при анализе дифракции.

Меняет свои свойства при зацикливании и нормальное (гауссово) распределение. Его циклический аналог уже не будет устойчивым, и не к нему будут стремиться суммы случайных величин, как следует из Центральной предельной теоремы. На окружности эту роль играет распределение фон Мизеса с такой функцией плотности вероятности:

$$p(x) = \frac{e^{\kappa \cos(x-\mu)}}{2\pi I_0(\kappa)},$$

здесь I_0 – *модифицированная функция Бесселя*, одна из целого семейства специальных функций. Функции Бесселя появляются, если в задаче есть осевая симметрия. Например, с их помощью описывается профиль круговых волн, разбегающихся по воде от упавшей капли.

Тут мы встречаемся с Центральной предельной теоремой, но в обобщённом виде: оно является наиболее ожидаемым распределением для случайной величины, определённой на окружности, если известны её среднее значение и стандартное отклонение. Впрочем, когда дисперсия данных мала и x незначительно отклоняется от среднего значения μ, косинус можно разложить в степенной ряд, в котором лидирующую роль играет квадратичный член. Таким образом, когда влияние цикличности становится незначительным, то и распределение фон Мизеса становится похожим на «нормальное» гауссово распределение. Никуда от него не денешься!

Я сказал «похожим на нормальное», имея в виду, что распределения случайных величин тоже образуют метрическое пространство. Их можно сравнивать между собой, выбирать похожие и «наилучшие», причём, в различных смыслах. И любопытное совпадение; среди них тоже есть некое «нормальное» распределение, и его, если вспомнить Центральную предельную теорему, можно рассматривать как результат усреднения множества случайных величин. Однако, в отличие от мифического «среднего пилота» или усреднённой «идеальной женщины», случайные величины, подчиняющиеся нормальному распределению, встречаются повсеместно.

* * *

Сравнивайте разумно, не ищите в жизни нормальности и не бойтесь ненормальности. Сама математика подсказывает нам, что в сложном мире людей говорить можно лишь о степени подобия, но не о сравнении. Так что нет резона вести нескончаемые споры, в поисках истины, вместо этого, стоит прислушаться и постараться услышать иное мнение, увидеть взгляд из другого, сопряжённого, пространства, обогащая тем самым своё восприятие мира.

Мудрецы правы: все мы уникальны и в своей уникальности абсолютно одинаковы.