ПРИКЛАДНАЯ СТАТИСТИКА

Посиделка 8: большая сила o-малых

Ульянкин Ппилиф [*]

> I am not in danger, Skyler. I am the danger! A guy opens his door and gets shot and you think that of me? No. I am the one who knocks!
>
> *Walter White, Breaking Bad (2008-2013)*

В этой посиделке мы будем варить распределения, а также выведем из аксиом распределения Пуассона и эксоненциальное распределение. Называйте меня до её конца мистером Хайзенбергом, а я буду называть каждого из вас Джесси.

1 Варка распределений

Давайте попробуем немного поработать с различными непрерывными распределениями. Пусть случайная величина X имеет экспоненциальное распределение, $Exp(\alpha)$. То есть её функция распределения:

$$F_X(x) = \begin{cases} 1 - e^{-\alpha x}, & \text{если } x \geqslant 0 \\ 0 & \text{если } x < 0 \end{cases}$$

А плотность:

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x}, & \text{если } x \geqslant 0 \\ 0, & \text{если } x < 0 \end{cases}$$

[*] https://github.com/FUlyankin/yet_another_matstat_course

Обычно с помощью такого распределения моделируют время между событиями. Например, время до следующего лайка под фотографией или время до прихода следующего человека в очередь. Почему именно с помощью него, мы узнаем чуть ниже.

Сейчас же давайте посмотрим на случайную величину $Y = \sqrt{X}$. Нам нужно найти её плотность распределения. Разберёмся с областью определения новой случайной величины. Все значения X лежат на полуинтервале $x \in [0; +\infty)$. Получается, что для Y будет выполнятся $y \in [\sqrt{0}; \sqrt{+\infty}) = [0; +\infty)$.

Чтобы найти распределение Y воспользуемся определением функции распределения:

$$F_Y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}(\sqrt{X} \leqslant y) = \mathbb{P}(X \leqslant y^2) = F_X(y^2) = \begin{cases} 1 - e^{-\alpha y^2}, & \text{если } y \geqslant 0 \\ 0, & \text{если } y < 0 \end{cases}$$

Вспоминаем, что $f_Y(y) = F'_Y(y)$ и добиваем задачу:

$$f_Y(y) = \begin{cases} 2 \cdot \alpha \cdot y \cdot e^{-y^2}, & \text{если } x \geqslant 0 \\ 0, & \text{если } x < 0 \end{cases}$$

Ровно по такой же схеме можно пойти при работе с абсолютно любой функцией от случайной величины.

Посмотрим на ещё один пример $Y = 1 - e^{-\lambda X}$. В качестве функции от X мы берём её функцию распределения. Найдём область распределения новой случайной величины. Значения X лежат на полуинтервале $x \in [0; +\infty)$. Найдём для нашей функции пределы при $x \to 0$ и $x \to +\infty$. Получим область определения для Y, $y \in [0; 1]$.

Воспользуемся определением функции распределения

$$F_Y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}(1 - e^{-X} \leqslant y).$$

Аккуратно раскроем неравенство под знаком вероятности

$$\begin{aligned} 1 - y &\leqslant e^{-X} \\ \ln(1 - y) &\leqslant -X \\ X &\leqslant -\ln(1 - y) \end{aligned}$$

Получим, что

$$\mathbb{P}(1 - e^{-X} \leqslant y) = \mathbb{P}(X \leqslant -\ln(1 - y)) = F(-\ln(1 - y)) = 1 - e^{\ln(1-y)} = 1 - 1 + y = y.$$

То есть

$$F_Y(y) = \begin{cases} 1, \text{ если } y > 1 \\ y, \text{ если } y \in [0; 1] \\ 0, \text{ если } y < 0. \end{cases}$$

Взяв производную получим плотность распределения

$$f_Y(y) = \begin{cases} 1, \text{ если } y \in [0; 1] \\ 0, \text{ если } x \notin [0; 1]. \end{cases}$$

Эта плотность соответствует равномерному распределению на отрезке $[0; 1]$. Выходит, что $Y = F_X(X) \sim U[0; 1]$. Это неслучайный факт. Его обычно называют **квантильным преобразованием**. Именно про него мы подробнее поговорим в следующем разделе.

Здесь остаётся сказать, что приём, который мы использовали для получения новых распределений, довольно часто помогает получать распределения разных случайных величин. С его помощью можно получить распределение максимума и минимума из n случайных величин. С помощью него можно вывести формулы для получения плотностей суммы, произведения и частного случайных величин.

2 Квантильное преобразование

Теорема 1. Пусть функция распределения $F_X(x)$ непрерывна. Тогда случайная величина $Y = F(X)$ имеет равномерное распределение на отрезке $[0; 1]$.

Доказательство. Чтобы доказать этот факт, просто-напросто воспроизведём логику из предыдущего раздела, но в более общем случае. Начнём с области определения Y. Это всегда отрезок $[0; 1]$, так как именно на этом отрезке $F_X(x)$ принимает свои значения.

Если функция распределения на всей своей области определения возрастает, значит она обратима, тогда:

$$F_Y(y) = \mathbb{P}(F(X) < y) = \mathbb{P}(X < F^{-1}(y)) = F(F^{-1}(y)) = y, \text{ если } x \in (0, 1).$$

При этом мы знаем, что функция распределения $F(y) = y$ соответствует равномерному на отрезке $[0; 1]$ распределению.

Если функция F не является всюду возрастающей, то у неё есть участки постоянства. В этом случае просто обозначим через $F^{-1}(y)$ самую левую точку из замкнутого множества $\{t \mid F(t) = x\}$. При таком понимании обратной функции все равенства, перечисленные выше, остаются справедливы. □

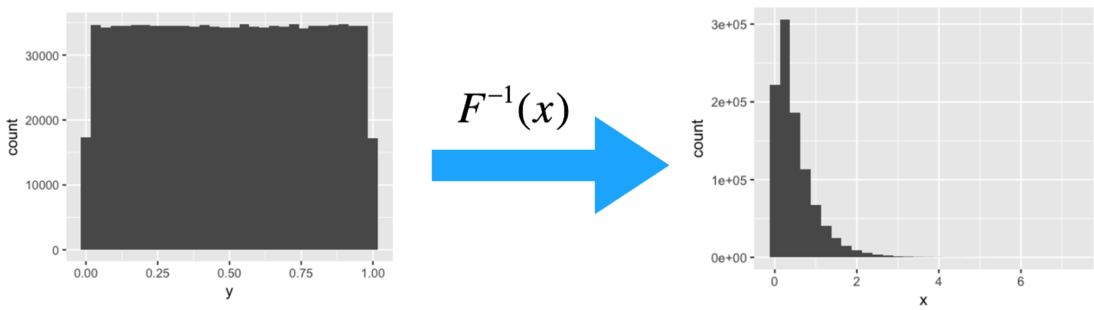
Следствие 1. Пусть $Y \sim U[0;1]$, а $F(x)$ произвольная функция распределения. Тогда случайная величина $X = F^{-1}(Y)$ будет иметь функцию распределения $F(x)$.

Это следствие разрешает нам варить на компьютере любые случайные величины, для которых мы знаем $F_X(x)$. Нужно просто сгенерировать выборку из равномерного распределения, а затем применить к ней преобразование $F^{-1}(y)$.

Например, если мы хотим получить выборку из экспоненциального распределения, $Exp(\alpha)$, мы можем сделать это следующим образом:

1. Генерируем выборку $y_1, \ldots, y_n \sim i.i.d. U[0;1]$
2. Находим выборку $x_1, \ldots, x_n$ по формуле:

$$x_i = F_X^{-1}(y_i) = -\frac{1}{\alpha} \ln(1 - y_i).$$



Для нормального распределения такой приём не сработает из-за того, что для него нет аналитической формулы для функции распределения из-за не берущегося интеграла. Для него приходится придумывать другие алгоритмы генерации. По аналогии для всех случайных величин, для которых функцию распределения нельзя записать в аналитическом виде, также придумывают свои методы генерации. На них мы посмотрим позже.

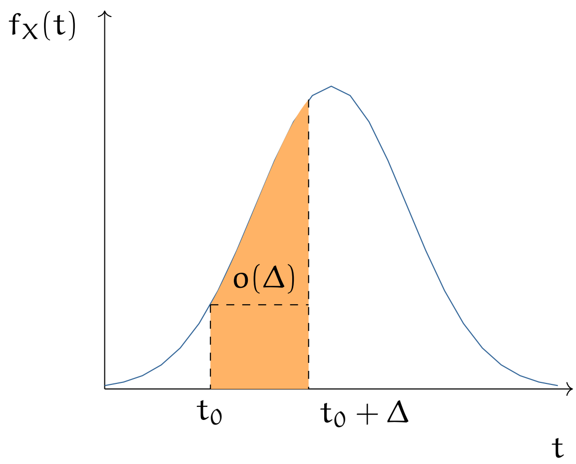
3 Большая сила о-малых

Символом о-малое обозначают любую бесконечно малую функцию $o(f(t))$ по сравнению с заданной функцией $f(t)$ при аргументе, стремящемся к некоторому конечному или бесконечному числу.

Фраза «функция $f(t)$ является о-малым от функции $g(t)$ в окрестности точки t_0» означает, что с приближением t к t_0 функция $f(t)$ уменьшается быстрее, чем $g(t)$. То есть отношение $\frac{g(t)}{f(t)}$ стремится к нулю.

Математический анализ — это тонкое искусство забивать. Обычно с помощью o-малых записывают то, на что можно забить. Например, пусть случайная величина X имеет плотность распределения $f(t)$. Нам нужно примерно найти вероятность $\mathbb{P}(X \in [t_0, t_0 + \Delta])$.

Предположим, что плотность выглядит как-то так:



Вероятность попасть в отрезок — это площадь под плотностью. Величина $\Delta \cdot f(t_0)$ — часть этой площади. Понятное дело, что это не вся площадь. Есть кусочек, который мы не учли. При маленьком Δ неучтённый кусочек будет очень маленьким и мы можем мм пренебречь. Обычно этот факт записывают с помощью o-малой

$$\mathbb{P}(X \in [t_0, t_0 + \Delta] = \Delta \cdot f(t_0) + o(\Delta).$$

Если мы откинем её, точность немного пострадает. Обращаю ещё раз ваше внимание на то, что так можно делать только при очень маленьких Δ.

Несколько важных свойств:

- Сколько o-малые не складывай, всё-равно получишь o-малую

$$o(\Delta) + o(\Delta) = o(\Delta).$$

- Сколько не умножай o-малую на число, больше она не станет

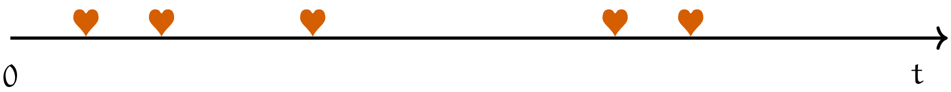
$$o(5\Delta) = o(\Delta), \quad -o(\Delta) = o(\Delta).$$

Дальше большая сила o-малых поможет нам аккуратно построить пуассоновский поток.

4 Пуассоновский поток из воздуха

Поговорим про распределение Пуассона. Обычно с помощью него моделируют простейший поток событий. Правда не очень понятно почему. Давайте по аналогии с нормальным распределением, выпишем несколько аксиом и затем попробуем получить из них распределение Пуассона.

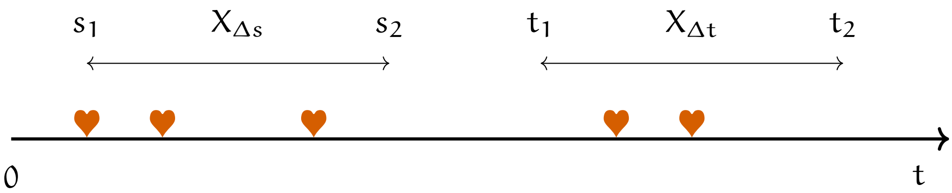
Даша выложила свою фотку и ждёт лайков. Лайки прилетают. Пусть X_t — число лайков, которое она получила за время $[0; t]$. Это дискретная случайная величина. В свою очередь, Y_n — время, которое прошло между $n-1$ и n лайками.



Чтобы понять как именно распределены X_t и Y_n введём несколько аксиом о том, как ведёт себя поток лайков.

PA-0 За нулевой промежуток времени мы всегда получаем ноль лайков, то есть $\mathbb{P}(X_0 = 0) = 1$.

PA-1 **Отсутствия после действия:** вероятность появления k событий на любом промежутке времени не зависит от того, сколько событий произошло до этого.
То есть, если у нас есть промежутки Δt и Δs, то $X_{\Delta t}$ и $X_{\Delta s}$ не зависят друг от друга. Случайная величина $X_{\Delta t}$ описывает сколько лайков мы получили за промежуток времени Δt.



PA-2 **Стационарность:** появление k событий на каком-то промежутке зависит только от числа k и от длины этого промежутка. Точка начального отсчёта не имеет значения.
То есть, интенсивность лайков постоянна во времени. Если у нас было два промежутка длиной в час, в разное время суток, то на этих двух промежутках появление k лайков равновероятно.

PA-3 **Ординарность:** наступление более одного события этого потока за бесконечно малый промежуток времени является практически невозможным.
Если Δt очень маленькое, в течение него вероятность получить 1 лайк пропорциональна Δt

$$\mathbb{P}(X_{\Delta t} = 1) = \lambda \cdot \Delta t + o(\Delta t).$$

Вероятность получить один лайк линейно зависит от промежутка времени.Она существенно выше вероятности получить 2 лайка и более

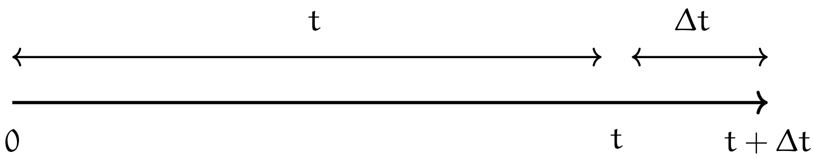
$$\mathbb{P}(X_{\Delta t} \geqslant 2) = o(\Delta t).$$

Под $o(\Delta t)$ в лучших традициях матана подразумевается что-то очень маленькое. То, на что можно легко забить. Величину λ обычно называют **интенсивностью потока событий.** Вероятность получить лайки определяется ей и промежутком времени. Поток событий, который подчиняется перечисленным аксиомам, называют **простейшим потоком событий.**

Давайте найдём отталкиваясь от этих аксиом $\mathbb{P}(X_t = k)$. Для начала выпишем вероятность того, что за очень маленький промежуток времени мы не получим ни одного лайка. В этом нам поможет аксиома $\boxed{\text{PA-3}}$

$$
\begin{aligned}
\mathbb{P}(X_{\Delta t} = 0) = 1 - \mathbb{P}(X_{\Delta t} = 1) - \mathbb{P}(X_{\Delta t} = 2) - \mathbb{P}(X_{\Delta t} = 3) - \ldots = \\
= 1 - (\lambda \cdot \Delta t + o(\Delta t)) - o(\Delta t) - o(\Delta t) - o(\Delta t) - \ldots = 1 - \lambda \cdot \Delta t + o(\Delta t).
\end{aligned}
$$

Давайте попробуем рассмотреть число событий, которое произошло за промежуток времени $[0; t]$ и число событий, которое произошло за промежуток времени Δt. **Обозначим вероятность того, что за промежуток $[0; t]$ не произойдёт ни одного лайка как $h_0(t) = \mathbb{P}(X_t = 0)$.** Нам хочется понять как связаны между собой $h_0(t)$, $h_0(\Delta t)$ и $h_0(t + \Delta t)$.



По аксиоме $\boxed{\text{PA-1}}$ число событий, произошедшее за промежутки t и Δt не зависят, так как промежутки не пересекаются.

- $h_0(t) = \mathbb{P}(X_t = 0)$ — вероятность того, что за время t мы не получили ни одного лайка.
- $h_0(\Delta t) = \mathbb{P}(X_{\Delta t} = 0)$ — вероятность того, что за время Δt мы не получили ни одного лайка.
- $h_0(t + \Delta t) = \mathbb{P}(X_{t+\Delta t} = 0)$ — вероятность того, что мы не за один из двух промежутков не получим лайк.

Получается, что в силу независимости событий, вероятность их произведения равно произведению их вероятностей:

$$h_0(t + \Delta t) = h_0(t) \cdot h_0(\Delta t).$$

Мы хотим понять как выглядит функция $h_0(t)$. У нас есть значение функции в точке t, а ещё есть приращение Δt. Это всё жутко напоминает производную. Давайте попробуем свести получившееся уравнение к дифуру.

Вычтем из обоих частей $h_0(t)$ и поделим их на Δt. Потом можно будет устремить Δt к нулю и получить дифур

$$
\begin{aligned}
\frac{h_0(t + \Delta t) - h_0(t)}{\Delta t} &= \frac{h_0(t) \cdot h_0(\Delta t) - h_0(t)}{\Delta t} \\
\frac{h_0(t + \Delta t) - h_0(t)}{\Delta t} &= \frac{h_0(t) \cdot (h_0(\Delta t) - 1)}{\Delta t}.
\end{aligned}
$$

В правой части $h_0(t)$ никак не зависит от Δt. Если мы устремим $\Delta t \to 0$, её можно вынести за знак предела. Для того, чтобы понять что происходит с оставшейся дробью, воспользуемся аксиомой $\boxed{\text{PA-3}}$ и распишем вероятность $h_0(\Delta t)$

$$\frac{h_0(\Delta t) - 1}{\Delta t} = \frac{1 - \lambda \Delta t + o(\Delta t) - 1}{\Delta t} = -\lambda + \frac{o(\Delta t)}{\Delta t} \to -\lambda.$$

Получаем уравнение

$$h_0'(t) = -\lambda \cdot h_0(t).$$

Хм... Производная функции с точностью до константы равна самой функции... Попахивает экспонентой. Решим дифур:

$$\begin{aligned} \frac{dh_0}{dt} &= -\lambda \cdot h_0 \\ \frac{dh_0}{h_0} &= -\lambda \cdot dt \\ \ln h_0 &= -\lambda t + \ln \text{const} \\ h_0(t) &= \text{const} \cdot e^{-\lambda \cdot t}. \end{aligned}$$

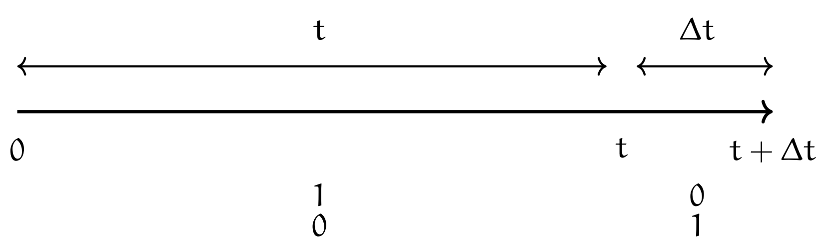
Найти константу можно из аксиомы $\boxed{\text{PA-0}}$

$$\mathbb{P}(X_0 = 0) = h_0(0) = \text{const} \cdot e^{-\lambda \cdot 0} = 1 \quad \Rightarrow \text{const} = 1.$$

Получается, что $\mathbb{P}(X_t = 0) = e^{-\lambda t}$. Обратите внимание, что здесь t — это время, например, часы. Параметр λ описывает интенсивность потока, то есть измеряется в единицах в час (1/час). Выходит, что вероятность безразмерная величина, как и положено.

Продолжаем искать нужные нам вероятности. Теперь посмотрим на $h_1(t) = \mathbb{P}(X_t = 1)$. По аналогии попробуем понять как между собой будут связаны $h_1(t), h_1(\Delta t)$ и $h_1(t + \Delta t)$.

- $h_1(t) = \mathbb{P}(X_t = 1)$ — вероятность того, что за время t мы получим ровно один лайк.
- $h_1(\Delta t) = \mathbb{P}(X_{\Delta t} = 1)$ — вероятность того, что за время Δt мы получим один лайк.
- $h_1(t + \Delta t) = \mathbb{P}(X_{t+\Delta t} = 1)$ — вероятность того, что мы в один из двух промежутков получим один лайк.



Получается, что мы должны получить лайк либо за первый промежуток, либо за второй. Отсюда получаем формулу:

$$h_1(t+\Delta t) = h_1(t)\cdot h_0(\Delta t) + h_0(t)\cdot h_1(\Delta t).$$

Снова постараемся превратить это всё в дифур. Вычтем из обеих частей $h(t)$ и поделим всё на Δt

$$\frac{h_1(t+\Delta t) - h_1(t)}{\Delta t} = \frac{h_1(t)\cdot(h_0(\Delta t)-1)}{\Delta t} + \frac{h_0(t)\cdot h_1(\Delta t)}{\Delta t}$$

Устремляем $\Delta t \to 0$. Слева от знака равенства получаем производную. Для первого слагаемого справа мы уже искали предел. Для второго получаем

$$\frac{h_0(t)\cdot h_1(\Delta t)}{\Delta t} = \frac{h_0(t)\cdot\lambda\cdot\Delta t + o(\Delta t)}{\Delta t} \to h_0(t)\cdot\lambda.$$

Получаем уравнение

$$h_1'(t) = -\lambda\cdot h_1(t) + \lambda\cdot h_0(t).$$

Значение $h_0(t)$ мы уже нашли

$$h_1'(t) = -\lambda\cdot h_1(t) + \lambda\cdot e^{-\lambda t}.$$

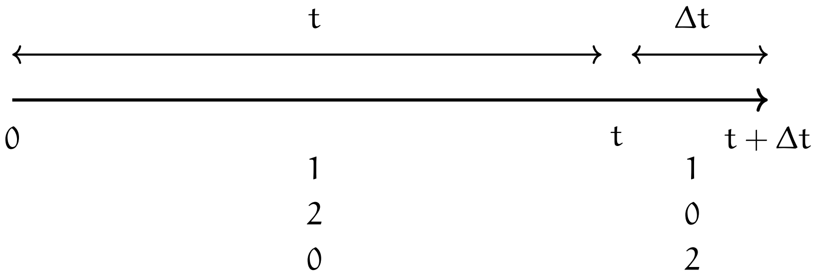
Домножим обе части уравнения на $e^{\lambda t}$

$$\begin{aligned}
&e^{\lambda t}\cdot h_1'(t) = -\lambda e^{\lambda t}\cdot h_1(t) + \lambda\\
&e^{\lambda t}\cdot h_1'(t) + \lambda e^{\lambda t}\cdot h_1(t) = \lambda\\
&(e^{\lambda t}\cdot h_1(t))' = \lambda\\
&e^{\lambda t}\cdot h_1(t) = \lambda t + \text{const}\\
&h_1(t) = \lambda t\cdot e^{-\lambda t} + \text{const}\cdot e^{-\lambda t}.
\end{aligned}$$

Значение константы мы можем получить из условия $h_1(0) = \mathbb{P}(X_0 = 1) = 0$. Оказывается, что $\text{const} = 0$. Выходит, что $\mathbb{P}(X_t = 1) = e^{-\lambda t}\cdot\lambda t$.

Кажется, что схема понятна. Попробуем понять как можно найти вероятность $h_2(t) = \mathbb{P}(X_t = 2)$. Снова посмотрим как между собой будут связаны $h_2(t)$, $h_2(\Delta t)$ и $h_2(t+\Delta t)$.

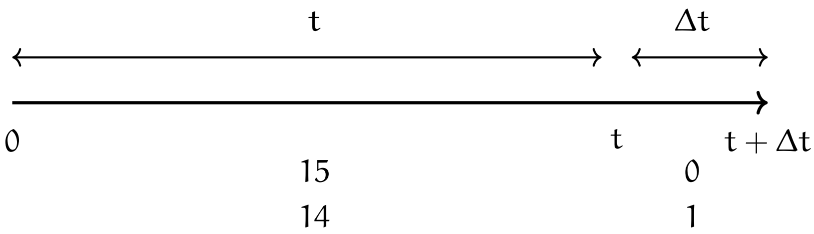
У нас есть два промежутка времени. За них в сумме должно произойти 2 события. Всего возможно три варианта:



По аксиоме $\boxed{\text{PA-3}}$ за маленький промежуток Δt точно не происходит больше одного события. Получается, что вероятность третьего варианта равна $o(\Delta t)$. Отсюда получаем уравнение

$$h_2(t + \Delta t) = h_1(t) \cdot h_1(\Delta t) + h_2(t) \cdot h_0(\Delta t) + o(\Delta t).$$

Дальше можно получить дифур и решить его. Как найти $h_{15}(t) = \mathbb{P}(X_t = 15)$? Да точно также! У нас может быть всего два варианта:



В остальных вариантах на промежуток времени Δt приходится больше одного события. Такое может произойти только с вероятностью $o(\Delta t)$, поэтому на них мы можем забить. Так для произвольного k мы можем аккуратно выписать уравнение, решить его, а затем по индукции мы можем получить общую формулу. Это дело техники, поэтому в этой лекции мы этим заниматься не будем.

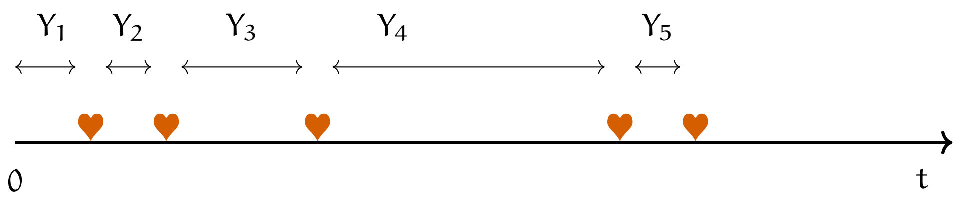
Итак, получается, что $X_t \sim \mathrm{Poiss}(\lambda t)$. Найти вероятность того, что мы получим за время t конкретное число лайков, можно по формуле

$$\mathbb{P}(X_t = k) = \frac{e^{-\lambda t} \cdot (\lambda t)^k}{k!}.$$

Обратите внимание на то, как можно проинтерпретировать параметр λ. Пусть у нас рассматривается промежуток в один час. Тогда $\mathbb{E}(X_1) = \lambda$. То есть это среднее число лайков, которое мы получаем в течение часа.

5 Время между событиями

До этого мы говорили, что время между событиями можно моделировать с помощью экспоненциального распределения. Число событий в потоке можно моделировать с помощью распределения Пуассона. Значит эти два распределения как-то взаимосвязаны между собой.



Пусть случайная величина Y_n — это промежуток времени между $(n-1)$-м и n—м событиями. Давайте попробуем найти функцию распределения для случайной величины Y_1

$$F_{Y_1}(t) = \mathbb{P}(Y_1 \leqslant t).$$

Вероятность $\mathbb{P}(Y_1 \leqslant t)$ можно прочитать, как вероятность возникновения хотя бы одного события в течение первых t минут. Вероятность $\mathbb{P}(Y_1 > t)$ можно проинтерпретировать, как вероятность того, что за t минут не произошло ни одного события. На языке распределения Пуассона её можно записать как $\mathbb{P}(X_t = 0)$.

Получается, что

$$F_{Y_1}(t) = \mathbb{P}(Y_1 \leqslant t) = 1 - \mathbb{P}(Y_1 > t) = 1 - \mathbb{P}(X_t = 0) = 1 - e^{-\lambda t}, \quad t \geqslant 0.$$

Получилась функция распределения экспоненциальной случайной величины. По аксиоме $\boxed{\text{PA-2}}$ случайные величины $X_{\Delta t}$ и $X_{\Delta s}$ независимы, если Δt и Δs не пересекаются. Получается, мы можем просто сдвинуть начало нашего потока в ту точку, где произошло первое событие и рассматривать её как старт нашего потока. Тогда случайная величина Y_2 в новых координатах станет Y_1. Выходит, что у неё такое же распределение. По аналогии получаем распределение всех остальных Y_n.

Прочувствовали связь? Давайте немного углубим её. Для Распределения Пуассона математическое ожидание составляет λ. Этот параметр интерпретируется, как интенсивность потока событий. Чем он больше, тем больше событий за единицу времени может произойти.

Для экспоненциального распределения математическое ожидание составляет $\frac{1}{\lambda}$. Чувствуете? Чем больше интенсивность потока событий, тем меньше в среднем времени проходит между ними.

6 Большая сила о-малых

Когда ты понимаешь, как выглядят аксиомы, которые лежат в основании Пуассоновского потока, ты можешь решать с их помощью задачки. Давайте сделаем это.

Задача 6.1. (Хозяйки кассы) В магазине две кассирши (ах, да! две хозяйки кассы). Допустим, что время обслуживания клиента распределено экспоненциально. Тетя Зина обслуживает в среднем 5 клиентов в час, а тетя Маша - 7. Два клиента подошли к кассам одновременно[1].

а) Какова вероятность того, что тетя Зина обслужит клиента быстрее?
б) Как распределено время обслуживания того клиента, который освободится быстрее?

а) Хозяйки работают независимо друг от друга. Обозначим время, за которые они обслуживают покупателя за $Y_z \sim \text{Exp}(\lambda_z)$ и $Y_m \sim \text{Exp}(\lambda_m)$.
Найдём $\mathbb{P}(Y_z < Y_m)$, то есть вероятность того, что тётя Зина справилась быстрее. Эту задачу можно решать двумя разными путями: путём простушки и путём хитрюшки. Опишем здесь обе дороги.
Путь простушки. У нас есть две случайные величины. Они независимы друг от друга. Получается совместная плотность распределения будет равна произведению частных плотностей. Нас интересует область, где $Y_z < Y_m$. Возьмём по ней интеграл от совместной плотности. Это и будет требуемая вероятность.

$$\mathbb{P}(Y_z < Y_m) = \int_0^{+\infty} \int_0^{y_m} \lambda_z e^{-\lambda_z y_z} \cdot \lambda_m e^{-\lambda_m y_m} \, dy_z \, dy_m.$$

Честно берём этот интеграл и получаем ответ.
Путь хитрюшки. Мы работаем с пуассоновским потоком. Давайте воспользуемся аксиомами, на которых он базируется. Рассмотрим маленький промежуток времени Δt. За него мы можем столкнуться с четырьмя разными ситуациями:

- Маша обслужила, Зина обслужила;
- Маша обслужила, Зина нет;
- Маша не успела, Зина успела;
- Обе не успели обслужить.

С точностью до Δ найдём вероятности всех четырёх ситуаций. Начнём с ситуации, когда обе хозяйки успели обслужить клиентов

$$\mathbb{P}(M^+Z^+) = (\lambda_Z \cdot \Delta t + o(\Delta t)) \cdot (\lambda_M \cdot \Delta t + o(\Delta t)) = \lambda_Z \cdot \lambda_M \cdot \Delta^2 t + o(\Delta t).$$

При $\Delta t \to 0$ величина $\Delta^2(t)$ уменьшается быстрее, чем Δt. Получается, что $\Delta^2(t) = o(\Delta t)$ и этой величиной можно пренебречь. Константа на бесконечно-малую — снова бесконечно малая. Получается, что вероятность того, что обе хозяйки за Δt обслужат клиента равна $o(\Delta t)$.
Находим вероятность ситуации, когда Маша обслужила, а Зина нет

$$\begin{aligned}\mathbb{P}(M^+Z^-) &= (\lambda_M \cdot \Delta t + o(\Delta t)) \cdot (1 - \lambda_Z \cdot \Delta t + o(\Delta t)) = \\ &= \lambda_M \cdot \Delta t - \lambda_Z \cdot \lambda_M \cdot \Delta^2 t + o(\Delta t) = \lambda_M \cdot \Delta t + o(\Delta t).\end{aligned}$$

Вероятность того, что Зина успела, а Маша нет находится по аналогии

[1] Ищи больше задач в сборнике Бориса Демешева. На пуассоновский поток можно порешать раздел 8.

$$\mathbb{P}(M^-Z^+) = (\lambda_Z \cdot \Delta t + o(\Delta t)) \cdot (1 - \lambda_M \cdot \Delta t + o(\Delta t)) = \lambda_Z \cdot \Delta t + o(\Delta t).$$

Осталась вероятность того, что обе хозяйки не успели обслужить клиентов

$$\begin{aligned}\mathbb{P}(M^-Z^-) &= (1-\lambda_Z \cdot \Delta t + o(\Delta t)) \cdot (1 - \lambda_M \cdot \Delta t + o(\Delta t)) = \\ &= 1 - \lambda_M \cdot \Delta t - \lambda_Z \cdot \Delta t + \lambda_Z \cdot \lambda_M \cdot \Delta^2 t + o(\Delta t) = 1 - (\lambda_Z + \lambda_M) \cdot \Delta t + o(\Delta t).\end{aligned}$$

Распишем $\mathbb{P}(Y_Z < Y_M)$ через найденные вероятности. Первое слагаемое отвечает за то, что Зина победила. Во втором случае борьба между хозяйками за время Δt не разрешилась. Игра в таком случае начинается заново и $\mathbb{P}(M^-Z^-)$ мы умножаем на $\mathbb{P}(Y_Z < Y_M)$. Третье слагаемое отвечает за всё, чем мы в рамках Δt можем пренебречь

$$\mathbb{P}(Y_Z < Y_M) = \lambda_Z \cdot \Delta t + (1 - (\lambda_Z + \lambda_M)) \cdot \mathbb{P}(Y_Z < Y_M) + o(\Delta t).$$

Немного перепишем получившееся уравнение

$$\mathbb{P}(Y_Z < Y_M) \cdot (\lambda_Z + \lambda_M) \cdot \Delta t = \lambda_Z \cdot \Delta t + o(\Delta t).$$

Поделим всё на Δt и устремим его к нулю

$$\mathbb{P}(Y_Z < Y_M) \cdot (\lambda_Z + \lambda_M) = \lambda_Z.$$

Получим ответ для искомой вероятности

$$\mathbb{P}(Y_Z < Y_M) = \frac{\lambda_Z}{\lambda_Z + \lambda_M} = \frac{5}{5+7}.$$

Получившийся ответ очень логичен и красив.

б) Чтобы ответить на второй пункт задачи, нам нужно немного поработать с функциями распределения. Нас интересует случайная величина $X = \min(Y_Z, Y_M)$. Найдём её функцию распределения

$$F_X(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}(\min(Y_Z, Y_M) \leqslant x) = 1 - \mathbb{P}(\min(Y_Z, Y_M) > x).$$

Если минимум оказался больше x, то и все случайные величины будут больше x, потому что они больше минимума. Также вспомним, что хозяйки обслуживают клиентов независимо друг от друга

$$\begin{aligned}F_X(x) = 1 - \mathbb{P}(\min(Y_Z, Y_M) > x) = 1 - \mathbb{P}(Y_Z > x \cap Y_M > x) = 1 - \mathbb{P}(Y_Z > x) \cdot \mathbb{P}(Y_M > x) = \\ = 1 - (1 - \mathbb{P}(Y_Z \leqslant x)) \cdot (1 - \mathbb{P}(Y_M \leqslant x)).\end{aligned}$$

Наконец мы получили функции распределения для Y_M и Y_Z. Дальше дело остаётся за малым

$$F_X(x) = 1-(1-F_{Y_z}(x))\cdot(1-F_{Y_M}(x)) = 1-(1-1+e^{-\lambda_Z x})\cdot(1-1+e^{-\lambda_M x}) = 1-e^{-(\lambda_M+\lambda_Z)x}.$$

Область определения по ходу вычислений нигде не менялась. Получается, что время обслуживания того клиента, который освободится быстрее, тоже имеет экспоненциальное распределение

$$X = \min(Y_Z, Y_M) \sim \mathrm{Exp}(\lambda_M + \lambda_Z).$$