

# Проклятие режиссёра и проклятые принтеры

*Мы порассуждаем о цейтнотах, дедлайнах и о невовремя ломающихся принтерах.*

## Стратегия балбеса

В предыдущей главе мы говорили о случайных процессах. Одним из самых простых процессов, требующих минимума дополнительных предположений, является *пуассоновский поток*. Напомню, что его можно реализовать, случайно распределяя известное количество независимых событий по временному интервалу. Хорошими примерами могут быть удары капель дождя по крыше, поток частных автомобилей на дороге, сильные землетрясения и т.п.

А что мы получим, если события перестанут быть независимыми, а будут образовывать упорядоченную цепочку? Скажем, в цепочке  $\{A, B, C\}$  событие  $B$  может случиться только после события  $A$  и перед событием  $C$ , хотя моменты в которые эти события произойдут пусть остаются случайными. Посмотрим, как разместятся такие упорядоченные цепочки на ограниченном временном интервале. Первое событие мы расположим в произвольной точке, второе — тоже случайно, но обязательно позднее первого, третье — после второго и так далее. Для каждого следующего этапа будет оставаться всё меньше и меньше времени, так что к правой части интервала (перед дедлайном) должно наблюдаться заметное увеличение интенсивности процесса. Рано или поздно, время для выполнения задач закончится и цепочка завершится. Назовём построенный нами процесс *стохастической цепочкой с дедлайном*, а выбранную безалаберную стратегию выполнения работы *стратегией балбеса*. На рисунке показан пример построенной таким образом цепочки из 5 этапов работы, на которую было отпущено 10 дней.



*Пример стохастической цепочки с дедлайном. В данном случае пять дел сделать удалось, можно ещё успеть сделать шестое, а на семь времени уже не хватит.*

Понятно, делая дела как попало, непросто успеть в срок, но можно ли как-то проанализировать это явление? Сформулируем задачу, взяв в качестве испытуемого, скажем, театрального режиссёра. Пусть в распоряжении режиссёра и труппы имеется  $n$  дней для постановки некоего действия. Подготовка разбивается на  $k$  последовательных репетиционных этапов, каждый из которых требует один день на выполнение. Какова вероятность не уложиться в срок, реализуя описанный нами процесс выполнения работ? Если подготовка мероприятия требует вовлечения разных людей и различных производственных процессов, то возможны накладки, болезни или попросту хандра — все предпосылки к реализации нашей стохастической цепочки с дедлайном.

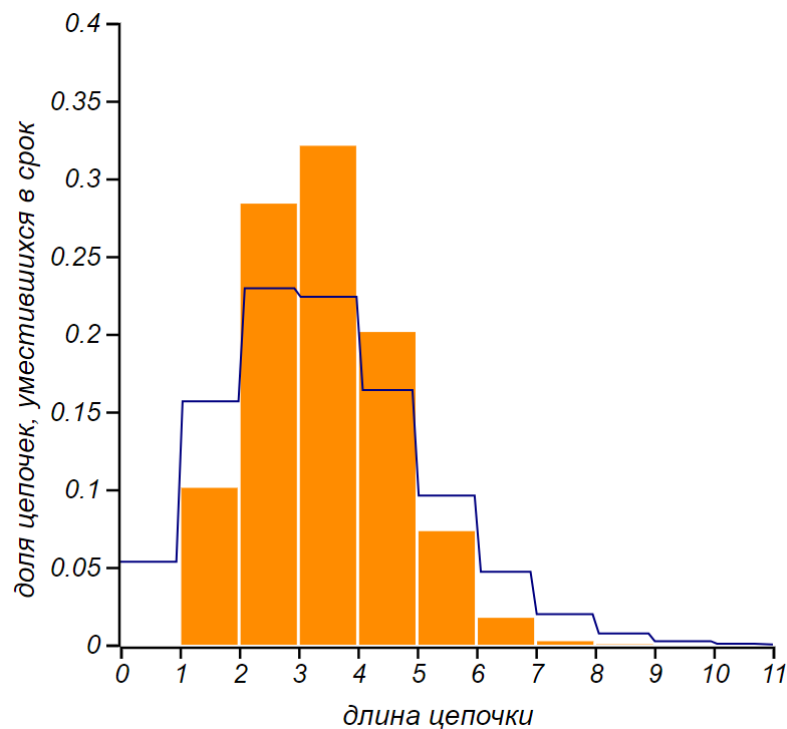
Для начала, я обратился к имитационному моделированию, чтобы выяснить, как распределяется длина цепочек, которые удаётся выполнить в ограниченный промежуток времени заданной длины, пользуясь стратегией балбеса. Вычисления состояли в генерации стохастических цепочек и в подсчёте их длин для различных ограничений по времени по следующему алгоритму:

```

Вход: число дней  $n$ 
Повторять, пока не набрано нужное число цепочек
     $x := n$ 
     $k := 0$ 
    Повторять, пока  $x > 0$ 
        выбрать случайное целое число  $x \sim \text{Uniform}([0, x])$ 
        увеличить счётчик  $k$ 
    конец
    добавить  $k$  в гистограмму
конец

```

Вот такая гистограмма получается, например, для  $n = 10$ :



*Гистограмма функции вероятности для длины цепочек, которые удаётся сделать в отведённый срок. Синей линией показано распределение Пуассона с интенсивностью, соответствующей наблюдаемой средней длине цепочек.*

Подсчитывая события в настоящем пуассоновском потоке с интенсивностью  $\lambda$ , мы придём к известному распределению Пуассона:

$$p(k) = \frac{e^{-\lambda} \lambda^k}{k!},$$

которое описывает вероятность получить ровно  $k$  событий в единичном интервале времени. Полученное мною распределение внешне похоже на распределение Пуассона, но оказалось всё же, что это не оно. Давайте разберёмся, откуда взялись именно такие доли.

Шансов не успеть сделать одно дело нет совсем, поскольку для одного дела время найдётся обязательно, пусть даже и в последний день. Короткие цепочки из двух дел устроены следующим образом: в последний день должно быть сделано второе дело (чтобы ограничить длину цепочки), а на расположение первого дела ограничений нет, так что вероятность для  $k = 2$  равна  $1/n$ . Дальше можно действовать индуктивно. Для произвольного  $k$ , последний этап обязательно должен быть выполнен в последний день, это может случиться с вероятностью  $1/n$ . После чего мы можем поместить предпоследнее дело в любой из свободных дней, скажем в день номер  $m$ , сведя при этом задачу к случаю  $k - 1$  дел и  $n - m$  дней. Выбор  $m$  ограничен снизу числом  $k-2$ , поскольку два дела – последнее и предпоследнее уже «сделаны». Таким образом, мы получаем способ получить точное решение искомой задачи, но для этого нужно знать решения всех входящих в неё подзадач:

$$\begin{aligned} p_n(1) &= 0 \\ p_n(2) &= \frac{1}{n} \\ p_n(k) &= \frac{1}{n} [p_{n-1}(k-1) + p_{n-2}(k-1) + \dots + p_{k-2}(k-1)]. \end{aligned}$$

Такое определение функции называется *рекуррентным*. Для того, чтобы им можно было воспользоваться, необходимо знать решение некоторых *базовых подзадач*, в нашем случае, это выражения для  $k = 0$  и  $1$ . Полученное нами рекуррентное соотношение позволяет вычислить распределение, но его трудно анализировать. Нужно превратить его в *конечную форму*, то есть формулу, содержащую конечное фиксированное число арифметических действий над хорошо известными функциями. Мне удалось получить такую форму, оказавшуюся весьма компактной:

$$p_n(k) = \frac{1}{n!} \begin{bmatrix} n \\ k \end{bmatrix}.$$

здесь конструкция  $\begin{bmatrix} n \\ k \end{bmatrix}$  обозначает так называемые *числа Стирлинга первого рода*, они возникают в комбинаторике при подсчёте циклических перестановок. Пользуясь правом первооткрывателя, назову это распределение именем Джеймса Стирлинга, который ввёл в обиход эти числа. По правде говоря, числа Стирлинга тоже вычисляются рекуррентным соотношением:

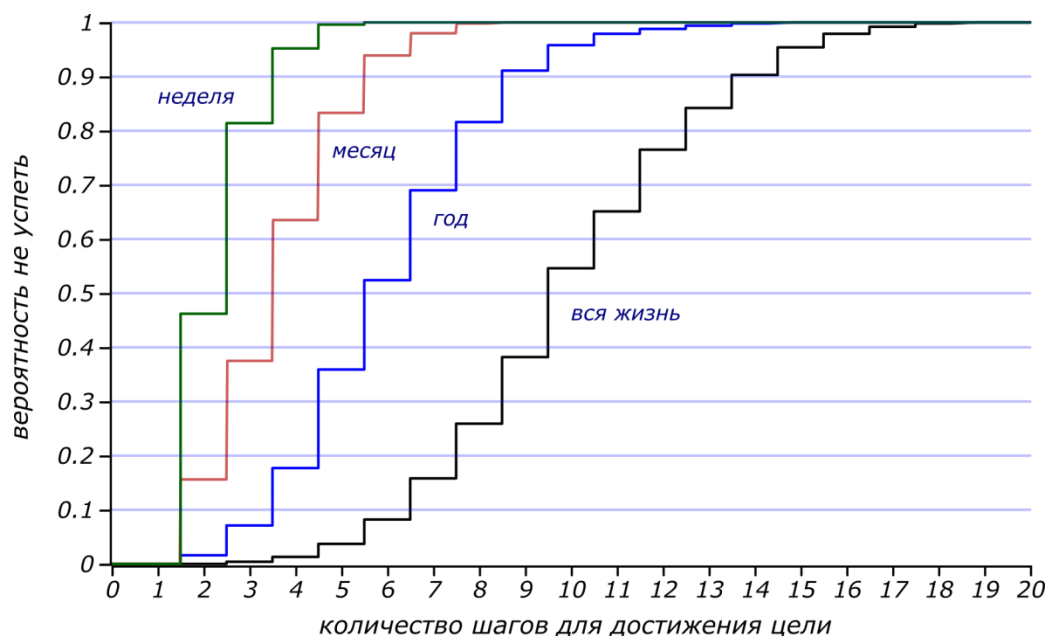
$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = 1, \quad \begin{bmatrix} 0 \\ k \end{bmatrix} = \begin{bmatrix} n \\ 0 \end{bmatrix} = 0, \quad \begin{bmatrix} n \\ k \end{bmatrix} = (n-1) \begin{bmatrix} n-1 \\ k \end{bmatrix} + \begin{bmatrix} n-1 \\ k-1 \end{bmatrix},$$

но они используются с середины XVIII века и достаточно широко, чтобы можно было счесть их «хорошо известными». Самое главное, хорошо известны их свойства, позволяющие анализировать полученное решение. Благодаря этому удалось получить точные выражения для математического ожидания длины цепочек и её дисперсии, собственно, ради вычисления этих значений я и исследовал получившееся распределение:

$$M[k] = H_n, \quad D[k] = H_n - H_{n,2}.$$

Эти величины выражаются через очень интересные *гармонические числа*:  $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n}$ , или в конечной форме  $H_n = \frac{1}{n!} \left[ \frac{n+1}{2} \right]$ , а  $H_{n,2} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \dots + \frac{1}{n^2}$ . Эти числа играют важную роль в такой удивительно сложной области математики, как *теория чисел*. Казалось бы, что может быть проще, чем изучение чисел, а тем более, целых чисел? Арифметику проходят в школе, со свойствами чисел, такими как делимость, мы знакомимся на личном опыте, пытаюсь честно разделить пять рублей на троих. Но именно эта область математики ставит перед исследователем чрезвычайно сложные проблемы. Одна великая теорема Ферма чего стоит! От гармонических чисел дорожка ведёт к *дзета-функции Римана*, а от неё — к великой загадке распределения простых чисел. Нам не потребуются результаты теории чисел явным образом, но свойства гармонических чисел мы используем. Средняя длина цепочек с ростом  $n$  растёт очень медленно, хоть и неограниченно: имея бесконечное время, можно в среднем успеть сделать бесконечное число дел. Не сильно ошибившись, можно сказать, что она растёт логарифмически. В свою очередь, дисперсия не сильно отличается от среднего, а добавочный коэффициент  $H_{n,2}$  стремится к константе  $\frac{\pi^2}{6}$ . Немного позже нам пригодится это наблюдение.

На наш вопрос: «Какова вероятность, не уложиться в  $n$  дней, имея перед собой  $k$  последовательных этапов выполнения задачи?» поможет ответить функция распределения, то есть, кумулятивная кривая для распределения Стирлинга. Построим такие кривые для  $n = 7, 30, 365$  и  $25000$ , соответствующие неделе, месяцу, году и (конечно, условно) всей жизни.



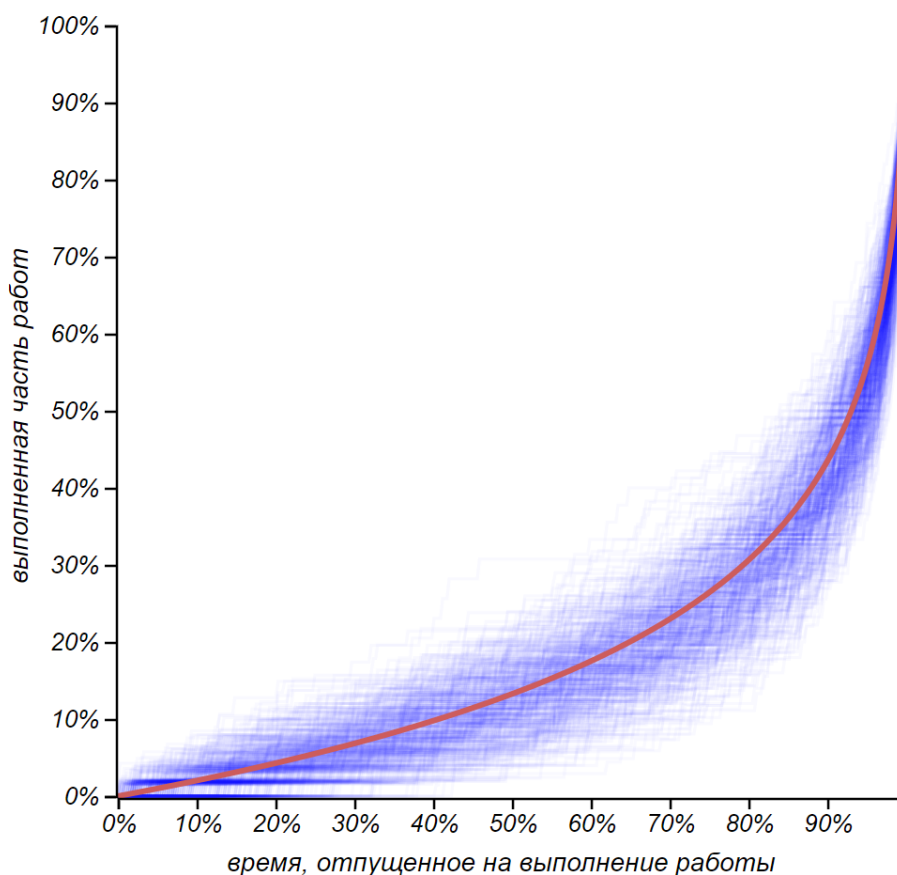
*Вероятность не успеть выполнить цепочки различной длины в тот или иной срок.*

Эти графики показывают, что вероятность не уложиться в месяц с заданием, имеющим 5 шагов, превышает 80%. И что неорганизованному балбесу в неделю лучше не планировать более трёх дел, ну, а десяток дел он не сделает, с вероятностью, превышающей

50%, и за всю жизнь! Мы убеждаемся в том, что при увеличении сроков на несколько порядков, число дел, выполнимых как попало, увеличивается незначительно. Жизнь так коротка!

### Быстрее, ещё быстрее!

Давайте теперь исследуем само явление цейтнота, его выматывающие свойства. Для этого обратимся к методу Монте-Карло и построим несколько тысяч стохастических цепочек, после чего усредним их, получив ожидаемый *темп выполнения работы*.



*Множество стохастических цепочек с дедлайном и ожидаемый темп выполнения работы.*

Обратите внимание на то, что оси графика приведены к общему числу дел и всему отпущенному времени. Это, с одной стороны, позволяет нам сравнивать, как разные сроки, так и различные по длине цепочки, а с другой — мы опять получили что-то подобное кривой Лоренца: некое формализованное отражение несправедливости.

Наблюдаемый темп, увы, сильно неравномерен: в первую половину срока будет сделано едва ли 10% работы, а добрую половину всех дел придётся выполнять, имея в своём распоряжении мене 10% времени. Но главная особенность: темп, вернее его наклон, стремительно увеличивается при приближении к дедлайну! Мы получили модель предновогоднего ража или паники в преддверии годового отчёта, а также нащупали закон подлости, знакомый всякому, кому приходилось организовывать концерт, костюмированный вечер или иное мероприятие:

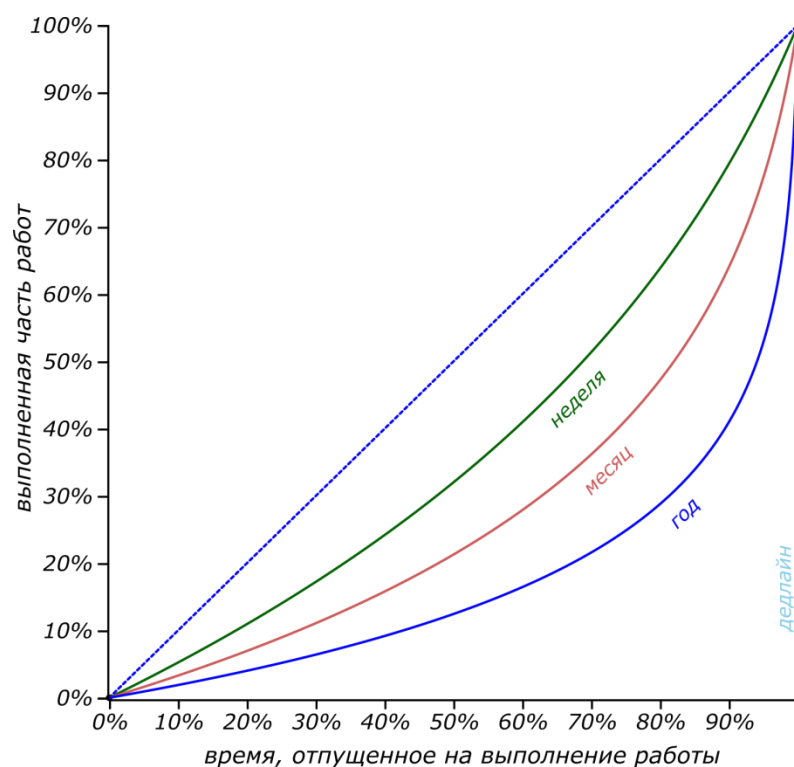
**Сколько бы времени ни было отпущено на подготовку мероприятия, большая часть дел останется на последнюю ночь!**

Прекрасные живые примеры таких процессов описаны, например, в рассказах Карела Чапека "Как делают газету" и "Как ставится пьеса". Неужели причина этого проклятия только в нашей неорганизованности и безалаберности? Это, конечно основные причины, но мы не настолько в ней виноваты, чтобы нельзя было попробовать оправдаться каким-либо математическим законом. Стратегия балбеса, конечно, выглядит глупо, но экспоненциальный рост темпа — это не шутки! Можно ли вообще с ним справиться?

Ожидаемый темп выполнения работы можно вычислить точно. Формула не слишком изящна, однако примечательно, что в неё входит число дней  $n$  и не входит число запланированных дел:

$$T_n(x) = -\frac{1}{H_n + 1} \log_2[1 - x(1 - 2^{-H_n - 1})].$$

Логарифм функция медленная, если только его не прижать к стенке. В последние дни перед дедлайном темп растёт катастрофически, с такой же скоростью, с которой логарифм проваливается в бездну при приближении к нулю. Однако, от числа выделенных дней он, всё же зависит. Можно посмотреть на то, как выглядит ожидаемый темп для недели, месяца, и года:



*Наиболее вероятный темп выполнения работы в ограниченный срок. Интересно, что жёсткое ограничение по времени влияет благотворно. Имя в запасе всего неделю, мы, скорее всего, станем выполнять работу равномернее (к половине срока будет готова треть работы), а если впереди целый год, то можно и расслабиться, ну, а потом об этом пожалеть.*

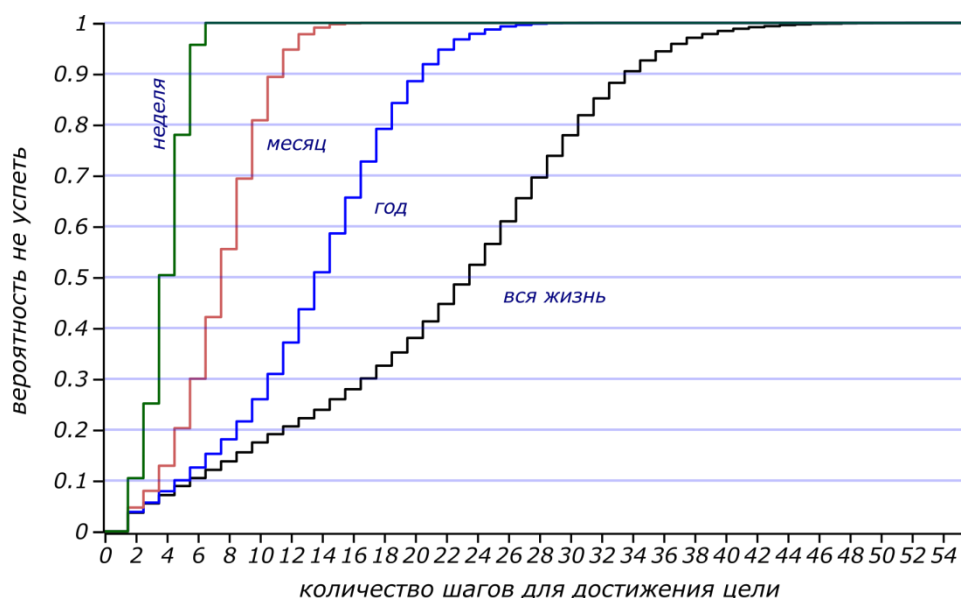
У идеального исполнителя-перфекциониста, который выполняет работу в точности равномерно, темп выполнения должен стремиться к диагонали (синяя пунктирная линия на рисунке). Это похоже на кривую равенства на диаграмме Лоренца, знаменующую справедливость. Подобно тому, как мы вычисляли коэффициент Джини для диаграммы Лоренца, мы можем, основываясь на площади между кривой темпа выполнения работ и идеальной кривой, вычислить некий коэффициент подлости, который покажет, насколько мы далеки от идеала. Он зависит от длины выделенного срока и потихоньку увеличивается с ростом  $n$ . В приведённых нами примерах для недели, месяца и года, коэффициент подлости равен, соответственно 0.37, 0.49 и 0.63.

Этот индекс растёт с ростом  $n$  очень медленно, но если устремить число дней к бесконечности, он будет стремиться к единице. Итак, мы приходим к парадоксальному, но по-своему, красивому результату: имея в своём распоряжении бесконечное время, балбес может запланировать выполнить бесконечное число дел, однако ожидаемый темп выполнения будет иметь вид дельта-функции. Это значит, что *почти наверняка* он не выполнит ничего из запланированного, отложив все дела на бесконечное будущее. Вспоминаются привычные сетования: *«Целое лето (каникулы, жизнь) пролетело, а я так ничего и не успел!»* Что же, даже этому есть математическое объяснение.

Даосы в Китае очень крепко размышляли о вечной жизни, и делали они это очень грамотно: наряду с упражнениями тела, необходимыми для решения такой задачи, они занимались упражнениями ума, дабы приспособить его к вечному существованию. Как видно, вечная жизнь требует большой дисциплины, иначе даже вечность весьма вероятно потратить впустую.

### Мостим дорогу благими намерениями

Но как же бороться с нарастающей волной забот и цейтнотом? Можно, например, взять себя в руки. Человек с синдромом отличника может стремиться сделать следующее дело как можно раньше, насколько это возможно, конечно. Правдоподобной моделью будет выбор момента для выполнения следующего дела, следуя экспоненциальному распределению с интенсивностью, обратно пропорциональной оставшемуся времени. Это не исключит некоторой неопределённости, присущей нашей жизни, но выразит благие стремления делать все дела как можно скорее. Назовём эту стратегию *стратегией благих намерений*. Вот какими будут распределения вероятностей выполнения заданий в срок для приверженца этой стратегии, который в половине случаев сделает очередное дело в первую четверть оставшегося времени:

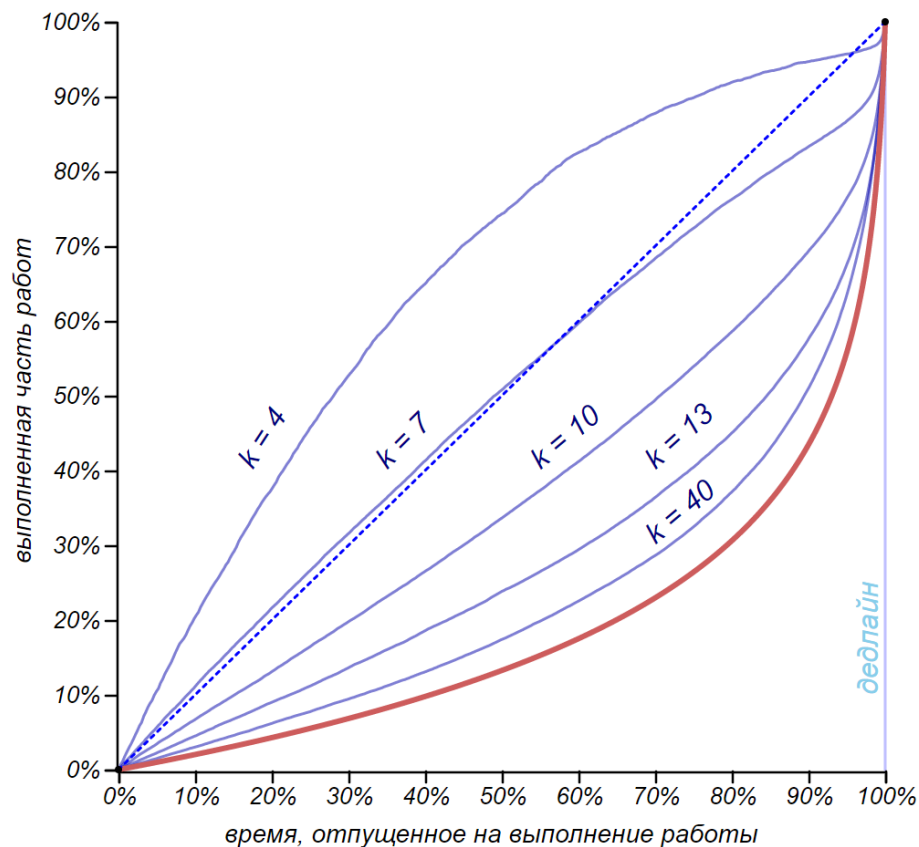


*Распределение вероятности не успеть в срок для стратегии благих намерений.*

Существенно лучше! В течение недели можно с неплохой вероятностью успеть сделать пять дел, и оставить себе два выходных дня. Но, всё же, для больших периодов увеличение возможностей не революционное. Проблема кроется в том, что ожидаемое число успешно завершаемых дел всё равно остаётся пропорциональным логарифму отпущенного времени, а логарифм растёт чрезвычайно медленно! Так что, планируя многое, нужно иметь в виду, что интенсивность процесса будет неизбежно возрастать, и времени в преддверии дедлайна, скорее всего, будет не хватать. В любом случае, необходимо помнить, что жизнь коротка и чтобы успеть реализовать задуманное, нужно действовать прямо сейчас!

Давайте полюбуемся на темп благонамеренного отличника.

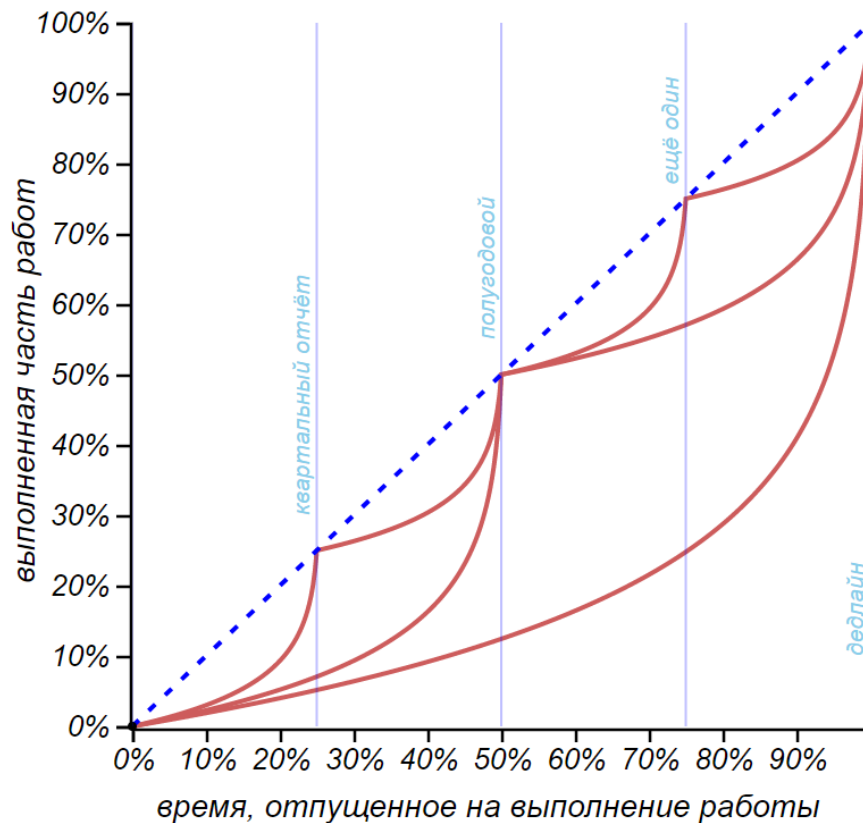




*Ожидаемый темп выполнения работы методичным человеком, старающимся приступить к следующему этапу работы как можно скорее. На графиках представлены результаты усреднения десятка тысяч численных экспериментов, моделирующих выполнение задания с фиксированным числом этапов. Красной линией обозначен предельный темп для большого числа задач.*

Нашему аккуратисту удалось более равномерно распределить работу, и сделать существенно больше дел, но его всё равно ожидает цейтнот. Короткие цепочки такой человек будет выполнять с существенным перевыполнением плана, а цепочку из семи дел — практически идеально. Однако по мере увеличения числа дел, ожидаемый темп быстро стремится к теоретическому темпу, полученному с помощью стратегии балбеса! Увеличилась общая производительность, но запарка перед самым дедлайном никуда не делись. Так что нагружая можно доканать и заправского зануду!

Впрочем, существует ещё один широко известный способ существенно дисциплинировать выполнение работ: *вместо одного дедлайна надо сделать их много*. Давайте разобьем срок выполнения работы на две равные части и будем придерживаться этого нового дедлайна, считая его, скажем, промежуточным отчётом. Для каждой из этих частей мы можем построить кривую ожидаемого темпа выполнения работ, как показано на рисунке.

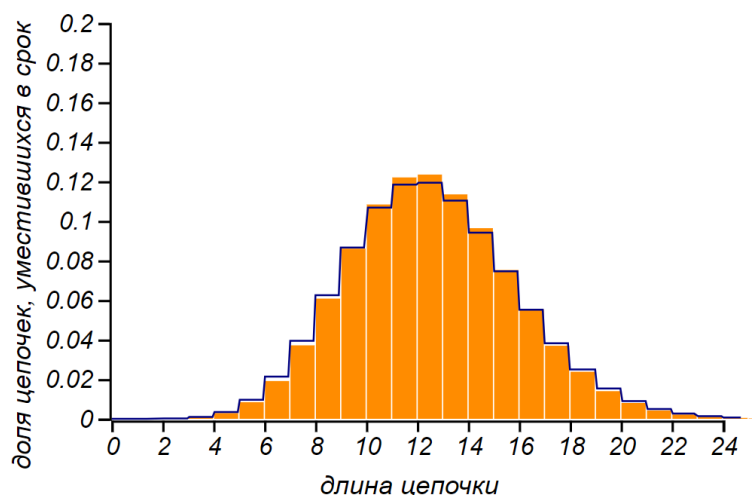


*Разбиение времени выполнения работы на несколько промежуточных отчётных периодов позволяет выполнить работу более равномерно, но добавляет стресс при приближении каждого нового отчёта.*

Несмотря на нервотрёпку с промежуточным отчётом, мы достигли своей цели: площадь под общей кривой темпа выполнения сократилась, и коэффициент подлости уменьшился от 0.65 до 0.3. Кроме того, сокращение срока (вместе с сокращением числа дел, разумеется) приближает ожидаемый темп выполнения работы к идеальному темпу, поэтому коэффициент подлости уменьшился более чем в два раза. Добавление ещё двух, скажем, квартальных отчётов, уменьшат его уже до 0.13, но тем самым мы вгоним наших исполнителей сразу в четыре стрессовых периода, и они всё равно станут громко страдать, жалуюсь на судьбу и на начальство! Что же, мы можем показать работникам наши выкладки и доказать, что введя ежеквартальную отчётность, в пять раз понизили коэффициент подлости их жизни, если это, конечно, послужит им утешением. Более того, при стремлении количества промежуточных дедлайнов к числу дней, отпущенных на работу, темп выполнения работы приблизится к идеальному, но очень занудному темпу.

### **Ну вот! Ещё и принтер сломался!**

Добавим ещё пару слов о стратегии балбеса. Числа Стирлинга при увеличении  $n$  имеют асимптотическое разложение, которое сводит распределение длин цепочек с дедлайном к смещённому распределению Пуассона. По существу, они становятся неотличимы друг от друга, когда, согласно центральной предельной теореме, стремятся к нормальному распределению.



*Распределение Стирлинга и Пуассона для  $n = 100000$  становятся очень близки друг к другу.*

Таким образом, наш стохастический процесс с дедлайном можно рассматривать либо как пуассоновский процесс на сгущающейся временной сетке, либо как неоднородный пуассоновский процесс, интенсивность которого монотонно и стремительно растёт. И хотя, строго говоря, наш процесс не является пуассоновским, поскольку события в нём не независимы, однако, нужные нам статистические свойства у них схожи. Об их схожести говорит и подмеченная ранее близость среднего значения и дисперсии распределения Стирлинга, характерная именно для пуассоновского распределения.

Этот вывод позволяет задать вопрос: что если добавить к построенному нами процессу выполнения цепочки дел какие-либо независимые от нас редкие неприятности: пургу, жуткую пробку, насморк, поломку принтера, или всенародный праздник?

Для пуассоновского процесса определён *процесс случайного прореживания*, заключающийся в удалении из потока событий, с какой-то известной вероятностью. Случайное прореживание с вероятностью  $(1 - p)$  оставляет процесс пуассоновским, но его интенсивность уменьшается, умножаясь на  $p$ . События, соответствующие совпадению неприятности и какого-либо этапа выполнения работы сами образуют пуассоновский процесс, с существенно меньшей интенсивностью, но в нашем случае, также, монотонно и стремительно растущей. Так стремительно, что какой бы малой ни была вероятность неприятности, для достаточно большого числа дел (или срока, отведённого на работу), ближе к дедлайну она может увеличиться до вполне наблюдаемой. И принтер забарахлит именно накануне сдачи курсовика! Разумеется, это работает для достаточно длинных цепочек.

\*\*\*

Не удивляйтесь, если автобус сломается именно тогда, когда вы уже опаздываете. Автобус не желает вам зла. Просто, если вы девушка, то последовательность дел: выбрать платье, съесть конфетку, умыться, надеть выбранное платье, накраситься, надеть цепочку, переложить вещи из сумочки в клатч, почистить туфли и прочее и прочее подходит к самому главному и волнительному дедлайну — к свиданию! И темп с которым вы летите навстречу судьбе уже такой сумасшедший, что начинают происходить самые маловероятные чудеса. В конце концов, а что же такое чудо, как не реализация невероятного!