

# Вероятностные модели и статистика случайных процессов

Конспекты лекций

ЛЕКТОР: А.В. АРТЁМОВ

НИУ ВШЭ, 2018



# Оглавление

<b>1</b>	<b>Лекции: чистовик</b>	<b>5</b>
1.1	Введение в курс . . . . .	5
1.2	Основы теории случайных процессов . . . . .	8
1.3	Непрерывность случайных процессов . . . . .	12
1.4	Лирическое отступление: гауссовские векторы . . . . .	14
1.5	Примеры случайных процессов . . . . .	19
1.5.1	Гауссовский и винеровский процессы . . . . .	19
1.5.2	Процесс Орнштейна-Уленбека . . . . .	20
1.5.3	Пуассоновский процесс . . . . .	21
1.6	Стационарность случайных процессов . . . . .	24
1.7	Эргодичность случайных процессов . . . . .	26
1.8	Генерирование реализаций случайных процессов . . . . .	27
1.8.1	Генерирование пуассоновских случайных процессов . . . . .	27
1.8.2	Метод стохастического интегрирования . . . . .	29
1.8.3	Метод гауссовских векторов . . . . .	30
1.8.4	Генерирование гауссовских случайных процессов . . . . .	30
1.9	Марковские цепи . . . . .	34
1.9.1	Основные понятия . . . . .	34
1.9.2	Пример применения марковских сетей: модель системы массового обслуживания . . . . .	41
1.10	Марковские процессы . . . . .	43
1.10.1	Непрерывное время, дискретные состояния . . . . .	43



# Глава 1

## Лекции: чистовик

### 1.1 Введение в курс

Как понятно из названия, этот пункт посвящён простой цели: вспомнить основные понятия, необходимые для дальнейшего чтения. Начнём с самого основного: *вероятностного пространства*.

**Определение 1.** Вероятностное пространство, или *тройка Колмогорова* — это тройка  $(\Omega, \mathcal{F}, \mathbb{P})$ , где

- $\Omega$  — пространство элементарных исходов.
- $\mathcal{F}$  —  $\sigma$ -алгебра над  $\Omega$ , пространство событий. Неформально говоря,  $\mathcal{F}$  определяет объекты, относительно которых делаются утверждения.
- $\mathbb{P} : \mathcal{F} \mapsto [0, 1]$  — вероятностная мера.

На всякий случай напомним определения сигма-алгебры и вероятностной меры.

**Определение 2.**  $\sigma$ -алгебра над множеством  $A$  — это множество  $\mathcal{F} \subseteq 2^A$ , обладающее следующими свойствами:

1.  $\emptyset \in \mathcal{F}$ .
2. Если  $X \in \mathcal{F}$ , то и дополнение  $A \setminus X \in \mathcal{F}$ .
3.  $\mathcal{F}$  замкнуто относительно счётного объединения, то есть объединение счётного подсемейства из  $\mathcal{F}$  лежит в  $\mathcal{F}$ .<sup>1</sup>

**Определение 3.** Пусть  $\Omega$  — пространство элементарных исходов, а  $\mathcal{F}$  —  $\sigma$ -алгебра его подмножеств (пространство событий). *Вероятностной мерой* или же вероятностью называется функция  $\mathbb{P} : \mathcal{F} \mapsto [0, 1]$ , удовлетворяющая двум свойствам:

1. (счётная аддитивность) Пусть  $\{A_n\}_{n=1}^\infty$  — последовательность попарно не пересекающихся событий. Тогда

$$\mathbb{P}\left(\bigsqcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

2. (нормированность)  $\mathbb{P}(\Omega) = 1$ .

Теперь посмотрим на несколько примеров вероятностных пространств.

---

<sup>1</sup>То есть это алгебра, к которой конечное объединение заменено на счётное.

**Пример 1.** Допустим, что человек бросает монетку. В данном случае элементарными исходами являются “орёл” и “решка”. Обозначим их за 0 и 1 соответственно. Тогда вероятностное пространство будет устроено следующим образом: пространство элементарных исходов  $\Omega = \{0, 1\}$ , пространство событий  $\mathcal{F} = \{\emptyset, \{0\}, \{1\}, \Omega\}$ , а вероятность  $P$  определяется следующим образом:

$$P(\emptyset) = 0, \quad P(\{1\}) = p, \quad P(\{0\}) = 1 - p, \quad P(\Omega) = 1, \quad p \in [0, 1].$$

**Пример 2.** Теперь подбросим монетку  $n$  раз. В таком случае  $\Omega = \{0, 1\}^n$ , то есть если  $\omega \in \Omega$ , то  $\omega = (\omega_1, \dots, \omega_n)$ , где  $\omega_i \in \{0, 1\}$ . Пространство событий же введём просто как множество всех подмножеств:  $\mathcal{F} = 2^\Omega$ . Введение же вероятностной меры оставим читателю.

В дискретном случае алгебра событий, как известно, очень проста: обычно это просто множество всех подмножеств. Теперь перейдём от дискретного случая к более общему. Пусть  $\Omega = \mathbb{R}$ . Как тогда ввести алгебру событий? Ведь, как известно, в  $\mathbb{R}$  есть неизмеримые подмножества. Для этого вводят *борелевскую  $\sigma$ -алгебру*  $\mathcal{B}(\mathbb{R})$  и говорят, что  $\mathcal{F} = \mathcal{B}(\mathbb{R})$ . Напомню определение:

**Определение 4.** Борелевская  $\sigma$ -алгебра над  $\mathbb{R}$   $\mathcal{B}(\mathbb{R})$  — это минимальная (по включению) сигма-алгебра над  $\mathbb{R}$ , содержащая все полуинтервалы вида  $(a, b]$ .<sup>2</sup>

Всё это вводилось ради того, с чем мы работаем на постоянной основе: ради *случайных величин*.

**Определение 5.** Пусть  $(\Omega, \mathcal{F}, P)$  — вероятностное пространство. Случайной величиной будем называть отображение  $\xi : \Omega \mapsto \mathbb{R}$  такое, что для любого  $x \in \mathbb{R}$  событие  $\{\xi \leq x\} = \{\omega \in \Omega : \xi(\omega) \leq x\}$  лежит в  $\mathcal{F}$ .<sup>3</sup>

Характеризовать поведение случайной величины помогает *функция распределения*:

**Определение 6.** Пусть  $\xi$  — случайная величина на вероятностном пространстве  $(\Omega, \mathcal{F}, P)$ . Функцией распределения случайной величины  $\xi$  называют функцию  $F_\xi : \mathbb{R} \mapsto [0, 1]$ , определяемую следующим образом:  $F_\xi(x) = P(\xi \leq x)$ .

Дальше по плану сходимости. Как известно, их четыре типа, и они не равноценны. Чтобы не повторяться, сразу же введём обозначения. Пусть  $(\Omega, \mathcal{F}, P)$  — вероятностное пространство, а  $\{\xi_n\}_{n=1}^\infty$  и  $\xi$  — случайные величины на нём.

**Определение 7.**  $\xi_n$  сходится к  $\xi$  почти наверное, если

$$P\left(\left\{\omega \in \Omega : \lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)\right\}\right) = 1.$$

Обозначение:  $\xi_n \xrightarrow{\text{п.н.}} \xi$  или же просто  $\xi_n \rightarrow \xi$  (если не указан вид сходимости и он не ясен из контекста, то сходимость идёт почти наверное).

**Определение 8.**  $\xi_n$  сходится к  $\xi$  по вероятности, если для любого  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P(\{\omega \in \Omega : |\xi_n(\omega) - \xi(\omega)| > \varepsilon\}) = 0.$$

Обозначение:  $\xi_n \xrightarrow{P} \xi$ .

<sup>2</sup>Хоть она и минимальна, но всё равно она огромна — найти неборелевское множество не так уж и просто.

<sup>3</sup>Это свойство принято называть измеримостью. Такое требование нужно для того, чтобы были осмысленны вопросы типа “Чему равна вероятность того, что значение случайной величины будет лежать в таком-то отрезке?”.

**Определение 9.**  $\xi_n$  сходится к  $\xi$  в среднем порядка  $p$ , если

$$\lim_{n \rightarrow \infty} E[|\xi_n - \xi|^p] = 0.$$

Обозначение:  $\xi_n \xrightarrow{L^p} \xi$ .

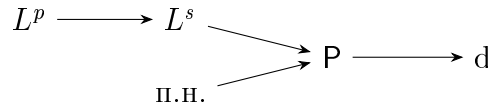
На практике обычно берётся  $p = 2$ . В таком случае говорят, что имеет место *сходимость в среднеквадратичном смысле*. У такой сходимости есть аж три разных обозначения:  $\xi_n \xrightarrow{L^2} \xi$ ,  $\xi_n \xrightarrow{\text{с.к.}} \xi$  или же вообще  $\xi = \text{l.i.m.} \xi_n$ .

**Определение 10.**  $\xi_n$  сходятся к  $\xi$  по распределению,<sup>4</sup> если для любой ограниченной непрерывной функции  $f : \mathbb{R} \mapsto \mathbb{R}$

$$\lim_{n \rightarrow \infty} E[f(\xi_n)] = E[f(\xi)].$$

Обозначение:  $\xi_n \xrightarrow{d} \xi$ .

Цепочка взаимосвязей сходимостей устроена следующим образом. На данном рисунке  $p > s \geq 1$  и стрелка из  $A$  в  $B$  означает “из  $A$  следует  $B$ ”:



Поехали дальше. Для случайных величин вводились такие вещи, как матожидание и дисперсия. Напомню:

**Определение 11.** Математическое ожидание случайной величины  $\xi$  с функцией распределения  $F_\xi$  и плотностью  $p_\xi$  — это интеграл

$$E[\xi] \equiv \int_{-\infty}^{+\infty} x dF_\xi(x) = \int_{-\infty}^{+\infty} x p_\xi(x) dx$$

Рядом с матожиданием вводятся *дисперсия*, *ковариация* и *корреляция*.

**Определение 12.** Дисперсией случайной величины  $\xi$  называется  $D[\xi] = E[(\xi - E[\xi])^2] = E[\xi^2] - (E[\xi])^2$ . Корень из дисперсии  $\sigma$  называется *среднеквадратичным отклонением*.

**Определение 13.** Ковариацией случайных величин  $\xi$  и  $\eta$  называется

$$\text{cov}(\xi, \eta) = E[(\xi - E[\xi])(\eta - E[\eta])] = E[\xi\eta] - E[\xi] E[\eta].$$

**Определение 14.** Корреляцией случайных величин  $\xi$  и  $\eta$  называется

$$\rho(\xi, \eta) = \frac{\text{cov}(\xi, \eta)}{\sqrt{D[\xi] D[\eta]}}.$$

В многомерном случае ситуация немного изменяется. Матожидание случайного вектора определяется, как вектор из матожиданий компонент. Ковариация (да и дисперсия тоже) случайных векторов  $\boldsymbol{\xi} \in \mathbb{R}^m$  и  $\boldsymbol{\eta} \in \mathbb{R}^n$  равна

$$\text{cov}(\boldsymbol{\xi}, \boldsymbol{\eta}) = E[(\boldsymbol{\xi} - E[\boldsymbol{\xi}])(\boldsymbol{\eta} - E[\boldsymbol{\eta}])^\top] = (\text{cov}(\xi_i, \eta_j))_{m \times n}.$$

При анализе некоторых вещей могут понадобиться условные матожидания. Введём их.

<sup>4</sup>Она так называется по той причине, что в одномерном случае её условие расносит сходимости функций распределения почти наверное.

**Определение 15.** Пусть  $\xi$  — случайная величина, а  $\boldsymbol{\eta}$  — случайный вектор из  $\mathbb{R}^n$ . Тогда условным математическим ожиданием  $\xi$  относительно  $\boldsymbol{\eta}$  называется случайная величина  $E[\xi | \boldsymbol{\eta}]$ , удовлетворяющая двум условиям:

1.  $E[\xi | \boldsymbol{\eta}] = \varphi(\boldsymbol{\eta})$ , где  $\varphi : \mathbb{R}^n \mapsto \mathbb{R}$  — некоторая борелевская функция (свойство измеримости).
2. Для любого  $B \in \mathcal{B}(\mathbb{R}^n)$   $E[\xi I\{\boldsymbol{\eta} \in B\}] = E[E[\xi | \boldsymbol{\eta}] I\{\boldsymbol{\eta} \in B\}]$  (интегральное свойство).

Но считать условное матожидание по определению весьма грустно. Поэтому введём две теоремы, которые упрощают жизнь. Они опираются на понятие *условного распределения* и *условной плотности*.

**Определение 16.** Условным распределением случайной величины  $\xi$  при условии, что  $\boldsymbol{\eta} = \mathbf{y}$  назовём функцию  $P(\xi \in B | \boldsymbol{\eta} = \mathbf{y}) \equiv E[I\{\xi \in B\} | \boldsymbol{\eta} = \mathbf{y}]$ , рассматриваемую, как функцию от  $B \in \mathcal{B}(\mathbb{R}^n)$  при фиксированном  $\mathbf{y} \in \mathbb{R}^k$ .

**Определение 17.** Если условное распределение имеет плотность  $p_{\xi|\boldsymbol{\eta}}(x | \mathbf{y})$ , то назовём его *условной плотностью  $\xi$  относительно  $\boldsymbol{\eta}$* . То есть для любого  $B \in \mathcal{B}(\mathbb{R})$

$$P(\xi \in B | \boldsymbol{\eta} = \mathbf{y}) = \int_B p_{\xi|\boldsymbol{\eta}}(x | \mathbf{y}) dx.$$

**Теорема 1** (о вычислении условного математического ожидания). Пусть  $\xi$  — случайная величина, а  $f(x)$  — некоторая борелевская функция. Если  $E[|f(\xi)|] < +\infty$  и существует плотность  $p_{\xi|\boldsymbol{\eta}}(x | \mathbf{y})$ , то

$$E[f(\xi) | \boldsymbol{\eta} = \mathbf{y}] = \int_{\mathbb{R}} f(x) p_{\xi|\boldsymbol{\eta}}(x | \mathbf{y}) dx.$$

**Теорема 2.** Пусть  $\xi$  и  $\eta$  таковы, что есть совместная плотность  $p_{\xi,\eta}(x, y)$ . Тогда существует условная плотность  $p_{\xi|\eta}(x | y)$  и она равна

$$p_{\xi|\eta}(x | y) = \frac{p_{\xi,\eta}(x, y)}{p_{\eta}(y)} I\{p_{\eta}(y) > 0\}$$

Обобщение на многомерный случай ровно такое же, как и для обычных матожиданий.

## 1.2 Основы теории случайных процессов

Теперь можно приступать к делу. Для начала введём понятие случайной функции.

**Определение 18.** Пусть  $(\Omega, \mathcal{F}, P)$  — вероятностное пространство, а  $T$  — произвольное неслучайное множество (множество индексов). Тогда любое множество случайных величин  $X = \{X_t | t \in T\}$  называется *случайной функцией*.

Случайные функции можно классифицировать по строению множества индексов.

- Если  $T = \{1\}$  (или любое другое одноэлементное множество), то  $X = X_1 : \Omega \mapsto \mathbb{R}$  — случайная величина.
- Если  $T = \{1, 2, \dots, N\}$ , то  $X = (X_1, \dots, X_N)$  — случайный вектор.



- Если  $T$  дискретно (то есть не более, чем счётно), то  $X$  называют *случайной последовательностью*. Например, возьмём  $T = \mathbb{N}$  и  $\Omega = \{0, 1\}$ . Тогда случайная последовательность  $X = (X_t)_{t \in T}$  соответствует броскам монетки.
- Если  $T \subseteq \mathbb{R}$ , то параметр  $t \in T$  можно характеризовать, как время. В этом случае  $X$  принято называть *случайным процессом*.
- Аналогично, если  $T \subseteq \mathbb{R}^d$ , где  $d \geq 1$ , то параметр можно характеризовать, как точку в  $d$ -мерном пространстве. Тогда  $X$  называют *случайным полем*. Например, если  $T = \mathbb{R}^3$ , то  $X_t(\omega)$  может соответствовать давлению в точке  $t = (x, y, z)$  в момент времени  $\omega$ .

Впрочем, данная классификация не является строгой. Например, понятие “случайный процесс” обычно считается безусловным синонимом термина “случайная функция”.

Случайные процессы ещё разбивают по мощности множества индексов. Если оно не более, чем счётно, то говорят, что случайный процесс *дискретен во времени*. Если же  $T$  континуально, то у случайного процесса *непрерывное время*.

Рассмотрим несколько примеров.

**Пример 3.** Пусть  $T = \mathbb{N}$ , а  $\{X_n\}_{n=1}^\infty$  — последовательность независимых и одинаково распределённых случайных величин с нулевым матожиданием.<sup>5</sup> Тогда  $X = (X_t)_{t \in T}$  — случайный процесс с дискретным временем, называемый *белым шумом*.

**Пример 4.** Пусть  $Y = (Y_t)_{t \in T}$  — последовательность iid случайных величин. Тогда построим новую последовательность случайных величин следующим образом:  $X_t = Y_1 + \dots + Y_t$ . Тогда  $X = (X_t)_{t \in T}$  — случайный процесс с дискретным временем, называемый *случайным блужданием*.

**Пример 5.** Пусть  $\xi_1, \dots, \xi_d$  — случайные величины. Тогда “случайный полином”

$$X = (X_t)_{t \in \mathbb{R}}, \quad X_t = \sum_{n=1}^d \xi_n t^n$$

образует случайный процесс.

**Пример 6.** Пусть некоторое устройство (например, жёсткий диск) начинает работу в момент времени 0 и ломается в момент времени  $T_1$ . В этот же момент его меняют на новое (временем замены пренебрегаем). Оно, в свою очередь, ломается спустя время  $T_2$ , после чего его снова заменяют и этот процесс продолжается. Введём случайные величины  $S_n = T_1 + \dots + T_n$ . Случайный процесс  $S = (S_t)_{t \in \mathbb{N}}$  называется процессом моментов восстановления. Сечения будут соответствовать моментам “восстановления”. Далее, введём следующий случайный процесс:

$$N = (N_t)_{t \in \mathbb{R}}, \quad N_t = \sum_{n=1}^{\infty} \mathbf{I}\{S_n \leq t\} = \#\{n \in \mathbb{N} \mid S_n \leq t\}.$$

Этот процесс принято называть *процессом восстановления*.

В дальнейшем нам понадобятся понятия *сечения* и *траектории* случайной функции. Введём их.

<sup>5</sup>Для удобства в дальнейшем будем иногда сокращать “независимые и одинаково распределённые” до iid (independent and identically distributed).

**Определение 19.** Пусть  $X = (X_t)_{t \in T}$  — случайная функция. Тогда  $X_t$  при фиксированном  $t$  называется *сечением* случайной функции.

**Определение 20.** Пусть  $X = (X_t)_{t \in T}$  — случайная функция и  $\omega \in \Omega$  — фиксированный исход. *Траекторией* случайной функции называется функция  $\varphi_\omega : T \mapsto \mathbb{R}$  такая, что  $\varphi_\omega(t) = X_t(\omega)$ .

Теперь вопрос: есть случайный процесс. Как описать вероятность того, что реализация удовлетворяет какому-то условию? Да и вообще, какие вопросы можно задавать относительно реализаций?

Посмотрим на пару частных случаев случайного процесса  $X = (X_t)_{t \in T}$

1. Пусть  $T = \{1\}$ . Тогда случайный процесс превращается в случайную величину  $\xi$ . Но её значения описываются функцией распределения  $F_\xi(x) = \mathbf{P}(\xi \leq x)$ .
2. Аналогично, пусть  $T = \{1, 2, \dots, n\}$ . Тогда  $X = (X_1, \dots, X_n)$  — это случайный вектор, а его значения описываются совместной функцией распределения

$$F_X(x_1, \dots, x_n) = \mathbf{P}(X_1 \leq x_1, \dots, X_n \leq x_n)$$

В конечном случае всё хорошо и можно обойтись совместной функцией распределения. А как перейти в бесконечный случай? Сделать прямое обобщение вряд ли получится: учитывать бесконечное число условий сложно и не факт, что возможно. Поэтому будем смотреть на конечные поднаборы. Тем самым мы пришли к следующему определению.

**Определение 21.** Пусть  $(\Omega, \mathcal{F}, \mathbf{P})$  — вероятностное пространство,  $T$  — множество индексов, а  $X = (X_t)_{t \in T}$  — случайный процесс. Тогда *семейством конечномерных распределений* случайного процесса  $X$  называется набор  $F$  его конечномерных функций распределения

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = \mathbf{P}(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n),$$

то есть

$$F = \{F_{t_1, \dots, t_n}(x_1, \dots, x_n) \mid n \in \mathbb{N}, t_1, \dots, t_n \in T\}.$$

**Пример 7.** Пусть  $X = (X_t)_{t \in \mathbb{N}}$  — случайный процесс, состоящий из iid случайных величин из распределения  $\mathcal{N}(\mu, \sigma^2)$ . Тогда

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = \prod_{k=1}^n F_{X_{t_k}}(x_k) = \prod_{k=1}^n \int_{-\infty}^{x_k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} dy.$$

Вообще говоря, конечномерное распределение  $F_{t_1, \dots, t_n}(x_1, \dots, x_n)$  — это распределение случайного вектора  $(X_{t_1}, \dots, X_{t_n})$ .

Как мы видим, по любому случайному процессу можно построить семейство конечномерных функций распределения. Теперь вопрос: а можно ли по семейству построить случайный процесс? Оказывается, что можно.

**Теорема 3** (Колмогорова о существовании случайного процесса). Пусть  $F$  — заданное семейство конечномерных функций распределения над множеством индексов  $T$ , которое удовлетворяет свойству согласованности: для  $k < n$

$$F_{t_1, \dots, t_k}(x_1, \dots, x_k) = F_{t_1, \dots, t_n}(x_1, \dots, x_k, +\infty, \dots, +\infty).$$

Тогда существует вероятностное пространство  $(\Omega, \mathcal{F}, \mathbf{P})$  и случайный процесс  $X = (X_t)_{t \in T}$  такой, что  $\mathbf{P}(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n) = F_{t_1, \dots, t_n}(x_1, \dots, x_n)$ .

Вообще, доказательство этой теоремы не особо сложное, но оно требует введения не самых простых вещей наподобие борелевских сигма-аглейб над пространством функций. Поэтому я лишь напишу идею.

Теперь вспомним, что независимость бесконечного набора — это независимость в совокупности любого конечного поднабора. Тогда из теоремы Колмогорова сразу же получаем следующее

**Следствие.** Пусть для каждого  $t \in T \subseteq \mathbb{R}$  определена одномерная функция распределения  $F_t(x)$ . Тогда существует вероятностное пространство  $(\Omega, \mathcal{F}, \mathbf{P})$  и случайный процесс  $X = (X_t)_{t \in T}$  с независимыми сечениями такой, что  $\mathbf{P}(X_t \leq x) = F_t(x)$ .

Теперь рассмотрим пример применения этой теоремы. Пусть  $T = [0, 1]$ , а

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = \prod_{k=1}^n \int_{-\infty}^{x_k} g(y) dy,$$

где  $g$  — симметричная относительно нуля плотность. Возьмём случайный процесс  $X = (X_t)_{t \in T}$  из теоремы Колмогорова. Несложно понять, что в таком случае

$$\mathbf{P}(X_t > \varepsilon, X_s < -\varepsilon) = \left( \int_{\varepsilon}^{\infty} g(y) dy \right)^2$$

Теперь возьмём последовательность событий  $A_n = \{X_t > \varepsilon, X_{t+1/n} < -\varepsilon\}$ . Если процесс является непрерывным (поточечно), то  $A_n \rightarrow \emptyset$  при  $n \rightarrow \infty$ . Но тогда нарушается непрерывность вероятностной меры в нуле, то есть  $\mathbf{P}(A_n) \not\rightarrow 0$ . Следовательно, непрерывность, независимость и одинаковая распределённость не совместимы.

Для описания случайных величин вводились такие вещи, как матожидание, дисперсия и так далее. Так вот: их можно ввести и для случайных процессов, хотя это уже будут не константы, а функции (неслучайные).

**Определение 22.** Математическое ожидание случайного процесса  $X = (X_t)_{t \in T}$  — это функция  $m : T \mapsto \mathbb{R}$ , устроенная следующим образом:  $m(t) = \mathbf{E}[X_t]$ .

**Определение 23.** Дисперсия случайного процесса  $X = (X_t)_{t \in T}$  — это функция  $D : T \mapsto \mathbb{R}$ , устроенная следующим образом:  $D(t) = \mathbf{D}[X_t] = \mathbf{E}[(X_t - \mathbf{E}[X_t])^2]$ .

**Определение 24.** Среднеквадратичное отклонение случайного процесса  $X = (X_t)_{t \in T}$  — это функция  $\sigma : T \mapsto \mathbb{R}$ , устроенная следующим образом:  $\sigma(t) = \sqrt{\mathbf{D}[X_t]}$ .

**Определение 25.** Ковариационная функция случайного процесса  $X = (X_t)_{t \in T}$  — это функция  $R : T^2 \mapsto \mathbb{R}$ , устроенная следующим образом:

$$R(t_1, t_2) = \text{cov}(X_{t_1}, X_{t_2}) = \mathbf{E}[(X_{t_1} - \mathbf{E}[X_{t_1}])(X_{t_2} - \mathbf{E}[X_{t_2}])]$$

Наряду с ковариационной функцией вводят корреляционную функцию

$$r(t_1, t_2) = \frac{R(t_1, t_2)}{\sigma(t_1)\sigma(t_2)}$$

Про неё нужно сказать, что по неравенству Коши-Буняковского-Шварца  $|r(t_1, t_2)| \leq 1$ .

Осталось ввести ещё одну функцию  $K : T^2 \mapsto \mathbb{R}$ , определяемую следующим образом:  $K(t_1, t_2) = \mathbf{E}[X_{t_1} X_{t_2}]$ .

Теперь выпишем несколько очевидных свойств:

- $D(t) = R(t, t) \geq 0$ , так как дисперсия неотрицательна.
- $K(t_1, t_2) = \mathbb{E}[(X_{t_1} - m(t_1) + m(t_1))(X_{t_2} - m(t_2) + m(t_2))] = R(t_1, t_2) + m(t_1)m(t_2)$  (проверьте!).
- $R(t_1, t_2) = R(t_2, t_1)$ .
- Для любого  $n \in \mathbb{N}$  и наборов  $(z_1, \dots, z_n)$ ,  $(t_1, \dots, t_n)$  матрица  $(R(t_i, t_j))_{n \times n}$  неотрицательно определена:

$$\sum_{i,j=1}^n R(t_i, t_j) z_i z_j \geq 0$$

Это следует из того, что  $(X_{t_1}, \dots, X_{t_n})$  — случайный вектор, а  $(R(t_i, t_j))_{n \times n}$  есть его матрица ковариаций.

Допустим, что у нас есть какой-то случайный процесс. Что произойдёт с его параметрами при линейном преобразовании?

Пусть  $X = (X_t)_{t \in T}$  — случайный процесс, а  $a, b : T \mapsto \mathbb{R}$  — ограниченные неслучайные функции. Построим новый случайный процесс  $Y = (Y_t)_{t \in T}$  следующим образом:  $Y_t = a(t)X_t + b(t)$ . Посмотрим на матожидание, дисперсию и ковариационную функцию данного процесса. Для того, чтобы различать их для разных процессов, будем указывать название процесса в качестве индекса: например,  $R_Y(t_1, t_2)$ .

$$\begin{aligned} m_Y(t) &= \mathbb{E}[Y_t] = \mathbb{E}[a(t)X_t + b(t)] = a(t)m_X(t) + b(t) \\ R_Y(t_1, t_2) &= \text{cov}(Y_{t_1}, Y_{t_2}) = \text{cov}(a(t_1)X_{t_1} + b(t_1), a(t_2)X_{t_2} + b(t_2)) = a(t_1)a(t_2)R_X(t_1, t_2) \\ D_Y(t) &= R_Y(t, t) = a^2(t)R_X(t, t) = a^2(t)D_X(t) \end{aligned}$$

Теперь сделаем небольшое обобщение и возьмём линейную комбинацию  $n$  процессов. Пусть  $X_i = (X_t^i)_{t \in T}$ ,  $i = 1, \dots, n$ , а  $a_1, \dots, a_n$  — набор ограниченных функций из  $T$  в  $\mathbb{R}$ . Далее, строим новый случайный процесс  $Y = (Y_t)_{t \in T}$  по следующему правилу:

$$Y_t = \sum_{i=1}^n a_i(t)X_t^i + b(t)$$

Для упрощения жизни введём *взаимную корреляционную функцию*  $R_{X_i, X_j}(t_1, t_2) = \text{cov}(X_{t_1}^i, X_{t_2}^j)$ . Тогда несложно показать, что

$$\begin{aligned} m_Y(t) &= \sum_{i=1}^n a_i(t)m_{X_i}(t) + b(t) \\ R_Y(t_1, t_2) &= \sum_{i=1}^n a_i(t_1)a_i(t_2)R_{X_i}(t_1, t_2) + \sum_{\substack{i,j=1 \\ i \neq j}}^n a_i(t_1)a_j(t_2)R_{X_i, X_j}(t_1, t_2) \\ D_Y(t) &= \sum_{i=1}^n a_i^2(t)D_{X_i}(t) + \sum_{\substack{i,j=1 \\ i \neq j}}^n a_i(t)a_j(t)R_{X_i, X_j}(t, t) \end{aligned}$$

## 1.3 Непрерывность случайных процессов

Вообще, случайные функции — тоже вполне себе функции, поэтому вполне осмысленно посмотреть на её непрерывность. Только в данном случае можно вводить разные виды непрерывности, так как последовательности сходятся не только поточечно.

**Определение 26.** Случайный процесс  $X = (X_t)_{t \in T}$  непрерывен в среднеквадратичном смысле в точке  $t \in T$ , если

$$X_{t+\varepsilon} \xrightarrow{\text{с.к.}} X_t \text{ при } \varepsilon \rightarrow 0.$$

Если это выполнено для любого  $t \in T$ , то процесс называют непрерывным в среднеквадратичном смысле.

Вот есть у нас процесс. Можем ли мы сказать, что он будет непрерывным в среднеквадратичном смысле, не вспоминая определение каждый раз? Можем.

**Теорема 4.** Случайный процесс  $X = (X_t)_{t \in T}$  непрерывен в среднеквадратичном смысле тогда и только тогда, когда непрерывны  $R(t_1, t_2)$  и  $m(t)$ .

*Доказательство.* Для начала проверим, что этого условия достаточно. Действительно, из непрерывности  $R(t_1, t_2)$  и  $m(t)$  следует непрерывность  $K(t, t)$  и

$$\mathbb{E}[(X_{t+\varepsilon} - X_t)^2] = \mathbb{E}[X_{t+\varepsilon}^2] + \mathbb{E}[X_t^2] - 2\mathbb{E}[X_t X_{t+\varepsilon}] = K(t + \varepsilon, t + \varepsilon) + K(t, t) - 2K(t, t + \varepsilon).$$

Устремляя  $\varepsilon$  к нулю, получаем, что  $\mathbb{E}[(X_{t+\varepsilon} - X_t)^2] \rightarrow 0$ , что и даёт непрерывность в среднеквадратичном смысле. Теперь покажем, что это условие необходимо. Для этого воспользуемся следующей леммой:

**Лемма.** Пусть  $\{X_n\}_{n=1}^\infty$ ,  $\{Y_n\}_{n=1}^\infty$ ,  $X$  и  $Y$  — случайные величины такие, что  $X_n \xrightarrow{\text{с.к.}} X$ ,  $Y_n \xrightarrow{\text{с.к.}} Y$ ,  $\mathbb{E}[X^2] < \infty$  и  $\mathbb{E}[Y^2] < \infty$ . Тогда

$$\lim_{m,n \rightarrow \infty} \mathbb{E}[X_n Y_m] = \mathbb{E}[XY].$$

*Доказательство.* Для начала покажем, что  $(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2] \mathbb{E}[Y^2]$ . Для этого рассмотрим  $f(t) = \mathbb{E}[(tX + Y)^2]$ . Понятно, что  $f(t)$  есть квадратный трёхчлен от  $t$  и он неотрицателен. Тогда его дискриминант отрицателен и мы получаем желаемое. По сути, это просто неравенство Коши-Буняковского-Шварца.

Заметим, что по нему

$$|\mathbb{E}[(X_n - X)(Y_m - Y)]| \leq \sqrt{\mathbb{E}[(X_n - X)^2] \mathbb{E}[(Y_m - Y)^2]} \rightarrow 0.$$

Теперь раскроем скобки:

$$\mathbb{E}[(X_n - X)(Y_m - Y)] = \mathbb{E}[X_n Y_m] - \mathbb{E}[X_n Y] - \mathbb{E}[X Y_m] + \mathbb{E}[XY]$$

Покажем, что  $\mathbb{E}[X_n Y] \rightarrow \mathbb{E}[XY]$ . Действительно,

$$\mathbb{E}[(X_n - X)Y] \leq \sqrt{\mathbb{E}[Y^2] \mathbb{E}[(X_n - X)^2]} \rightarrow 0.$$

Аналогично,  $\mathbb{E}[X Y_m] \rightarrow \mathbb{E}[XY]$ . Тем самым мы получаем желаемое.  $\square$

Из неё сразу же следует, что при  $\varepsilon_1, \varepsilon_2 \rightarrow 0$

$$\begin{cases} X_{t_1+\varepsilon_1} \xrightarrow{\text{с.к.}} X_{t_1} \\ X_{t_2+\varepsilon_2} \xrightarrow{\text{с.к.}} X_{t_2} \end{cases} \implies \mathbb{E}[X_{t_1+\varepsilon_1} X_{t_2+\varepsilon_2}] \rightarrow \mathbb{E}[X_{t_1} X_{t_2}] \implies K(t_1 + \varepsilon_1, t_2 + \varepsilon_2) \rightarrow K(t_1, t_2)$$

Это означает непрерывность  $K(t_1, t_2)$ . Теперь вспомним, что

$$\mathbb{D}[|X_{t_1} - X_{t_2}|] = \mathbb{E}[(X_{t_1} - X_{t_2})^2] - (\mathbb{E}[|X_{t_1} - X_{t_2}|])^2 \geq 0$$

Отсюда сразу же следует, что при  $t_1 \rightarrow t_2$   $\mathbb{E}[|X_{t_1} - X_{t_2}|] \rightarrow 0$ . Следовательно,  $\mathbb{E}[X_{t_1}] \rightarrow \mathbb{E}[X_{t_2}]$  и  $m(t)$  непрерывна. Отсюда получаем, что и  $R(t_1, t_2)$  тоже непрерывна.  $\square$

Но не среднеквадратичным случаем единым! У него весьма жёсткие требования, так что введём ещё пару видов.

**Определение 27.** Будем говорить, что  $X = (X_t)_{t \in T}$  — случайный процесс с непрерывными траекториями, если для любого  $\omega \in \Omega$  траектория  $\varphi_\omega$  непрерывна, как функция от  $t$ .

**Пример 8.** Пусть  $T = [0, 1]$ ,  $f : [0, 1] \mapsto \mathbb{R}$  — какая-то непрерывная функция, а  $\xi$  — случайная величина. Тогда случайный процесс  $X = (X_t)_{t \in T}$ , построенный по правилу  $X_t = f(t)\xi$ , будет иметь непрерывные траектории.

**Определение 28.** Будем говорить, что процесс  $X = (X_t)_{t \in T}$  стохастически непрерывен, если выполняется сходимость по вероятности: для любого  $t \in X_s \xrightarrow{P} X_t$  при  $s \rightarrow t$ .

**Пример 9.** Пусть для любого натурального  $n$   $T_n$  — это iid случайные величины с распределением  $\text{Exp}(\lambda)$ , то есть их плотность равна  $p(z) = \lambda e^{-\lambda z} \mathbf{I}\{z \geq 0\}$ . Далее,  $S_n = T_1 + \dots + T_n$ , а для любого  $t \geq 0$

$$N_t = \sum_{n=1}^{\infty} \mathbf{I}\{S_n \leq t\} = \#\{n \in \mathbb{N} : S_n < t\}.$$

По сути, это процесс восстановления для экспоненциального распределения. Далее мы докажем, что для него выполнено следующее свойство: для всех  $0 \leq s < t < +\infty$   $N_t - N_s \sim \text{Pois}(\lambda(t-s))$ . Но из него сразу же получается стохастическая непрерывность:

$$\mathbf{P}(|N_t - N_s| > \varepsilon) \leq \mathbf{P}(|N_t - N_s| > 0) = \sum_{k=1}^{\infty} \frac{(\lambda(t-s))^k}{k!} e^{-\lambda(t-s)} = 1 - e^{-\lambda(t-s)} \xrightarrow{s \rightarrow t} 0.$$

Вообще говоря, между непрерывностями случайных процессов и видами сходимостей есть прямая связь. Например, если процесс непрерывен в среднеквадратичном смысле, то он будет стохастически непрерывен.

Ранее мы показали, что непрерывность исключает независимость. Можем ли мы сказать, что стохастическая непрерывность тоже исключает независимость? Можем. Пусть  $X = (X_t)_{t \in T}$  — случайный процесс такой, что для какой-то окрестности  $t_0$   $X_t$  независимо с  $X_{t_0}$  и  $X_t$  имеет плотность  $g(x)$ . Посмотрим на вероятность отклониться на  $\varepsilon$ :

$$\mathbf{P}(|X_t - X_{t_0}| > \varepsilon) = \iint_{|x-y| > \varepsilon} g(x)g(y) dx dy.$$

Но этот интеграл не стремится к нулю при  $t_0 \rightarrow t$ . Тогда стохастической непрерывности нет. Впрочем, как и непрерывных траекторий.

## 1.4 Лирическое отступление: гауссовские векторы

Сделаю небольшое лирическое отступление и вспомним гауссовские векторы. Для этого вспомним, что такое *характеристическая функция* случайной величины и вектора.

**Определение 29.** Пусть  $\xi$  — случайная величина с плотностью  $p_\xi$ . Тогда её характеристической функцией называется функция  $\varphi_\xi : \mathbb{R} \mapsto \mathbb{C}$ , определяемая следующим образом:

$$\varphi_\xi(t) = \mathbf{E}[e^{it\xi}] = \int_{-\infty}^{+\infty} e^{itx} p_\xi(x) dx.$$

**Определение 30.** Пусть  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$  — случайный вектор с совместной плотностью  $p_{\boldsymbol{\xi}}$ . Тогда её характеристической функцией называется функция  $\varphi_{\boldsymbol{\xi}} : \mathbb{R}^n \mapsto \mathbb{C}$ , определяемая следующим образом:  $\varphi_{\boldsymbol{\xi}}(\mathbf{t}) = \mathbb{E}[e^{i\langle \boldsymbol{\xi}, \mathbf{t} \rangle}]$ , где  $\langle \cdot, \cdot \rangle$  — скалярное произведение в  $\mathbb{R}^n$ .

По сути, характеристическая функция — это преобразование Фурье функции распределения.

Далее, из курса теории вероятности известно, что если  $\xi \sim \mathcal{N}(\mu, \sigma^2)$ , то

$$\varphi_{\xi}(t) = \exp \left\{ i\mu t - \frac{1}{2}\sigma^2 t^2 \right\}.$$

Поэтому гауссовский вектор вводят следующим образом:

**Определение 31.** Случайный вектор  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$  подчиняется *многомерному нормальному распределению*, если его характеристическая функция равна

$$\varphi_{\boldsymbol{\xi}}(\mathbf{t}) = \exp \left\{ i\langle \boldsymbol{\mu}, \mathbf{t} \rangle - \frac{1}{2}\langle \boldsymbol{\Sigma} \mathbf{t}, \mathbf{t} \rangle \right\},$$

где  $\boldsymbol{\mu} \in \mathbb{R}^n$  — некоторый фиксированный вектор, а  $\boldsymbol{\Sigma}$  — некоторая симметрическая и неотрицательно определённая матрица. В таком случае пишут, что  $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

Это определение не очень удобно. Докажем одну теорему, которая даст несколько более удобное определение.

**Теорема 5.** Пусть  $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$  — случайный вектор. Он будет гауссовским тогда и только тогда, когда для любого неслучайного вектора  $\boldsymbol{\lambda} \in \mathbb{R}^n$   $\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle$  имеет нормальное распределение.

*Доказательство.* Пусть  $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Тогда посмотрим на характеристическую функцию  $\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle$ . Заметим, что она равна

$$\varphi_{\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle}(t) = \mathbb{E}[e^{it\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle}] = \mathbb{E}[e^{i\langle t\boldsymbol{\lambda}, \boldsymbol{\xi} \rangle}] = \varphi_{\boldsymbol{\xi}}(t\boldsymbol{\lambda}) = \exp \left\{ it\langle \boldsymbol{\mu}, \boldsymbol{\lambda} \rangle - \frac{t^2}{2}\langle \boldsymbol{\Sigma} \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle \right\}.$$

Это означает, что  $\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle \sim \mathcal{N}(\langle \boldsymbol{\mu}, \boldsymbol{\lambda} \rangle, \langle \boldsymbol{\Sigma} \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle)$ .

Теперь предположим, что  $\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle$  имеет нормальное распределение для любого  $\boldsymbol{\lambda}$ . Тогда посмотрим на характеристическую функцию  $\boldsymbol{\xi}$ :

$$\varphi_{\boldsymbol{\xi}}(\boldsymbol{\lambda}) = \mathbb{E}[e^{i\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle}] = \varphi_{\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle}(1) = \exp \left\{ i\mathbb{E}[\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle] - \frac{1}{2}\mathbb{D}[\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle] \right\}.$$

Теперь заметим, что

$$\begin{aligned} \mathbb{E}[\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle] &= \mathbb{E} \left[ \sum_{k=1}^n \lambda_k \xi_k \right] = \sum_{k=1}^n \lambda_k \mathbb{E}[\xi_k] = \langle \boldsymbol{\lambda}, \mathbb{E}[\boldsymbol{\xi}] \rangle, \\ \mathbb{D}[\langle \boldsymbol{\lambda}, \boldsymbol{\xi} \rangle] &= \text{cov} \left( \sum_{k=1}^n \lambda_k \xi_k, \sum_{k=1}^n \lambda_k \xi_k \right) = \sum_{i,j=1}^n \text{cov}(\xi_i, \xi_j) \lambda_i \lambda_j = \langle \mathbb{D}[\boldsymbol{\xi}] \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle. \end{aligned}$$

Отсюда получаем, что  $\boldsymbol{\xi} \sim \mathcal{N}(\mathbb{E}[\boldsymbol{\xi}], \mathbb{D}[\boldsymbol{\xi}])$ . □

Теперь выпишем следствия из этой теоремы.

**Следствие** (Смысл параметров). Если случайный вектор  $\boldsymbol{\xi} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , то  $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{\xi}]$ ,  $\boldsymbol{\Sigma} = \mathbb{D}[\boldsymbol{\xi}]$ .

**Следствие** (Линейные преобразования). Любое линейное преобразование гауссовского вектора также является гауссовским вектором.

Теперь вспомним про плотности. У гауссовских векторов она есть не всегда.

**Теорема 6** (о плотности гауссовских векторов). Пусть  $\xi \sim \mathcal{N}(\mu, \Sigma)$  —  $n$ -мерный гауссовский вектор. Тогда, если  $\Sigma$  положительно определена, то существует плотность  $p_\xi(\mathbf{t})$  и она равна

$$p_\xi(\mathbf{t}) = \frac{1}{(2\pi)^{n/2} \sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} \langle \Sigma^{-1}(\mathbf{t} - \mu), (\mathbf{t} - \mu) \rangle \right\}.$$

**Задача 1.** Пусть  $\mathbf{X} = (\xi, \eta)$  — гауссовский вектор, для которого:<sup>6</sup>

$$\mathbb{E}[\mathbf{X}] = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \mathbb{D}[\mathbf{X}] = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad |\rho| < 1.$$

Докажите, что плотность случайного вектора  $\mathbf{X}$  равна

$$p_{\mathbf{X}}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left( \left( \frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left( \frac{x_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho \frac{x_1 - \mu_1}{\sigma_1} \frac{x_2 - \mu_2}{\sigma_2} \right) \right\}$$

Порой хочется сказать, что любой вектор, состоящий из нормальных случайных величин, является гауссовским. Но это неверно.

**Пример 10.** Пусть  $\xi_1$  и  $\xi_2$  — это независимые стандартные нормальные случайные величины. Построим случайный вектор  $(X_1, X_2)$  следующим образом:

$$(X_1, X_2) = \begin{cases} (\xi_1, |\xi_2|), & \xi_1 \geq 0 \\ (\xi_1, -|\xi_2|), & \xi_1 < 0 \end{cases}$$

Данный случайный вектор не будет гауссовским (почему?).

У гауссовских векторов есть одно уникальное свойство, связанное с некоррелированностью.

**Теорема 7.** Пусть  $\xi$  — гауссовский вектор. Тогда его компоненты независимы тогда и только тогда, когда они некоррелированы.

Вопрос: допустим, что у нас есть вектор, состоящий из некоррелированных нормальных случайных величин. Можно ли сказать, что он гауссовский? Оказывается, что нет.

**Пример 11.** Пусть  $X \sim \mathcal{N}(0, 1)$  и  $c \geq 0$ . Построим по ней новую случайную величину  $Y$  следующим образом:

$$Y = \begin{cases} X, & |X| \leq c \\ -X, & |X| > c \end{cases}$$

Оказывается, что  $Y \sim \mathcal{N}(0, 1)$ . Тогда  $\text{cov}(X, Y) = \mathbb{E}[XY]$ , что, в свою очередь, равно  $\mathbb{E}[X^2 \mathbb{I}\{|X| \leq c\}] - \mathbb{E}[X^2 \mathbb{I}\{|X| > c\}]$ . Что мы можем сказать про ковариацию?

<sup>6</sup>Несложно понять, что  $\rho$  есть коэффициент корреляции между  $\xi$  и  $\eta$ .



- Она является непрерывной функцией от  $c$ .
- Если  $c = 0$ , то  $\text{cov}(X, Y) = -1$ , а если  $c = +\infty$ , то  $\text{cov}(X, Y) = 1$ .

В таком случае можно сказать, что есть  $c$  такая, что  $\text{cov}(X, Y) = 0$ . Зафиксируем её. Можно ли сказать, что  $(X, Y)$  — это гауссовский вектор? Увы, но нет. Если бы это было так, то  $X$  и  $Y$  были бы независимы. Но  $P(X > c, Y > c) = 0 \neq P(X > c)P(Y > c)$ .

Теперь расскажем два факта, которые могут понадобиться в дальнейшем и связаны с нормальным распределением.

**Пример 12.** Пусть  $(\xi, \eta)$  — случайный вектор с совместной плотностью

$$p_{\xi, \eta}(x, y) = C \exp\{-(1+x^2)(1+y^2)\}, \text{ где } C = \iint_{\mathbb{R}^2} \exp\{-(1+x^2)(1+y^2)\} dx dy$$

Попробуем найти условную плотность  $p_{\xi|\eta}(x | y)$ . Сразу же заметим, что процесс аналогичен для  $p_{\eta|\xi}(y | x)$ . Для этого найдём плотность  $\eta$ :

$$\begin{aligned} p_{\eta}(y) &= \int_{-\infty}^{+\infty} C e^{-(1+x^2)(1+y^2)} dx = C e^{-(1+y^2)} \int_{-\infty}^{+\infty} e^{-x^2(1+y^2)} dx = \\ &= \frac{C e^{-(1+y^2)}}{\sqrt{y^2+1}} \int_{-\infty}^{+\infty} e^{-u^2} du = \frac{C \sqrt{\pi} e^{-(1+y^2)}}{\sqrt{y^2+1}}. \end{aligned}$$

Теперь несложно посчитать условную плотность:

$$p_{\xi|\eta}(x | y) = \frac{g(x, y)}{p_{\eta}(y)} = \frac{\sqrt{y^2+1}}{\sqrt{\pi}} e^{-x^2(1+y^2)} = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{x^2}{2\sigma_y^2}}, \text{ где } \sigma_y^2 = \frac{1}{2+2y^2}.$$

Из математической статистики вам известны такие понятия, как оценки. Следующая теорема в некоторой степени связана с ними, но это станет ясней позднее, когда дело дойдёт до фильтров Калмана.

**Теорема 8** (о нормальной корреляции, одномерный случай). Пусть  $(\xi, \eta)$  — двумерный гауссовский вектор. Тогда

$$\begin{aligned} E[\eta | \xi] &= E[\eta] + \frac{\text{cov}(\xi, \eta)}{D[\xi]}(\xi - E[\xi]), \\ \Delta &= E[(\eta - E[\eta | \xi])^2] = D[\eta] - \frac{\text{cov}^2(\xi, \eta)}{D[\xi]}. \end{aligned}$$

*Доказательство.* Докажем эту теорему в лоб. Для этого посчитаем условную плотность  $p_{\eta|\xi}(y | x)$ . Формула для совместной плотности была выведена в задаче 1. Пользуясь теми же обозначениями, получим (проверьте!), что

$$p_{\eta|\xi}(y | x) = \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}} \exp \left\{ -\frac{1}{2\sigma_2^2(1-\rho^2)} \left( y - \mu_2 - \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1) \right)^2 \right\}.$$

Из этого можно сделать следующий вывод:

$$(\eta | \xi = x) \sim \mathcal{N} \left( \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1), \sigma_2^2(1-\rho^2) \right) \implies E[\eta | \xi] = \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (\xi - \mu_1).$$

Заменяя, получим первую часть утверждения. Для доказательства второго утверждения подставим полученный результат и раскроем скобки как квадрат разности:

$$\begin{aligned} \Delta = E \left[ \left( \eta - E[\eta] - \frac{\text{cov}(\xi, \eta)}{D[\xi]} (\xi - E[\xi]) \right)^2 \right] &= E[(\eta - E[\eta])^2] - \\ &- 2 \frac{\text{cov}(\xi, \eta)}{D[\xi]} E[(\eta - E[\eta])(\xi - E[\xi])] + \frac{\text{cov}^2(\xi, \eta)}{D^2[\xi]} E[(\xi - E[\xi])^2] = D[\eta] - \frac{\text{cov}^2(\xi, \eta)}{D[\xi]}. \quad \square \end{aligned}$$

У данной теоремы есть многомерное обобщение.

**Теорема 9** (о нормальной корреляции, общий случай). Пусть  $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  —  $n$ -мерный гауссовский вектор. Сделаем следующее разбиение:

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{pmatrix} \text{ с размерами } \begin{pmatrix} q \times 1 \\ (n - q) \times 1 \end{pmatrix}$$

Соответственным образом вводятся разбиения матожидания и матрицы ковариаций:

$$\begin{aligned} \boldsymbol{\mu} &= \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \text{ с размерами } \begin{pmatrix} q \times 1 \\ (n - q) \times 1 \end{pmatrix} \\ \boldsymbol{\Sigma} &= \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \text{ с размерами } \begin{pmatrix} q \times q & q \times (n - q) \\ (n - q) \times q & (n - q) \times (n - q) \end{pmatrix} \end{aligned}$$

Тогда верны следующие формулы:

$$\begin{aligned} E[\boldsymbol{\xi} \mid \boldsymbol{\eta}] &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\boldsymbol{\eta} - \boldsymbol{\mu}_2), \\ \Delta &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}. \end{aligned}$$

*Доказательство.* Для начала найдём линейную комбинацию  $\boldsymbol{\delta} = \boldsymbol{\xi} + \mathbf{A}\boldsymbol{\eta}$  такую, что она некоррелирована (а следовательно, и независима) с  $\boldsymbol{\eta}$ . Для этого распишем корреляцию:

$$\text{cov}(\boldsymbol{\delta}, \boldsymbol{\eta}) = \text{cov}(\boldsymbol{\xi}, \boldsymbol{\eta}) + \text{cov}(\mathbf{A}\boldsymbol{\eta}, \boldsymbol{\eta}) = \text{cov}(\boldsymbol{\xi}, \boldsymbol{\eta}) + \mathbf{A} D[\boldsymbol{\eta}] = \boldsymbol{\Sigma}_{12} + \mathbf{A} \boldsymbol{\Sigma}_{22} \implies \mathbf{A} = -\boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1}.$$

Отсюда получаем, что

$$\begin{aligned} E[\boldsymbol{\xi} \mid \boldsymbol{\eta}] &= E[\boldsymbol{\delta} - \mathbf{A}\boldsymbol{\eta} \mid \boldsymbol{\eta}] = E[\boldsymbol{\delta} \mid \boldsymbol{\eta}] - E[\mathbf{A}\boldsymbol{\eta} \mid \boldsymbol{\eta}] = E[\boldsymbol{\delta}] - \mathbf{A}\boldsymbol{\eta} = \boldsymbol{\mu}_1 + \mathbf{A}(\boldsymbol{\mu}_2 - \boldsymbol{\eta}) = \\ &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\boldsymbol{\eta} - \boldsymbol{\mu}_2). \end{aligned}$$

Теперь посмотрим на то, как себя ведёт  $\Delta$ . Как известно, по формуле полной вероятности:

$$\Delta = E[(\boldsymbol{\xi} - E[\boldsymbol{\xi} \mid \boldsymbol{\eta}])^2] = E[E[(\boldsymbol{\xi} - E[\boldsymbol{\xi} \mid \boldsymbol{\eta}])^2 \mid \boldsymbol{\eta}]] = E[D[\boldsymbol{\xi} \mid \boldsymbol{\eta}]].$$

Теперь посмотрим на условную дисперсию:

$$D[\boldsymbol{\xi} \mid \boldsymbol{\eta}] = D[\boldsymbol{\delta} - \mathbf{A}\boldsymbol{\eta} \mid \boldsymbol{\eta}] = D[\boldsymbol{\delta} \mid \boldsymbol{\eta}] + D[\mathbf{A}\boldsymbol{\eta} \mid \boldsymbol{\eta}] - \text{cov}(\boldsymbol{\delta}, \mathbf{A}\boldsymbol{\eta} \mid \boldsymbol{\eta}) - \text{cov}(\mathbf{A}\boldsymbol{\eta}, \boldsymbol{\delta} \mid \boldsymbol{\eta}).$$

Как известно, матожидание вектора, умноженного на матрицу слева/справа, равно матожиданию вектора, умноженного на эту же матрицу слева/справа, а условная дисперсия случайной величины, не зависящей от условия, равна обычной дисперсии. Тогда это равно

$$D[\boldsymbol{\delta} \mid \boldsymbol{\eta}] + \mathbf{A} D[\boldsymbol{\eta} \mid \boldsymbol{\eta}] \mathbf{A}^\top - \text{cov}(\boldsymbol{\delta}, \boldsymbol{\eta} \mid \boldsymbol{\eta}) \mathbf{A}^\top - \mathbf{A} \text{cov}(\boldsymbol{\eta}, \boldsymbol{\delta} \mid \boldsymbol{\eta}) = D[\boldsymbol{\delta} \mid \boldsymbol{\eta}] = D[\boldsymbol{\delta}].$$

Осталось посчитать эту дисперсию. Для этого распишем дисперсию суммы, пользуясь свойствами матриц  $\Sigma_{ij}$ :

$$\begin{aligned}
 D[\delta] &= D[\xi + A\eta] = D[\xi] + D[A\eta] + \text{cov}(\xi, A\eta) + \text{cov}(A\eta, \xi) = \\
 &= D[\xi] + A D[\eta] A^\top + \text{cov}(\xi, \eta) A^\top + A \text{cov}(\eta, \xi) = \\
 &= \Sigma_{11} + (-\Sigma_{12} \Sigma_{22}^{-1}) \Sigma_{22} (-\Sigma_{12} \Sigma_{22}^{-1})^\top + \Sigma_{12} (-\Sigma_{12} \Sigma_{22}^{-1})^\top + (-\Sigma_{12} \Sigma_{22}^{-1}) \Sigma_{21} = \\
 &= \Sigma_{11} + \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{22} \Sigma_{22}^{-1} \Sigma_{21} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = \\
 &= \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}
 \end{aligned}$$

Так как матожидание константы есть сама константа, то  $\Delta = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ .  $\square$

## 1.5 Примеры случайных процессов

Наше небольшое введение закончилось. Теперь можно посмотреть на несколько основных примеров случайных процессов.

### 1.5.1 Гауссовский и винеровский процессы

Многие процессы, которые попадают на практике, обладают так называемыми независимыми приращениями. Что это значит?

**Определение 32.** Пусть  $X = (X_t)_{t \geq 0}$  — некоторый случайный процесс. Будем говорить, что  $X$  есть *процесс с независимыми приращениями*, если для любых  $t_0, t_1, \dots, t_n \in T$  таких, что  $0 = t_0 < t_1 < \dots < t_n$  случайные величины  $X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$  независимы в совокупности.

В 1827 году Роберт Броун открыл движение пылевых зёрен в жидкости. Исследуя пыльцу под микроскопом, он установил, что в растительном соке плавающие пылевые зёрна двигаются совершенно хаотически зигзагообразно во все стороны. В дальнейшем это хаотическое движение называли *броуновским*. Для его математического описания используется так называемый *винеровский процесс*. Как он вводится?

**Определение 33.** Случайный процесс  $B = (B_t)_{t \geq 0}$  называется винеровским, если для него выполнены следующие условия:

1.  $B_0 = 0$  почти наверное.
2.  $B$  — процесс с независимыми приращениями.
3.  $B_t - B_s \sim \mathcal{N}(0, t - s)$  для любых  $0 \leq s < t < +\infty$ .
4.  $B$  имеет непрерывные почти наверное траектории, то есть с вероятностью 1  $B_t$  непрерывна, как функция от  $t$ .

Обычно наряду с винеровскими процессами вводят *гауссовские процессы*.

**Определение 34.** Случайный процесс  $X = (X_t)_{t \in T}$  называется гауссовским, если все его конечномерные функции распределения являются гауссовскими, то есть задают гауссовский вектор.

**Свойство 1.** Гауссовский процесс однозначно определяется своим математическим ожиданием и ковариационной функцией.

**Пример 13.** Оказывается, что винеровский процесс является гауссовским. Действительно, возьмём произвольные  $t_0, t_1, \dots, t_n$  таким образом, что  $0 = t_0 < t_1 < \dots < t_n$ . Как известно, у винеровского процесса независимые приращения. Следовательно,  $B_{t_0}, B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}}$  независимы в совокупности и образуют гауссовский вектор. Теперь поймём, какие распределения имеют  $B_{t_0}, B_{t_1}, \dots, B_{t_n}$ . Для этого заметим, что

$$\begin{pmatrix} B_{t_0} \\ B_{t_1} \\ B_{t_2} \\ \vdots \\ B_{t_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} B_{t_0} \\ B_{t_1} - B_{t_0} \\ B_{t_2} - B_{t_1} \\ \vdots \\ B_{t_n} - B_{t_{n-1}} \end{pmatrix}$$

Следовательно,  $B_{t_i} \sim \mathcal{N}(0, t_i)$  и они образуют гауссовский вектор.

Из этого сразу же получаем, что  $\mathbb{E}[B_t] = 0$ . Теперь покажем, что  $R_B(t_1, t_2) = \min(t_1, t_2)$ . Без ограничения общности будем считать, что  $t_1 < t_2$ . Тогда

$$R_B(t_1, t_2) = \mathbb{E}[B_{t_1} B_{t_2}] = \mathbb{E}[B_{t_1} ((B_{t_2} - B_{t_1}) + B_{t_1})] = \mathbb{E}[B_{t_1} (B_{t_2} - B_{t_1})] + \mathbb{E}[B_{t_1}^2].$$

Так как  $B_{t_1}$  и  $B_{t_2} - B_{t_1}$  независимы, то  $\mathbb{E}[B_{t_1} (B_{t_2} - B_{t_1})] = \mathbb{E}[B_{t_1}] \mathbb{E}[B_{t_2} - B_{t_1}] = 0$ . Тем самым мы получаем, что  $R_B(t_1, t_2) = t_1$ .

### 1.5.2 Процесс Орнштейна-Уленбека

Этот процесс пошёл из теории стохастических дифференциальных уравнений. Пусть  $B = (B_t)_{t \geq 0}$  — винеровский процесс. Построим по нему новый процесс  $X = (X_t)_{t \geq 0}$  по следующему правилу:  $X_t = e^{-t} B_{e^{2t}}$ . Полученный процесс называется *процессом Орнштейна-Уленбека*. Каковы его свойства?

**Свойство 1.**  $\mathbb{E}[X_t] = 0$ ,  $R_X(t, s) = e^{-|t-s|}$ .

*Доказательство.* Первая часть очевидна:  $\mathbb{E}[X_t] = \mathbb{E}[e^{-t} B_{e^{2t}}] = 0$ . Теперь рассмотрим ковариационную функцию:

$$R_X(t, s) = \mathbb{E}[X_t X_s] = e^{-(s+t)} \mathbb{E}[B_{e^{2t}} B_{e^{2s}}] = e^{2 \min(t, s) - (s+t)} = e^{-|t-s|}. \quad \square$$

**Свойство 2.** Процесс Орнштейна-Уленбека является гауссовским.

*Доказательство.* Без ограничения общности зафиксируем числа  $t_1 < t_2 < \dots < t_n$ . Рассмотрим случайный вектор  $(X_{t_1}, \dots, X_{t_n})$ . Как он устроен? Распишем последний член:

$$\begin{aligned} X_{t_n} &= e^{-t_n} B_{e^{2t_n}} = e^{-t_n} (B_{e^{2t_n}} - B_{e^{2t_{n-1}}}) + e^{-t_n} B_{e^{2t_{n-1}}} = \dots = \\ &= e^{-t_n} \sum_{k=1}^{n-1} (B_{e^{2t_{k+1}}} - B_{e^{2t_k}}) + e^{-t_n} B_{e^{2t_1}} \end{aligned}$$

Далее, нам известно, что  $B_{e^{2t_1}}, B_{e^{2t_2}} - B_{e^{2t_1}}, \dots, B_{e^{2t_n}} - B_{e^{2t_{n-1}}}$  независимы в совокупности и имеют следующие распределения:

$$B_{e^{2t_1}} \sim \mathcal{N}(0, e^{2t_1}), \quad B_{e^{2t_k}} - B_{e^{2t_{k-1}}} \sim \mathcal{N}(0, e^{2t_k} - e^{2t_{k-1}}).$$

Осталось заметить, что

$$\begin{pmatrix} X_{t_1} \\ X_{t_2} \\ X_{t_3} \\ \vdots \\ X_{t_n} \end{pmatrix} = \begin{pmatrix} e^{-t_1} & 0 & 0 & \dots & 0 \\ e^{-t_2} & e^{-t_2} & 0 & \dots & 0 \\ e^{-t_3} & e^{-t_3} & e^{-t_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{-t_n} & e^{-t_n} & e^{-t_n} & \dots & e^{-t_n} \end{pmatrix} \begin{pmatrix} B_{e^{2t_1}} \\ B_{e^{2t_2}} - B_{e^{2t_1}} \\ B_{e^{2t_3}} - B_{e^{2t_2}} \\ \vdots \\ B_{e^{2t_n}} - B_{e^{2t_{n-1}}} \end{pmatrix}$$

Отсюда получаем, что  $(X_{t_1}, \dots, X_{t_n})$  — действительно гауссовский вектор.  $\square$

Теперь посчитаем совместную плотность такого вектора. Для этого выпишем совместную плотность  $\xi = (B_{e^{2t_1}}, B_{e^{2t_2}} - B_{e^{2t_1}}, \dots, B_{e^{2t_n}} - B_{e^{2t_{n-1}}})$ . Так как компоненты независимы, то она равна произведению плотностей каждой компоненты:

$$p_{\xi}(x_1, \dots, x_n) = \frac{1}{\sqrt{2\pi}e^{t_1}} \exp\left\{-\frac{x_1^2}{2e^{2t_1}}\right\} \prod_{k=2}^n \frac{1}{\sqrt{2\pi}(e^{2t_k} - e^{2t_{k-1}})} \exp\left\{-\frac{x_k^2}{2(e^{2t_k} - e^{2t_{k-1}})}\right\}$$

Теперь выразим компоненты вектора  $\xi$  через компоненты вектора  $\eta = (X_1, X_2, \dots, X_n)$ :

$$\begin{cases} B_{e^{2t_1}} = e^{t_1} X_1 \\ B_{e^{2t_2}} = e^{t_2} X_2 \\ \dots \\ B_{e^{2t_n}} = e^{t_n} X_n \end{cases} \Rightarrow \begin{cases} B_{e^{2t_1}} = e^{t_1} X_1 \\ B_{e^{2t_2}} - B_{e^{2t_1}} = e^{t_2} X_2 - e^{t_1} X_1 \\ \dots \\ B_{e^{2t_n}} - B_{e^{2t_{n-1}}} = e^{t_n} X_n - e^{t_{n-1}} X_{n-1} \end{cases}$$

В матричном виде это можно записать следующим образом:

$$\begin{pmatrix} B_{e^{2t_1}} \\ B_{e^{2t_2}} - B_{e^{2t_1}} \\ B_{e^{2t_3}} - B_{e^{2t_2}} \\ \vdots \\ B_{e^{2t_n}} - B_{e^{2t_{n-1}}} \end{pmatrix} = \begin{pmatrix} e^{t_1} & 0 & 0 & \dots & 0 \\ -e^{t_1} & e^{t_2} & 0 & \dots & 0 \\ 0 & -e^{t_2} & e^{t_3} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & e^{t_n} \end{pmatrix} \begin{pmatrix} X_{t_1} \\ X_{t_2} \\ X_{t_3} \\ \vdots \\ X_{t_n} \end{pmatrix}$$

Несложно понять, что матрица перехода служит матрицей Якоби и её определитель равен  $e^{t_1 + \dots + t_n}$ . Следовательно,

$$p_{\eta}(x_1, \dots, x_n) = e^{t_1 + \dots + t_n} p_{\xi}(e^{t_1} x_1, e^{t_2} x_2 - e^{t_1} x_1, \dots, e^{t_n} x_n - e^{t_{n-1}} x_{n-1}).$$

Подставляя, получаем, что

$$\begin{aligned} p_{\eta}(x_1, \dots, x_n) &= \frac{e^{t_1}}{\sqrt{2\pi}e^{t_1}} \exp\left\{-\frac{x_1^2 e^{2t_1}}{2e^{2t_1}}\right\} \prod_{k=2}^n \frac{e^{t_k}}{\sqrt{2\pi}(e^{2t_k} - e^{2t_{k-1}})} \exp\left\{-\frac{(e^{t_k} x_k - e^{t_{k-1}} x_{k-1})^2}{2(e^{2t_k} - e^{2t_{k-1}})}\right\} = \\ &= \left(2\pi \prod_{k=2}^n (1 - e^{2(t_{k-1} - t_k)})\right)^{-\frac{1}{2}} \exp\left\{-\frac{x_1^2}{2} - \frac{1}{2} \sum_{k=2}^n \frac{(x_k - x_{k-1} e^{t_{k-1} - t_k})^2}{1 - e^{2(t_{k-1} - t_k)}}\right\}. \end{aligned}$$

### 1.5.3 Пуассоновский процесс

Перейдём к одному из самых простых для исследования процессов — к *пуассоновскому потоку*.

**Определение 35.** Случайный процесс  $X = (X_t)_{t \geq 0}$  называется пуассоновским потоком с *интенсивностью*  $\lambda$ , если он удовлетворяет трём условиям:

1.  $X_0 = 0$  почти наверное.
2.  $X$  — процесс с независимыми приращениями.
3. Для всех  $0 \leq s < t < +\infty$   $X_t - X_s \sim \text{Pois}(\lambda(t-s))$ , то есть для любого  $n \in \mathbb{Z}_+$

$$\mathbb{P}(X_t - X_s = n) = \frac{(\lambda(t-s))^n}{n!} e^{-\lambda(t-s)}.$$

Теперь построим пример такого процесса. Для этого вспомним процесс восстановления, описанный в **примере 6**.

Пусть для любого натурального  $n$   $T_n$  — это iid случайные величины с распределением  $\text{Exp}(\lambda)$ , то есть их плотность равна  $p(z) = \lambda e^{-\lambda z} \mathbf{I}\{z \geq 0\}$ . Дальше,  $S_n = T_1 + \dots + T_n$ , а для любого  $t \geq 0$

$$N_t = \sum_{n=1}^{\infty} \mathbf{I}\{S_n \leq t\} = \#\{n \in \mathbb{N} : S_n < t\}.$$

Полученный случайный процесс обладает некоторыми интересными свойствами.

**Свойство 1.** Для любого  $t > 0$   $N_t \sim \text{Pois}(\lambda t)$ .

*Доказательство.* Понятно, что  $N_t$  принимает значения в  $\mathbb{Z}_+$ . Следовательно, достаточно показать, что вероятности принять нужное значение будут именно такими, какими они должны быть.

Пусть  $n = 0$ . Тогда

$$\mathbf{P}(N_t = 0) = \mathbf{P}(T_1 > t) = \int_t^{\infty} \lambda e^{-\lambda x} dx = e^{-\lambda t} \int_t^{\infty} \lambda e^{-\lambda(x-t)} dx = e^{-\lambda t} = \frac{(\lambda t)^0}{0!} e^{-\lambda t}.$$

Теперь посмотрим на вероятность события  $N_t = n$ . Она равна

$$\mathbf{P}(N_t = n) = \mathbf{P}(N_t \geq n) - \mathbf{P}(N_t \geq n+1) = \mathbf{P}(S_n \leq t) - \mathbf{P}(S_{n+1} \leq t).$$

Для того, чтобы посчитать полученные вероятности, вспомним один факт: сумма  $n$  iid случайных величин с распределением  $\text{Exp}(\lambda)$  имеет гамма-распределение  $\Gamma(1/\lambda, n)$  с плотностью

$$p_n(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)} = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!}.$$

Тогда эта вероятность равна

$$\mathbf{P}(N_t = n) = \int_0^t \left( \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} - \frac{\lambda^{n+1} x^n e^{-\lambda x}}{n!} \right) dx = \int_0^t \left( \frac{\lambda^n x^{n-1} e^{-\lambda x}}{n!} \right)' dx = \frac{(\lambda t)^n}{n!} e^{-\lambda t}. \quad \square$$

**Следствие.**  $\mathbf{E}[N_t] = \lambda t$ ,  $\mathbf{D}[N_t] = \lambda t$ .

**Свойство 2.**  $N = (N_t)_{t \geq 0}$  — это пуассоновский поток с интенсивностью  $\lambda$ .

*Доказательство.* Для начала покажем, что  $N_0 = 0$  почти наверное. Действительно, если  $N_0 \neq 0$ , то  $T_1 = 0$ , что происходит с нулевой вероятностью. Тем самым  $N_0 = 0$  почти наверное.

Теперь докажем, что  $(N_t)_{t \geq 0}$  удовлетворяет двум последним свойствам. Для этого заметим, что

$$\begin{pmatrix} S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_n \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} T_1 \\ T_2 \\ T_3 \\ \vdots \\ T_n \end{pmatrix}$$

Так как у матрицы единичный определитель и замена линейна, то

$$p_{S_1, \dots, S_n}(x_1, x_2, \dots, x_n) = p_{T_1, \dots, T_n}(x_1, x_2 - x_1, \dots, x_n - x_{n-1})$$

Пользуясь независимостью  $T_n$ , получаем, что

$$p_{S_1, \dots, S_n}(x_1, x_2, \dots, x_n) = \lambda e^{-\lambda x_1} \prod_{k=2}^n \lambda e^{-\lambda(x_k - x_{k-1})} \mathbb{I}\{x_k - x_{k-1} \geq 0\}$$

После преобразования получаем, что

$$p_{S_1, \dots, S_n}(x_1, x_2, \dots, x_n) = \lambda^n e^{-\lambda x_n} \mathbb{I}\{x_n \geq x_{n-1} \geq \dots \geq x_1 \geq 0\}$$

Дальше, зафиксируем какие-либо числа  $0 \leq t_1 < t_2 < \dots < t_n$ ,  $k_1 \leq k_2 \leq \dots \leq k_n$  и посмотрим на следующую вероятность:

$$\mathbb{P}(N_{t_1} = k_1, N_{t_2} - N_{t_1} = k_2 - k_1, \dots, N_{t_n} - N_{t_{n-1}} = k_n - k_{n-1})$$

Поймём, как связать это с  $S_n$ . Для этого поймём, как устроено первое условие. Оно означает, что  $S_1, \dots, S_{k_1} \leq t_1$ , а  $S_{k_1+1} > t_1$ . Аналогично, получаем, что эта вероятность равна

$$\mathbb{P}(S_1, \dots, S_{k_1} \in (0, t_1], S_{k_1+1}, \dots, S_{k_2} \in (t_1, t_2], \dots, S_{k_{n-1}+1}, \dots, S_{k_n} \in (t_{n-1}, t_n], S_{k_n+1} > t_n)$$

Пользуясь плотностью случайного вектора из  $S_k$ , запишем это в виде интеграла:

$$\int \dots \int \lambda^{k_n+1} e^{-\lambda x_{k_n+1}} \mathbb{I}\{x_{k_n+1} \geq x_{k_n} \geq \dots \geq x_1 \geq 0\} dx_1 \dots dx_{k_n} dx_{k_n+1}$$

$0 < x_1, \dots, x_{k_1} \leq t_1$   
 $t_1 < x_{k_1+1}, \dots, x_{k_2} \leq t_2$   
 $\dots$   
 $t_{n-1} < x_{k_{n-1}+1}, \dots, x_{k_n} \leq t_n$   
 $x_{k_n+1} > t_n$

Разобьём его в произведение интегралов, положив  $k_0 = t_0 = 0$ :

$$\lambda^{k_n} \int_{t_n}^{\infty} \lambda e^{-\lambda x_{k_n+1}} dx_{k_n+1} \prod_{j=1}^n \int_{t_{j-1}}^{t_j} dx_{k_{j-1}+1} \dots dx_{k_j}$$

Интегралы в произведении берутся достаточно просто: это объём симплекса. Рассуждая по аналогии с трёхмерным случаем, получаем, что интеграл равен

$$\lambda^{k_n} e^{-\lambda t_n} \prod_{j=1}^n \frac{(t_j - t_{j-1})^{k_j - k_{j-1}}}{(k_j - k_{j-1})!}$$

Теперь сделаем подгон: заметим, что

$$k_n = k_n - k_0 = \sum_{j=1}^n (k_j - k_{j-1}), \quad t_n = t_n - t_0 = \sum_{j=1}^n (t_j - t_{j-1})$$

Тогда интеграл равен

$$\prod_{j=1}^n \frac{(\lambda(t_j - t_{j-1}))^{k_j - k_{j-1}}}{(k_j - k_{j-1})!} e^{-\lambda(t_j - t_{j-1})}.$$

Какая красота. Отсюда мы сразу получаем оба свойства. Тем самым процесс восстановления для экспоненциального распределения является пуассоновским потоком с интенсивностью  $\lambda$ .  $\square$

Следующее свойство связано с понятием стационарных приращений.

**Определение 36.** Случайный процесс  $X = (X_t)_{t \geq 0}$  имеет *стационарные приращения*, если для любых  $0 \leq t_0 < t_1 < t_2 < \dots < t_n < +\infty$  и  $\forall h \geq 0$

$$(X_{t_1+h} - X_{t_0+h}, \dots, X_{t_n+h} - X_{t_{n-1}+h}) \stackrel{d}{=} (X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}).$$

**Свойство 3.** Пуассоновский поток имеет стационарные приращения.

*Доказательство.* Следует из того, что  $X_{a+h} - X_{b+h} \sim \text{Pois}(a-b) \sim X_a - X_b$  и приращения независимы.  $\square$

**Свойство 4.** Пусть  $N^1, \dots, N^k$  — независимые пуассоновские потоки с интенсивностями  $\lambda_1, \dots, \lambda_k$ . Тогда случайный процесс  $N_t = N_t^1 + \dots + N_t^k$  — это пуассоновский процесс с интенсивностью  $\lambda = \lambda_1 + \dots + \lambda_k$ .

## 1.6 Стационарность случайных процессов

На практике часто попадаются процессы, которые неизменны во времени. Их принято называть *стационарными*. Это понятие было введено и для случайных процессов, хоть и не в одной ипостаси.

Буквальный перевод вышесказанного на математический язык даёт *стационарность в узком смысле*.

**Определение 37.** Случайный процесс  $X = (X_t)_{t \in T}$  называется *сильно стационарным* (strong sense stationary, SSS, стационарным в узком смысле), если для любого натурального  $n$ , любых индексов  $t_1, t_2, \dots, t_n \in T$  и любого сдвига  $h \geq 0$

$$F_{t_1+h, \dots, t_n+h}(x_1, \dots, x_n) = F_{t_1, \dots, t_n}(x_1, \dots, x_n).$$

**Пример 14.** Последовательность iid случайных величин — это стационарная в узком смысле случайная последовательность.

**Пример 15.** Винеровский процесс не является стационарным в узком смысле. Действительно, пусть  $n = 1$  и  $h > 0$ . Тогда  $B_t \sim \mathcal{N}(0, t)$ , а  $B_{t+h} \sim \mathcal{N}(0, t+h)$  и  $F_{t+h}(x) \neq F_t(x)$ .

У стационарных в узком смысле процессов есть одно полезное свойство. Но у него есть требование — процесс должен быть второго порядка. Что это значит?

**Определение 38.** Случайный процесс  $(X_t)_{t \in T}$  называется *процессом второго порядка*, если  $E[X_t^2]$  конечно для всех  $t$  (то есть это ограниченная функция от  $t$ ).

**Свойство 1.** Если  $(X_t)_{t \in T}$  — стационарный в узком смысле процесс второго порядка, то

$$1. m(t) = \mu = \text{const}, D(t) = \sigma^2 = \text{const}.$$

$$2. \text{ Для любых } t_1, t_2, h \in T \quad R_X(t_1, t_2) = R_X(t_1 + h, t_2 + h).$$

Данное свойство очевидным образом следует из определения стационарного в узком смысле процесса и того, что матожидание, дисперсия и ковариация конечны. Последнее свойство позволяет свести ковариационную функцию к одному аргументу:  $R_X(t_1, t_2) = R_X(0, t_2 - t_1) \equiv R_X(t_2 - t_1)$ .

Вообще говоря, выполнение этих трёх свойств — это тоже в некоторой степени стационарность. Только в широком смысле.



**Определение 39.** Случайный процесс  $X = (X_t)_{t \in T}$  называется слабо стационарным (стационарным в широком смысле, wide sense stationary, WSS, ковариационно стационарным, стационарным второго порядка),<sup>7</sup> если выполнены следующие условия:

1.  $m(t) = \mu = \text{const}$ ,  $D(t) = \sigma^2 = \text{const}$ .
2. Для любых  $t_1, t_2, h \in T$   $R_X(t_1, t_2) = R_X(t_1 + h, t_2 + h)$ .

Обычно сильная и слабая стационарности идут вместе (если это не так, то что-то пошло не туда). Например, если семейство конечномерных функций распределения полностью задаётся матожиданием и ковариационной функцией, то сильная стационарность равносильна слабой стационарности.

У ковариационной функции стационарного в широком смысле случайного процесса есть несколько свойств:

1. Она неотрицательна в нуле:  $R_X(0) = R_X(t, t) = D(t) = \sigma^2 \geq 0$ .
2. Она чётна:  $R_X(-\tau) = R_X(0, -\tau) = R_X(-\tau, 0) = R_X(0, \tau) = R_X(\tau)$ .
3. Она ограничена по модулю дисперсией. Действительно, по неравенству Коши-Буняковского-Шварца  $|R_X(\tau)| = |R_X(t, t + \tau)| = |\text{cov}(X_t, X_{t+\tau})| \leq \sqrt{D[X_t] D[X_{t+\tau}]} = \sigma^2$ .
4. Аналог неотрицательной определённости: для любого натурального  $n$ , любого неслучайного вектора  $(z_1, \dots, z_n)$  и любого набора индексов  $t_1, \dots, t_n$

$$\sum_{i,j=1}^n R_X(t_i - t_j) z_i z_j \geq 0.$$

5. Если  $R_X(\tau)$  непрерывна в нуле, то она непрерывна для любого  $\tau$ .

*Доказательство.* Пусть  $\xi_t = X_t - \mathbb{E}[X_t]$ . Пользуясь этим обозначением, распишем разность ковариационных функций.

$$\begin{aligned} R_X(t_1 + h_1, t_2 + h_2) - R_X(t_1, t_2) &= \mathbb{E}[\xi_{t_1+h_1} \xi_{t_2+h_2} - \xi_{t_1} \xi_{t_2}] = \\ &= \mathbb{E}[\xi_{t_1+h_1} (\xi_{t_2+h_2} - \xi_{t_2})] + \mathbb{E}[(\xi_{t_1+h_1} - \xi_{t_1}) \xi_{t_2}] \end{aligned}$$

Далее, по неравенству треугольника:

$$|R_X(t_1 + h_1, t_2 + h_2) - R_X(t_1, t_2)| \leq |\mathbb{E}[\xi_{t_1+h_1} (\xi_{t_2+h_2} - \xi_{t_2})]| + |\mathbb{E}[(\xi_{t_1+h_1} - \xi_{t_1}) \xi_{t_2}]|$$

Теперь воспользуемся неравенством Коши-Буняковского-Шварца:

$$|R_X(t_1 + h_1, t_2 + h_2) - R_X(t_1, t_2)| \leq \sqrt{\mathbb{E}[\xi_{t_1+h_1}^2] \mathbb{E}[(\xi_{t_2+h_2} - \xi_{t_2})^2]} + \sqrt{\mathbb{E}[(\xi_{t_1+h_1} - \xi_{t_1})^2] \mathbb{E}[\xi_{t_2}^2]}$$

Осталось показать, что эта сумма стремится к нулю. Покажем, что первый член уходит в ноль (второй рассматривается аналогично). Действительно, по непрерывности  $R_X(t, t)$

$$\begin{aligned} \mathbb{E}[(\xi_{t_2+h_2} - \xi_{t_2})^2] &= \mathbb{E}[\xi_{t_2+h_2}^2] - 2\mathbb{E}[\xi_{t_2+h_2} \xi_{t_2}] + \mathbb{E}[\xi_{t_2}^2] = R_X(t_2 + h_2, t_2 + h_2) + \\ &+ R_X(t_2, t_2) - 2R_X(t_2 + h_2, t_2) \xrightarrow{h_2 \rightarrow 0} 0. \end{aligned}$$

Следовательно,  $R_X(t_1 + h_1, t_2 + h_2) \xrightarrow{h_1, h_2 \rightarrow 0} R_X(t_1, t_2)$  и  $R_X(t_1, t_2)$  непрерывна везде.  $\square$

<sup>7</sup>Название ковариационной стационарности появилось из-за того, что ковариационная функция зависит только от разности индексов. Стационарность второго порядка же означает постоянство второго момента.

Теперь посмотрим на три класса случайных процессов: IID, SSS и WSS. Есть ли между ними какая-либо связь? Есть. Начнём с очевидной цепочки вложений:  $\text{IID} \subseteq \text{SSS} \subseteq \text{WSS}$ . Хотя второе вложение не совсем корректно — не все сильно стационарные процессы являются процессами второго порядка. Если добавить это требование, то вложение станет корректным. Теперь покажем, что все вложения строгие.

- Начнём с  $\text{SSS} \setminus \text{IID}$ . Возьмём случайную последовательность  $X = (X_t)_{t \in T}$ , устроенную следующим образом: для всех  $t$   $X_t = \xi$ , где  $\xi$  — это какая-то фиксированная случайная величина. Она очевидно является стационарной в сильном смысле и все её сечения одинаково распределены, но она не задаёт последовательность независимых случайных величин.
- Теперь посмотрим на  $\text{WSS} \setminus \text{SSS}$ . Суть примера в том, что мы будем брать разные распределения, у которых совпадают матожидание и дисперсия. Например, пусть  $X = (X_t)_{t \geq 0}$  — случайная последовательность независимых случайных величин такая, что  $X_{2n} \sim \text{Bern}(p)$ , а  $X_{2n+1} \sim \mathcal{N}(p, p(1-p))$ . Понятно, что ни о каком равенстве распределений и речи быть не может, а вот слабая стационарность выполнена (почему?).
- Приведём ещё один пример процесса из  $\text{SSS} \setminus \text{IID}$ . Пусть  $\{X_n\}_{n=1}^\infty$  — последовательность iid случайных величин. Построим по ней новую последовательность  $\{Y_n\}_{n=1}^\infty$  по правилу  $Y_n = X_n + X_{n+1}$ . Понятно, что полученная последовательность не будет состоять из независимых случайных величин, но она будет сильно стационарной.

## 1.7 Эргодичность случайных процессов

Пусть  $X = (X_t)_{t \geq 0}$  — какой-то случайный процесс. Как оценить  $E[X_t]$ ? Достаточно просто: берём  $N$  реализаций  $X_t(\omega_1), \dots, X_t(\omega_N)$  и берём их среднее арифметическое. Получается *среднее по ансамблю траекторий*:

$$\hat{\mu}_\omega = \frac{1}{N} \sum_{k=1}^N X_t(\omega_k).$$

Теперь предположим, что у случайного процесса  $X$  постоянное матожидание  $\mu$ . Можно ли тогда заменить усреднение по ансамблю траекторий на *усреднение по времени*

$$\hat{\mu}_t = \frac{1}{n} \sum_{k=1}^n X_{t_k}(\omega)?$$

Если окажется так, что  $\hat{\mu}_t \xrightarrow{\text{с.к.}} \mu$  при  $n \rightarrow \infty$ , то процесс  $X$  называют *эргодическим в среднеквадратичном смысле*.

**Пример 16.** Возьмём последовательность iid случайных величин. Она эргодична в среднеквадратичном смысле. Действительно,

$$E \left[ \left( \frac{1}{n} \sum_{k=1}^n X_k - \mu \right)^2 \right] = D \left[ \frac{1}{n} \sum_{k=1}^n X_k \right] = \frac{D[X_1]}{n} \xrightarrow{n \rightarrow \infty} 0.$$

**Пример 17.** Возьмём случайную последовательность  $X = (X_n)_{n \in \mathbb{N}}$ , устроенную следующим образом: для всех  $n$   $X_n = \xi$ , где  $\xi$  — это какая-то фиксированная случайная величина. Она сильно стационарна, но не эргодична, так как все траектории — константы.

Проверять эргодичность по определению — это явно занятие на любителя. Поэтому сформулируем без доказательства одну теорему, которая даёт критерий эргодичности (но для сходимости в среднем, а не в среднеквадратичном):

**Теорема 10** (условие Слущкого). *Слабо стационарный случайный процесс  $X = (X_t)_{t \geq 0}$  является эргодическим в среднем тогда и только тогда, когда*

$$\lim_{T \rightarrow \infty} \frac{1}{T^2} \int_0^T \int_0^T R_X(t_1, t_2) dt_1 dt_2 = 0.$$

*Если же  $X$  является процессом второго порядка, то он будет эргодическим тогда и только тогда, когда  $R_X(t) \rightarrow 0$  при  $t \rightarrow \infty$ .*

## 1.8 Генерирование реализаций случайных процессов

Теперь посмотрим, как симулировать различные процессы. Начнём с самого простого — с пуассоновского потока.

### 1.8.1 Генерирование пуассоновских случайных процессов

#### Однородный пуассоновский поток событий

Для тех, кто забыл — определение пуассоновского потока дано в [примере 6](#). Единственная сложность в генерации реализации состоит в том, что нужно уметь генерировать случайные величины из экспоненциального распределения. Но мы можем свободно генерировать случайные величины из  $U(0, 1)$ . Как получить из него экспоненциальное распределение? Для этого докажем одно утверждение:

**Лемма** (Метод обратного преобразования). *Пусть  $X$  — случайная величина с неубывающей функцией распределения  $F : \mathbb{R} \mapsto [0, 1]$ . Введём обратную функцию  $F^{-1} : [0, 1] \mapsto \mathbb{R}$  следующим образом:  $F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}$ . Тогда, если  $U \sim U(0, 1)$ , то  $F^{-1}(U)$  имеет функцию распределения  $F$ .*

*Доказательство.* Действительно,  $P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$ . □

Теперь покажем, как генерировать случайные величины из распределения  $\text{Exp}(\lambda)$ . Рассмотрим функцию распределения:

$$F(x) = 1 - e^{-\lambda x} \implies x = -\frac{1}{\lambda} \ln(1 - F(x)) \implies F^{-1}(x) = -\frac{1}{\lambda} \ln(1 - x).$$

Тогда по методу обратного преобразования  $-\ln(1 - U)/\lambda$  будет иметь распределение  $\text{Exp}(\lambda)$ . Теперь заметим, что  $1 - U \stackrel{d}{=} U$ . Тогда получаем, что  $-\frac{1}{\lambda} \ln U$  будет иметь нужное распределение.

Теперь несложно написать алгоритм генерации реализации однородного пуассоновского потока.

---

**Алгоритм 1** Алгоритм генерации реализации однородного пуассоновского потока

---

**Вход:** Интенсивность  $\lambda$ , максимальное время  $T$ .

- 1:  $t \leftarrow 0, I \leftarrow 0, S \leftarrow \emptyset$
  - 2: сгенерировать  $U \sim U(0, 1)$
  - 3:  $t \leftarrow t - \ln(U)/\lambda$
-

---

```

4: while  $t \leq T$  do
5:    $I \leftarrow I + 1, S(I) \leftarrow t$ 
6:   сгенерировать  $U \sim U(0, 1)$ 
7:    $t \leftarrow t - \ln(U)/\lambda$ 

```

---

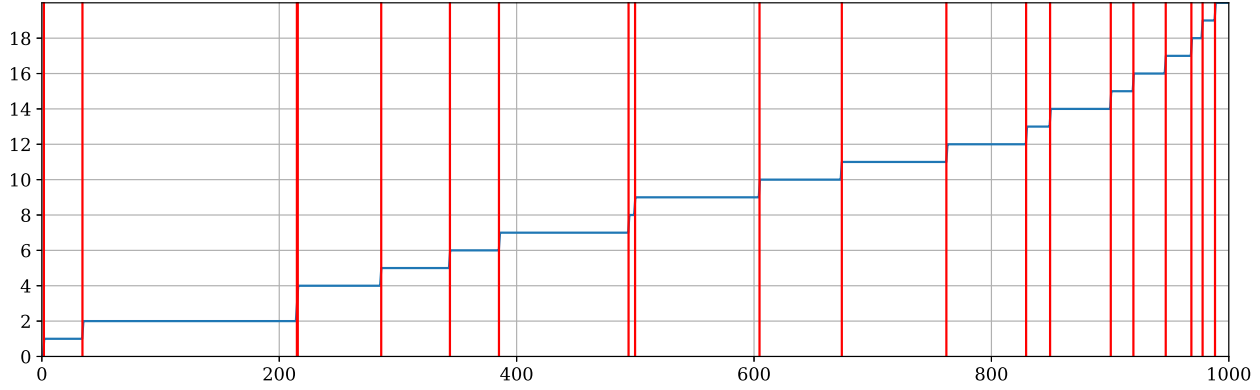


Рис. 1.1: Пример реализации пуассоновского потока с параметрами  $T = 1000$ ,  $\lambda = 0.02$

### Неоднородный пуассоновский поток событий

Ранее мы смотрели на однородный пуассоновский процесс. Однородный он по той простой причине, что его интенсивность постоянна. Теперь скажем, что  $\lambda$  — это какая-то функция от  $t$ . В таком случае получим *неоднородный пуассоновский поток*. Определяется он почти так же, как и **однородный**, только немного изменяется третье свойство:

$$X_t - X_s \sim \text{Pois} \left( \int_s^t \lambda(x) dx \right).$$

Но считать интегралы не очень приятно. Можно ли обойтись без них? Можно. Рассмотрим однородный пуассоновский поток  $N_t$  с интенсивностью  $\lambda$ . Пусть событие, появляющееся в момент времени  $t$  “засчитывается” с некоторой вероятностью  $p(t)$ , то есть

$$\mathbf{P}(N_t = N_{t-\varepsilon} + 1) = p(t), \quad \mathbf{P}(N_t = N_{t-\varepsilon}) = 1 - p(t)$$

Оказывается, что  $N_t$  — неоднородный пуассоновский поток событий с интенсивностью  $\lambda(t) = \lambda p(t)$ . Этот результат называется *теоремой Льюиса-Шедлера*.

Пусть  $\tilde{\lambda} = \max_{t \in [0, T]} \lambda(t)$ . Тогда алгоритм будет выглядеть так:

---

**Алгоритм 2** Алгоритм генерации реализации неоднородного пуассоновского потока

---

**Вход:** Интенсивность  $\lambda(t)$ , максимальное время  $T$ .

```

1:  $t \leftarrow 0, I \leftarrow 0, S \leftarrow \emptyset$ ,
2: сгенерировать  $U_1 \sim U(0, 1)$ 
3:  $t \leftarrow t - \ln(U_1)/\tilde{\lambda}$ 
4: while  $t \leq T$  do
5:   сгенерировать  $U_2 \sim U(0, 1)$ 
6:   if  $U_2 \leq \lambda(t)/\tilde{\lambda}$  then
7:      $I \leftarrow I + 1, S(I) \leftarrow t$ 
8:   сгенерировать  $U_1 \sim U(0, 1)$ 
9:    $t \leftarrow t - \ln(U_1)/\tilde{\lambda}$ 

```

---

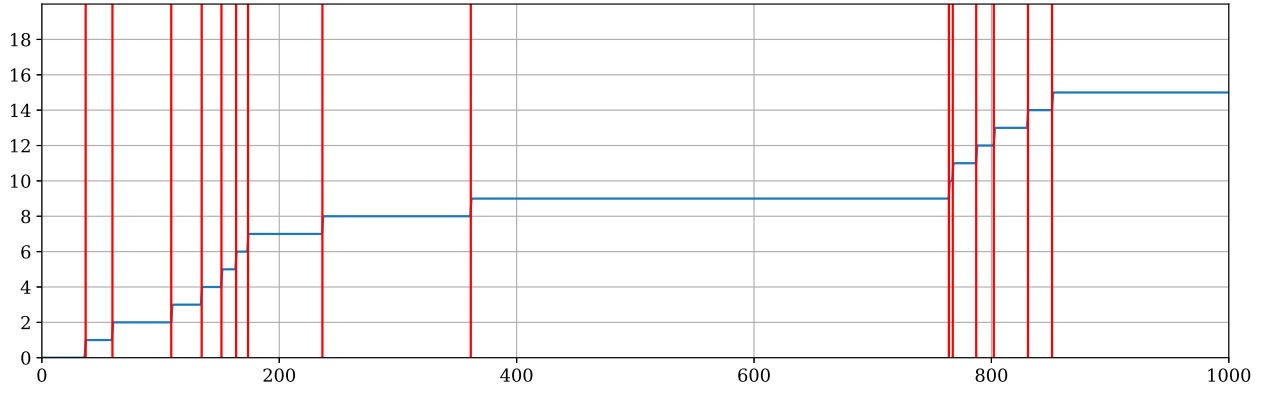


Рис. 1.2: Пример реализации неоднородного пуассоновского потока с параметрами  $T = 1000$ ,  $\lambda(t) = (\sin(t/100) + 1)/100$ .

### 1.8.2 Метод стохастического интегрирования

Как известно, дифференциальные уравнения описывают очень многое. Но, оказывается, их можно приспособить и для описания случайных процессов. Основное отличие состоит в том, что в данном случае функция, относительно которой решается уравнение, является случайной величиной. Такие дифференциальные уравнения называют *стохастическими*.

Оказывается, что многие процессы, которые изучаются на практике, “управляются” броуновским движением. Однако есть проблема: траектории винеровского процесса нигде не дифференцируемы почти наверное. Поэтому манипулирование с процессами такого типа потребовало создания собственного исчисления, называемого теорией *стохастических интегралов*. Дадим определение:

**Определение 40.** Пусть  $T = \{t_k\}_{k=0}^n$  — некоторое разбиение отрезка  $[0, t]$ . Далее, выберем точки  $\tau = \{\tau_k\}_{k=1}^n$  по правилу  $\tau_k \in [t_{k-1}, t_k]$ . Составим по этому разбиению интегральную сумму:

$$S_n = \sum_{k=1}^n b(\tau_k)(B_{t_k} - B_{t_{k-1}})$$

Стохастическим интегралом от неслучайной функции  $b(t)$  по броуновскому движению  $B = (B_t)_{t \geq 0}$  называют предел интегральных сумм при диаметре разбиения, стремящемся к нулю:<sup>8</sup>

$$\int_0^t b(x) dB_x = \lim_{\Delta T \rightarrow 0} S_n$$

*Примечание.* Не стоит забывать, что  $B_{t_i} - B_{t_{i-1}} \sim \mathcal{N}(0, t_i - t_{i-1})$ . Это поможет при симуляции процесса.

Многие стохастические процессы могут быть записаны в виде

$$X_t = \int_0^t a(x) dx + \int_0^t b(x) dB_x,$$

<sup>8</sup>Вопрос о том, почему этот предел вообще существует и каким образом последовательность частичных сумм сходится к нему, оставим за кадром.

где  $a(x)$  и  $b(x)$  — некоторые неслучайные функции. Это же выражение можно записать в дифференциалах:  $dX_t = a(t)dt + b(t)dB_t$ .

Как использовать этот метод? Примерно так же, как и в численном интегрировании: заменить дифференциал на малое изменение и суммировать.

$$X_{t+\varepsilon} - X_t \approx a(t)\varepsilon + b(t)(B_{t+\varepsilon} - B_t).$$

### 1.8.3 Метод гауссовских векторов

Если нужно сгенерировать не слишком большую реализацию гауссовского процесса, то ситуация становится несколько проще. Как известно, у них все конечномерные функции распределения являются гауссовскими, поэтому реализация будет являться гауссовским вектором. Далее, нам известны математическое ожидание и ковариационная функция процесса. Из этого можно вытащить математическое ожидание и матрицу ковариаций нужного вектора.

Но есть проблема: как генерировать случайный гауссовский вектор с заданным распределением? Сходу это сделать не получится. Для этого проведём одно рассуждение.

Допустим, что вектор одномерный, то есть это просто случайная величина  $\xi \sim \mathcal{N}(\mu, \sigma^2)$ . Как из стандартного нормального распределения получить нужное распределение? Легко:  $\xi \stackrel{d}{=} \mu + \sigma\eta$ , где  $\eta \sim \mathcal{N}(0, 1)$ .

Оказывается, что для общего случая верно нечто похожее. Пусть  $\mu$  — некоторый фиксированный вектор,  $\Sigma$  — квадратная матрица, а  $\xi \sim \mathcal{N}(\mathbf{0}, \mathbf{E})$ . Какое распределение будет у случайного вектора  $\eta = \mu + \Sigma\xi$ ? Так как преобразование линейно, то это гауссовский вектор с параметрами

$$\begin{aligned} \mathbb{E}[\eta] &= \mathbb{E}[\mu + \Sigma\xi] = \mathbb{E}[\mu] + \Sigma \mathbb{E}[\xi] = \mu \\ \mathbb{D}[\eta] &= \mathbb{E}[\Sigma\xi(\Sigma\xi)^\top] = \Sigma \mathbb{E}[\xi\xi^\top] \Sigma^\top = \Sigma\Sigma^\top. \end{aligned}$$

Теперь вспомним один факт из линейной алгебры.

**Теорема 11** (Разложение Холецкого). Пусть  $\Sigma$  — неотрицательно определённая симметричная матрица. Тогда существует нижнетреугольная матрица  $\mathbf{C}$  с неотрицательными членами на диагонали такая, что  $\Sigma = \mathbf{C}\mathbf{C}^\top$ . Если же  $\Sigma$  положительно определена, то все члены  $\mathbf{C}$  на диагонали строго положительны.

Выпишем формулы для вычисления матрицы  $\mathbf{C}$ :

$$C_{kk} = \sqrt{\Sigma_{kk} - \sum_{j=1}^{k-1} C_{ki}^2}, \quad C_{ij} = \frac{1}{C_{jj}} \left( \Sigma_{ij} - \sum_{k=1}^{j-1} C_{ik} C_{jk} \right)$$

Отсюда понятно, как генерировать гауссовский вектор с заданным распределением  $\mathcal{N}(\mu, \Sigma)$ . Для этого мы независимо генерируем  $n$  случайных величин с распределением  $\mathcal{N}(0, 1)$  (это можно сделать с помощью того же преобразования Бокса-Мюллера) и получаем гауссовский вектор  $\xi$ . Далее, берём матрицу  $\mathbf{C}$  из разложения Холецкого и строим новый вектор  $\eta = \mu + \mathbf{C}\xi$ . Полученный вектор будет иметь нужное распределение.

### 1.8.4 Генерирование гауссовских случайных процессов

#### Винеровский процесс

Сначала разберёмся, как генерировать его с помощью гауссовских векторов. Для этого достаточно сгенерировать матрицу ковариаций и вектор матожиданий. Как известно,  $m(t) = 0$ , а  $R(s, t) = \min(s, t)$ .

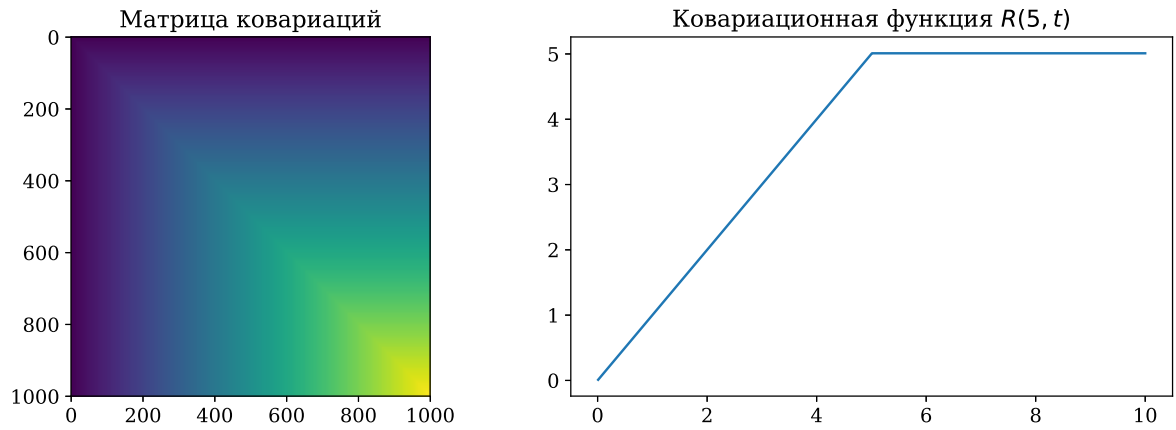


Рис. 1.3: Визуализация матрицы ковариаций и  $R(5, t)$  для  $t \in (0, 10)$ .

Интереснее генерация с помощью стохастического интегрирования. Хотя и данном случае всё достаточно очевидно:  $B_t$  можно приблизить суммой достаточно большого числа нормальных случайных величин:

$$B_t = \int_0^t dB_x \approx \sum_{k=1}^N (B_{t_k} - B_{t_{k-1}}).$$

Для примера посмотрим на первые десять секунд. Для этого разобьём отрезок  $[0, 10]$  на 1000 равных кусков и посчитаем эту сумму. Это даст приемлемую точность.

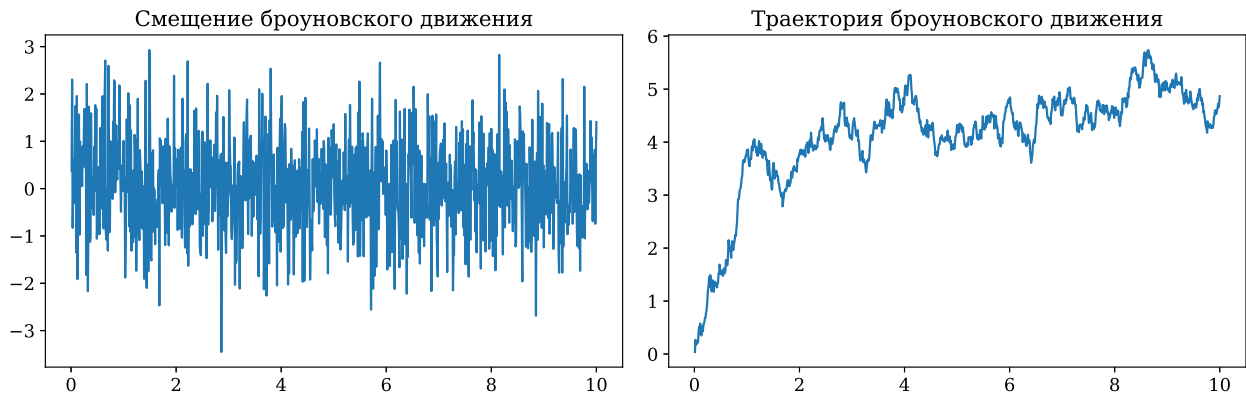


Рис. 1.4: Пример реализации первых десяти секунд броуновского движения.

## Процесс Орнштейна-Уленбека

Ранее мы обсуждали процесс Орнштейна-Уленбека. Однако на самом деле он определяется немного по-другому:

**Определение 41.** Случайный процесс  $X = (X_t)_{t \geq 0}$ , удовлетворяющий стохастическому дифференциальному уравнению

$$dX_t = \theta(\mu - X_t) dt + \sigma dB_t, \quad X_0 = x_0,$$

называется процессом Орнштейна-Уленбека.

Для того, чтобы получить ранее описанный процесс, нужно подставить  $\sigma = \sqrt{2}$ ,  $\theta = 1$ ,  $\mu = 0$  и  $x_0 = 0$ . Для генерации его реализации с помощью гауссовского вектора достаточно вспомнить, что  $E[X_t] = 0$  и  $\text{cov}(X_t, X_s) = e^{-|t-s|}$ .

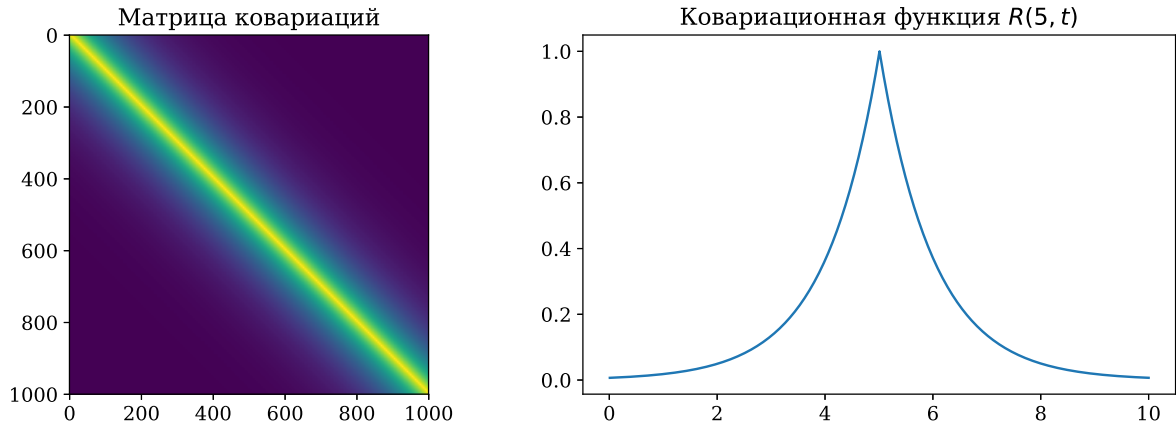


Рис. 1.5: Визуализация матрицы ковариаций и  $R(5, t)$  для  $t \in (0, 10)$ .

Разностная схема устроена следующим образом:

$$X_{t_{i+1}} - X_{t_i} = \theta(\mu - X_{t_i})(t_{i+1} - t_i) + \sigma(B_{t_{i+1}} - B_{t_i}).$$

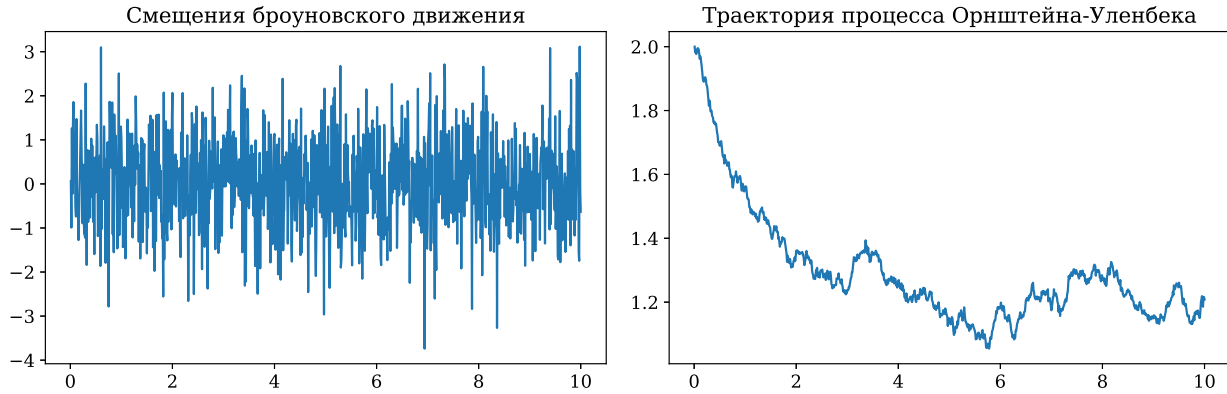


Рис. 1.6: Пример реализации первых десяти секунд процесса Орнштейна-Уленбека с параметрами  $x_0 = 2$ ,  $\sigma = 0.1$ ,  $\mu = 1.2$ ,  $\theta = 1$ .

### Фрактальное броуновское движение

Напоследок рассмотрим ещё один случайный процесс, называемый *фрактальным броуновским движением*.

**Определение 42.** Фрактальное броуновское движение с параметром Хёрста  $H \in (0, 1)$  — это гауссовский случайный процесс с непрерывным временем  $B^H = (B_t^H)_{t \in [0, T]}$ , удовлетворяющий следующим условиям:

- $B_0^H = 0$  почти наверное,
- $E[B_t^H] = 0$  для всех  $t \in [0, T]$ ,
- $\text{cov}(B_t^H, B_s^H) = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$ .



Оказывается, что если подставить  $H = 1/2$ , то получится обычное броуновское движение.<sup>9</sup> В остальных случаях получается некоторый гауссовский процесс.

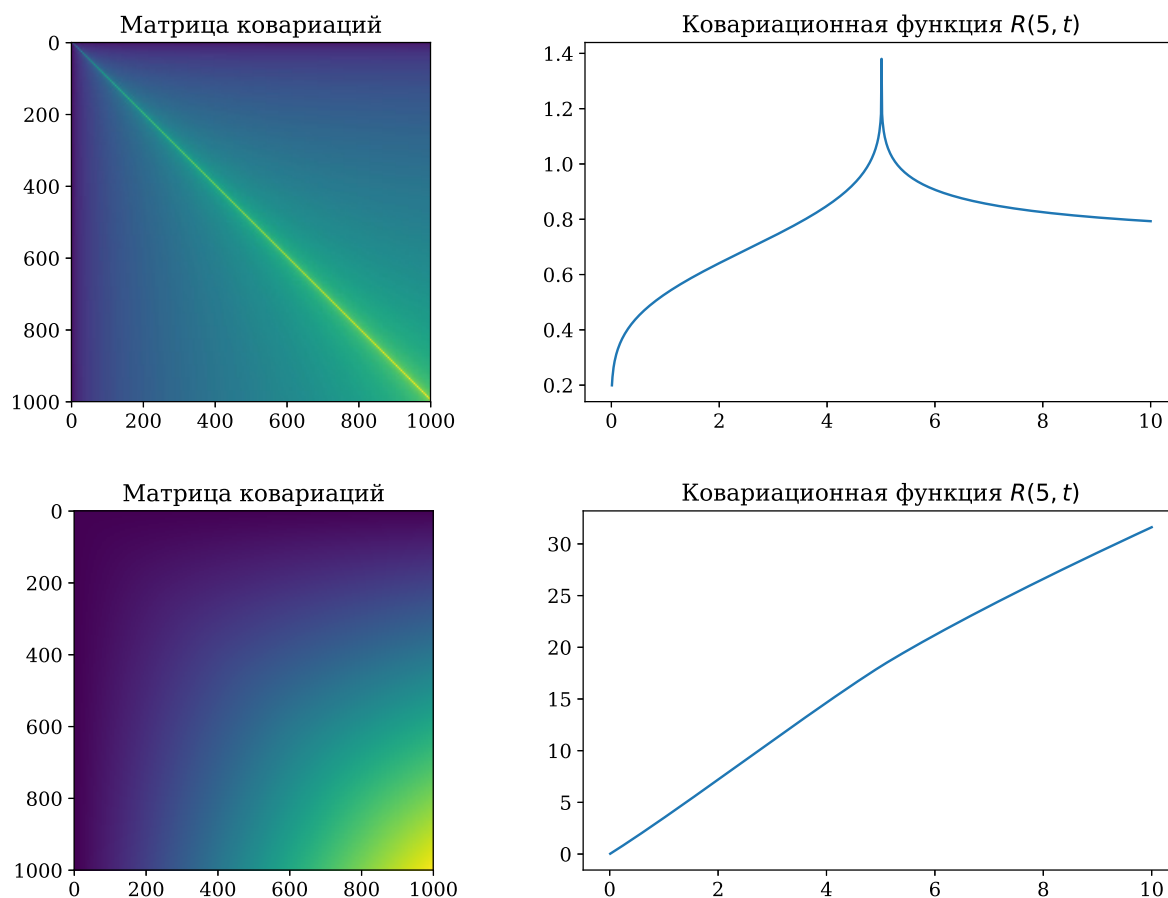


Рис. 1.7: Визуализация матрицы ковариаций и  $R(5, t)$  для  $t \in (0, 10)$  при  $H = 0.1$  и  $H = 0.9$ .

После того, как была получена матрица ковариаций, дело остаётся за малым: получить нужную реализацию с помощью разложения Холецкого. Я не буду описывать технические детали, ибо они и так очевидны. Теперь посмотрим, как себя ведут траектории в зависимости от параметра Хёрста.

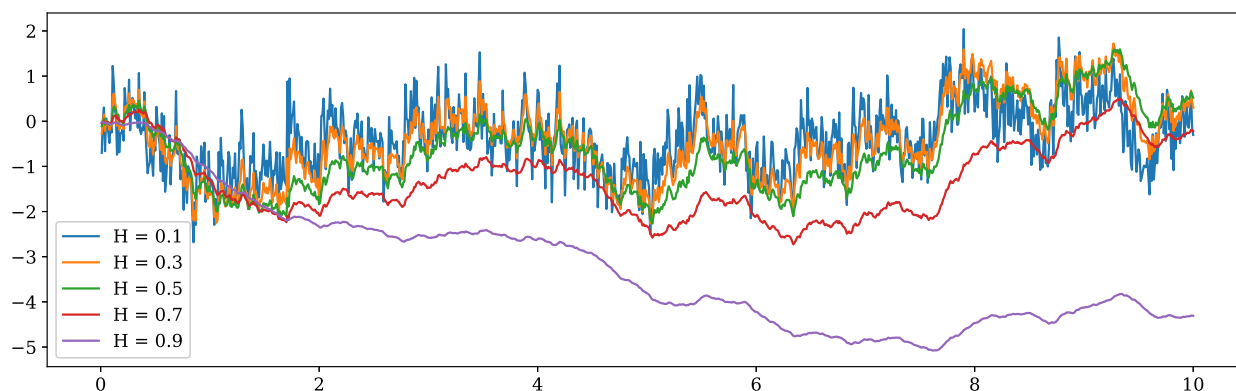


Рис. 1.8: Пример реализаций первых десяти секунд фрактального броуновского движения при разных параметрах Хёрста.

<sup>9</sup>Возникает вопрос о том, что делать с независимостью приращений. Но есть теорема, которая гласит, что фрактальное броуновское движение имеет независимые приращения только при  $H = 1/2$ .

## 1.9 Марковские цепи

### 1.9.1 Основные понятия

Наверное, многие слышали про такое понятие, как марковские цепи. Что это такое? Перед этим дадим несколько необходимых понятий.

Пусть  $E$  — это некоторое дискретное (конечное или счётное) множество, которое называют *пространством состояний*. Если система находится в состоянии  $i \in E$  в момент времени  $n$ , то в момент времени  $n + 1$  она может перейти в состояние  $j \in E$  с *переходной вероятностью*  $p_{ij}$ . Это сразу даёт два свойства переходной вероятности:

$$\forall i, j \in E \quad p_{ij} \geq 0 \text{ и } \forall i \in E \quad \sum_{j \in E} p_{ij} = 1.$$

Переходные вероятности образуют *матрицу переходных вероятностей*  $P = (p_{ij})_{i,j \in E}$ . Теперь можно дать определение марковской цепи.

**Определение 43.** *Марковская цепь* с пространством состояний  $E$  и матрицей переходных вероятностей  $P$  — это случайный процесс с дискретным временем  $X = (X_n)_{n \in \mathbb{N}}$ ,  $X_n \in E$ , для которого

- известны начальные распределения  $\alpha_i \equiv P(X_0 = i)$ ,
- верно *марковское свойство*: для любого натурального  $n$  и любых  $i_0, i_1, \dots, i_{n-1}, i, j$

$$P(X_{n+1} = j \mid X_n = i) = P(X_{n+1} = j \mid X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = p_{ij},$$

если условные вероятности хорошо определены, то есть  $P(X_0 = i_0, \dots, X_n = i) > 0$ .

Неформально говоря, марковское свойство означает, что то, как система будет развиваться в текущий момент, не зависит от того, что было в прошлом и зависит только от настоящего.

Теперь вопрос: допустим, что у нас есть какая-то траектория (последовательность состояний). Какова её вероятность? Ответ на этот вопрос даст одна простая теорема.

**Теорема 12** (о состояниях марковской цепи). *Для любого натурального  $n$  и любых  $i_0, i_1, \dots, i_{n-1}, i, j$*

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \alpha_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}.$$

*Доказательство.* Индукция по количеству состояний. Пусть  $n = 0$ . Тогда по определению марковской цепи  $P(X_0 = i_0) = \alpha_{i_0}$ .

Теперь предположим, что утверждение верно для  $n$  состояний. Покажем, что оно верно и для  $n + 1$  состояний. Действительно, по определению условной вероятности и марковскому свойству

$$\begin{aligned} P(X_0 = i_0, \dots, X_{n+1} = i_{n+1}) &= P(X_0 = i_0, \dots, X_n = i_n) P(X_{n+1} = i_{n+1} \mid X_n = i_n) = \\ &= \alpha_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n} p_{i_n i_{n+1}}. \quad \square \end{aligned}$$

**Следствие.** *Для любого натурального  $n$  и любого  $i_n \in E$*

$$P(X_n = i_n) = \sum_{i_0, \dots, i_{n-1} \in E} \alpha_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}.$$

*Доказательство.* Прямое следствие из формулы полной вероятности и теоремы о состояниях марковской цепи.  $\square$

Но обычно нас не интересует полный путь, а лишь начало и конец. Поэтому вводят вероятность перейти из состояния  $i$  в состояние  $j$  за  $n$  шагов:

$$p_{ij}^{(n)} = P(X_n = j \mid X_0 = i)$$

Чему равна эта вероятность? Воспользуемся теоремой о состояниях:

$$\begin{aligned} P(X_n = j \mid X_0 = i) &= \frac{P(X_n = j, X_0 = i)}{P(X_0 = i)} = \\ &= \sum_{i_1, \dots, i_{n-1} \in E} \frac{P(X_0 = i, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = j)}{P(X_0 = i)} = \sum_{i_1, \dots, i_{n-1} \in E} p_{ii_1} \dots p_{i_{n-1}j}. \end{aligned}$$

Если мы посмотрим на случай  $n = 2$ , то полученное выражение очень похоже на скалярное произведение строк матрицы переходной вероятности. Оказывается, что это не так уж и далеко от истины.

**Теорема 13.** Пусть  $P^{(n)} = (p_{ij}^{(n)})_{i,j \in E}$ . Тогда  $P^{(n)} = P \cdot P \cdot \dots \cdot P = P^n$ .

*Доказательство.* Индукция по количеству шагов. База ( $n = 1$ ) очевидна, так как  $p_{ij}^{(1)} \equiv p_{ij}$ . Теперь предположим, что утверждение выполнено для  $n$  шагов. Тогда  $P^{(n)} = P^n$ . Посмотрим на  $P^{(n+1)}$ :

$$\begin{aligned} p_{ij}^{(n+1)} &= P(X_{n+1} = j \mid X_0 = i) = \sum_{i_n \in E} P(X_{n+1} = j \mid X_n = i_n, X_0 = i) P(X_{n+1} = i_n \mid X_0 = i) = \\ &= \sum_{i_n \in E} P(X_{n+1} = j \mid X_n = i_n) P(X_{n+1} = i_n \mid X_0 = i) = \sum_{i_n \in E} p_{ii_n}^{(n)} p_{i_n j}. \end{aligned}$$

Отсюда получаем, что  $P^{(n+1)} = P^{(n)}P = P^{n+1}$ .  $\square$

Однако это доказательство работает не всегда. Почему же? Потому что никто не обещал, что переходная вероятность не зависит от шага. Если она действительно не зависит, то говорят, что марковская цепь *однородна*.

Теперь поговорим про состояния марковских цепей. В зависимости от переходных вероятностей поведение цепи в этом состоянии может кардинально различаться. Поэтому их классифицируют.

Первая классификация связана с важностью состояния. Может оказаться так, что из состояния можно выйти за конечное число шагов, но вернуться назад уже невозможно. Такие состояния не слишком влияют на долговременное поведение марковской цепи, поэтому их считают несущественными. Формализуем это:

**Определение 44.** Пусть  $X = (X_n)_{n \in \mathbb{N}}$  — марковская цепь с матрицей переходных вероятностей  $P$  и дискретным множеством состояний  $E$ . Будем называть состояние  $i \in E$  *несущественным*, если существует состояние  $j$  и натуральное  $n$  такое, что  $p_{ij}^{(n)} > 0$ , но  $\forall m \in \mathbb{N} p_{ji}^{(m)} = 0$ . В противном случае состояние называют *существенным*.

Вторая классификация связана с возвращением.

**Определение 45.** Состояние  $i$  марковской цепи называется *возвратным* (recurrent), если

$$P(X_n = i \text{ для бесконечно многих } n) = 1.$$

Если же эта вероятность равна нулю, то состояние называют *невозвратным* (transient).

Так как на данный момент мы рассматриваем только дискретный случай, то у нас есть роскошь: можно смотреть на марковскую цепь, как на ориентированный граф, где вершины — это события, а вес ребра — переходная вероятность. Теперь вспомним, что в теории графов вводилось такое понятие, как связность. Для марковских цепей вводится похожее понятие.

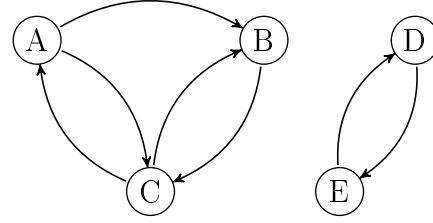


Рис. 1.9: Пример изображения марковской цепи в виде графа. В данном случае  $E = \{A, B, C, D, E\}$ . Отсутствие ребра означает, что переходная вероятность равна нулю.

**Определение 46.** Состояние  $j$  марковской цепи называются *достижимым* из состояния  $i$ , если существует такое натуральное  $n$ , что  $p_{ij}^{(n)} > 0$ . Обозначение:  $i \rightarrow j$ .

**Определение 47.** Существенные состояния  $i$  и  $j$  марковской цепи называются *сообщающимися*, если  $i \rightarrow j$  и  $j \rightarrow i$ . Обозначение:  $i \leftrightarrow j$ .

У этого понятия есть одно полезное свойство:

**Свойство 1.** *Сообщаемость задаёт отношение эквивалентности.*

*Доказательство.* Для начала нужно показать, что отношение сообщаемости рефлексивно. Пусть  $i \in E$  существенно. Это означает, что существуют такие натуральные  $m$  и  $n$ , что  $p_{ij}^{(m)} > 0$  и  $p_{ji}^{(n)} > 0$ . А это и означает, что  $i \leftrightarrow i$ .

Коммутативность очевидна. Теперь покажем, что выполнена транзитивность. Для этого достаточно показать, что если  $i \rightarrow j$  и  $j \rightarrow k$ , то и  $i \rightarrow k$ . Действительно, если  $p_{ij}^{(m)} > 0$  и  $p_{jk}^{(n)} > 0$ , то  $p_{ik}^{(m+n)} \geq p_{ij}^{(m)} p_{jk}^{(n)} > 0$ .  $\square$

В итоге мы получаем, что марковскую цепь можно разбить на классы сообщающихся вершин, коих будет не более, чем счётное число. Если такой класс один, то марковскую цепь называют *неприводимой* (или *неразложимой*).

Допустим, что мы разбили множество состояний марковской цепи на классы сообщаемости  $S_1, \dots, S_n, \dots$  и класс несущественных состояний  $S_0$ . Как будет выглядеть матрица переходных вероятностей  $P^{(n)}$ ? Оказывается, что она будет иметь блочный вид:

$$P^{(n)} = \begin{pmatrix} S_0 \rightarrow S_0 & S_0 \rightarrow S_1 & S_0 \rightarrow S_2 & \dots & S_0 \rightarrow S_n & \dots \\ 0 & S_1 \rightarrow S_1 & 0 & \dots & 0 & \dots \\ 0 & 0 & S_2 \rightarrow S_2 & \dots & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \\ 0 & 0 & 0 & \dots & S_n \rightarrow S_n & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \end{pmatrix},$$

где  $S_i \rightarrow S_j$  — блочная матрица переходных вероятностей из состояний класса  $S_i$  в состояния класса  $S_j$ .

Последняя классификация вершин связана с возвращением в состояние. Введём понятие периода состояния.

**Определение 48.** Пусть  $X = (X_n)_{n \in \mathbb{N}}$  — однородная марковская цепь с матрицей переходных вероятностей  $P$  и дискретным множеством состояний  $E$ . Число  $d_i = \gcd\{n \in \mathbb{N} \mid p_{ii}^{(n)} > 0\}$  будем называть *периодом* состояния  $j$ . Если  $d_j > 1$ , то состояние  $j$  *периодическое*, иначе же  $j$  *аперодическое*.

Осталось ввести одно понятие, которое может понадобиться в дальнейшем.

**Определение 49.** Состояние  $i$  называется *нулевым*, если  $p_{ii}^{(n)} \rightarrow 0$  при  $n \rightarrow \infty$ .

Теперь посмотрим на какой-нибудь жизненный пример, который можно описать марковской цепью — например, простейшее случайное блуждание. Строится оно так же, как в **примере 4**, только берётся распределение, принимающее значения 1, 0 и  $-1$  с какими-то вероятностями. Марковская цепь, симулирующая такой процесс, устроена просто. Множеством состояний является  $\mathbb{Z}$ , а матрица переходных вероятностей задаётся так:

$$p_{ij} = \begin{cases} 0, & |i - j| > 1 \\ p_0, & j = i \\ p_1, & j = i + 1 \\ p_2, & j = i - 1 \end{cases}$$

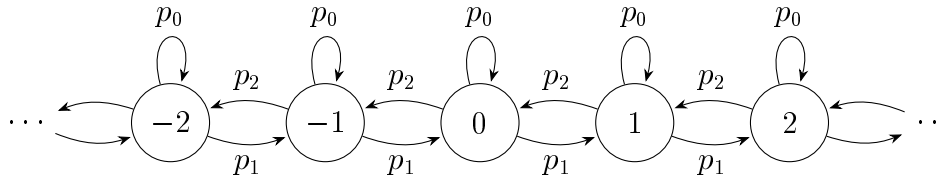


Рис. 1.11: Графическое изображение марковской цепи, соответствующей простому случайному блужданию.

**Пример 18.** Пусть блуждание начинается в нуле и  $p_0 = p_1 = 1/2$ . Отсюда сразу же получаем, что  $p_2 = 0$ . Что мы можем сказать про такое блуждание? А ничего хорошего. Все состояния несущественны, невозвратны и нулевые.

**Пример 19.** Теперь изменим вероятности:  $p_1 = p_2 = 1/2$ . Получится так называемое *симметричное случайное блуждание*. В таком случае мы можем сказать, что нулевое состояние периодически с  $d_0 = 2$ , так как есть ненулевая вероятность вернуться в начало за  $2k$  шагов (и понятно, что вернуться за нечётное число шагов невозможно). Вообще, если и  $p_1$ , и  $p_2$  больше нуля, то все состояния существенны и сообщаются.

**Пример 20.** Сузим множество состояний до  $\{-k, \dots, -1, 0, 1, \dots, k\}$  и изменим вероятности перехода на границах:  $p_{kk} = p_{-k, -k} = 1$ . Получится *случайное блуждание с поглощающими экранами*. В таком случае существенными состояниями будут только  $\pm k$ .

Вернёмся к возвратности. Как можно определить, является ли состояние возвратным? По определению это сделать весьма непросто. Но есть теорема, которая упрощает жизнь.

**Теорема 14.** Пусть  $f_i = P(\exists n \in \mathbb{N} : X_n = i)$  — вероятность того, что мы хотя бы раз попали в состояние  $i$ . Тогда состояние  $i$  будет возвратным тогда и только тогда, когда  $f_i = 1$ .

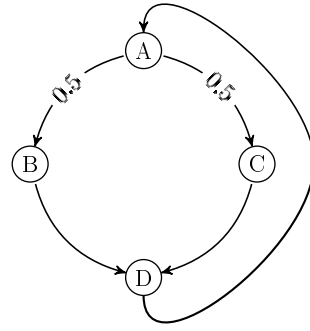


Рис. 1.10: В данной марковской цепи периодическими будут состояния B и C.

Как посчитать  $f_i$ ? Пусть  $A_i^{(n)} = \{\text{вернуться в } i \text{ ровно за } n \text{ шагов}\}$ . Вероятность этого события равна:

$$f_i(n) \equiv P\left(A_i^{(n)}\right) = P(X_n = i, X_k \neq i, k \in \{1, \dots, n-1\} \mid X_0 = i)$$

Отсюда несложно получить, что  $f_i$  равна сумме  $f_i(n)$  по всем натуральным  $n$ .

*Доказательство.* Зафиксируем состояние  $i$ . Пусть  $B_k = \{X_n = i \text{ хотя бы } k \text{ раз}\}$ . Это можно записать по-другому:

$$B_k = \{\exists n_1, \dots, n_k \in \mathbb{N} : X_{n_1} = i, \dots, X_{n_k} = i\}.$$

Что мы можем сказать про такие события? Во-первых,  $B_{k+1} \subseteq B_k$ , так как если мы посетили состояние  $i$   $k+1$  раз, то мы его посетили и  $k$  раз. Далее, марковское свойство даёт нам “отсутствие памяти”. Тогда  $P(B_k) = f_i^k$ .

Теперь заметим, что вероятность возвратности можно записать следующим образом:

$$P(X_n = i \text{ для бесконечно многих } n) = P\left(\bigcap_{n=1}^{\infty} B_n\right).$$

По непрерывности вероятностной меры

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n \rightarrow \infty} P(B_n) = \lim_{n \rightarrow \infty} f_i^n = \begin{cases} 1, & f_i = 1 \\ 0, & f_i < 1 \end{cases} \quad \square$$

**Теорема 15.** Пусть  $i$  — состояние марковской цепи. Оно будет возвратным тогда и только тогда, когда расходится ряд  $\sum p_{ii}^{(n)}$ .

*Доказательство.* Зафиксируем состояние  $i$  и рассмотрим случайную величину

$$V_i = \sum_{k=1}^{\infty} I\{X_k = i\} = \#\{n \in \mathbb{N} : X_n = i\}.$$

Что мы можем сказать про неё? Во-первых,  $P(V_i \geq k) = P(B_k) = f_i^k$ . Далее, мы можем посчитать матожидание двумя способами:

$$\begin{aligned} E[V_i] &= E\left[\sum_{k=1}^{\infty} I\{X_k = i\}\right] = \sum_{k=1}^{\infty} P(X_k = i) = \alpha_i \sum_{k=1}^{\infty} p_{ii}^{(n)}. \\ E[V_i] &= \sum_{k=1}^{\infty} k P(V_i = k) = \sum_{k=1}^{\infty} \left(\sum_{j=k}^{\infty} P(V_i = j)\right) = \sum_{k=1}^{\infty} P(V_i \geq k) = \sum_{k=1}^{\infty} f_i^k. \end{aligned}$$

Теперь заметим одну вещь: если матожидание конечно, то вероятность того, что  $V_i = +\infty$ , нулевая. Тогда мы сразу получаем желаемое: если  $i$  возвратно, то  $f_i = 1$  и ряд  $\sum p_{ii}^{(n)}$  расходится. Иначе же ряд  $\sum f_i^n$  сходится и матожидание конечно. Следовательно, ряд  $\sum p_{ii}^{(n)}$  сходится.  $\square$

Теперь применим эту теорему к простейшим случайным блужданиям.

**Теорема 16.** Пусть простейшее случайное блуждание задаётся случайными величинами, принимающими значения 1 и  $-1$  с вероятностями  $p$  и  $q \equiv 1 - p$  соответственно. Если  $p = 1/2$ , то все состояния возвратны. Иначе же все состояния невозвратны.

*Доказательство.* Без ограничения общности рассмотрим состояние  $i = 0$ . Как известно, соответствующее случайное блуждание задаётся так:

$$X_n = \xi_1 + \dots + \xi_n, \quad \{\xi_n\}_{n=1}^\infty \text{ — iid}, \quad \mathbf{P}(\xi_k = 1) = p, \quad \mathbf{P}(\xi_k = -1) = 1 - p.$$

Заметим, что  $\mathbf{E}[\xi_k] = \mathbf{P}(\xi_k = 1) - \mathbf{P}(\xi_k = -1) = 2p - 1$ . Тогда оно не ноль при  $p \neq 1/2$ . Далее, по усиленному закону больших чисел

$$\frac{X_n}{n} \xrightarrow{\text{п.н.}} \mathbf{E}[\xi_1] \implies \begin{cases} X_n \xrightarrow{\text{п.н.}} +\infty, & p > 1/2 \\ X_n \xrightarrow{\text{п.н.}} -\infty, & p < 1/2 \end{cases}$$

Из этого следует, что если  $p \neq 1/2$ , то все состояния невозвратны. Теперь покажем, что если  $p = 1/2$ , то состояние возвратно. Для этого посчитаем вероятность вернуться в ноль:

$$p_{ii}^{(n)} = \begin{cases} 0, & n \neq 2k, k \in \mathbb{Z} \\ 2^{-2k} C_{2k}^k, & n = 2k, k \in \mathbb{Z} \end{cases}$$

Теперь оценим поведение  $p_{ii}^{(n)}$ . Для этого можно вспомнить формулу Стирлинга:

$$p_{ii}^{(2k)} \sim \frac{1}{2^{2k}} \frac{\sqrt{4\pi k} (2k)^{2k} e^{-2k}}{2\pi k \cdot k^{2k} e^{-2k}} = \frac{1}{\sqrt{\pi k}} \implies \sum_{n=1}^{\infty} p_{ii}^{(n)} = +\infty.$$

Следовательно, состояние  $i = 0$  возвратно.  $\square$

Пусть у нас есть какая-то марковская цепь и  $i$  — её состояние. Можно ли как-то оценить, сколько времени цепь проводит в этом состоянии? Пусть  $T_1, T_2, \dots$  — это упорядоченные моменты попадания в состояние  $i$ . Далее, введём “времена вне состояния  $i$ ”  $T^{(1)} = T_1, T^{(2)} = T_2 - T_1, \dots$ . Если мы рассматриваем однородную марковскую цепь, то эти случайные величины независимы в совокупности и одинаково распределены.

Нередко возникает вопрос об относительном времени пребывания в состоянии  $i$ , то есть рассматривается отношение  $n/N$ , где  $n$  — количество шагов, на которых цепь находилась в состоянии  $i$ , а  $N$  — общее количество шагов.

Теперь рассмотрим матожидание времени вне состояния  $i$  для однородной цепи:

$$\mathbf{E}[T^{(n)}] = \mathbf{E}[T^{(1)}] = \sum_{k=1}^{\infty} k \mathbf{P}(T^{(1)} = k).$$

Если состояние  $i$  возвратно, то по определению  $\mathbf{P}(T^{(1)} < \infty) = 1$  и мы ничего не можем сказать про сходимость полученного ряда. Однако мы можем сказать две вещи:

- Если матожидание конечно, то  $\mathbf{P}(T^{(1)} < \infty) = 1$  и состояние  $i$  возвратно.
- Если состояние невозвратно, то ряд точно расходится. Действительно, в таком случае  $\mathbf{P}(T^{(1)} = \infty) = 1 - \mathbf{P}(T^{(1)} < \infty) > 0$ .

Далее, введём обозначение  $m_T \equiv \mathbf{E}[T^{(1)}]$ . Что мы можем сказать про предел отношения  $n/N$  при  $N \rightarrow \infty$ ?

- Пусть  $m_T < \infty$ . В таком случае состояние  $i$  возвратно и  $n$  неограниченно возрастает вместе с ростом  $T_n \equiv N$ . Далее, по усиленному закону больших чисел:

$$\frac{n}{N} = \frac{T_n}{N} = \frac{1}{n} \sum_{k=1}^n T^{(k)} \xrightarrow{\text{п.н.}} \mathbf{E}[T^{(1)}] = m_T \implies \frac{n}{N} \xrightarrow{\text{п.н.}} \frac{1}{m_T}.$$

Как известно, если  $Z_n \xrightarrow{\text{P}} c$ , где  $c$  — неслучайная величина, то  $\mathbf{E}[Z_n] \rightarrow c$ . Тогда  $\mathbf{E}[n/N] \rightarrow 1/m_T$ .

- Теперь предположим, что  $m_T = \infty$ . В таком случае мы уже не можем сказать, что состояние возвратно или не возвратно. Поэтому рассмотрим оба случая. Пусть  $i$  не возвратно. Тогда всё просто: мы посещаем состояние конечное число раз с вероятностью 1 и  $n/N \xrightarrow{P} 0$ . Оказывается, что и для возвратного состояния выполнена та же сходимост.

Теперь докажем одну теорему, связанную с классами сообщаемости. Оказывается, у состояний в одном классе весьма немало общего.

**Теорема 17.** Пусть  $X$  — это неразложимая однородная марковская цепь с множеством состояний  $E$  и матрицей переходных вероятностей  $P$ . Тогда

- Если одно из состояний цепи нулевое, то все состояния нулевые.
- Если одно из состояний цепи возвратное, то все состояния возвратные.
- Если одно из состояний цепи имеет период  $d$ , то все состояния имеют тот же период.

*Доказательство.* Пусть  $i \leftrightarrow j$ , то есть существуют натуральные  $M$  и  $N$  такие, что  $\alpha \equiv p_{ij}(M) > 0$  и  $\beta \equiv p_{ji}(N) > 0$ . Теперь возьмём произвольное натуральное  $n$  и распишем  $p_{ii}(M + n + N)$ :

$$p_{ii}(M + n + N) = \sum_{k,l \in E} p_{ik}(M)p_{kl}(n)p_{li}(N) = \sum_{\substack{k,l \in E \\ k \neq j, l \neq j}} p_{ik}(M)p_{kl}(n)p_{li}(N) + p_{ij}(M)p_{jj}(n)p_{ji}(N) \geq \alpha\beta p_{jj}(n).$$

Теперь можно приступить к доказательству:

1. Пусть  $i$  нулевое. Тогда и  $j$  нулевое:

$$0 = \lim_{n \rightarrow \infty} p_{ii}(M + n + N) = \lim_{n \rightarrow \infty} \alpha\beta p_{jj}(n) \implies \lim_{n \rightarrow \infty} p_{jj}(n) = 0.$$

2. Пусть  $i$  возвратное. По **теореме 15**:

$$\infty = \sum_{n=1}^{\infty} p_{ii}(M + n + N) = \alpha\beta \sum_{n=1}^{\infty} p_{jj}(n) \implies \sum_{n=1}^{\infty} p_{jj}(n) = \infty.$$

Отсюда получаем, что и  $j$  является возвратным состоянием.

3. Теперь предположим, что  $i$  и  $j$  имеют периоды  $d_i$  и  $d_j$  соответственно. Докажем, что  $d_i = d_j$ . Введём два множества:

$$W_i = \{n \in \mathbb{N} : p_{ii}(n) > 0\}, \quad d_i = \gcd W_i, \\ W_j = \{n \in \mathbb{N} : p_{jj}(n) > 0\}, \quad d_j = \gcd W_j.$$

Так как  $p_{ii}(M + N) \geq \alpha\beta > 0$ , то  $M + N \in W_i$ . Аналогично,  $M + N \in W_j$ . Теперь построим новое множество:

$$W = \{M + N + n \mid n \in W_j\} \subset W_i.$$

По построению  $W$  имеет общий делитель  $d_j$ . Однако для любого  $n \in W_i$   $p_{jj}(M + n + M) \geq \alpha\beta p_{ii}(n) > 0$ . Тогда  $W \subseteq W_j$  и любой элемент из  $W$  делится на  $d_j$ . Тогда можно сказать, что  $d_j \leq d_i$ . Рассуждая аналогично, получим желаемое.



Рассуждая таким образом для всех состояний, получаем желаемое.  $\square$

Эргодичность вводится и для марковских цепей, хоть и немного по-другому.

**Определение 50.** Пусть  $X = (X_n)_{n \in \mathbb{N}}$  — марковская цепь с множеством состояний  $E$  и матрицей переходных вероятностей  $\mathbf{P}$ . Будем называть её *эргодической*, если существует независимое от начального распределения ненулевое предельное распределение вероятностей и состояния  $\{p_j^*\}_{j \in E}$  такие, что

$$\forall i, j \in E \lim_{n \rightarrow \infty} p_{ij}(n) = p_j^* > 0.$$

Когда марковская цепь является эргодической и как определить предельное распределение? Для этого введём понятие стационарного распределения:

**Определение 51.** Набор состояний  $\{\bar{p}_j\}_{j \in E}$  называется *стационарным распределением вероятностей* (или *инвариантной мерой*) дискретной марковской цепи, если он не изменяется со временем.

Найти стационарное распределение  $\mathbf{p}$  не так уж и сложно. Заметим, что

$$p_j(n+1) = \sum_{i \in E} p_i(n) p_{ij} \implies \bar{p}_j = \sum_{i \in E} \bar{p}_i p_{ij} \implies \mathbf{p} = \mathbf{pP}.$$

Теперь можно сформулировать теорему (к сожалению, без доказательства):

**Теорема 18** (первая эргодическая). *Марковская цепь является эргодической тогда и только тогда, когда она неразложима и периодична. При этом её предельное распределение равно стационарному распределению, которое единственно.*

### 1.9.2 Пример применения марковских сетей: модель системы массового обслуживания

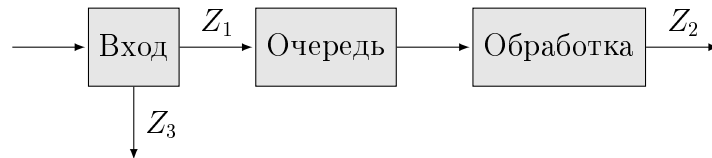


Рис. 1.12: Графическое изображение модели системы массового обслуживания.  $Z_1$  — поток объектов, принятых на обработку,  $Z_2$  — поток обработанных объектов,  $Z_3$  — поток объектов, отклонённых от обработки из-за переполненности очереди.

Рассмотрим модель некоторой организационной системы, предназначенной для массовой обработки однотипных объектов и состоящей из *устройства обработки* и *очереди*, способной хранить  $N$  объектов, ожидающих обработки.

- Будем считать, что время дискретно: что-либо происходит только в моменты времени  $t + k\tau$ , где  $k \in \mathbb{Z}$ , а  $t$  и  $\tau$  — какие-то фиксированные отрезки времени.
- В каждый момент времени с вероятностью  $p$  появляется новый вызов.
- Если в момент времени  $t$  ещё не была закончена обработка, то она заканчивается в момент времени  $t + \tau$  с вероятностью  $q$ .

- Процессы поступления и обработки независимы.
- Объект проходит три стадии: приём (в случае, если в очереди есть место), ожидание и обработка, после чего направляется на выходной поток.
- Очередь имеет объём  $N < \infty$ . Из этого следует, что система всего вмещает в себя  $N + 1$  объект и множество состояний можно представить в виде  $E = \{0, 1, \dots, N, N + 1\}$ .  $N + 1$  возникает из-за того, что на обработке может находиться 0 или 1 объект.

Теперь введём три события:

$$A_k(t) = \{\text{в момент времени } t \text{ в системе ровно } k \text{ объектов}\}$$

$$B = \{\text{в момент времени } t \text{ в системе появился новый объект}\}$$

$$C = \{\text{в момент времени } t + \tau \text{ система закончила обрабатывать объект}\}$$

Теперь, как выразить  $A_k(t + \tau)$  через них? Начнём с нуля. Заметим, что в системе может не оказаться объектов в двух случаях:

1. В системе в предыдущий момент времени не было объектов. Тогда либо в систему не подали новый объект, либо подали, но система закончила обрабатывать другой объект.
2. В системе был один объект, обработка которого закончилась. При этом нового объекта не поступило.

Формально это можно записать так:

$$A_0(t + \tau) = (A_0(t) \cap (\bar{B} \cup (B \cap C))) \cup (A_1(t) \cap \bar{B} \cap C).$$

Теперь посмотрим на крайний случай —  $A_N(t + \tau)$ . Рассуждая аналогичным образом, получаем, что

$$A_N(t + \tau) = (A_{N-1}(t) \cap B \cap \bar{C}) \cup (A_N(t) \cap ((B \cap C) \cup (\bar{B} \cap \bar{C}))) \cup (A_{N+1}(t) \cap C).$$

Отсутствие  $\bar{B}$  в последней скобке объясняется тем, что  $A_{N+1}(t)$  уже влечёт это событие. В остальных случаях же

$$A_k(t + \tau) = (A_{k-1}(t) \cap B \cap \bar{C}) \cup (A_k(t) \cap ((B \cap C) \cup (\bar{B} \cap \bar{C}))) \cup (A_{k+1}(t) \cap \bar{B} \cap C).$$

Осталось заметить, что  $A_0(t) \cup A_1(t) \cup \dots \cup A_{N+1}(t) = V$ , где  $V$  — достоверное событие.

Теперь можно записать вероятности этих событий, пользуясь независимостью:

$$\begin{cases} P_0(t + \tau) = P_0(t)(1 - p + pq) + P_1(t)(1 - p)q \\ P_k(t + \tau) = P_{k-1}(t)p(1 - q) + P_k(t)(pq + (1 - p)(1 - q)) + P_{k+1}(t)(1 - p)q \\ P_N(t + \tau) = P_{N-1}(t)p(1 - q) + P_N(t)(pq + (1 - p)(1 - q)) + P_{N+1}(t)q \\ P_0(t + \tau) + \dots + P_{N+1}(t + \tau) = P_0(t) + \dots + P_{N+1}(t) = 1. \end{cases}$$

Теперь предположим, что мы хотим найти стационарное состояние. Тогда система превращается в не очень элегантную систему линейных уравнений:

$$\begin{cases} P_0 = P_0(1 - p + pq) + P_1(1 - p)q \\ P_k = P_{k-1}p(1 - q) + P_k(pq + (1 - p)(1 - q)) + P_{k+1}(1 - p)q \\ P_N = P_{N-1}p(1 - q) + P_N(pq + (1 - p)(1 - q)) + P_{N+1}q \\ P_0 + \dots + P_{N+1} = 1. \end{cases}$$

В качестве упражнения оставим то, что решение этой системы выглядит так:

$$\forall k \in \{1, 2, \dots, N\} P_k = \left[ \frac{p(1-q)}{q(1-p)} \right]^k P_0, \quad P_{N+1} = (1-p) \left[ \frac{p(1-q)}{q(1-p)} \right]^{N+1} P_0$$

$$P_0 = \left( \sum_{k=1}^N \left[ \frac{p(1-q)}{q(1-p)} \right]^k + (1-p) \left[ \frac{p(1-q)}{q(1-p)} \right]^{N+1} \right)^{-1}.$$

Оказывается, что данная модель задаёт эргодическую марковскую цепь, поэтому стационарные вероятности являются предельными и неплохо описывают поведение системы, устаканивающееся после достаточно большого отрезка времени. Эти соотношения позволяют решать такие практические задачи, как вычисление вероятности отказа от обслуживания поступающего объекта (вероятность  $P_{N+1}(t)$ ), среднего времени ожидания обслуживания, среднего числа объектов, находящихся в бункере и так далее.

## 1.10 Марковские процессы

### 1.10.1 Непрерывное время, дискретные состояния

Постепенно будем переходить от дискретного случая к общему.