# Введение

В этой главе мы начнём наш разговор о неприятностях, и на примере велосипедиста рассмотрим нужные нам инструменты для измерения несправедливости: кривую Лоренца и индекс Джини, а также упомянем пресловутого Парето и грозного инспектора.

## О чем эта книжка

Здесь речь пойдёт о различных неприятностях. Привычных, ожидаемых и настолько предсказуемых, что они получили статус законов. Их уже сформулировано множество: это и закон падающего бутерброда, и закон Мерфи: "*Если какая-нибудь неприятность может произойти, она случится*" и законы Чизхолма на тему: "*Когда дела идут хорошо, что-то должно случиться в самом ближайшем будущем.*" и наблюдение Этторе: "*Соседняя очередь всегда движется быстрее.*" Большая их часть вполне тривиальна, но согласно закону Муира "*Когда мы пытаемся вытащить что-нибудь одно, оказывается, что оно связано со всем остальным.*" Мы постараемся найти рациональное зерно этих закономерностей, но не для того, чтобы с ними бороться, а для удовольствия. И поскольку при этом мы будем использовать математику, то удовольствие будет своеобразным и полезным, в отличие от самого результата. Ну, а если наши рассуждения заведут нас слишком далеко, мы можем взять на вооружение постулат Персига: "*Число разумных гипотез, объясняющих любое данное явление, бесконечно*". В конце концов, Гроссман, ошибочно цитируя Х. Л. Менкина верно указал, что "*Сложные проблемы всегда имеют простые, легкие для понимания неправильные решения.*" Со всеми этими глубокомысленными фразами и законами мы и станем разбираться, опираясь при этом на язык математики и, по-возможности, строгие выкладки.

Современная математика – это огромная страна со сложным «ландшафтом». В ней есть и свои цветущие долины и древние памятники, развлекательные центры и пряничные городки, даже супермаркеты с готовыми решениями на все случаи жизни. Всё это соединено друг с другом хорошо оборудованными дорогами с указателями и путеводителями. Но есть там и глухие участки с густыми непроходимыми лесами, горами и топкими болотами, через которые проходят внезапно исчезающие тропинки и шаткие мостики гипотез и предположений. Наконец, эта страна повсюду окружена неизведанными землями, куда если и осмеливался ступить человек, то лишь очень отважный и часто одинокий в своих поисках. Я не случайно так увлёкся этой аллегорией. Она гораздо ближе к пониманию того, что такое наука, чем может показаться на первый взгляд. Дело в том, что в любом городе можно ходить от одной площади до другой, от одного здания до другого по-разному. Наконец, в любом городе по-разному можно жить.

Если вы впервые попадаете в новый интересный для вас город, то, скорее всего, вы выберете для ознакомления экскурсионный маршрут, который уже отработан годами и представляет собой своеобразное произведение искусства. Так за какие-нибудь пару часов вы получите яркие впечатления об этом городе, которые останутся с вами на всю жизнь. Но вы не сможете сказать, что узнали этот город по-настоящему.

Может быть вас в этот город занесёт по работе, скажем, случится более или менее длинная командировка. В таком случае, вам неплохо удастся изучить основные полезные маршруты и появятся навыки мастерски пользоваться общественным транспортом, перемещаясь быстро, эффективно и удобно. Но и после нескольких недель такой жизни, город вам может быть незнаком.

Наконец, может случиться так, что город станет вашим по-настоящему. Вы полюбите его и будете бродить по его улочкам бесцельно, получая удовольствие от самих прогулок. Вы будете отыскивать новые проходы от одной площади до другой через закоулки и дворы, садами и тропинками. Эти пути могут оказаться на удивление короткими, а могут завести бог знает куда, но это не страшно, ведь вы знаете этот город, вы никогда в нём не заблудитесь.

Общий школьный и университетеский курс математики ближе к экскурсии. Вам покажут главные древние памятники и знаменитые площади, к которым ведут большие проспекты. Погружение в ту или иную прикладную задачу похоже на командировку, тут не до блужданий и важно чётко понять на какую линию садиться и на какой остановке пересаживаться каждый день, чтобы не терять драгоценных сил и времени. Но с математикой у вас может случиться и настоящая любовь. И тогда вы уже не остаётесь в рамках лишь практической пользы или удобства, вам становится важным понять, почувствовать, что математика, как большой город, это не только дома и не только площади и не линии метро, а единая система, соединяющая всё, что в ней есть не только взаимным расположением, но и смыслами, контекстами, историями.

Эта книжка не совсем про математику. С её помощью я приглашаю вас на прогулку по некоторым её местечкам, хорошо известным и имеющим большую практическую пользу. Но двигаться мы будем несколько необычным маршрутом, не прямым, как в учебнике и не сложным и запутанным, как в научной работе, а лёгким, как бесцельное шатание в хорошей компании под интересный разговор. То и дело мы будем оказываться на развилках и площадях с хорошо обозначенными названиями, соответствующих разделам математики. Оглянувшись, мы отправимся дальше, но читатель может отметить про себя, что пересечённый нами проспект или бульвар это целое направление, в которое можно углубиться самостоятельно, будь на то интерес или необходимость.

В стране математики говорят на своём языке и не все указатели и надписи бывает легко перевести на русский язык. Иногда я буду приводить цитаты на языке аборигенов, это значит, что в книжке будут формулы. Но формулы вовсе не единственный алфавит языка математики, их можно выразить графически и я всегда буду сопровождать уравнения графиками, которые можно понимать интуитивно. Почему же я не отказался от формул совсем? В нашей математической стране не принято верить каждому встречному, не принято полагаться на интуицию, чутьё и даже на опыт. Да, опыт, в отличие от физики или психологии здесь имеет сравнительно невысокую цену. В ходу здесь только доказательство – самая твёрдая валюта, которой неведомы ни девальвация, ни инфляция, ни мода, ни конъюнктура. Она не обесценивается тысячелетиями (и это не фигура речи, мы используем доказательства тысячелетней давности, каждый день). Таким образом, всё что я вам здесь наговорю не должно приниматься на веру. Любое моё утверждение, вывод, даже самый завиральный, можно проверить строгими доказательствами, и потому, везде, где возможно оставлены ключевые заметки в виде формул, которыми я руководствовался. Это, впрочем, не лишает меня возможности давать математическим закономерностям не очень серьёзные и даже фривольные житейские интерпретации. Ведь так гораздо интереснее!

## Разновидности неприятностей

Какие-то случающиеся с нами неприятности закономерны, как говорят математики, детерминированы, то есть, не зависят от случайностей. Например, если вам понизили зарплату на 10%, а потом извинились и повысили на 10%, то в итоге этих махинаций, вы останетесь в убытке, поскольку

Более того, если зарплату сначала повысят, а потом, не извинившись даже, понизят на те же самые 10%, результат выйдет такой же, поскольку неважно в каком порядке перемножать коэффициенты. Это очень просто, обидно, но к удаче отношения не имеет.

Другой пример детерминированной неприятности — волшебство, случающееся в наших карманах с наушниками: кладём аккуратно сложенные наушники в карман, а через полчаса там происходит чудо, и из кармана мы вынимаем дикий узел проводов. В 2007 году вышла серьёзная научная статья двух учёных из солнечного и безмятежного Сан-Диего «Спонтанное образование узлов на возбуждаемой нити»[[1]](#footnote-2), в которой детально анализируется и моделируется запутывание наушников в кармане. Авторы, основываясь на теории узлов, теории вероятностей и физических экспериментах, убедительно показывают, что при стандартном способе сматывания, наушники, действительно, должны запутываться, причём, спустя лишь несколько секунд тряски. Впрочем, это мы и так наблюдаем практически каждый день. Неожиданной здесь может оказаться только ожидаемая скорость запутывания. С этой неприятностью вполне можно бороться математическим способом: нужно поменять способ складывания наушников — не кольцами, которым свойственно образовывать узлы, а чередой петель взаимно-обратного направления, как например, показано на рисунке. При таком способе складывания петли разных «знаков» взаимно уничтожают друг друга, и узлы не формируются. Уже много лет я складываю наушники именно таким образом, чувствуя себя крутым топологом, и всякий раз радуюсь, как фокусу, когда они разматываются сами от одного небрежного встряхивания рукой.



Один из способов складывания проводов, не приводящий к их запутыванию. Он хорош ещё и тем, что попутно вы складываете пальцы в мудру любви.

Но и среди стохастических по своей природе законов не все одинаково интересны. Например, закон Бука: «ключи всегда находишь в последнем кармане» не имеет под собой какого-либо рационального основания. Простой подсчёт показывает, что при равной вероятности отыскать ключи для всех карманов, последний ничем не отличается от прочих. Впрочем, этот закон можно трактовать разве что как забавный софизм: утверждение Бука верно всегда, поскольку тот карман, в котором ключи будут обнаружены, завершит процесс поиска и, следовательно, будет последним. Однако и здесь есть о чем поговорить. В процессе перебора карманов, так называемая условная вероятность того, что ключи лежат в последнем кармане, действительно повышается. Но это, уже нельзя просто трактовать, как вероятность того, что ключи находятся в последнем кармане, это уже другая задача. С тем, что же здесь имеется в виду, мы разберёмся в третьей главе, когда и понятие вероятности станет нам ближе и можно будет его несколько усложнить.

Нас будут интересовать законы парадоксальные и поучительные, законы, которые выглядят злым роком, выбирающим из множества вариантов самые досадные и неприятные, наперекор интуиции подсказывающей, что этот выбор не должен быть самым вероятным. И прежде чем приступить к детальным и точным рассуждениям о случайностях и о вероятностях, давайте предположим, что какая-то интуиция в отношении случайных процессов и вероятностей у нас уже есть. Это вполне допустимо даже в математической книжке -- до какого-то момента использовать интуитивное представление о предмете, а потом дать строгое определение. Тем самым, в первую очередь, мы определяем границы применимости нашей интуиции, а во-вторых, расширяем эти границы в правильном с научной точки зрения направлении. Но не будем забывать о законе Вертерна: «Предположение — мать любой неразберихи», и все наши предположения и даже строгие выводы постараемся, где это возможно, проверить с помощью имитационного моделирования.

## А при чём тут математика?

Петли, наушники, законы подлости, неприятности… причём тут математика? Почему, вообще, имеет смысл рассуждать о законах подлости не так, как это сделал Артур Блох, просто посмеявшись и найдя меткий афоризм.

С математикой знакомы все, но мало кто готов ответить на вопрос: что делают математики? Считают и вычисляют? Рисуют треугольники и круги на бумаге в клеточку? Передвигают буквы туда-сюда в уравнениях? Придумывают странные значки и закорючки, чтобы потом писать совершенно непонятные тексты? Если вы никогда этого не делали, загляните в математический журнал, просто из любопытства. Сейчас это легко сделать, не выходя из дома в сети, поищите что-нибудь на тему «гомологическая теория типов» или «топология». Вы поразитесь тому насколько то, что вы там обнаружите, окажется не похоже на школьный образ математики! Но вот что важно, эта колоссальная разница не говорит о том, что есть какая-то одна «простая» математика и другая «сложная». Математику часто называют языком, и как на любом живом человеческом языке можно писать анекдоты и незамысловатые детские стишки, а можно написать неуловимо тонкую поэзию, тяжеловесный роман или многостраничный договор, так и с помощью математики можно рассуждать о числах и отрезках, а можно и петлях и поверхностях, о многомерных пространствах и даже об основах самой математики. И не нужно думать, что числа и отрезки это просто! Теория чисел и геометрия это огромные и во многом неизведанные области современной математики в которых ведутся очень интенсивные исследования.

Что же, все-таки изучают математики? Для чего им этот язык? Чаще всего, речь идёт о тех или иных моделях. Например, что может быть моделью количества? Число, скажете вы. Но любое ли число годится для этого? Школьники сталкиваясь с отрицательными числами испытывают замешательство, ведь модель оказывается шире привычного им понятия количества. Переход от количества к шагам, например, помогает понять, что числа годятся и как модель движения по прямой, тогда отрицательные и многие другие числа обретают, наконец, смысл. А чем можно моделировать скорость? Тоже числом. Но если я скажу я двигаюсь со скоростью 60, этого будет достаточно для описания того, что со мной происходит? Точно нет! Остаётся неясным ни куда я двигаюсь, ни собственно, с какой скоростью, поскольку 60 может означать как 60 км/ч, так и 60 мм/год. Значит для моделирования скорости или силы одного числа не достаточно. Зато стрелка – ориентированный отрезок для этого годится лучше, они показывают направление, а сравнивая их с единичной стрелкой можно определить их масштаб. Более того, стрелки можно складывать и умножать на числа и вновь получать стрелки! Причём, если мне удастся придумать как однозначно сопоставлять скорости стрелкам, и окажется, что v1 соответствует стрелка a, а скорости v2 – стрелка b, то сумме скоростей 3v1+v2 будут соответствовать стрелка 3a+b, никакая иная. А чем можно моделировать стрелки? – упорядоченным набором чисел, или вектором. Так постепенно математики приходят к мысли о чём-то что моделирует и стрелки и скорости и вектора: линейные векторные пространства. Изучая свойства этих моделей (изучая, а не придумывая, об этом мы поговорим позже), они вырабатывают единый язык для разговора о таких разных вещах, как, например, цвета, движения и преобразования предметов и их изображений в пространстве, который называется линейной алгеброй. Пользуясь этим языком уже можно найти оптимальную стратегию в экономической игре или научить компьютер узнавать буквы или лица людей.

Математики работают со *структурами*, которые являются универсальными моделями всего, с чем имеет дело человеческий разум. Группы, поля, решётки, графы, петли, косы, слова и целые бесконечномерные пространства… всё это структуры с чётко определёнными свойствами. И вот уже тысячи лет математики исследуют взаимосвязи между этими структурами, обнаруживают что ими можно моделировать и при каких условиях.

В нашей книжке мы преимущественно будем иметь дело с двумя структурами: с*о случайными величинами* и *случайными функциями*. Но изучая их мы встретимся со многими другими моделями, понятиями и поищем связи между ними.

А начнём мы прямо сейчас с некоторого простого инструментария, который нам будет полезен на протяжении всей книжки. И для этого нам потребуется… велосипед.

## Если долго, долго, долго, если долго по тропинке

Я большой энтузиаст велосипедного любительского спорта. Многие задачи, ставшие примерами для этой книжки я обмозговывал в седле, вертя их мысленно и так и эдак, пытаясь найти наиболее наглядный и простой подход к ним. Но, вообще, что может быть лучше, чем мчаться по трассе ранним утром, по холодку, скатываясь с лёгкого склона… это ощущение стоит того, чтобы ради него преодолевать бесконечные подъёмы или сопротивление встречному ветру! Правда, порой кажется, что подъёмов как-будто бы больше, чем спусков, а ветер норовит быть встречным, куда ни поверни. В книгах по мерфологии в этой связи приводится **закон велосипедиста**:

Независимо от того, куда вы едете — это в гору и против ветра.

Живу я на Камчатке. В Петропавловске много горок, и катаясь по городу, их не миновать. Однако меня должна успокаивать мысль, что начиная свой путь из дома, я возвращаюсь снова домой, а это значит, что суммарный спуск должен быть равен суммарному подъёму. Особенно честным будет радиальный маршрут, в котором прямой и обратный пути совпадают.

Давайте представим себе 2-километровую трассу, состоящую из одной симметричной горки: километр вверх, километр вниз. Вверх по склону я могу достаточно долго ехать со скоростью 10 км/ч, а на спуске стараюсь держать скорость в 40 км/ч (не переживайте, я осторожный и езжу в шлеме). Исходя из этих положений, на подъём я буду тратить в четыре раза больше времени, чем на спуск, и общая картина получится такой: 4/5 времени путешествия уйдёт на тягучий подъём, и лишь 1/5 — на приятный спуск. Получается обидно — 80% времени прогулки составит сложный участок пути! И этот результат не будет зависеть от длины горок, а лишь от соотношения скоростей. Если я выкачусь из нашего холмистого города, в сторону океана или в долину реки Авачи, горок почти не будет, но в моём распоряжении остаются встречный и попутный ветер, или участки с плохой дорогой, которые также занимают значительную часть времени путешествия.

Давайте взглянем на закон велосипедиста с несколько иной точки зрения. Если я сделаю множество фотографий-селфи на протяжении своей велопрогулки, в случайные моменты времени а потом подсчитаю их, то обнаружу, что значительная часть картинок покажет мне согбенную фигуру в оранжевом шлеме, упорно ползущую вверх по склону либо сопротивляющуюся встречному ветру. Число снимков с летящим и сияющим велосипедистов, как с рекламной картинки, увы, составит лишь около 20%. А что скажет статистика? Если мы выпустим на холмистую трассу большую толпу велосипедистов, подождём немного, и пронаблюдаем за их плотностью, то увидим, как большая часть спортсменов толпится на трудных участках, и доля безмятежно улыбающихся лиц в общей массе окажется не так уж и велика!

## Измеряем уровень подлости

Давайте, как когда-то в школе, покажем на графике зависимость перемещения велосипедиста от времени, при движении по симметричной треугольной горке. Только сделаем всё «по-взрослому», в так называемых *собственных масштабах задачи[[2]](#footnote-3)*: расстояние будем измерять не в километрах, а в долях общего пути, так же поступим и со временем путешествия. Первую половину пути (отрезок AB) велосипедист двигался медленно и долго — 4/5 всего времени, а вторую (отрезок BC) он преодолел быстро — за 1/5 времени.

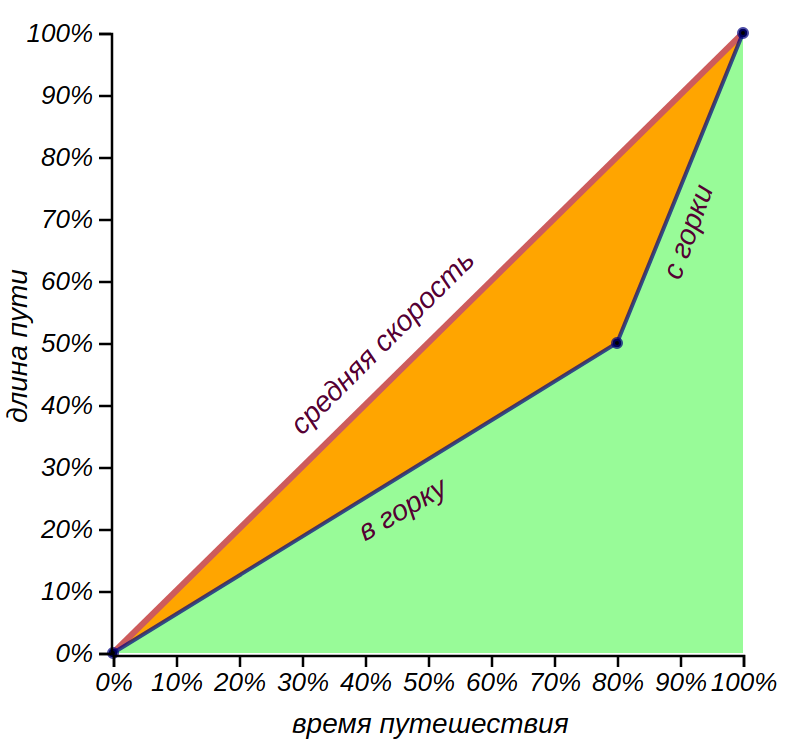


Диаграмма перемещения велосипедиста в долях от общего пути и времени.

Что же нам показывает график, который мы получили? Во-первых, мы можем сравнить скорости на разных участках пути (наклоны) со средней скоростью, которая соответствует диагональной линии. Во-вторых, становится наглядным соотношение: 80/50 — 80% времени путешествия заняла трудная половина маршрута. Кроме того, из графика можно заключить, что за первую половину расчётного времени путешествия велосипедист преодолеет лишь треть всего пути. Пока всё предельно просто и понятно.

А что если маршрут велосипедиста усложнится и перестанет быть симметричным. Если участков с подъёмами и спусками будет несколько и все они будут разными? Можно изобразить путешествие и на этот раз, например так, как показано на рисунке:

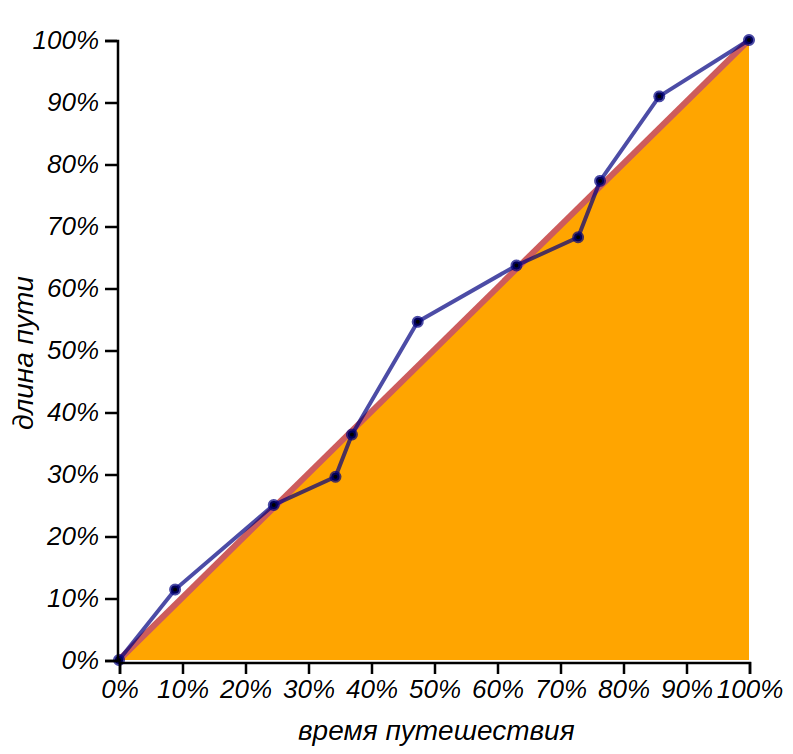


Диаграмма перемещения велосипедиста для более сложного маршрута.

Такая диаграмма хорошо отражает характер пути, но не даёт представления об общем соотношении легких и трудных участков, иными словами, о распределении скоростей. О том, какой смысл мы вкладываем в слово «распределение», речь пойдёт в следующей главе, пока же доверимся интуиции и тому, что это слово мы используем достаточно часто, и порою, не вкладывая в него точный математический смысл. Чтобы увидеть это самое распределение давайте теперь упорядочим отрезки пути по скорости от самых медленных до самых быстрых, после чего вновь нанесём их на диаграмму:

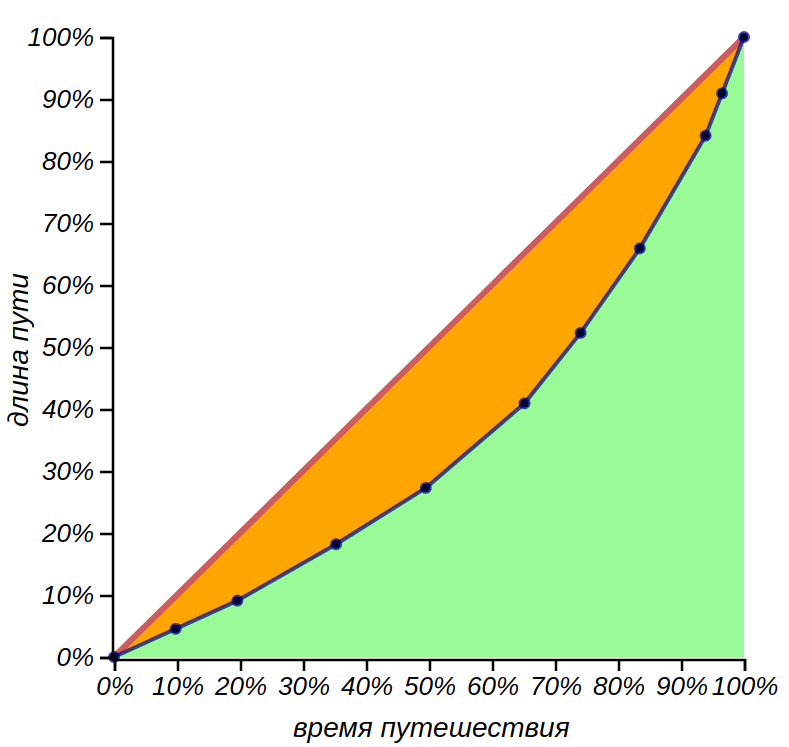
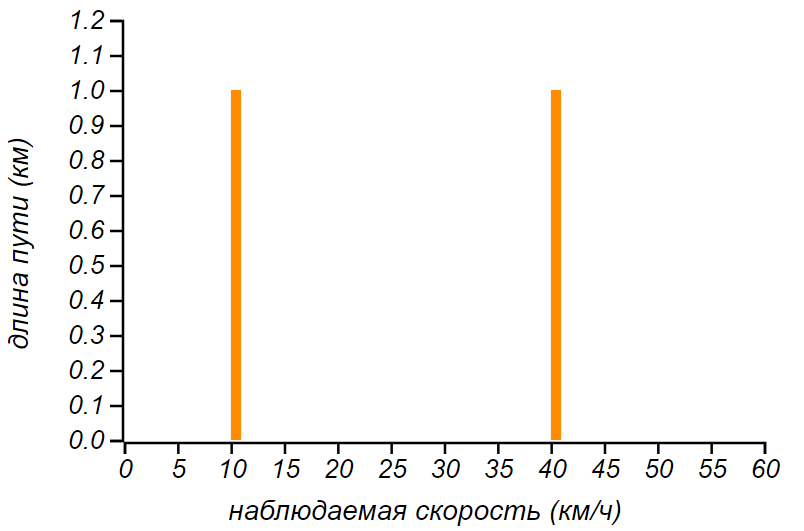


Диаграмма перемещения велосипедиста для распределения скоростей.

Мы потеряем при этом информацию о последовательности участков, но зато получим некоторую обобщающую картину, отражающую то, что можно было бы условно назвать «справедливостью» распределения. Более того, если от одного велосипедиста мы переключимся к целой группе, ездящей по этому маршруту в произвольном направлении, то это диаграмма практически не изменится, разве что несколько сгладится из-за разброса скоростей. Её смысл останется прежним — она покажет насколько этот маршрут отклоняется от самого справедливого маршрута на котором время преодоления участка не не зависит от его «трудности», а определяется только его длиной. Сейчас я поясню откуда взялась такая странная терминология.

С начала XX века у эконометристов, демографов, экологов или маркетологов появились вполне универсальные способы суждения о несправедливости этого мира — кривая Лоренца и связанный с ней индекс Джини. Для известного распределения чего-нибудь ценного, например, денег, в некоторой популяции, можно, предварительно отсортировав членов множества по возрастанию уровня богатства, построить кумулятивную кривую. Такая кривая покажет как, по мере добавления новых членов, растёт общее благосостояние популяции. Далее, нужно нормировать ось X на численность популяции, а ось Y — на общее её благосостояние, перейдя от конкретных чисел к долям. Получится кривая, носящая имя американского экономиста Макса Отто Лоренца. Когда мы строили график перемещения велосипедиста по симметричной горке, мы, по существу, построили кривую Лоренца для распределения скоростей по отрезкам пути, состоящего всего из двух столбцов.



Распределение скорости велосипедиста по пройденному пути.

Конечно же, не всякий график перемещения можно воспринимать, как кривую Лоренца. Перед тем как её строить, нужно отсортировать периоды путешествия по возрастанию скорости, после чего уже приступать к построению. Иными словами, сначала нужно построить гистограмму скоростей, после чего последовательно складывать вклады всех столбиков гистограммы, начиная со вклада малых значений, заканчивая самыми большими. Результатом должна явиться всюду вогнутая кривая, которая проходит ниже диагонали (AC). Эта диагональ называется кривой равенства, она, в нашем случае, соответствует постоянной (средней) скорости на всём пути, или гистограмме с одним единственным столбиком, (такое распределение называется вырожденным), а в экономическом смысле — всеобщему равенству благосостояния. Чем больше кривая Лоренца отклоняется от кривой равенства, тем менее «справедливым» можно считать распределение. Коль скоро мы изучаем законы подлости и несправедливости нашего мира, разумно использовать и терминологию и инструменты, используемые для исследования справедливости.

Площадь под кривой Лоренца для любого распределения, отличного от вырожденного, будет меньше площади под кривой равенства. Их разница может служить формальной характеристикой неравенства или «несправедливости» распределения. Эту характеристику отражает индекс Джини. Он вычисляется, как удвоенная площадь замкнутой фигуры, образованной кривой равенства и кривой Лоренца. Для идеального вырожденного мира индекс Джини равен 0, в самом кошмарном варианте он стремится к единице. В рассмотренном нами примере, он равен 0.35. Это вполне неплохой показатель. Скажем, распределение богатства среди населения в России сейчас имеет индекс Джини 0.39, в США — 0.49, в Австрии и Швеции он не превышает 0.3, а для всего Мира он в 2017 г. составил 0.66. Так что ситуация с велосипедистами, конечно, обидна и несправедлива, но вполне терпима.

Обратите внимание, с помощью некоторого формального индекса мы стали сравнивать совершенно разные и не сравнимые вещи, это одновременно и заманчиво и опасно. Нужно отдавать себе отчёт в том, что формальные индексы и критерии всегда чему-то равны, не зависимо от того есть в этом смысл, или нет. Мы сравниваем распределение богатства среди населения стран и распределение времени затрачиваемого на преодоление пути с точки зрения отличия от некоторого варианта, которое сочли бы справедливым. Покуда мы ведём фривольные и, подчас, хулиганские разговоры о законах подлости, пожалуй, это оправданное сравнение, но в математике так, конечно же, делать нельзя. Кривую Лоренца, а по ней и индекс Джини можно формально рассчитать и для гистограммы яркости пикселей на картинке или для частотности слов в живой речи, к справедливости это не будет иметь никакого отношения, да и смысла останется совсем немного. Поэтому, имея в виду индекс Джини для чего попало, мы будем его называть индексом подлости, чтобы не вводить читателя в заблуждение наукообразностью терминов.

Понимая, что это не козни судьбы, а простейшая математика, с которой бороться смысла нет, можно научиться получать удовольствие и от затяжных подъёмов и от нудных, но неизбежных этапов работы, хотя бы, решая в уме задачки, или медитируя. Даосы стремились жить вечно, и правильно рассудили, что вместе с работой над телом, для достижения их цели, требуется подготовка ума. Ведь для вечной жизни нужно не только умение отпускать привязанности, но и терпение, а также умение получать удовольствие от затяжных участков.

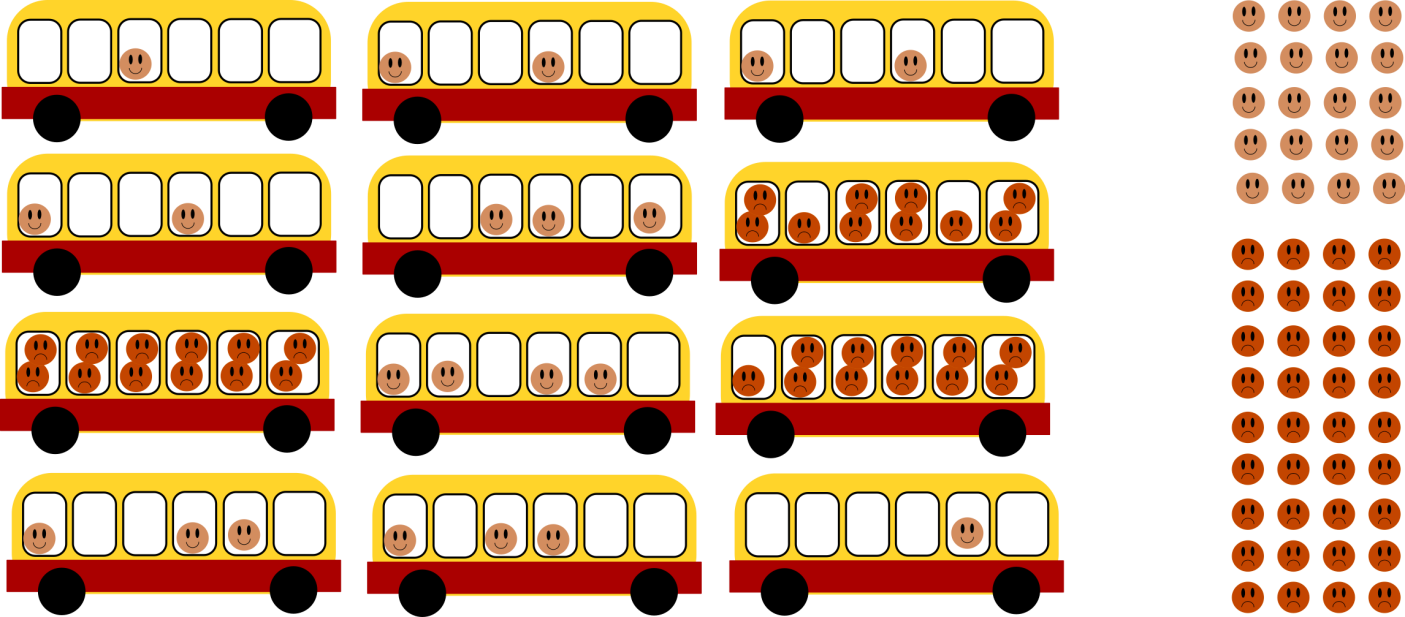
## От велосипедиста к инспектору

Вывод, который делает велосипедист, пыхтя на пониженной передаче: «мир несправедлив и большую часть сил отнимает самая дурацкая часть работы», часто именуют принципом Парето или принципом «80/20». Это абсолютная эмпирика, принцип Парето никто не доказывал, но его так часто цитируют, что он уже производит впечатление истины. Его используют, как оправдание и как инструкцию, обнаруживают в самых разных проявлениях и иногда это работает, например, принципу «80/20» соответствует индекс подлости порядка 0.6 — как для распределения богатства во всем мире.

У принципа Парето есть полезное для понимания более строгое обобщение. Закон подлости, названный в честь безымянного велосипедиста, имеет официальное научное звание: парадокс инспекции. Это хорошо известное явление встречается в самых разных исследованиях, связанных с социологическими опросами, тестированием в теории отказов (разделе прикладной математики, занимающемся надёжностью сложных систем), неявно, но систематически смещая наблюдаемые результаты в сторону более часто наблюдаемых явлений.

Приведём классический пример, с опросом пассажиров общественного транспорта. На линии в день работает множество автобусов. В относительно короткий час пик автобусы переполняются, а всё остальное время они ходят почти пустыми. Если мы станем опрашивать пассажиров, то значительная их часть окажется именно в переполненном автобусе (там попросту больше людей), и получим выражение общего недовольства. Если же мы опросим водителей, то они пожалуются на незаполненность значительной части маршрутов и неразумность начальства, гоняющего их попусту. Гибкий график сгладит ситуацию, но, в любом случае, кривая Лоренца будет отклоняться от кривой равенства, соответствующей невероятной ситуации всегда одинакового числа пассажиров во всех автобусах.

В учебниках по теории вероятностей часто встречается специальный непрозрачный мешок, в который математики складывают разнообразные объекты, а потом наугад вытаскивают, делая, подчас, очень глубокомысленные выводы. Разрешение нашего парадокса состоит в том, что анализируя систему пассажиропотока в целом, мы кладём в мешок автобусы, а проводя опрос, мы достаём из него наугад (инспектируем) пассажиров, и по их данным пытаемся делать выводы. Картинка показывает в чём тут разница:

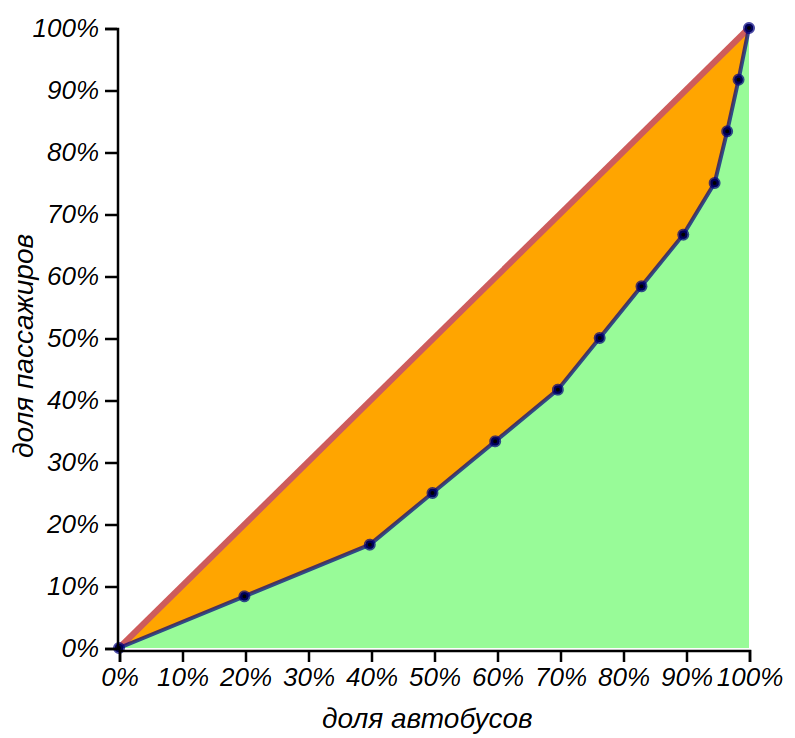


Статистика по автобусам говорит, что 75% из них свободна и ездит впустую. В то же время, опрос пассажиров обнаружит, что 64% пассажиров, проехавших в этот день, оказались в переполненном транспорте.

Давайте рассмотрим эту ситуацию подробнее, построив кривую Лоренца, (на этот раз, настоящую), для числа пассажиров в автобусах, показанных на рисунке. Для этого нужно отсортировать автобусы по числу пассажиров и последовательно суммировать вклад каждого из них в общий пассажиропоток.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Число пассажиров в автобусе | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 4 | 10 | 11 | 12 |
| Кумулятивная сумма пассажиров | 1 | 2 | 4 | 6 | 8 | 11 | 14 | 17 | 21 | 31 | 42 | 54 |
| Кумулятивная сумма автобусов | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |

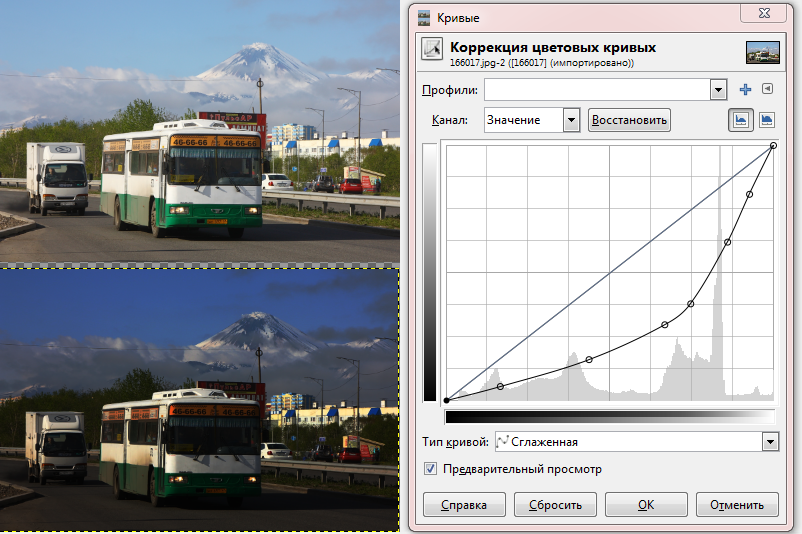
Полученные кумулятивные суммы следует разделить на их максимальные значения, чтобы получить доли, например, в процентах, после чего их можно нанести на диаграмму:



Кривая Лоренца хорошо иллюстрирует наблюдаемую несправедливость ситуации с автобусами: половина автобусов возит лишь четверть всего пассажиропотока, в то время как на 20% перегруженных автобусов приходится половина всех пассажиров.

Кривая Лоренца, в данном случае, показывает как распределение числа элементов в некоторых группах (горизонтальная ось) смещаются при анализе распределения элементов по принадлежности к группам (вертикальная ось). В этом, собственно, и состоит парадокс инспекции: картинка, которую наблюдает инспектор, оказывается искажённой, ведь он анализирует не группы, а элементы групп, а при этом наблюдаемые значения смещаются в сторону более «весомого хвоста» распределения.

Сам по себе, наш закон велосипедиста очень прост, но он то и дело будет усугублять другие законы подлости, добавляя им угрюмую эмоциональную окраску. Размышляя об этом, мне нравится представлять искажение в восприятия мира инспектором в терминах изменения цветовых кривых какого-либо изображения. В растровых графических редакторах мы с помощью инструмента «Кривые» изменяем картинки, смещая распределение числа пикселов по яркости. Вот, например, как меняет восприятие реальности кривая Лоренца, полученная нами для автобусов. Картина мира становится мрачнее, как мы и ожидаем.



Кривая Лоренца из примера, применённая в качестве фильтра «Кривая» в растровом графическом редакторе, делает видимую картину камчатского автобуса мрачнее. Сетуя на то, что автобусы «вечно опаздывают» и «вечно полны народу», утешайтесь тем, что, это всего лишь иллюзия, связанная с парадоксом инспекции!

Парадокс инспекции может проявляться в своей крайности: если среди групп, помещённых в наш теоретический мешок, есть такие, элементы которых не просто редки, но ненаблюдаемы вовсе. В таком случае мы получаем то, что статистики, демографы и публицисты называют систематической ошибкой выжившего.

Чаще всего, эту ошибку демонстрируют на примере с дельфинами, спасающими людей, оказавшихся волею несчастного случая в открытом море. Дельфины, обнаружив любопытный несъедобный объект (человека), играют с ним, подталкивая носом. При этом, они не обязательно толкают его в сторону ближайшего берега. Разумно предположить, что для дельфина берег, да ещё населённый людьми представляет опасность. Однако, если, всё же случается так, что дельфины толкают потерпевшего именно к берегу, в сторону спасения, и он благодаря этому выживает, то весь мир узнаёт: дельфины спасли человека! О поведении дельфинов во всех прочих печальных случаях, увы, мы не узнаем ничего. Так что, если судить лишь по выборке новостей, мы получим существенно искажённую картину.

Об этом явлении часто рассказывают в различных демотивирующих статьях, для начинающих бизнесменов, уверяя их в том, что успешный путь, описываемый в мотивирующих книгах, скорее всего не для них, ибо: «неудачники книг не пишут». Впрочем, к законам подлости это отношения не имеет, тут мы касаемся психологии. Парадокс инспектора и ошибка выжившего, действительно способны искажать восприятие действительности, омрачая её, либо придавая излишне радужную окарску. Но с научной точки зрения, это методические ошибки, допускаемые при получении и обработке данных. К сожалению, они приводят к расхожему мнению о статистике, как о нечестном манипулировании фактическими данными, среди людей весьма далёких от этих методик. О них знать полезно, чтобы избегать их в своей работе и что критически относиться к новостям, слухам и недобросовестным исследованиям.

Мы встретимся с законом велосипедиста и его влиянием ещё не раз: стоя в очереди или на автобусной остановке, наблюдая несправедливость распределения богатства. А кривые Лоренца и индекс подлости позволят нам смело сравнивать между собой возмутительно разные вещи. Математика — точная наука, но никто не запрещает математикам хулиганить. В своём, конечно, кругу и без драк.

1. Dorian M. Raymer and Douglas E. Smith, Spontaneous knotting of an agitated string, PNAS October 16, 2007 104 (42) 16432-16437; [↑](#footnote-ref-2)
2. Подробнее о собственных масштабах и обезразмеривании задачи мы поговорим во второй главе, когда речь пойдёт о бутербродах. [↑](#footnote-ref-3)