# Лекция 14. Процесс Бернулли

19 мая 2022 г.

## 1 Процесс Бернулли

После сложных мартингалов мы рассмотрим более простой процесс Бернулли (для подготовки к рассмотрению более сложного процесса Пуассона). Процессом Бернулли называется случайный процесс $\{X_t\}_{t\in\mathbb{N}}$, в котором все $X_t$ независимы и следуют одинаковому распределению Бернулли с одинаковым параметром $p$. С его помощью часто моделируют разные реальные штуки:

- Браки на производстве
- Прибытие посетителя в магазин в интервалы времен
- Поступление запросов на сервер

Как вообще стоит рассматривать случайный процесс, в том числе процесс Бернулли? В случае, когда у нас конечное множество с.в. $\{X_t\}$, то нам достаточно задать совместную функцию вероятности или совместную плотность вероятности для всех возможных подмножеств $\{X_{t_1}, \dots, X_{t_k}\}$. Это необходимо, так как с.в. могут быть и зависимы. Но в случае с бесконечными последовательностями даже если мы определим совместные функции/плотности вероятности на всех конечных подмножествах, мы все равно не сможем определить, например, вероятность того что все $X_t = 1$.

Поэтому в случае со случайными процессами стоит рассматривать $\Omega$ как множество всех возможных последовательностей, которые могут получиться из значений случайных величин. В случае с процессом Бернулли — это все возможные последовательности из нолей и единиц. В таком случае легко найти вероятность события, что все $X_t = 1$ как в случае, когда $p < 1$:

$$\begin{aligned} &\Pr(\forall t \in \mathbb{N}\ X_t = 1) \le \Pr(\forall t \in [1..n]\ X_t = 1) = p^n \\ \Rightarrow &\Pr(\forall t \in \mathbb{N}\ X_t = 1) = 0, \end{aligned}$$

так и для $p = 1$:

$$\Pr(\forall t \in \mathbb{N}\ X_t = 1) \ge 1 - \sum_{t=1}^{+\infty} \Pr[X_t = 0] = 1.$$

В дальнейшем случай с $p = 1$ (как и с $p = 0$) мы рассматривать не будем, так как ничего интересного в этом нет.

## 1.1 Свойства процесса Бернулли

Рассмотрим несколько свойств процесса Бернулли. Напомним, что так как каждый $X_t$ следует распределению Бернулли независимо от других с.в., то

- $E[X_t] = p$
- $\mathrm{Var}(X_t) = p(1-p)$

Рассмотрим следующие две случайные величины.

$S_n$ **— число единиц в первые $n$ единиц времени.** Так как $X_t$ независимы, то $S_n \sim \mathrm{Bin}(n, p)$, причем

- $\Pr[S_n = k] = \binom{n}{k} p^k (1-p)^{n-k}$
- $E[S_n] = np$
- $\mathrm{Var}(S_n) = np(1-p)$

$T_1$ **— время первой единицы.** Так как $X_t$ независимы, то $T_1 \sim \mathrm{Geom}(p)$, причем

- $\Pr[T_1 = k] = p(1-p)^{k-1}$
- $E[T_n] = \frac{1}{p}$
- $\mathrm{Var}(T_n) = \frac{1-p}{p^2}$

**Беспамятство.** Если мы рассматриваем процесс $Y_t = X_{t+n}$ для какого-то фиксированного $n$, то это опять будет процесс Бернулли, так как:

- Все $Y_t$ независимы
- Все $Y_t$ имеют распределение $\mathrm{Bern}(p)$

Более того, если мы говорим, что $Y_t = X_{t+N}$, где $N$ — какая-то случайная величина, то в некоторых случаях мы также получаем процесс Бернулли. Например:

- Пусть $N$ — время первого успеха в процессе $X_t$. Тогда все $X_{t+N}$ независимы друг от друга и от $N$, поэтому являются процессом Бернулли.
- Пусть теперь $N$ — время перед первым успехом. Тогда $X_{N+1}$ уже не независим от $N$, а точно равен единице.

В общем случае $Y_t$ будет процессом Бернулли, если $N$ является временем останова, то есть событие $N \geq n$ зависит только от $X_1, \ldots, X_{n-1}$.

Беспамятство дает нам много других полезных свойств процесса Бернулли.

**Продолжительность занятого периода.** Рассмотрим сервер, на который в каждую единицу времени приходит или не приходит запрос. Это можно описать процессом Бернулли, где событие "во время $t$ пришел запрос" соответсвует $X_t = 1$. Скажем, что сервер занят во время $t$, если в это время на него пришел запрос. Занятый период — это отрезок $[t_1..t_2]$, такой, что во все $t \in [t_1..t_2]$ сервер занят, но в моменты $t_1 - 1$ и $t_2 + 1$ — свободен или еще не был запущен. Нас интересует продолжительность первого занятого периода.

Пусть $T_1$ — время первого запроса на сервер. Благодаря беспамятству мы можем сказать,что $X_{t+T_1}$ задают новый процесс Бернулли с тем же параметром $p$. В этом процессе первый ноль появится во время $T_2 \sim \text{Geom}(1-p)$. При этом $T_2$ в таком случае и будет равно длине занятого периода.

**Время $k$-ого успеха.** Пусть $Y_k$ — время, в которое на сервер прибыл $k$-ый по счету запрос. Пусть $T_k = Y_k - Y_{k-1}$, то есть интервал между $k-1$-ым и $k$-ым запросами. При этом $Y_k = T_1 + \cdots + T_k$.

Благодаря беспамятству мы можем сказать, что в каждый момент $Y_k$ мы запускаем процесс заново. Поэтому каждый $T_k \sim \text{Geom}(p)$. Отсюда мы сразу можем сказать про $Y_k$, что

- $E[Y_k] = \sum_{i=1}^{k} E[T_k] = \frac{k}{p}$
- $\text{Var}(Y_k) = \sum_{i=1}^{k} \text{Var}(T_k) = \frac{k(1-p)}{p^2}$.

Также уже было в разборе контрольной, что функция распределения имеет следующий вид:

$$p_{Y_k}(t) = \binom{t-1}{k-1} p^k (1-p)^{t-k},$$

так как мы точно знаем, что во время $t$ произошел успех, и нам надо расставить остальные $k-1$ успех по оставшимся $t-1$ ячейкам времени.

# 2 Слияние и разделение процесса Бернулли

Рассмотрим случай, когда у нас на сервер могут поступать запросы с двух источников. Поступление запросов от каждого из источников описывается процессом Бернулли, причем эти два процесса (назовем их $X_t$ и $Y_t$) независимы. Рассмотрим процесс $Z_t$, где $Z_t$ — индикаторная с.в. события "в момент времени $t$ поступил запрос от хотя бы одного источника". То есть $Z_t = \max\{X_t, Y_t\}$. Заметим, что $Z_t$ не зависит от любого $Z_{s \neq t}$, так как является функцией от двух величин $X_t$ и $Y_t$, не зависящих от любых других $X_{s \neq t}$ и $Y_{s \neq t}$. А также

$$\Pr[Z_t = 1] = \Pr[X_t = 1 \cup Y_t = 1] = 1 - \Pr[X_t = 0 \cap Y_t = 0] = p + q - pq,$$

где $p$ и $q$ — параметры процессов $X_t$ и $Y_t$ соответственно. Таким образом, $Z_t$ также является процессом Бернулли с параметром $p+q-pq$. Заметьте, что вместо логического "И" мы можем применить любую другую операцию к двум поступающим запросам. Например, одновременный запрос от двух источников мы можем считать коллизией, и не считать его за запрос. Тогда $Z_t$ также будет процессом Бернулли, но с другим параметром $p+q-2pq$.

Мы также можем разделять процесс Бернулли на два. Пусть у нас есть процесс Бернулли $Z_t$ с параметром $p$, описывающий запросы на сервер. Каждый раз, когда на сервер поступает запрос, он решает, в какой из двух выходов его отправить дальше, причем он отправляет его в первый выход с вероятностью $q$ и во второй — с вероятностью $1-q$, принимая это решение независимо от всего остального. В таком случае запросы, отправленные в первый выход, $X_t$, являются процессом Бернулли с параметром $pq$. Очевидно, что $X_t$ не зависит от любого $X_{s\neq t}$, так как является функцией от величины $Z_t$, которая, в свою очередь, не зависит ни от каких $Z_{s\neq t}$. Аналогично, процессом Бернулли являются и запросы на втором выходе $Y_t$, но с параметром $p(1-q)$. Однако процессы $X_t$ и $Y_t$ не являются независимыми, так как если мы знаем, что $X_t=1$, то мы точно знаем, что $Y_t=0$.

# 3 Аппроксимация процесса Бернулли

Рассмотрим процесс Бернулли с параметром $p$. Чему равна вероятность того, что за время $n$ будет $k$ успехов? По тому, что мы считали, это будет

$$p_X(k)=\binom{n}{k}p^k(1-p)^{n-k}.$$

Но что будет, если мы раздробим время на меньше интервалы (при этом вероятность успеха в каждый интервал будет падать обратно пропорционально длине интервала)? То есть для каждого $n$ мы будем брать $p=\lambda n$ для какой-то фиксированной $\lambda$ и устремим $n\to\infty$. В таком случае будет:

$$\begin{aligned}
p_X(k) &= \binom{n}{k}p^k(1-p)^{n-k}\\
&= \frac{n(n-1)\dots(n-k+1)}{k!}\cdot\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^{n-k}\\
&= \frac{n}{n}\cdot\frac{n-1}{n}\cdot\ldots\cdot\frac{n-k+1}{n}\cdot\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-k}\\
&\to 1\cdot 1\cdot\ldots\cdot 1\cdot\frac{\lambda^k}{k!}e^{-\lambda}\cdot 1=\frac{\lambda^k}{k!}e^{-\lambda}.
\end{aligned}$$

И это подводит нас к процессу Пуассона.

# 4 Процесс Пуассона: определение

Процессом Пуассона называется некоторый процесс с вещественным временем, и в каждый момент времени происходит или не происходит событие. Однако в силу непрерывности вероятность, что в каждый конкретный момент времени произойдет событие равна нулю, зато на временном интервале может произойти более одного события. Для того, чтобы такой процесс назывался процессом Пуассона, мы требуем:

1. Независимость непересекающихся временных интервалов: число событий, произошедших в интервал времени $\tau_1$ не зависит от числа событий в другом интервале $\tau_2$.

2. Неизменяемость по времени. Вероятность $P_\tau(k)$ того, что в интервал длиной $\tau$ произойдет $k$ событий не зависит от того, в какой именно момент времени начался интервал. Заметим, что для любого $\tau$ выполняется $\sum_{k=0}^{+\infty} P_\tau(k) = 1$.

3. Для маленьких интервалов длиною $\delta$ мы также требуем

$$P_\delta(k) \approx \begin{cases} 1 - \lambda\delta, & k = 0 \\ \lambda\delta, & k = 1 \\ 0, & \text{иначе.} \end{cases}$$

   То есть мы хотим, чтобы для достаточно маленького интервала вероятность того, что в нем уместилось более двух событий была ничтожно маленькая, а вероятность одного события была пропорционально длине интервала. Более точное требование: при $\delta \to 0$ должно выполняться

$$P_\delta(k) = \begin{cases} 1 - \lambda\delta + O(\delta^2), & k = 0 \\ \lambda\delta + O(\delta^2), & k = 1 \\ O(\delta^2), & \text{иначе.} \end{cases}$$

   В данном случае $\lambda$ называется *интенсивностью* процесса.

Примеры из реальной жизни:

1. Смерти от удара лошади в Прусской армии (самое первое применение, причем не у Пуассона, а позже Борткевичем)

2. Испускание радиоактивной частицы

3. Детекция фотона от слабого источника

4. Потрясения рынка (финансовые кризисы и т.п.)

5. Сфера обслуживания: звонки в колл-центры, приход покупателя, ...

# 5 Распределение Пуассона

Пусть случайная величина $N_\tau$ — число событий в процессе Пуассона с интенсивностью $\lambda$, которые произошли в интервал длиной $\tau$. Попробуем оценить ее распределение. Для этого разобьем весь интервал длиной $\tau$ на отрезки (слоты) длиной $\delta$, всего получится $\frac{\tau}{\delta}$ отрезков. При этом вероятность $p$ того, что есть хотя бы один отрезок, куда попали два или более событий не больше, чем сумма вероятностей по всем отрезкам того, что в этот отрезок попало два и более событий. По третьему свойству процесса Пуассона имеем.

$$p \le \sum_{i=1}^{\tau/\delta} \Pr[N_\delta \ge 2] = \frac{\tau}{\delta} O(\delta^2) = O(\tau\delta).$$

Это значит, что мы можем добиться сколь угодно маленькой вероятности того, что у нас есть слот с двумя и более событиями, а значит, мы можем считать, что $N_\tau \sim \mathrm{Bin}(\frac{\tau}{\delta}, \lambda\delta)$. При этом в начале лекции мы показали, что

$$\Pr[N_\tau = k] = \lim_{\delta \to 0} \Pr[X \sim \mathrm{Bin}(\frac{\tau}{\delta}, \lambda\delta + O(\delta^2)) = k] = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau}.$$

Говорят, что $N_\tau$ в таком случае следует распределению Пуассона с параметром $(\lambda\tau)$:

$$N_\tau \sim \mathrm{Poisson}(\lambda\tau).$$

# 6 Свойства распределения Пуассона

**Матожидание и дисперсия.**

Матожидание распределения Пуассона можно посчитать напрямую

$$\begin{aligned} E[N_\tau] &= \sum_{k=0}^{+\infty} k \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau} \\ &= e^{-\lambda\tau} \lambda\tau \sum_{k=0}^{+\infty} \frac{(\lambda\tau)^k}{(k-1)!} \\ &= \lambda\tau. \end{aligned}$$

А можно сказать, что оно близко к биномиальному распределению $\mathrm{Bin}(\frac{\tau}{\delta}, \lambda\delta + O(\delta^2))$, то есть его матожидание есть

$$E[N_\tau] = \lim_{\delta \to 0} \frac{\tau}{\delta} (\lambda\delta + O(\delta^2)) = \lambda\tau.$$

Дисперсию тоже посчитаем этим методом:

$$\mathrm{Var}(N_\tau) = \lim_{\delta \to 0} \frac{\tau}{\delta} (\lambda\delta + O(\delta^2))(1 - \lambda\delta + O(\delta^2)) = \lambda\tau.$$

То есть для распределения Пуассона с параметром $\lambda\tau$

$$E[N_\tau] = \mathrm{Var}(N_\tau) = \lambda\tau$$

Интенсивность процесса Пуассона при этом можно интерпретировать как среднее число событий, происходящих в единицу времени:

$$\lambda = \frac{E[N_\tau]}{\tau}.$$

**Время первого события $T_1$.**

Вероятность того, что первое событие произошло через время большее, чем $t$, есть вероятность того, что за время $t$ произошло ноль событий, поэтому можем посчитать функцию распределения $T_1$.

$$F_{T_1}(t) = \Pr[T_1 \le t] = 1 - \Pr[T_1 > t] = 1 - P_t(0) = 1 - e^{-t\lambda}.$$

Получается, время $T_1$ следует распределению $\mathrm{Exp}(\lambda)$. Заметим, что в силу беспамятства экспоненциального распределения, при условии $T_1 > t$ величина $T_1 - t$ также следует $\mathrm{Exp}(\lambda)$.

**Время $k$-ого события $T_k$.**

Можно попытаться посчитать функцию распределения $T_k$ тем же образом:

$$\Pr[T_k \le t] = \sum_{i=k}^{+\infty} P_t(i),$$

но это чревато дебрями сложной математики (остаток ряда экспоненты выглядит не очень мило). Поэтому давайте попробуем посчитать плотность вероятности следующим образом:

$$f_{T_k}(t)\delta \approx \Pr[t \le T_k \le t+\delta] = P_t(k-1)(\lambda\delta + O(delta^2)) + (\sum_{i=0}^{k-2} P_t(i))O(\delta(2)) \approx P_t(k-1)\lambda\delta.$$

Таким образом,

$$f_{T_k}(t) \approx \lambda P_t(k-1) = \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-(k-1)t}.$$

Это, кстати, называется распределением Эрланга (частный случай Гамма-распределения, как некоторые из вас знают с контрольной).

**Беспамятство.**

Как и в случае с процессом Бернулли, у процесса Пуассона есть беспамятство. Рассмотрим случай, когда мы начинаем смотреть процесс в фиксированное время $t$. Тогда для наблюдаемого нами процесса по-прежнему выполняются все три свойства процесса Пуассона. То же верно, когда $T$ — время останова.

Поэтому если мы начинаем смотреть в любой фиксированный момент времени, то время до следующего события следует экспоненциальному распределению. Также если мы начинаем смотреть в момент $T_k$, когда произошло событие $k$, то время до следующего события $Y_{k+1} = T_{k+1} - T_k$ также следует $\text{Exp}(\lambda)$. Поэтому существует альтернативное определение процесса Пуассона.

Пусть есть последовательность независимых с.в. $\{Y_k\}_{k\in\mathbb{N}}$, следующих распределению $\text{Exp}(\lambda)$. Тогда последовательность с.в. $T_k = \sum_{i=1}^{k} Y_k$ образует процесс Пуассона.

Отсюда следуют свойства $T_k$:

- $E[T_k] = \frac{k}{\lambda}$
- $\text{Var}(T_k) = \frac{k}{\lambda^2}$

# 7 Сравнение с процессом Бернулли

| | Пуассон | Бернулли |
|---|---|---|
| Время | Непрерывное | Дискретное |
| Интенсивность | $\lambda$ в единицу времени | $p$ в тайм-слот |
| Распределение числа событий в интервале | $\text{Poisson}(\lambda\tau)$ | $\text{Bin}(n, p)$ |
| Интервалы между событиями | $\text{Exp}(\lambda)$ | $\text{Geom}(p)$ |
| $k$-е событие | Эрланг | Паскаль (inverse Bin) |

# 8 Сумма двух с.в. Пуассона

Рассмотрим процесс Пуассона с $\lambda = 1$. Рассмотрим два последовательных временных интервала длиной $\mu$ и $\nu$. Пусть $X$ — число событий в первом интервале, а $Y$ — число событий во втором. Тогда $X \sim \text{Poisson}(\mu)$ и $Y \sim \text{Poisson}(\nu)$, причем они независимы.

С.в. $Z = X + Y$ есть число событий в интервале длиной $\mu + \nu$, то есть $Z \sim \text{Poisson}(\mu + \nu)$. Таким образом, для двух независимых с.в. Пуассона выполняется

$$\text{Poisson}(\mu) + \text{Poisson}(\nu) = \text{Poisson}(\mu + \nu),$$

что есть равенство распределений (то есть совпадение функции распределения во всех точках).

# 9 Слияние двух процессов Пуассона

Аналогично с процессами Бернулли, мы можем сливать процесс Пуассона из двух независимых процессов. Пусть у нас есть две лампочки разных цветов, зеленая и красная. Они независимо мигают с интенсивностью $\lambda_1$ и $\lambda_2$ (то есть мигания каждой лампочки — процессы Пуассона). У нас есть датчик, регистрирующий вспышку, но не воспринимающий цвет. Тогда регистрации на датчике тоже будут процессом Пуассона.

1. Выполняется независимость непересекающихся интервалов
2. Выполняется неизменность во времени
3. Рассмотрим вероятность того, то в интервале длины $\delta$ произошло $k$ событий

| | | $1-\lambda_1\delta$ | $\lambda_1\delta$ | $O(\delta^2)$ |
|---|---|---|---|---|
| | | $0$ | $1$ | $\geq 2$ |
| $1-\lambda_2\delta$ | $0$ | $1-(\lambda_1+\lambda_2)\delta$ | $\lambda_1\delta$ | |
| $\lambda_2\delta$ | $1$ | $\lambda_2\delta$ | | |
| $O(\delta^2)$ | $\geq 2$ | | | $O(\delta^2)$ |

То есть интенсивность нового процесса есть $\lambda_1+\lambda_2$.

Также мы можем показать вероятность того, из какого процесса произошло $k$-е событие. Возьмем очень маленький интервал $\delta$, в котором это событие произошло, и по формуле условной вероятности и по таблице выше почитаем, что вероятность того, что событие пришло из первого процесса есть $\frac{\lambda_1}{\lambda_1+\lambda_2}$.

Причем источники различных событий в слитом процессе независимы. То есть, последовательность источников событий образует процесс Бернулли.

**Пример.**

Пусть у нас есть три лампочки, время работы которых следует $\mathrm{Exp}(\lambda)$. Когда перегорит первая лампочка и когда перегорит последняя?

Насчет распределения времени перегорания первой можно посчитать в тупую тройной интеграл функции $\min\{T_1, T_2, T_3\}$, но это скучно (можно написать этот страшный интеграл на доске).

Можно посчитать через функцию распределения:

$$\Pr[\min\{T_1,T_2,T_3\} \geq t] = \Pr[T_1 \geq t]\Pr[T_2 \geq t]\Pr[T_3 \geq t] = e^{-3\lambda t}$$

А можно просто сказать, что перегорание каждой лампочки — это первое событие в процессе Пуассона с интенсивностью $\lambda$. Поэтому слитый процесс имеет интенсивность $3\lambda$, и первая лампочка перегорает через время $T \sim \mathrm{Exp}(3\lambda)$.

Для оценки того, когда перегорает последняя, можем заметить, что если мы смотрим на процесс после с момента перегорания первой лампочки, то мы смотрим на слияние двух процессов, которое имеет интенсивность $2\lambda$, то есть время между первой и второй следует $\mathrm{Exp}(2\lambda)$. Ну а последняя перегорает еще через $\mathrm{Exp}(\lambda)$. Таким образом, время перегорания последней лампочки следует распределению $\mathrm{Exp}(3\lambda) + \mathrm{Exp}(2\lambda) + \mathrm{Exp}(\lambda)$, где все три распределения — независимы.

# 10 Разделение процесса Пуассона

Рассмотрим процесс Пуассона с интенсивностью $\lambda$. Можем при каждом событии подкидывать нечестную монетку и с вероятностью $q$ отправлять событие в поток $A$ и с вероятностью $1 - q$ — в поток $B$. Каждый из потоков будет процессом Пуассона, причем поток $A$ — с интенсивностью $\lambda q$, а поток $B$ — с интенсивностью $\lambda(1 - q)$. Будут ли они независимы? В отличие от процессов Бернулли — да, будут, так как мы находимся в ситуации непрерывного времени.

# 11 Парадокс наблюдателя

Допустим, вы знаете, что автобусы ходят с интенсивностью 4 автобуса в час. То есть интервал в среднем должен быть 15 минут. Вы приходите на остановку, интересуетесь у продавца в ларьке, как давно ушел последний автобус и ждете следующего. Таким образом вы замеряете интервал между автобусами и уходите. Так вы делаете много дней подряд и обнаруживаете, что средний интервал 30 минут, а не 15. Почему так?

Когда мы пришли на остановку — матожидание времени до следующего автобуса есть 15 минут из-за беспамятства процесса Пуассона. Однако ожидаемое время до предыдущего автобуса — тоже 15 минут, так как мы можем рассматривать автобусы в прошлом как независимый от будущего процесс Пуассона, в котором время течет обратно. В чем подвох?

Подсказка. Пусть есть процесс, в котором события происходят через какие-то интервалы. Но при этом каждый интервал либо 10, либо 5 минут, причем равновероятно. То есть среднее время между событиями есть 7.5 минут. Однако если мы хотим прийти в случайный момент времени и замерить длину интервала, в который мы пришли, то получится, что мы с вероятностью $\frac{2}{3}$ придем в 10-минутный интервал, так как они занимают треть от всей числовой оси. То есть для нас средняя длина отрезка будет $\frac{5}{3} + 203 \approx 8.3$.

Также и с процессом Пуассона: мы намного вероятнее приходим на остановку в длинные интервалы.

Тут также стоит сказать, что подобный эффект может влиять на опросы.

- Если мы опрашиваем случайных людей о размере их семьи, то наш средний результат будет выше, чем реальный размер семьи.
- Если мы опрашиваем случайных людей о заполненности автобуса, на котором они сегодня ехали — средний ответ тоже будет выше среднего значения.