

# Seminar 2: Процессы с независимыми приращениями. Пуассоновский процесс

Carlos Buitrago

## Table of contents

|          |                                                                                            |           |
|----------|--------------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Процессы с независимыми приращениями. Пуассоновский процесс</b>                         | <b>1</b>  |
| 1.1      | Процессы с независимыми приращениями . . . . .                                             | 2         |
| 1.2      | Критерий существования процесса с независимыми приращениями . . . . .                      | 2         |
| <b>2</b> | <b>Пуассоновский процесс</b>                                                               | <b>3</b>  |
| 2.1      | Симуляция: как выглядит пуассоновский процесс? . . . . .                                   | 4         |
| 2.1.1    | Влияние интенсивности $\lambda$ . . . . .                                                  | 5         |
| 2.2      | Существование пуассоновского процесса . . . . .                                            | 6         |
| 2.3      | Траектории пуассоновского процесса . . . . .                                               | 6         |
| 2.3.1    | Целочисленность . . . . .                                                                  | 6         |
| 2.3.2    | Монотонность . . . . .                                                                     | 6         |
| 2.4      | Симуляция: распределение $N_t$ . . . . .                                                   | 7         |
| <b>3</b> | <b>Явная конструкция пуассоновского процесса</b>                                           | <b>7</b>  |
| 3.1      | Симуляция: экспоненциальные времена ожидания $\rightarrow$ пуассоновский процесс . . . . . | 8         |
| 3.2      | Совместное распределение моментов событий . . . . .                                        | 8         |
| 3.3      | Размер скачков пуассоновского процесса . . . . .                                           | 10        |
| 3.4      | Моменты скачков . . . . .                                                                  | 11        |
| 3.5      | Симуляция: распределение момента $n$ -го скачка . . . . .                                  | 11        |
| 3.6      | Времена между последовательными событиями . . . . .                                        | 12        |
| 3.7      | Симуляция: экспоненциальные промежутки между событиями . . . . .                           | 13        |
| 3.8      | Две точки зрения на пуассоновский процесс . . . . .                                        | 13        |
| 3.8.1    | 1. Считать число событий . . . . .                                                         | 13        |
| 3.8.2    | 2. Рассматривать времена ожидания . . . . .                                                | 14        |
| 3.9      | Симуляция: две эквивалентные точки зрения на одной траектории . . . . .                    | 14        |
| <b>4</b> | <b>Основные результаты</b>                                                                 | <b>15</b> |
| <b>5</b> | <b>Где возникает пуассоновский процесс?</b>                                                | <b>15</b> |
| 5.0.1    | Обучение модели: оценивание интенсивности . . . . .                                        | 16        |
| 5.0.2    | Обнаружение аномалий . . . . .                                                             | 18        |
| 5.0.3    | Compound Poisson Process . . . . .                                                         | 18        |
| 5.1      | Когда пуассоновский процесс — плохая модель? . . . . .                                     | 21        |

## 1 Процессы с независимыми приращениями. Пуассоновский процесс

В этой теме мы изучим один из важнейших классов случайных процессов — **процессы с независимыми приращениями**.

Основным примером будет **пуассоновский процесс**, описывающий количество событий, произошедших к моменту времени  $t$ .

Основные вопросы:

- что означает независимость приращений процесса;
- при каких условиях существует процесс с заданными распределениями приращений;
- как определяется пуассоновский процесс;
- как явно построить пуассоновский процесс;
- как устроены его траектории;
- какое распределение имеют моменты скачков и промежутки между ними.

## 1.1 Процессы с независимыми приращениями

### Определение

Случайный процесс  $(X_t)_{t \geq 0}$  называется **процессом с независимыми приращениями**, если для любого  $n \in \mathbb{N}$  и любых моментов времени

$$0 \leq t_1 < t_2 < \dots < t_n$$

случайные величины

$$X_{t_1}, \quad X_{t_2} - X_{t_1}, \quad \dots, \quad X_{t_n} - X_{t_{n-1}}$$

независимы в совокупности.

### Замечание

Если множество моментов времени дискретно,

$$T = \mathbb{N},$$

то процессы с независимыми приращениями представляют собой **случайные блуждания**. Действительно, если положить

$$\xi_k = X_k - X_{k-1},$$

то

$$X_n = X_0 + \sum_{k=1}^n \xi_k,$$

где приращения  $\xi_1, \xi_2, \dots$  независимы.

## 1.2 Критерий существования процесса с независимыми приращениями

Предположим, что для каждой пары

$$0 \leq s < t$$

задано распределение вероятностей  $Q_{s,t}$  на  $\mathbb{R}$ , которое мы хотим получить в качестве распределения приращения

$$X_t - X_s.$$

Обозначим через

$$\varphi_{s,t}(\tau)$$

характеристическую функцию распределения  $Q_{s,t}$ .

**! Теорема — критерий существования**

Процесс  $(X_t)_{t \geq 0}$  с независимыми приращениями такой, что

$$X_t - X_s \stackrel{d}{=} Q_{s,t}, \quad 0 \leq s < t,$$

существует тогда и только тогда, когда для любых

$$0 \leq s < u < t$$

выполнено

$$\varphi_{s,t}(\tau) = \varphi_{s,u}(\tau) \varphi_{u,t}(\tau).$$

При этом распределение начального значения  $X_0$  может быть выбрано произвольным.

**💡 Почему возникает условие  $\varphi_{s,t} = \varphi_{s,u} \varphi_{u,t}$ ?**

Заметим, что

$$X_t - X_s = (X_u - X_s) + (X_t - X_u).$$

Если процесс имеет независимые приращения, то случайные величины

$$X_u - X_s \quad \text{и} \quad X_t - X_u$$

независимы.

Для независимых случайных величин характеристическая функция суммы равна произведению характеристических функций. Поэтому

$$\begin{aligned} \varphi_{s,t}(\tau) &= \mathbb{E} \exp(i\tau(X_t - X_s)) \\ &= \mathbb{E} \exp(i\tau(X_u - X_s)) \mathbb{E} \exp(i\tau(X_t - X_u)) \\ &= \varphi_{s,u}(\tau) \varphi_{u,t}(\tau). \end{aligned}$$

Таким образом, это условие выражает согласованность распределений приращений на вложенных временных интервалах.

## 2 Пуассоновский процесс

Пуассоновский процесс используется для моделирования **числа событий, произошедших к моменту времени  $t$** .

Параметр

$$\lambda > 0$$

называется **интенсивностью процесса**.

### Определение

Случайный процесс  $(N_t)_{t \geq 0}$  называется **пуассоновским процессом интенсивности  $\lambda > 0$** , если:

1. начальное значение равно нулю:

$$N_0 = 0 \quad \text{п.н.};$$

2. процесс имеет независимые приращения;
3. для любых  $0 \leq s < t$

$$N_t - N_s \sim \text{Pois}(\lambda(t - s)).$$

В частности,

$$N_t \sim \text{Pois}(\lambda t),$$

поэтому

$$\mathbb{E}N_t = \lambda t, \quad \text{Var}(N_t) = \lambda t.$$

### Интерпретация интенсивности

Так как

$$N_t - N_s \sim \text{Pois}(\lambda(t - s)),$$

то

$$\mathbb{E}(N_t - N_s) = \lambda(t - s).$$

Следовательно,  $\lambda$  можно интерпретировать как **среднее число событий в единицу времени**.

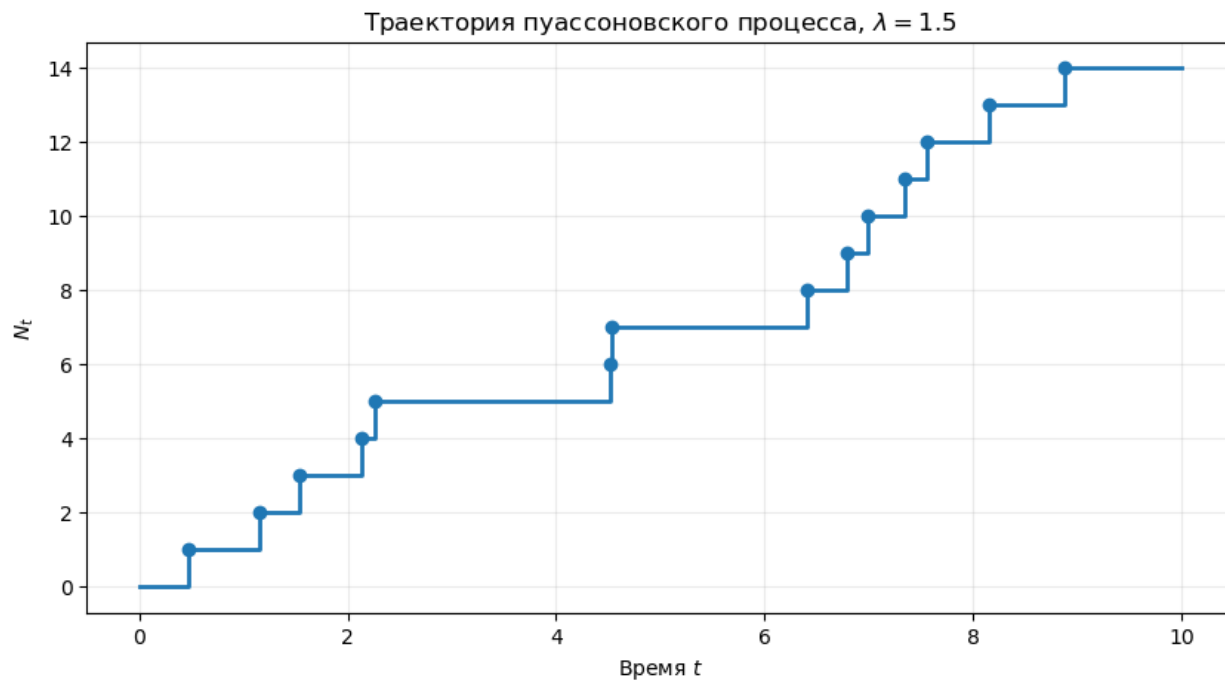
Например, на промежутке длины  $h$

$$\mathbb{E}(N_{t+h} - N_t) = \lambda h.$$

При этом распределение числа событий зависит только от **длины промежутка  $h$** , а не от его положения на временной оси.

## 2.1 Симуляция: как выглядит пуассоновский процесс?

Сначала визуализируем одну траекторию. Скачки происходят в случайные моменты времени, а между скачками процесс остаётся постоянным.

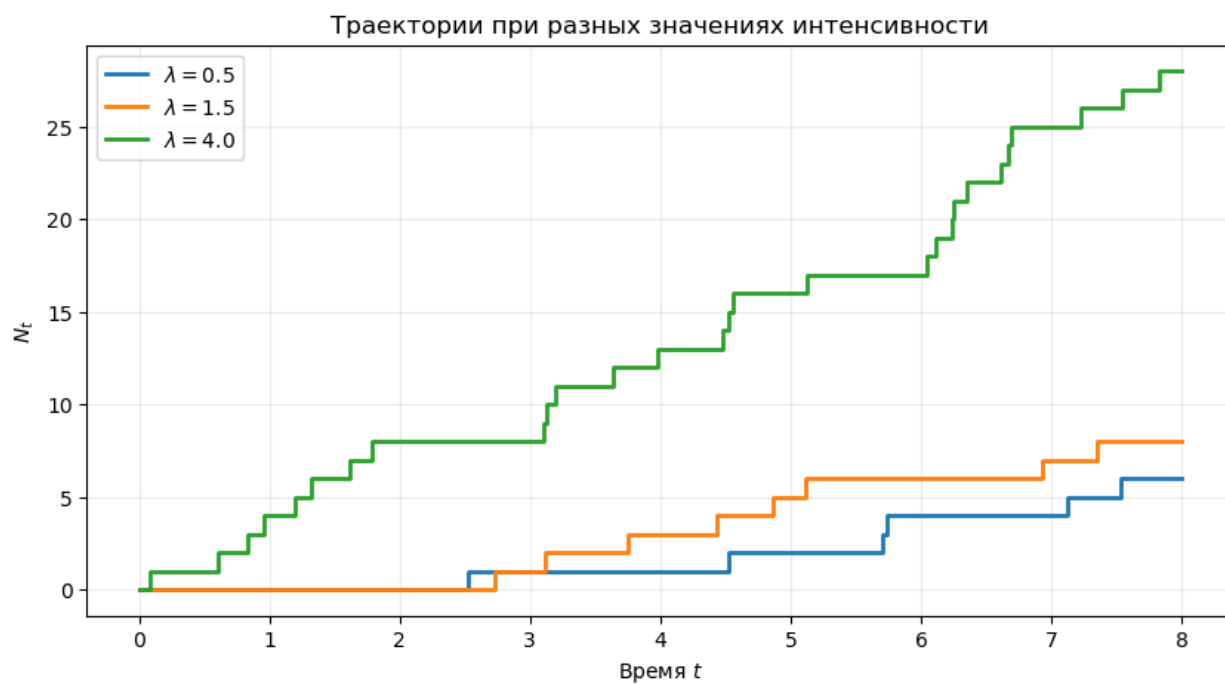


Число событий к моменту  $T=10$ : 14

Первые моменты скачков: [0.472 1.155 1.534 2.131 2.269 4.524 4.531 6.404]

### 2.1.1 Влияние интенсивности $\lambda$

Чем больше  $\lambda$ , тем меньше среднее время ожидания между событиями и тем быстрее растёт процесс.



## 2.2 Существование пуассоновского процесса

### ! Утверждение

Пуассоновский процесс интенсивности  $\lambda > 0$  существует.

### Доказательство

Рассмотрим распределение

$$Q_{s,t} = \text{Pois}(\lambda(t-s)).$$

Характеристическая функция пуассоновского распределения имеет вид

$$\varphi_{s,t}(\tau) = \exp \{ \lambda(t-s)(e^{i\tau} - 1) \}.$$

Для  $s < u < t$  имеем

$$\begin{aligned} \varphi_{s,u}(\tau) \varphi_{u,t}(\tau) &= e^{\lambda(u-s)(e^{i\tau}-1)} e^{\lambda(t-u)(e^{i\tau}-1)} \\ &= e^{\lambda(t-s)(e^{i\tau}-1)} \\ &= \varphi_{s,t}(\tau). \end{aligned}$$

Следовательно, выполнен критерий существования процесса с независимыми приращениями. Таким образом, пуассоновский процесс существует.

## 2.3 Траектории пуассоновского процесса

Из определения сразу следуют два важных свойства.

### 2.3.1 Целочисленность

Поскольку

$$N_t \sim \text{Pois}(\lambda t),$$

то для каждого фиксированного  $t$

$$N_t \in \mathbb{Z}_+.$$

То есть пуассоновский процесс является **счётным процессом**.

### 2.3.2 Монотонность

Для  $s < t$

$$N_t - N_s \sim \text{Pois}(\lambda(t-s)),$$

поэтому

$$N_t - N_s \geq 0.$$

Следовательно, траектории пуассоновского процесса являются **неубывающими**.

Возникают два естественных вопроса:

1. Каков размер скачков процесса?
2. Как распределены моменты этих скачков?

Для ответа на них построим пуассоновский процесс явно.

## 2.4 Симуляция: распределение $N_t$

Для фиксированного момента  $t$  должно выполняться

$$N_t \sim \text{Pois}(\lambda t).$$

Проверим это методом Монте-Карло и сравним эмпирические частоты с точной пуассоновской вероятностью.

## 3 Явная конструкция пуассоновского процесса

Пусть

$$\xi_1, \xi_2, \dots$$

— независимые одинаково распределённые случайные величины такие, что

$$\xi_i \sim \text{Exp}(\lambda).$$

Определим

$$S_0 = 0, \quad S_n = \xi_1 + \dots + \xi_n.$$

Величину  $S_n$  можно интерпретировать как **момент наступления  $n$ -го события**.

Теперь положим

$$X_t = \sup\{n : S_n \leq t\}.$$

То есть  $X_t$  — число событий, успевших произойти к моменту времени  $t$ .

**! Теорема — явная конструкция пуассоновского процесса**

Пусть

$$\xi_1, \xi_2, \dots \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\lambda),$$

и

$$S_n = \xi_1 + \dots + \xi_n.$$

Тогда процесс

$$X_t = \sup\{n : S_n \leq t\}$$

является пуассоновским процессом интенсивности  $\lambda$ .

Таким образом, пуассоновский процесс можно построить следующим образом:

- генерируем независимые экспоненциальные времена ожидания  $\xi_1, \xi_2, \dots$ ;

- определяем моменты событий

$$S_n = \sum_{k=1}^n \xi_k;$$

- считаем число событий, произошедших до момента  $t$ :

$$X_t = \#\{n : S_n \leq t\}.$$

### 3.1 Симуляция: экспоненциальные времена ожидания $\longrightarrow$ пуассоновский процесс

Эта визуализация буквально реализует явную конструкцию:

$$\xi_i \sim \text{Exp}(\lambda), \quad S_n = \xi_1 + \dots + \xi_n, \quad N_t = \#\{n : S_n \leq t\}.$$

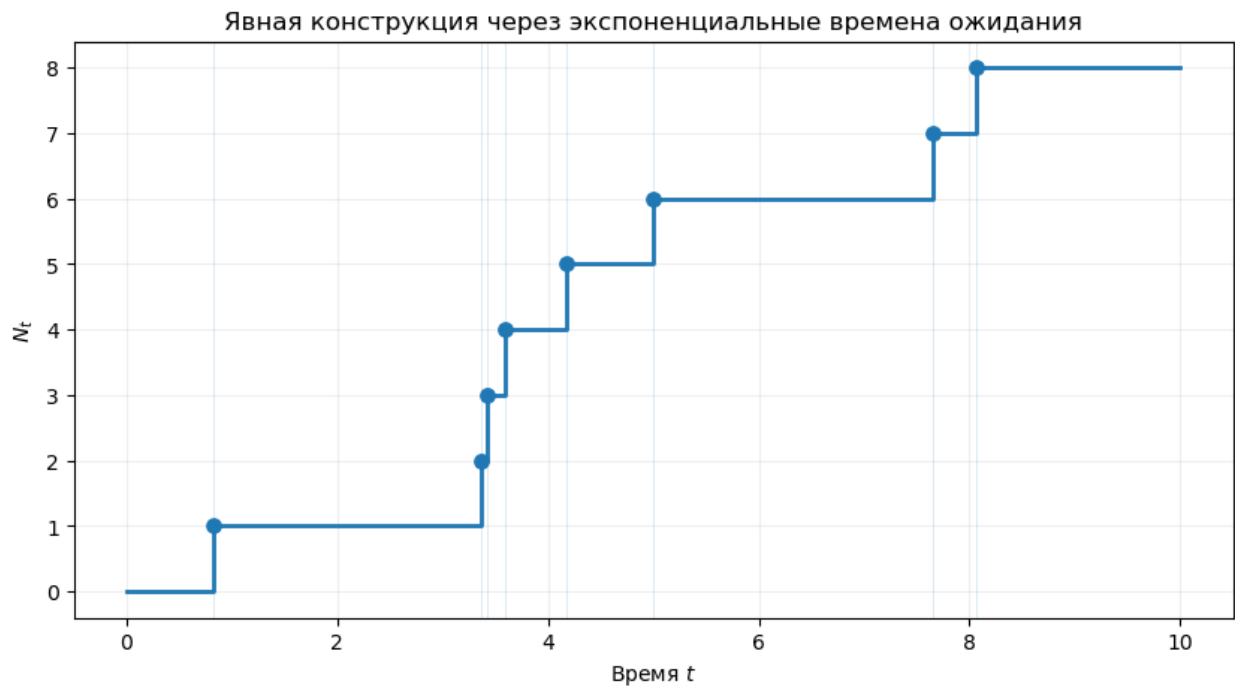
Сначала генерируются времена ожидания  $\xi_i$ , затем накапливаются моменты событий  $S_n$ , после чего строится ступенчатая траектория  $N_t$ .

Первые времена ожидания  $\xi_i$ :

[0.82 2.544 0.056 0.171 0.581 0.833 2.646 0.416]

Соответствующие моменты событий  $S_n$ :

[0.82 3.364 3.42 3.591 4.172 5.004 7.65 8.066]



### 3.2 Совместное распределение моментов событий

Для доказательства рассмотрим случайный вектор

$$(S_1, \dots, S_m).$$

Поскольку

$$S_1 = \xi_1, \quad S_2 - S_1 = \xi_2, \quad \dots, \quad S_m - S_{m-1} = \xi_m,$$

получаем

$$\begin{aligned} p_{S_1, \dots, S_m}(x_1, \dots, x_m) &= p_{\xi_1, \dots, \xi_m}(x_1, x_2 - x_1, \dots, x_m - x_{m-1}) \\ &= \prod_{k=1}^m \lambda e^{-\lambda(x_k - x_{k-1})} \mathbf{1}_{\{x_k > x_{k-1}\}}, \end{aligned}$$

где  $x_0 = 0$ .

После сокращения экспонент получаем

$$p_{S_1, \dots, S_m}(x_1, \dots, x_m) = \lambda^m e^{-\lambda x_m} \mathbf{1}_{\{0 < x_1 < \dots < x_m\}}.$$

💡 Почему полученный процесс является пуассоновским?

Возьмём

$$0 = t_0 < t_1 < \dots < t_n$$

и целые числа

$$0 = k_0 \leq k_1 \leq \dots \leq k_n.$$

Событие

$$X_{t_i} - X_{t_{i-1}} = k_i - k_{i-1}$$

означает, что на промежутке

$$(t_{i-1}, t_i]$$

произошло ровно  $k_i - k_{i-1}$  событий.

Используя совместную плотность моментов событий  $S_1, S_2, \dots$ , можно получить

$$\begin{aligned} \mathbb{P}(X_{t_1} = k_1, X_{t_2} - X_{t_1} = k_2 - k_1, \dots, X_{t_n} - X_{t_{n-1}} = k_n - k_{n-1}) \\ = \prod_{i=1}^n \frac{(\lambda(t_i - t_{i-1}))^{k_i - k_{i-1}}}{(k_i - k_{i-1})!} e^{-\lambda(t_i - t_{i-1})}. \end{aligned}$$

Правая часть является произведением вероятностей пуассоновских распределений.

Следовательно,

$$X_{t_i} - X_{t_{i-1}} \sim \text{Pois}(\lambda(t_i - t_{i-1})),$$

и эти приращения независимы.

Значит,  $(X_t)_{t \geq 0}$  является пуассоновским процессом интенсивности  $\lambda$ .

**i** Откуда появляется фактор  $\frac{(t_i - t_{i-1})^m}{m!}$ ?

В доказательстве возникает интеграл по области

$$t_{i-1} < x_1 < x_2 < \dots < x_m \leq t_i.$$

Куб

$$(t_{i-1}, t_i]^m$$

имеет объём

$$(t_i - t_{i-1})^m.$$

Все  $m!$  возможных порядков координат имеют одинаковый объём.  
Поэтому область

$$t_{i-1} < x_1 < \dots < x_m \leq t_i$$

занимает  $\frac{1}{m!}$  объёма куба:

$$\int_{t_{i-1} < x_1 < \dots < x_m \leq t_i} dx_1 \dots dx_m = \frac{(t_i - t_{i-1})^m}{m!}.$$

Именно отсюда в вычислениях возникает фактор  $m!$  в знаменателе пуассоновской вероятности.

### 3.3 Размер скачков пуассоновского процесса

Из явной конструкции видно, что процесс увеличивается в моменты

$$S_1, S_2, S_3, \dots$$

Поскольку

$$S_{n+1} - S_n = \xi_{n+1},$$

где

$$\xi_{n+1} \sim \text{Exp}(\lambda),$$

и экспоненциальное распределение абсолютно непрерывно, имеем

$$\mathbb{P}(\xi_{n+1} = 0) = 0.$$

Следовательно,

$$\mathbb{P}(S_n = S_{n+1}) = 0.$$

Таким образом, с вероятностью 1 два события не происходят одновременно.

**! Размер скачков**

С вероятностью 1 все скачки пуассоновского процесса имеют размер ровно

$$1.$$

### 3.4 Моменты скачков

Обозначим через  $Y_n$  момент  $n$ -го скачка пуассоновского процесса.

Из явной конструкции

$$Y_n = S_n = \xi_1 + \dots + \xi_n,$$

где

$$\xi_i \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\lambda).$$

Сумма  $n$  независимых экспоненциальных случайных величин имеет гамма-распределение.

**! Распределение момента  $n$ -го скачка**

$$Y_n \sim \Gamma(\lambda, n).$$

Его плотность равна

$$f_{Y_n}(x) = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} \mathbf{1}_{\{x>0\}}.$$

### 3.5 Симуляция: распределение момента $n$ -го скачка

По теории

$$Y_n = \xi_1 + \dots + \xi_n \sim \Gamma(\lambda, n).$$

Сравним эмпирическое распределение  $Y_n$  с теоретической гамма-плотностью.



Теоретическое  $E[Y_n] = n/\lambda = 4.2857$   
 Эмпирическое среднее  $= 4.2868$   
 Теоретическая  $\text{Var}(Y_n) = n/\lambda^2 = 3.0612$   
 Эмпирическая дисперсия  $= 3.0521$

### 3.6 Времена между последовательными событиями

Рассмотрим разности

$$Y_n - Y_{n-1}.$$

Из явной конструкции

$$Y_n - Y_{n-1} = S_n - S_{n-1} = \xi_n.$$

Следовательно,

$$Y_n - Y_{n-1} \sim \text{Exp}(\lambda).$$

Кроме того, случайные величины

$$Y_1, \quad Y_2 - Y_1, \quad Y_3 - Y_2, \quad \dots$$

независимы.

**! Important**

Таким образом, **времена ожидания между последовательными событиями пуассоновского процесса независимы и имеют экспоненциальное распределение с параметром  $\lambda$ .**

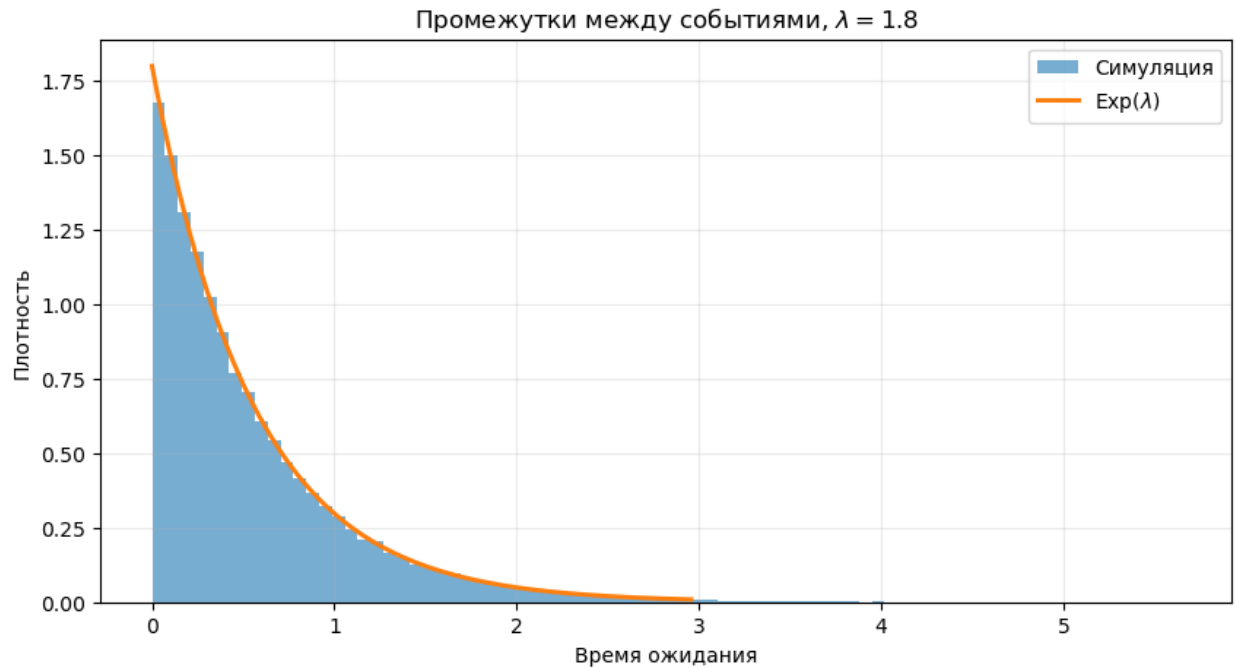
### 3.7 Симуляция: экспоненциальные промежутки между событиями

Проверим распределение времени между последовательными скачками:

$$Y_n - Y_{n-1} \sim \text{Exp}(\lambda).$$

Экспоненциальное распределение особенно важно благодаря свойству отсутствия памяти:

$$\mathbb{P}(\xi > s + t \mid \xi > s) = \mathbb{P}(\xi > t).$$



Проверка свойства отсутствия памяти:

$$\mathbb{P}(\xi > s+t \mid \xi > s) \approx 0.1992$$

$$\mathbb{P}(\xi > t) \approx 0.1968$$

$$\text{Точное значение} = 0.1979$$

### 3.8 Две точки зрения на пуассоновский процесс

Пуассоновский процесс можно описывать двумя эквивалентными способами.

#### 3.8.1 1. Считать число событий

Для любого  $0 \leq s < t$

$$N_t - N_s \sim \text{Pois}(\lambda(t - s)).$$

То есть число событий на промежутке длины  $h$  имеет распределение

$$\text{Pois}(\lambda h).$$

### 3.8.2 2. Рассматривать времена ожидания

Пусть

$$Y_1 < Y_2 < \dots$$

— моменты событий. Тогда

$$Y_n - Y_{n-1} \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\lambda).$$

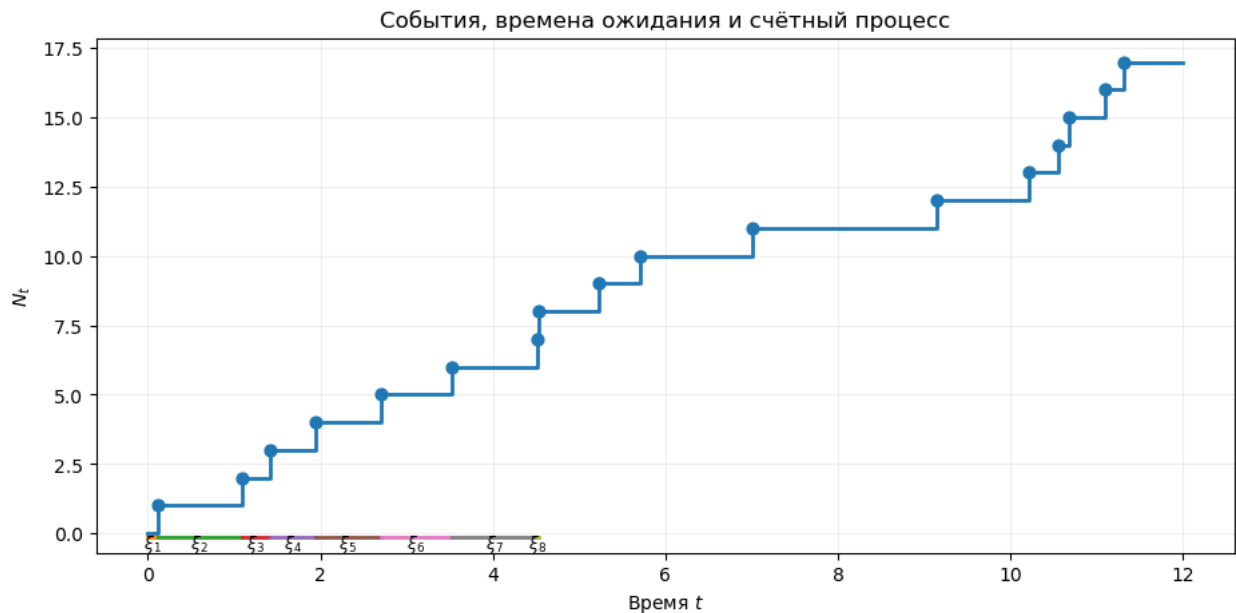
Таким образом,

$$\text{пуассоновское число событий} \iff \text{экспоненциальные времена ожидания.}$$

### 3.9 Симуляция: две эквивалентные точки зрения на одной траектории

В одной и той же реализации можно одновременно смотреть на процесс двумя способами:

- по горизонтали — на случайные времена ожидания между событиями;
- по вертикали — на число событий, накопленное к моменту  $t$ .



💡 Что важно увидеть на симуляциях?

1. Траектория пуассоновского процесса ступенчатая и неубывающая.
2. Каждый скачок имеет размер 1.
3. Приращение на интервале длины  $h$  имеет распределение  $\text{Pois}(\lambda h)$ .
4. Приращения на непересекающихся интервалах независимы.
5. Времена между скачками имеют распределение  $\text{Exp}(\lambda)$ .
6. Момент  $n$ -го скачка является суммой  $n$  экспоненциальных величин и имеет гамма-распределение.

## 4 Основные результаты

В этой теме мы получили следующие результаты.

1. **Процесс с независимыми приращениями** удовлетворяет условию независимости изменений процесса на непересекающихся временных промежутках.
2. Для существования процесса с заданными распределениями приращений необходимо и достаточно условие согласованности

$$\varphi_{s,t}(\tau) = \varphi_{s,u}(\tau)\varphi_{u,t}(\tau).$$

3. **Пуассоновский процесс** интенсивности  $\lambda$  характеризуется свойствами

$$N_0 = 0,$$

независимостью приращений и

$$N_t - N_s \sim \text{Pois}(\lambda(t - s)).$$

4. Пуассоновский процесс можно явно построить с помощью независимых случайных величин

$$\xi_i \sim \text{Exp}(\lambda)$$

и моментов событий

$$S_n = \xi_1 + \dots + \xi_n.$$

5. Все скачки пуассоновского процесса с вероятностью 1 имеют размер 1.
6. Момент  $n$ -го скачка имеет распределение

$$Y_n \sim \Gamma(\lambda, n).$$

7. Времена между последовательными скачками независимы и

$$Y_n - Y_{n-1} \sim \text{Exp}(\lambda).$$

## 5 Где возникает пуассоновский процесс?

Пуассоновский процесс — одна из базовых моделей для **случайных событий, происходящих во времени**.

Основная идея очень проста:

если события происходят независимо друг от друга с некоторой приблизительно постоянной интенсивностью  $\lambda$ , то число событий за время  $t$  часто можно моделировать с помощью пуассоновского процесса.

В такой модели

$$N_t = \text{число событий, произошедших к моменту } t,$$

и

$$N_t - N_s \sim \text{Pois}(\lambda(t - s)).$$

При этом время между последовательными событиями имеет распределение

$$Y_n - Y_{n-1} \sim \text{Exp}(\lambda).$$

Эта простая конструкция возникает во множестве прикладных задач:

- машинное обучение;
- анализ данных;
- финансы;
- физика;
- биология;
- телекоммуникации;
- теория массового обслуживания;
- анализ надёжности.

#### **i** Machine Learning: данные о событиях

Во многих задачах машинного обучения наблюдения имеют форму **последовательности событий**

$$T_1 < T_2 < T_3 < \dots$$

вместо обычной таблицы независимых наблюдений.

Например:

- пользователь нажал на рекламу;
- пользователь совершил покупку;
- пришёл новый запрос к серверу;
- произошло мошенническое действие;
- пользователь открыл приложение.

Можно определить

$$N_t = \text{число событий, произошедших до момента } t.$$

Простейшая модель предполагает

$$(N_t)_{t \geq 0} \sim \text{Poisson Process}(\lambda).$$

Тогда

$$N_t - N_s \sim \text{Pois}(\lambda(t - s)).$$

Параметр  $\lambda$  характеризует **частоту появления событий**.

#### **5.0.1 Обучение модели: оценивание интенсивности**

Предположим, что за промежуток времени  $[0, T]$  мы наблюдали  $n$  событий.

Для пуассоновского процесса

$$N_T \sim \text{Pois}(\lambda T).$$

Следовательно,

$$\mathbb{P}(N_T = n) = e^{-\lambda T} \frac{(\lambda T)^n}{n!}.$$

Функция правдоподобия имеет вид

$$L(\lambda) = e^{-\lambda T} \frac{(\lambda T)^n}{n!}.$$

Логарифмическое правдоподобие:

$$\ell(\lambda) = -\lambda T + n \log \lambda + n \log T - \log(n!).$$

Дифференцируя,

$$\ell'(\lambda) = -T + \frac{n}{\lambda}.$$

Поэтому

$$\hat{\lambda}_{MLE} = \frac{N_T}{T}.$$

То есть естественная оценка интенсивности — просто **число наблюдаемых событий в единицу времени**.

### **i** От пуассоновского процесса к современным ML-моделям

Обычный пуассоновский процесс предполагает постоянную интенсивность

$$\lambda(t) = \lambda.$$

Но в реальных данных частота событий обычно зависит от времени и признаков.  
Например, вероятность покупки может зависеть от

$$x = (\text{время суток, день недели, возраст пользователя, цена, реклама, } \dots).$$

Поэтому вместо постоянной интенсивности рассматривают

$$\lambda(t, x).$$

Например,

$$\lambda(t, x) = \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p).$$

В более сложных моделях  $\lambda(t, x)$  может задаваться нейронной сетью.

Таким образом, пуассоновский процесс является отправной точкой для изучения **point processes** и современных моделей событийных данных.

### **i** Data Science: трафик и действия пользователей

Рассмотрим веб-сервис.

Пусть

$$N_t = \text{число запросов к серверу до момента } t.$$

Если запросы приходят приблизительно независимо с интенсивностью

$$\lambda = 120 \text{ запросов в минуту,}$$

то за интервал длины 30 секунд модель даёт

$$N_{t+0.5} - N_t \sim \text{Pois}(60).$$

Следовательно,

$$\mathbb{E}[N_{t+0.5} - N_t] = 60.$$

Такая модель позволяет оценивать:

- ожидаемую нагрузку;
- вероятность необычно большого числа запросов;
- необходимую пропускную способность;
- вероятность отсутствия запросов;
- время до следующего запроса.

### 5.0.2 Обнаружение аномалий

Предположим, что обычно сервер получает в среднем  $\lambda = 5$  запросов в секунду. Тогда число запросов за одну секунду можно моделировать как

$$X \sim \text{Pois}(5).$$

Если внезапно наблюдается

$$X = 20,$$

можно вычислить

$$\mathbb{P}(X \geq 20).$$

Если эта вероятность очень мала, наблюдение можно считать **аномальным**.

Таким образом, пуассоновская модель даёт простой статистический baseline для задач:

- обнаружения аномального сетевого трафика;
- fraud detection;
- мониторинга серверов;
- обнаружения необычной активности пользователей.

## **i** Финансы: случайные моменты сделок

Финансовые рынки естественным образом генерируют событийные данные. Например, можно определить

$$N_t = \text{число сделок с активом до момента } t.$$

В простейшей модели предполагается, что сделки происходят согласно пуассоновскому процессу:

$$N_t - N_s \sim \text{Pois}(\lambda(t - s)).$$

Тогда время между последовательными сделками

$$\tau_n = Y_n - Y_{n-1}$$

удовлетворяет

$$\tau_n \sim \text{Exp}(\lambda).$$

Параметр  $\lambda$  в этом случае можно интерпретировать как **интенсивность торгов**.

### 5.0.3 Compound Poisson Process

Часто нас интересует не только **количество событий**, но и их случайный размер.

Пусть

$$N_t$$

— число событий до момента  $t$ , а

$$Z_1, Z_2, \dots$$

— независимые случайные величины, описывающие размер каждого события.  
Тогда процесс

$$X_t = \sum_{k=1}^{N_t} Z_k$$

называется **составным пуассоновским процессом** (Compound Poisson Process).

Например, в финансах  $Z_k$  может описывать изменение цены в момент  $k$ -го скачка.

Тогда цена может моделироваться как

$$S_t = S_0 + \sum_{k=1}^{N_t} Z_k.$$

В отличие от непрерывных моделей типа броуновского движения, такая модель позволяет цене совершать **резкие случайные скачки**.

#### **i** Страхование: классическая модель риска

Одно из важнейших применений пуассоновского процесса возникает в страховании.

Пусть

$$N_t = \text{число страховых случаев до момента } t,$$

а

$$Z_1, Z_2, \dots$$

— размеры страховых выплат.

Тогда суммарный размер выплат к моменту  $t$  равен

$$C_t = \sum_{k=1}^{N_t} Z_k.$$

Если  $N_t$  — пуассоновский процесс, то  $C_t$  является составным пуассоновским процессом.

Если компания получает страховые премии с постоянной скоростью  $c$ , её капитал можно моделировать как

$$R_t = R_0 + ct - \sum_{k=1}^{N_t} Z_k.$$

Это одна из базовых моделей **теории риска**.

### **i** Физика: счётчики частиц и радиоактивный распад

Пуассоновский процесс естественно возникает при подсчёте редких физических событий. Например,

$N_t$  = число частиц, зарегистрированных детектором до момента  $t$ .

Если частицы регистрируются независимо с приблизительно постоянной интенсивностью  $\lambda$ , то

$$N_t \sim \text{Pois}(\lambda t).$$

Следовательно,

$$\mathbb{E}N_t = \text{Var}(N_t) = \lambda t.$$

Именно поэтому в экспериментах, основанных на подсчёте событий, часто возникает характерное соотношение

дисперсия  $\approx$  среднее.

А стандартное отклонение числа событий имеет порядок

$$\sqrt{\lambda t}.$$

### **i** Биология: мутации как случайные события

Пуассоновские модели также возникают в биологии. Например, можно рассматривать

$N_t$  = число мутаций, возникших к моменту  $t$ .

В простейшей модели предполагается, что мутации:

- возникают независимо;
- происходят с постоянной интенсивностью  $\mu$ .

Тогда

$$N_t \sim \text{Pois}(\mu t).$$

В частности,

$$\mathbb{P}(N_t = 0) = e^{-\mu t},$$

то есть вероятность того, что за время  $t$  не произошло ни одной мутации, экспоненциально убывает с  $t$ .

### **i** Теория массового обслуживания

Рассмотрим очередь:

- клиенты приходят в банк;
- запросы поступают на сервер;
- машины приезжают на заправку;
- звонки поступают в call-центр.

Пусть

$N_t$  = число клиентов, пришедших к моменту  $t$ .

Одна из самых простых моделей предполагает

$$(N_t)_{t \geq 0} \sim \text{PoissonProcess}(\lambda).$$

Тогда времена между последовательными клиентами независимы и

$$Y_n - Y_{n-1} \sim \text{Exp}(\lambda).$$

Если дополнительно время обслуживания имеет экспоненциальное распределение с параметром  $\mu$ , возникает классическая модель очереди

$$M/M/1.$$

Эта модель связывает пуассоновские процессы с анализом производительности компьютерных систем, сетей и сервисов.

## 5.1 Когда пуассоновский процесс — плохая модель?

Пуассоновский процесс является очень полезной **baseline-моделью**, но его предположения достаточно сильны.

Классическая модель предполагает:

1. независимость событий;
2. постоянную интенсивность  $\lambda$ ;
3. отсутствие одновременных событий;
4. экспоненциальные времена ожидания.

В реальных данных эти предположения часто нарушаются.

Например:

- пользователи чаще совершают покупки вечером, чем ночью;
- интенсивность торгов меняется в течение торгового дня;
- землетрясение может увеличивать вероятность последующих толчков;
- активность нейрона зависит от предыдущих импульсов;
- эпидемические события могут усиливать вероятность новых событий.

Поэтому пуассоновский процесс следует рассматривать как **базовую модель**, от которой переходят к более сложным моделям случайных событий.

### 💡 От пуассоновского процесса к более сложным моделям

Пуассоновский процесс — это начало большой теории **случайных точечных процессов**.

Если отказаться от некоторых его предположений, возникают более сложные модели:

$$\text{Poisson process} \longrightarrow \text{non-homogeneous Poisson process} \longrightarrow \text{Cox process} \longrightarrow \text{Hawkes process}.$$

Например:

- **Non-homogeneous Poisson process:** интенсивность зависит от времени,

$$\lambda = \lambda(t);$$

- **Cox process:** сама интенсивность  $\lambda(t)$  является случайным процессом;
- **Hawkes process:** произошедшее событие временно увеличивает интенсивность будущих событий,

$$\lambda(t) = \mu + \sum_{T_i < t} g(t - T_i).$$

Такие модели используются для анализа

- финансовых сделок;

- активности пользователей;
- событий в социальных сетях;
- нейронных импульсов;
- землетрясений;
- сетевого трафика.

Таким образом, простой пуассоновский процесс является естественной отправной точкой для многих современных моделей событийных данных.