



МЕХАНИКО-  
МАТЕМАТИЧЕСКИЙ  
ФАКУЛЬТЕТ  
МГУ ИМЕНИ  
М.В. ЛОМОНОСОВА

*teach-in*  
ЛЕКЦИИ УЧЕНЫХ МГУ

# ТЕОРИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ. СЕМИНАРЫ

ШКЛЯЕВ  
АЛЕКСАНДР ВИКТОРОВИЧ

МЕХМАТ МГУ

КОНСПЕКТ ПОДГОТОВЛЕН  
СТУДЕНТАМИ, НЕ ПРОХОДИЛ  
ПРОФ. РЕДАКТУРУ И МОЖЕТ  
СОДЕРЖАТЬ ОШИБКИ.  
СЛЕДИТЕ ЗА ОБНОВЛЕНИЯМИ  
НА [VK.COM/TEACHINMSU](https://vk.com/teachinmsu).

ЕСЛИ ВЫ ОБНАРУЖИЛИ  
ОШИБКИ ИЛИ ОПЕЧАТКИ,  
ТО СООБЩИТЕ ОБ ЭТОМ,  
НАПИСАВ СООБЩЕСТВУ  
[VK.COM/TEACHINMSU](https://vk.com/teachinmsu).



БЛАГОДАРИМ ЗА ПОДГОТОВКУ КОНСПЕКТА  
СТУДЕНТА ФАКУЛЬТЕТА ФКИ МГУ  
**БАРМИНА МАКСИМА АЛЕКСАНДРОВИЧА**



## Содержание

<b>Семинар 1. Случайные последовательности</b>	<b>5</b>
Борелевские сигма-алгебры . . . . .	5
Случайные величины, векторы, последовательности . . . . .	6
Распределение случайной последовательности . . . . .	7
Как задавать конечномерные распределения? . . . . .	8
<b>Семинар 2. Ветвящиеся процессы</b>	<b>10</b>
Производящие функции . . . . .	10
Ветвящиеся процессы Гальтона–Ватсона. Задача о вырождении . . .	10
Общее представление о ветвящихся процессах . . . . .	12
Предельные теоремы для процессов Гальтона–Ватсона . . . . .	13
<b>Семинар 3. Цепи Маркова</b>	<b>14</b>
Определение и способы задания . . . . .	14
Классификация состояний цепи . . . . .	15
Эргодическая теорема . . . . .	17
<b>Семинар 4. Стационарные в узком смысле последовательности</b>	<b>19</b>
Преобразования, сохраняющие меру . . . . .	19
Инвариантные множества и виды их преобразований . . . . .	19
Теорема Биркгофа–Хинчина для преобразований, сохраняющих меру	21
Эргодическая теория для стационарных последовательностей . . .	22
<b>Семинар 5. Мартингалы</b>	<b>24</b>
Введение в мартингалы . . . . .	24
Остановленный мартингал . . . . .	25
Теорема о сходимости субмартингалов . . . . .	26
<b>Семинар 6. Случайные процессы в общем виде</b>	<b>28</b>
Сигма-алгебры . . . . .	28
Случайный процесс . . . . .	30
Виды процессов . . . . .	32
<b>Семинар 7. Марковские процессы</b>	<b>34</b>
Представление марковского процесса . . . . .	34
Матрица переходных интенсивностей . . . . .	35
<b>Семинар 8. Пуассоновский процесс</b>	<b>41</b>
Определения . . . . .	41

Обобщенный пуассоновский поток . . . . .	43
Теоремы о пуассоновском потоке . . . . .	44
Экссесс пуассоновского потока . . . . .	45
<b>Семинар 9. Процессы восстановления</b>	<b>47</b>
Процесс восстановления . . . . .	47
Функция восстановления . . . . .	47
Теорема восстановления . . . . .	48
Уравнение восстановления . . . . .	49
<b>Семинар 10. Гауссовские процессы</b>	<b>52</b>
Броуновское движение и броуновский мост . . . . .	52
Свойства броуновского движения . . . . .	53
Распределение максимума . . . . .	55
<b>Семинар 11. Слабая сходимость. Принцип инвариантности</b>	<b>57</b>
Сходимость в смысле конечномерных распределений . . . . .	57
Слабая сходимость . . . . .	58
Связь слабой сходимости со сходимостью в конечномерных распределениях . . . . .	59
Принцип инвариантности и аналогичные теоремы . . . . .	61

## Семинар 1. Случайные последовательности

### Борелевские сигма-алгебры

Рассмотрим отрезки  $[a, b]$ . Они порождают минимальную сигма-алгебру, содержащую все отрезки, мы её называем *борелевской сигма-алгеброй* на прямой  $\mathbb{R}$ . Она же порождается всеми интервалами, полуинтервалами и др.

$$\mathcal{B}(\mathbb{R}) = \sigma([a, b], a, b) = \sigma((a, b)) = \sigma([a, b))$$

Чтобы задать *борелевскую сигма-алгебру* в  $\mathbb{R}^n$ , рассмотрим минимальную сигма-алгебру, порожденную параллелепипедами, то есть декартовыми произведениями  $n$  отрезков (либо любых других борелевских множеств на прямой):

$$\mathcal{B}(\mathbb{R}^n) = \sigma(I_1 \times \dots \times I_n, I_i = [a_i, b_i]) = \sigma(B_1 \times \dots \times B_n, B_i \in \mathcal{B}(\mathbb{R})).$$

В случае  $\mathbb{R}^\infty = \{a_n\}_{n=1}^\infty, a_i \in \mathbb{R}\}$ , возьмём  $n$  отрезков:

$$\begin{aligned} I_1 &= [b_1, c_1] \\ \dots & \dots \dots \\ I_n &= [b_n, c_n] \end{aligned}$$

Возьмём множество  $I_1 \times I_2 \times \dots \times I_n \times \mathbb{R}^\infty = \{a_n\}: a_1 \in I_1, a_2 \in I_2, \dots, a_n \in I_n\}$ . Такие последовательности называют *цилиндрами*. Они порождают *цилиндрическую (борелевскую) сигма-алгебру*:

$$\begin{aligned} \mathcal{B}(\mathbb{R}^\infty) &= \sigma(I_1 \times \dots \times I_n \times \mathbb{R}^\infty, n, b_i, c_i) = \\ &= \sigma(B_1 \times \dots \times B_n \times \mathbb{R}^\infty, B_i \in \mathcal{B}(\mathbb{R})) = \sigma(B \times \mathbb{R}^\infty, B \in \mathcal{B}(\mathbb{R}^n)). \end{aligned}$$

**Утверждение 1.1.**  $A = \{a_n\}: \exists \lim a_n \in \mathcal{B}(\mathbb{R}^\infty)$ .

**Доказательство.** В доказательстве будем пользоваться тем фактом, что счётные объединения и пересечения борелевских множеств также являются борелевскими.

Распишем по критерию Коши, что значит, что у последовательности  $\{a_n\}$  существует предел:

$$\forall \varepsilon > 0 \exists N \forall m, n > N: |a_m - a_n| < \varepsilon.$$

Сначала будем считать все параметры  $m$ ,  $n$ ,  $N$  и  $\varepsilon$  фиксированными. В терминах множеств последовательностей это запишется так:

$$\{\{a_n\}: |a_m - a_n| < \varepsilon\}$$

Такое множество лежит в  $\mathcal{B}(\mathbb{R}^\infty)$ , потому что оно прямо попадает в наше порождающее. Чтобы получить множество со всеми возможными  $m$  и  $n$ , возьмём счётное пересечение по ним:

$$\bigcap_{m,n>N} \{\{a_n\}: |a_m - a_n| < \varepsilon\}$$

Теперь берём счётное объединение по  $N$ , чтобы показать что  $N$  существует:

$$\bigcup_N \bigcap_{m,n>N} \{\{a_n\}: |a_m - a_n| < \varepsilon\}$$

Теперь нужно расписать любой  $\varepsilon$ . В общем случае их несчётное количество, поэтому заменим  $\varepsilon$  на  $\frac{1}{k}$  и возьмём счётное пересечение по  $k$ :

$$A = \bigcap_k \bigcup_N \bigcap_{m,n>N} \left\{ \{a_n\}: |a_m - a_n| < \frac{1}{k} \right\}$$

Так как на каждом шаге борелевость множества сохранялась, мы показали, что  $A \in \mathcal{B}(\mathbb{R}^\infty)$ . ■

## Случайные величины, векторы, последовательности

В данном разделе под  $\mathcal{F}$  подразумевается сигма-алгебра общего вероятностного пространства  $(\Omega, \mathcal{F}, P)$ .

**Определение 1.1.** Случайной величиной называют такое измеримое отображение  $X: \Omega \rightarrow \mathbb{R}$ , что прообраз любого множества  $B \in \mathcal{B}(\mathbb{R})$  лежит в сигма-алгебре  $\mathcal{F}$ , то есть  $X^{-1}(B) \in \mathcal{F}$ .

Зачем необходимо условие  $X^{-1}(B) \in \mathcal{F}$ ? Вспомним, что вероятность — мера на нашей сигма-алгебре ( $P: \mathcal{F} \rightarrow [0, 1]$ ). Хотим посчитать вероятность  $P(\omega: X(\omega) \in B)$ , которая равна  $P(X^{-1}(B))$ . Поэтому без условия измеримости мы не сможем посчитать вероятность попадания в множество.

**Определение 1.2.** Случайным вектором называют такое измеримое отображение  $X = (X_1, \dots, X_n): \Omega \rightarrow \mathbb{R}^n$ , что прообраз любого множества  $B \in \mathcal{B}(\mathbb{R}^n)$

лежит в сигма-алгебре  $\mathcal{F}$ , то есть  $X^{-1}(B) \in \mathcal{F}$ .

Это эквивалентно тому, что прообразы порождающих  $\mathcal{B}(\mathbb{R}^n)$  лежат в  $\mathcal{F}$ :

$$\begin{aligned} X^{-1}(I_1 \times I_2 \times \dots \times I_n) = X_1^{-1}(I_1) \cap \dots \cap X_n^{-1}(I_n) \in \mathcal{F} &\iff \\ &\iff \forall i X_i^{-1}(I_i) \in \mathcal{F} \iff X_i \text{ — случайная величина.} \end{aligned}$$

То есть случайный вектор — любой набор случайных величин.

**Определение 1.3.** Случайной последовательностью называют такое измеримое отображение  $X = (X_1, \dots, X_n, \dots): \Omega \rightarrow \mathbb{R}^\infty$ , что прообраз любого множества  $B \in \mathcal{B}(\mathbb{R}^\infty)$  лежит в сигма-алгебре  $\mathcal{F}$ , то есть  $X^{-1}(B) \in \mathcal{F}$ .

Это определение эквивалентно следующему:

$$\begin{aligned} X^{-1}(I_1 \times I_2 \times \dots \times I_n \times \mathbb{R}^\infty) \in \mathcal{F} \forall I_1, \dots, I_n &\iff \bigcap_{i=1}^n X_i^{-1}(I_i) \in \mathcal{F} \iff \\ &\iff \forall i X_i^{-1}(I_i) \in \mathcal{F} \iff X_i \text{ — случайная величина.} \end{aligned}$$

Таким образом, случайной последовательностью называется набор случайных величин  $X_1, \dots$ , определённых на общем вероятностном пространстве.

## Распределение случайной последовательности

**Определение 1.4.** Распределение случайной последовательности  $\{X_n\}$  — это набор вероятностей  $P(\{X_n\} \in B)$  для всех  $B \in \mathcal{B}(\mathbb{R}^\infty)$ . Другими словами, это мера, порождаемая последовательностью на  $\mathcal{B}(\mathbb{R}^\infty)$ .

Чтобы задать распределение случайной последовательности, достаточно задать распределение каждого вектора:

$$P(\{X_n\} \in B_1 \times \dots \times B_n \times \mathbb{R}^\infty) = P(X_1 \in B_1, \dots, X_n \in B_n).$$

**Определение 1.5.** Конечномерными распределениями последовательности называют набор вероятностей:

$$p_{t_1, \dots, t_n}(B_1, \dots, B_n) = P(\omega: X_{t_1} \in B_1, \dots, X_{t_n} \in B_n)$$

при  $t_i \in \mathbb{N}$ ,  $B_i \in \mathcal{B}(\mathbb{R})$ .

Условия согласованности:  $\{p_{t_1, \dots, t_n}\}$  задаёт последовательность  $\iff$  выполнены два условия:

- 1)  $p_{t_1, \dots, t_n}$  — мера  $\forall t_1 \leq \dots \leq t_n$ ;
- 2)  $p_{t_1, \dots, t_n}(B_1, \dots, B_{i-1}, \mathbb{R}, B_{i+1}, \dots, B_n) =$   
 $= p_{t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n}(B_1, \dots, B_{i-1}, B_{i+1}, \dots, B_n).$

## Как задавать конечномерные распределения?

**Пример.** *Последовательность независимых величин.* Конечномерное распределение зададим следующим образом:

$$\begin{aligned} P_{t_1, \dots, t_n}(B_1, \dots, B_n) &= P(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n) = \\ &= P(X_{t_1} \in B_1) \dots P(X_{t_n} \in B_n) = P_{t_1}(B_1) \dots P_{t_n}(B_n), \end{aligned}$$

где  $P_{t_i}(B_i)$  — заданное семейство одномерных мер при каждом  $t_i$ .

**Пример.** *Последовательность одинаковых величин  $X_i = X$ .* Конечномерное распределение зададим следующим образом:

$$P_{t_1, \dots, t_n}(B_1, \dots, B_n) = P(X \in B_1, \dots, X \in B_n) = P_X(B_1 \cap \dots \cap B_n).$$

Точно также задаётся любая другая последовательность: нужно понять, как устроено её конечномерное распределение, а дальше мера продлится на любую последовательность.

Введём несколько классов случайных величин, с которыми будем в дальнейшем работать:

### 1. Цепи Маркова

Рассмотрим последовательность  $P(X_1 \in B_1, \dots, X_n \in B_n)$  — наше конечномерное распределение.  $X_i \in S$ ,  $|S| \leq +\infty$ . Тогда  $P(X_1 = x_1, \dots, X_n = x_n)$  задаёт конечномерное распределение. Можем расписать её следующим образом (теорема о произведении):

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n) &= \\ &= P(X_1 = x_1)P(X_2 = x_2 | X_1 = x_1) \dots P(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_1 = x_1) \end{aligned}$$

**Определение 1.6.** Последовательность  $X_n$  называется **цепью Маркова**, если

$$P(X_n = j | X_{n-1} = i, X_{n-2} = i_{n-2}, \dots, X_1 = i_1) = P(X_n = j | X_{n-1} = i).$$

Распределение цепи Маркова задаётся вектором *начального распределения*  $P(X_1 = x_1)$ ,  $x_1 \in S$  и набором матриц  $P(X_n = j | X_{n-1} = i) = P_n(i, j)$ .

Если  $P_n(i, j)$  не зависит от  $n$ , то цепь Маркова называется *однородной*.



## 2. Стационарные в узком смысле последовательности

**Определение 1.7.**  $X_n$  — стационарная в узком смысле последовательность, если

$$(X_1, \dots, X_k) \stackrel{d}{=} (X_{m+1}, \dots, X_{m+k}) \quad \forall k \in \mathbb{N}, m \in \mathbb{N}.$$

**Определение 1.8.**  $X_n$  — стационарная в широком смысле последовательность, если

$$EX_i = \text{const} = \mu, \quad \text{cov}(X_i, X_{i+m}) = \text{cov}(X_1, X_{m+1}) = R(m).$$

Функцию  $R(m)$  называют *ковариационной функцией*.

**Пример.** Последовательность независимых одинаково распределённых случайных величин:

- цепь Маркова;
- стационарна в узком смысле;
- стационарна в широком смысле.

**Пример.** Последовательность одинаковых величин:

- цепь Маркова;
- стационарна в узком смысле;
- стационарна в широком смысле.

**Пример.** *Случайное блуждание.*  $X_i$  — независимые одинаково распределённые случайные величины,  $S_n = X_1 + \dots + X_n$ . Эта последовательность

- цепь Маркова, так как

$$\begin{aligned} P(S_n = j \mid S_{n-1} = i, S_{n-2} = i_{n-2}, \dots, S_1 = i_1) &= \\ &= P(S_{n-1} + X_n = j \mid S_{n-1} = i, S_{n-2} = i_{n-2}, \dots, S_1 = i_1) = \\ &= P(X_n = j - i) = P(S_n = j \mid S_{n-1} = i); \end{aligned}$$

- не стационарна в узком смысле, так как  $S_1 \stackrel{d}{\neq} S_2$  (за исключением случая  $X_i \equiv 0$ );
- не стационарна в широком смысле, так как  $ES_n = nEX_i$ ,  $DS_n = nDX_i$  не совпадают с  $ES_{n-1}$  и  $DS_{n-1}$ , если  $X_i \not\equiv 0$ .

## Семинар 2. Ветвящиеся процессы

### Производящие функции

Наиболее удобным аппаратом для работы с ветвящимися процессами являются *производящие функции*  $\varphi(s)$ .

Напомним, что для целочисленных и неотрицательных случайных величин, производящая функция определяется как ряд  $\varphi(s) = \mathbb{E}s^X = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k)$ ,  $s \in [0, 1]$ .

Вспомним также *базовые свойства производящей функции*  $\varphi(s)$ :

- 1) производящая функция кодирует распределение:  $\mathbb{P}(X = k) = \frac{\varphi^{(k)}(0)}{k!}$ ;
- 2)  $\varphi(s)$  монотонно не убывает и выпукла вниз на  $[0, 1]$ ;
- 3)  $\varphi(1) = 1$ ;
- 4)  $\varphi^{(k)}(1) = \mathbb{E}X(X-1)\dots(X-k+1)$ . Зная их, можем найти  $\mathbb{E}X^k$ .  
В частности,  $\mathbb{E}X = \varphi'(1)$ ,  $\mathbb{D}X = \varphi''(1) + \varphi'(1) - (\varphi'(1))^2$ ;
- 5) если  $X_i$  независимы, то  $\varphi_{X_1+\dots+X_n} = \prod_{i=1}^n \varphi_{X_i}(s)$ ;
- 6)  $\varphi_{X_1+\dots+X_N}(s) = \varphi_N(\varphi_X(s))$  — сумма случайного количества  $N$  независимых одинаково распределенных величин  $X_i$ , если  $N$  не зависит от  $X_i$ .  
Можно убедиться, что эта формула работает в том числе для нулевого числа слагаемых.

### Ветвящиеся процессы Гальтона–Ватсона. Задача о вырождении

Будем считать, что  $X_{i,j}$  — набор независимых одинаково распределенных неотрицательных и целочисленных случайных величин.

$X_{1,1}$  — случайная величина с каким-то дискретным распределением и заданной производящей функцией  $\varphi(s)$ :

$$X = X_{1,1} \sim \varphi(s)$$

Рассмотрим ветвящийся процесс, задаваемый уравнениями:

$$\begin{aligned} Z_0 &= 1, \\ Z_1 &= X_{1,1}, \\ Z_2 &= X_{2,1} + X_{2,2} + \dots + X_{2,Z_1}, \\ &\dots \\ Z_n &= X_{n,1} + X_{n,2} + \dots + X_{n,Z_{n-1}}. \end{aligned}$$

Этот процесс характеризует размер популяции в  $n$ -ом поколении и называется **ветвящимся процессом Гальтона–Ватсона**.

Используя свойство б), найдём производящую функцию  $Z_n$ :

$$\varphi_{Z_n}(s) = \varphi_{Z_{n-1}}(\varphi_{X_1}(s)) = \varphi(\varphi_{Z_{n-1}}(s)) \implies \varphi_{Z_n}(s) = \underbrace{\varphi(\varphi(\dots \varphi(s)\dots))}_{n \text{ раз}}$$

В задаче о вырождении мы хотим узнать *вероятность вырождения* — вероятность того, что существует  $n$ , при котором  $Z_n = 0$ :

$$P(\exists n: Z_n = 0) = \lim_{n \rightarrow \infty} P(Z_n = 0)$$

Можем перейти к пределу по  $n$ , так как если  $\exists n: Z_n = 0$ , то  $\forall m \geq n: Z_m = 0$ . Тогда свойство непрерывности меры позволяет нам перейти от вероятности объединения событий  $\{Z_n = 0\}$  к пределу вероятности этого события.

В терминах производящих функций задача о вырождении переписывается в виде

$$P(Z_n = 0) = \varphi(\varphi(\dots \varphi(0)\dots))$$

Поймём, как ведет себя найденное выражение для выпуклых функций  $\varphi(s)$ :

1. если  $\varphi(s) \geq s \forall s \in [0, 1]$  (рис. 2.1a), выражение принимает значение 1, так как предел существует по монотонности, а единственный кандидат на роль предела — это единица, ведь предел должен удовлетворять соотношению  $\varphi(s) = s$ ;
2. Если  $\varphi(s)$  бывает как больше, так и меньше  $s$  (рис. 2.1b), то выражение примет значение  $q^*$ , где  $q^*$  — решение  $\varphi(q^*) = q^*$  такое, что  $0 \leq q^* < 1$ . В силу выпуклости  $\varphi(s)$  большего количества точек пересечений графиков  $y = \varphi(s)$  и  $y = s$  быть не может.

Классифицируем найденное решение в удобных для нас терминах:

- Если  $\varphi'(1) > 1$ , то мы оказались в нижней полуплоскости относительно  $y = s$  и  $P(Z_n = 0) = q^*$ .

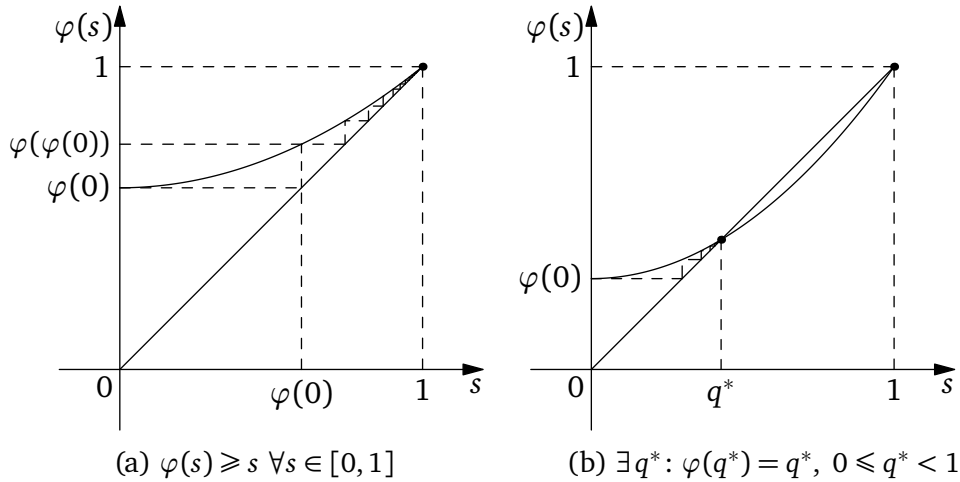


Рис. 2.1: Поведение  $\varphi(\varphi \dots (\varphi(0) \dots))$  для выпуклых функций

- Если  $\varphi'(1) \leq 1$ , то  $P(Z_n = 0) = 1$ .

Формализуем найденное решение в одном выражении:

$$P(Z_n = 0) = \min \{s \in [0, 1] : \varphi(s) = s\}$$

## Общее представление о ветвящихся процессах

Из свойства 4) производящих функций,  $\varphi'(1) = EX$ . В зависимости от его значения, ветвящиеся процессы имеют следующие названия:

- 1)  $EX < 1$  — *докритический процесс*, он вырождается;
- 2)  $EX = 1$  — *критический процесс*, вырождается, если  $\varphi(s) \neq s$  (процесс  $\varphi(s) = s$  называется *строго критическим*, он не вырождается и не размножается);
- 3)  $EX > 1$  — *надкритический процесс*, с положительной вероятностью такой процесс не вырождается никогда.

Посчитаем  $EZ_n$  через условные математические ожидания, воспользовавшись телескопическим свойством (используем обозначение  $\mu = EX$ ):

$$EZ_n = E\left(E\left(\sum_{i=1}^{Z_{n-1}} X_{n,i} \mid Z_{n-1}\right)\right) = \mu EZ_{n-1} = \dots = \mu^n$$

## Предельные теоремы для процессов Гальтона–Ватсона

**Теорема 2.1** (о надкритических процессах). Пусть  $\mu > 1$  (процесс надкритический),  $DX < +\infty$ . Тогда  $\frac{Z_n}{\mu^n} \xrightarrow{\text{п.н.}} W$ . При этом  $P(W = 0) = q^*$ ,  $EW = 1$ ,

$$DW = \frac{DX}{EX(EX - 1)}.$$

Закон больших чисел гласит, что  $\frac{Z_{n+1}}{Z_n} \approx \mu$ . По существу происходит следующее: на каждом шаге размер популяции увеличивается примерно в  $\mu > 1$  раз — экспоненциальный рост. Прервать этот процесс можно массовым вымиранием, но шансы этого экспоненциально уменьшаются с каждым шагом.

**Теорема 2.2** (о критических процессах). Пусть  $\mu = 1$  (процесс критический),  $DX < +\infty$ . Тогда  $P(Z_n > 0) \sim \frac{2}{nDX}$ , а  $P\left(\frac{2Z_n}{nDX} \leq x \mid Z_n > 0\right) \rightarrow 1 - e^{-x}$ ,  $x > 0$ .

Другими словами, в критических процессах работает принцип «Что не убивает меня — делает меня сильнее». Немногие выжившие к моменту  $n$  популяции станут более многочисленными (линейный рост), однако каждая популяция в какой-то момент пройдёт точку максимума и вымрет.

**Теорема 2.3** (о докритических процессах). Пусть  $\mu < 1$  (процесс докритический),  $DX < +\infty$ . Тогда  $P(Z_n = k \mid Z_n > 0) \rightarrow p_k^*$ , а  $P(Z_n > 0) \sim C \mu^n$ .

Шансы, что популяция доживёт до момента  $n$ , экспоненциально малы. Даже если это произойдёт, её размер будет небольшой.

## Семинар 3. Цепи Маркова

### Определение и способы задания

**Определение 3.1.**  $X_n$  — цепь Маркова, если

$$P(X_n = j | X_{n-1} = i, X_{n-2} = i_{n-2}, \dots, X_0 = i_0) = P(X_n = j | X_{n-1} = i) = p_{n,i,j}.$$

Достаточно, чтобы левая часть не зависела от  $i_{n-2}, \dots, i_0$ .

**Определение 3.2.** Цепь Маркова является *однородной*, если  $p_{n,i,j}$  не зависит от  $n$ :  $p_{n,i,j} = p_{i,j}$ .

Соответственно, можно задать *матрицу вероятностей перехода*:  $\mathbb{P} = (p_{i,j})$  (если цепь конечная, то это матрица; если цепь бесконечная, то это линейный оператор на пространстве последовательностей).

*Задание конечномерного распределения цепи* через матрицу вероятности перехода и начальное распределение:

$$P(X_0 = i_0, \dots, X_n = i_n) = P(X_0 = i_0) p_{i_0, i_1} p_{i_1, i_2} \dots p_{i_{n-1}, i_n}$$

Для одномерного распределения (формула полной вероятности):

$$P(X_n = j) = (p_0 \mathbb{P}^n)_j$$

*Построение графа по матрице и обратно.*

Если мы задаемся только вопросом «откуда куда можно попасть», то можно задать не саму матрицу  $\mathbb{P}$ , а только ориентированный граф, который порождает эта матрица  $\mathbb{P}$ : нарисуем состояния и соединим направленными ребрами те вершины, куда можно перейти за один ход, это соответствует ненулевым элементам матрицы.

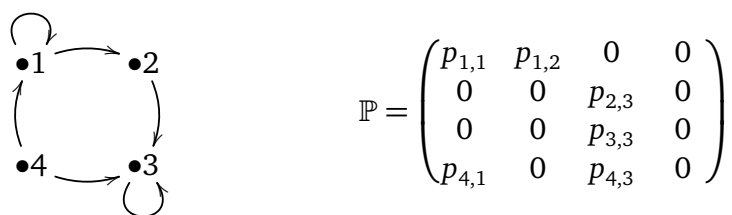


Рис. 3.1: Пример графа и соответствующей ему матрицы перехода

Если вся матрица  $\mathbb{P}$  состоит из ненулевых чисел, граф будет полным, включая петли.

Граф удобен тем, что мы можем мыслить на несколько ходов вперед и легко отслеживать длинные пути. По матрице мы видим только то, что происходит за ход, более длинные варианты необходимо перебирать вручную.

## Классификация состояний цепи

Введём несколько понятий:

- 1)  $i \rightarrow j$  — из  $i$  **следует**  $j$ , если  $\exists$  путь из  $i$  в  $j$  на графе, или (в терминах матрицы перехода)  $\exists n: (\mathbb{P}^n)_{ij} > 0$ ;
- 2)  $i \leftrightarrow j$  —  $i$  и  $j$  **сообщаются**, если  $i \rightarrow j, j \rightarrow i$ ;
- 3)  $i$  — **поглощающее** состояние, если  $p_{i,i} = 1$  ( $i \rightarrow j \implies j = i$ );
- 4)  $i$  — **существенное** состояние, если  $i \rightarrow j \implies j \rightarrow i$ .

**Пример.** Для графа на рис. 3.1 состояние 3 — существенное и поглощающее; состояния 1, 2 и 4 — несущественные.

**Пример.** Для графа на рис. 3.2 состояние 6 — существенное и поглощающее; состояния 1 и 2 — несущественные; 3, 4 и 5 — три сообщающихся существенных состояния.

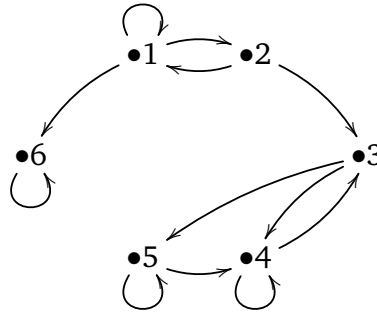


Рис. 3.2: Пример графа посложнее

**Теорема 3.1 (о разложении цепи).** Все состояния цепи Маркова разбиваются на какое-то количество классов  $E_0, E_1, \dots$ :  $E_0$  — класс несущественных состояний;  $E_1, E_2, \dots$  — замкнутые классы сообщающихся состояний (в  $E_1$  все сообщаются и из  $E_1$  не следует ничего, не лежащее в  $E_1$ ; в  $E_2$  все сообщаются и из  $E_2$  не следует ничего, не лежащее в  $E_2$  и так далее).

Схематически устройство графа и матрицы перехода любой цепи Маркова показаны на рис. 3.3.

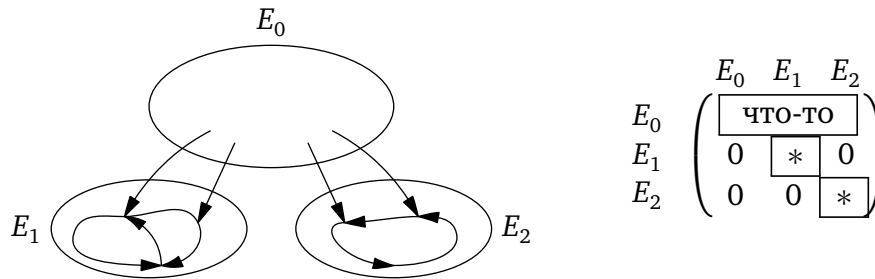


Рис. 3.3: Устройство графа и матрицы перехода произвольной цепи Маркова

**Определение 3.3.** Цепь называется **неразложимой**, если в её разложении присутствует только  $E_1$ , то есть несущественных состояний нет вообще, а существенные все со всеми сообщаются.

5) Существенное состояние  $i$  **периодично** с периодом  $d > 1$ , если  $(\mathbb{P}^n)_{i,i} = p_{i,i}^{(n)} > 0 \implies d \mid n$ , причём  $d$  — максимальное такое число. С точки зрения графа любой цикл из  $i$  в  $i$  имеет длину, кратную  $d$ . Если  $d > 1$  не существует, состояние **непериодично**.

**Пример.** Для графа на рис. 3.4 все состояния — непериодические (есть циклы длины 3 и 5). Если бы ребра  $3 \rightarrow 2$  не было, все состояния были бы периодичны с периодом 3.

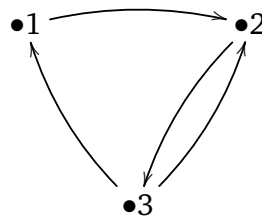


Рис. 3.4: Пример графа с непериодичными состояниями

**Теорема 3.2 (о солидарности).** Если  $i \leftrightarrow j$ , то

- а) оба существенны или оба несущественны;
- б) у обоих одинаковый период или оба непериодичны.

У каждого неразложимого класса у всех состояний один период, поэтому мы можем говорить о *периоде цепи (подцепи)*.



**Теорема 3.3 (о периоде цепи).** Если неразложимая цепь (подцепь) имеет период  $d$ , то состояния распадаются на классы  $S_1, \dots, S_d: S_1 \rightarrow S_2 \rightarrow \dots \rightarrow S_d \rightarrow S_1$ .

## Эргодическая теорема

Начнём с вопроса, как ведёт себя  $X_n$  при  $n \rightarrow \infty$ . Сойтись к конкретному состоянию он может только если мы окажемся в поглощающем состоянии. А вот по распределению сойтись можно. Так как  $X_n$  принимает конечное или счётное число значений, то сходимость по распределению — это просто сходимость  $X_n$ . Хотим понять, когда сходимость будет иметь место.

- $P(X_n = j | X_0 = i)$  — вряд ли сходится для периодических цепей;
- если цепь разложима, предел вероятности также будет зависеть от начального распределения.

**Теорема 3.4 (эргодическая для конечных цепей).** Если цепь Маркова неразложима, неперiodична и конечна, то

$$P(X_n = j | X_0 = i) \rightarrow \pi_j > 0, n \rightarrow \infty,$$

где  $\pi$  — единственное решение уравнения  $\pi P = \pi$  с  $\sum \pi_i = 1$  (левый собственный вектор или собственный вектор транспонированной матрицы с собственным значением 1).

**Пример.** Рассмотрим цепь, заданную матрицей перехода

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{pmatrix}$$

Она неразложима и неперiodическая. Выполнено условие эргодической теоремы, значит,  $\pi P = \pi$  должно иметь единственное вероятностное решение.

$$(P^T - E) \pi = 0$$

$$\begin{pmatrix} -\frac{1}{2} & 1 \\ \frac{1}{2} & -1 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} = 0$$

Отсюда видно, что  $\pi_2 = \frac{\pi_1}{2}$ . С учётом того, что  $\pi_1 + \pi_2 = 1$ , находим решение

$$\pi = \left( \frac{2}{3}, \frac{1}{3} \right).$$

- б)  $i$  — **возвратное** состояние, если  $P(\exists n: X_n = i | X_0 = i) = 1$ , иначе состояние **невозвратное**.

Из определения понятно, что любое несущественное состояние невозвратно. Для конечных цепей существенные состояния возвратны. Примером бесконечных цепей с невозвратными состояниями может служить несимметричное случайное блуждание.

7)  $i$  — **нулевое возвратное** состояние, если  $i$  возвратно, а  $\mathbb{P}_{ii}(n) \rightarrow 0$ ,  $n \rightarrow \infty$ , иначе возвратное состояние называется **положительно возвратным**.

Примером нулевой возвратности является симметричное случайное блуждание — мы обязательно вернёмся в состояние  $i$ , но на  $n$ -ом шаге это маловероятно. Если поставить барьеры с двух сторон, то возвратность станет положительной.

Теорема о солидарности распространяется на возвратность, невозвратность, нулевую возвратность и положительную возвратность.

**Теорема 3.5 (эргодическая).** Пусть цепь неразложима, непериодична и положительно возвратна. Тогда

$$P(X_n = j | X_0 = i) \rightarrow \pi_j > 0, n \rightarrow \infty,$$

где  $\pi$  — решение системы  $\pi P = \pi$  с  $\sum \pi_i = 1$ .

**Замечание 3.1.** Если неразложимая, непериодическая цепь нулевой возвратности или невозвратна, то

$$\nexists \pi: \sum \pi_i = 1, \pi_i \geq 0, \pi P = \pi;$$

$$P(X_n = j | X_0 = i) \rightarrow 0.$$

**Теорема 3.6 (критерий возвратности).**  $i$  — возвратно  $\iff \sum_n p_{i,i}(n) = +\infty$ .

Если общий член ряда к нулю не стремится, то возвратность положительная; если общий член ряда стремится к нулю, то возвратность нулевая; если ряд сходится, то состояние невозвратно.

## Семинар 4. Стационарные в узком смысле последовательности

### Преобразования, сохраняющие меру

**Определение 4.1.** Измеримое отображение  $T: \Omega \rightarrow \Omega$  называется *сохраняющим меру преобразованием*, если  $P(T^{-1}A) = P(A) \forall A \in \mathcal{F}$ .

**Пример.** Рассмотрим  $\Omega = \{0, 1\}$ . На этом множестве есть четыре различных преобразования:

$$T_1(\omega) = \omega, \quad T_2(\omega) = 1 + \omega, \quad T_3(\omega) = 1, \quad T_4(\omega) = 0. \quad (1)$$

Для произвольной меры сохранять меру будет только преобразование  $T_1$ ;  $T_2$  сохраняет только равномерную меру;  $T_3$  сохраняет меру, концентрированную на 1;  $T_4$  сохраняет меру, концентрированную на 0.

**Пример.** Возьмём преобразование  $T\omega = (\omega + a) \bmod 1$ ,  $a \in \mathbb{R}$ ,  $\Omega = [0, 1]$ ,  $P$  — мера Лебега (рис. 4.1). Оно будет сохранять меру.

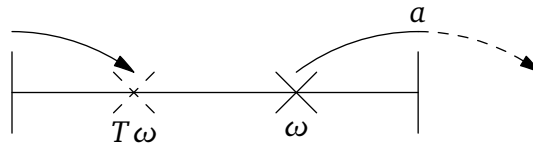


Рис. 4.1: Действие преобразования  $T\omega = (\omega + a) \bmod 1$  на точку  $\omega \in [0, 1]$

### Инвариантные множества и виды их преобразований

**Определение 4.2.** Измеримое множество  $A$  — **инвариантно**, если  $T^{-1}A = A$ . Измеримое множество  $A$  — **почти инвариантно**, если мера симметрической разности множеств нулевая:  $P(A \Delta T^{-1}A) = 0$ . Символом  $\mathcal{I}$  будем обозначать сигма-алгебру инвариантных множеств преобразования  $T$ .

**Утверждение 4.1.** Если  $A$  почти инвариантно, то  $\exists B$  — инвариантное, такое, что  $P(A \Delta B) = 0$ .

**Определение 4.3.** Преобразование  $T$  — **эргодическое**, если

$$\forall A \in \mathcal{I}: P(A) \in \begin{cases} 0; \\ 1. \end{cases}$$

Все множества меры 0 или 1 по определению почти инварианты. У эргодических преобразований других инвариантных множеств нет.

**Определение 4.4.** Преобразование  $T$  — перемешивающее, если

$$P(A \cap T^{-n}B) \rightarrow P(A)P(B) \quad \forall A, B$$

Если хорошо перемешать  $B$ , то оно станет почти независимым от  $A$ , даже если изначально оно совпадало с  $A$ .

Все перемешивающие преобразования эргодичны, обратное неверно.

**Пример.** Вернёмся к преобразованиям (1).

- У  $T_1(\omega) = \omega$  инвариантные множества все, при этом не все меры 0 или 1. Поэтому это преобразование будет эргодичным только если мера сосредоточена в 0 или 1;
- $T_2(\omega)$  — эргодическое при равномерной мере, так как его инвариантные множества —  $\Omega$  и  $\emptyset$ . При этом оно неперемешивающее, так как если  $n$  — чётное, а  $A = \{1\}$ ,  $B = \{1\}$ , то  $P(A \cap T^{-n}B) = \frac{1}{2} \neq \frac{1}{4}$ ;
- $T_3(\omega)$  и  $T_4(\omega)$  сохраняют меру, сосредоточенную в 0 или 1 и все события имеют меру 0 или 1, поэтому эти преобразования эргодические.

**Пример** (Задача про перестановку).

Рассмотрим перестановку  $\sigma: \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ .

Это преобразование сохраняет меру, так как

$$P(\sigma^{-1}(A)) = \frac{|\sigma^{-1}(A)|}{n} = \frac{|A|}{n} = P(A).$$

- Если  $\sigma$  — не цикл, то у неё есть цикл меньшей длины  $(i_1, \dots, i_k)$ . Этот цикл образует инвариантное множество, потому что все его элементы переходят в себя. Так как этот цикл не полной длины, он имеет меру не 0 и не 1, значит, эргодичность нарушается.
- Если  $\sigma$  — цикл. Пусть  $\sigma^{-1}A = A$ . Если  $a \in A$ , то и предыдущий элемент цикла также лежит в  $A$  и все последующие элементы. Значит,  $A$  может быть или пустым или  $\Omega$ , а  $\sigma$  эргодична.

Таким образом, у перестановок эргодичность означает, что перестановка имеет единую цикловую структуру.

## Теорема Биркгофа–Хинчина для преобразований, сохраняющих меру

**Теорема 4.1** (Биркгофа–Хинчина для преобразований, сохраняющих меру). Пусть  $T$  — преобразование, сохраняющее меру, тогда для любой случайной величины  $X : \Omega \rightarrow \mathbb{R}$  выполнено соотношение

$$\frac{\sum_{k=0}^n X(T^k \omega)}{n+1} \xrightarrow{\text{п.н.}} E(X | \mathcal{I}).$$

**Пример.**  $T : \{1, \dots, m\} \rightarrow \{1, \dots, m\}$  — перестановка чисел от 1 до  $n$ . Мера равномерная, а  $X(\omega) = \omega$ .

- Если у нас цикл, то перестановка будет перемещать  $X$  по всем числам от 1 до  $n$ . За  $m$  шагов встретим каждого числа по одной штуке, поэтому при  $n \rightarrow \infty$  наша дробь будет стремиться к  $EX = \frac{m+1}{2}$  по теореме Биркгофа–Хинчина.
- Если у нас перестановка с цикловой структурой, то всё будет зависеть от того, где мы начали. Если стартовая  $\omega$  была в каком-то блоке, она всегда в этом цикле жить и будет и  $E(X | \mathcal{I})$  будет зависеть от того, где именно лежит  $\omega$ .

Справедлива такая замечательная конструкция:

- 1) Если  $T$  сохраняет меру, а  $X$  — любая случайная величина, то последовательность  $X_n(\omega) = X(T^n \omega)$  — стационарная в узком смысле последовательность.

**Доказательство.** Проверим, что распределение вектора  $(X_1, \dots, X_d)$  такое же, как и вектора  $(X_{n+1}, \dots, X_{n+d})$ :

$$P((X_1, \dots, X_d) \in A) \stackrel{?}{=} P((X_{n+1}, \dots, X_{n+d}) \in A)$$

Обозначим множество  $B = \{\omega : (X_1, \dots, X_d) \in A\}$  и распишем выражения в левой и правой части:

$$P((X_1, \dots, X_d) \in A) = P(\omega \in B) = P(B);$$

$$P((X_{n+1}, \dots, X_{n+d}) \in A) = P(T^n \omega \in B) = P(\omega \in T^{-n}B) = P(T^{-n}B).$$

Так как  $T$  сохраняет меру, то  $P(B) = P(T^{-n}B)$ , поэтому наше равенство верное и последовательность стационарна в узком смысле. ■

## Эргодическая теория для стационарных последовательностей

- 2) Если  $X_n$  — стационарная в узком смысле последовательность, то  $\exists (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{P}), \tilde{T}$  — сохраняющая меру,  $\tilde{X}: \tilde{\Omega} \rightarrow \mathbb{R}$  — случайная величина, такие, что

$$(\tilde{X}(\tilde{\omega}), \tilde{X}(\tilde{T}\tilde{\omega}), \dots, \tilde{X}(\tilde{T}^k\tilde{\omega})) \stackrel{d}{=} (X_1(\omega), \dots, X_{k+1}(\omega)) \quad \forall k,$$

то есть можно поменять вероятностное пространство, что распределение последовательности останется тем же, но сама последовательность будет иметь форму, которая описывается в теореме Биркгофа–Хинчина.

**Теорема 4.2** (Биркгофа–Хинчина для стационарных последовательностей). Если  $X_n$  — стационарная в узком смысле последовательность, то

$$\frac{\sum_{i=0}^n X_i}{n+1} \xrightarrow{\text{п.н.}} \mathbb{E}(X_1 | \mathcal{I}),$$

где  $\mathcal{I}$  — некоторая сигма-алгебра инвариантных множеств:

$$\mathcal{I} = \{A \in \mathcal{F} : A = \{(X_k, \dots) \in B \in \mathcal{B}(R^\infty)\} \quad \forall k\}$$

**Определение 4.5.** Последовательность  $X_n$  — **эргодическая**, если  $\mathcal{I}$  состоит из множеств, имеющих меру 0 или 1, то есть  $\forall A \in \mathcal{I} \quad \mathbb{P}(A) \in \left\{ \begin{matrix} 0; \\ 1. \end{matrix} \right.$

**Определение 4.6.** Последовательность  $X_n$  — **перемешивающая**, если  $\mathbb{P}((X_1, \dots, X_k) \in A, (X_n, \dots, X_{n+l}) \in B) \rightarrow \mathbb{P}((X_1, \dots, X_k) \in A) \mathbb{P}((X_1, \dots, X_l) \in B)$  при  $n \rightarrow \infty$ .

**Пример.** Последовательность независимых одинаково распределённых случайных величин перемешивающая. Также из эргодической теоремы следует, что неразложимая, непериодическая, конечная цепь Маркова также является перемешивающей.

Из перемешиваемости, как и раньше, следует эргодичность, а значит в теореме Биркгофа–Хинчина можем заменить  $\mathbb{E}(X_1 | \mathcal{I})$  на  $\mathbb{E}X_1$ .

**Теорема 4.3** (УЗБЧ Биркгофа–Хинчина). Если  $X_n$  — стационарная в узком смысле эргодическая последовательность, то

$$\frac{X_1 + \dots + X_{n+1}}{n+1} \xrightarrow{\text{п.н.}} \mathbb{E}X_1.$$

Так как последовательность независимых одинаково распределённых случайных величин эргодическая, УЗБЧ Колмогорова является частным случаем УЗБЧ Биркгофа–Хинчина.

**Утверждение 4.2.** Пусть  $h: \mathbb{R}^\infty \rightarrow \mathbb{R}$  — измеримое отображение,  $X_n$  — стационарная в узком смысле последовательность.

Рассмотрим  $Y_n = h(X_n, X_{n+1}, \dots)$ :

- а)  $Y_n$  — стационарна в узком смысле.
- б) Если  $X_n$  — эргодическая, то  $Y_n$  — эргодическая.

**Пример.**  $X_i$  — независимые одинаково распределённые случайные величи-

ны;  $h(x, y) = x \cdot y \implies \frac{\sum_{i=1}^n X_i X_{i+1}}{n} \xrightarrow[n \rightarrow \infty]{\text{п.н.}} EX_1 EX_2$ .

## Семинар 5. Мартингалы

### Введение в мартингалы

Вспомним свойства условного математического ожидания:

- 1)  $E(c | \mathcal{A}) = c$  — математическое ожидание константы;
- 2)  $E(aX + bY | \mathcal{A}) = aE(X | \mathcal{A}) + bE(Y | \mathcal{A})$  — линейность;
- 3)  $E(E(X | \mathcal{A}_1) | \mathcal{A}_2) = E(E(X | \mathcal{A}_2) | \mathcal{A}_1) = E(X | \mathcal{A}_1)$ , если  $\mathcal{A}_1 \subseteq \mathcal{A}_2$  — теле-скопическое свойство;
- 4)  $E(\varphi(X)Y | X) = \varphi(X)E(Y | X)$ , где  $\varphi(X)$  — измеримая функция;
- 5)  $E(X | Y) = EX$ , если  $X, Y$  независимы.

**Пример (Казино).** Есть игрок в казино с  $X_0$  денег, которые он может распределить по разным автоматам (рис. 5.1) или оставить на руках. Автомат выдаёт выигрыш, который не зависит от всего, что происходит, кроме того, что в него засунули и в среднем выигрыш не больше того, что в него засунули. Если автомат получил на вход  $x$ , он выдаст на выходе случайную величину  $x + Y$ :  $EY \leq 0$ .

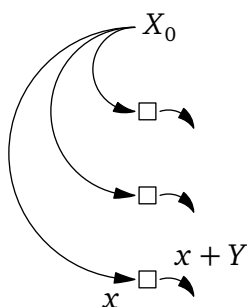


Рис. 5.1: Устройство казино

Капитал в момент времени  $n$  обладает следующим свойством:

$$E(X_n | X_{n-1}, \dots, X_0) \leq X_{n-1}$$

#### Определение 5.1.

- $X_n$  — **мартингал**, если  $E(X_n | X_{n-1}, \dots, X_0) = X_{n-1}$  п.н.  
 $X_n$  — **супермартингал**, если  $E(X_n | X_{n-1}, \dots, X_0) \leq X_{n-1}$  п.н.  
 $X_n$  — **субмартингал**, если  $E(X_n | X_{n-1}, \dots, X_0) \geq X_{n-1}$  п.н.



Приведённая модель не совсем точно описывает происходящее в реальном казино, так как решения игрока могут основываться на дополнительной информации об автоматах. Поэтому удобнее рассматривать не сигма-алгебру, порождённую  $X_i$ , а сигма-алгебру всего прошлого.

**Определение 5.2.** Пусть  $\mathcal{F}_i \subseteq \mathcal{F}_{i+1}$  — фильтрация (поток) сигма-алгебр, причём  $X_n$  удовлетворяет следующим условиям:

- 1)  $X_i \sim \mathcal{F}_i$ -измеримо;
- 2)  $E|X_i| < +\infty$ ;
- 3)  $E(X_{n+1} | \mathcal{F}_n) = X_n$  п.н.

Тогда  $(X_n, \mathcal{F}_n)$  — **мартингал**. Заменой равенства на неравенство в последнем условии определяются *субмартингал* и *супермартингал*.

Разница приведённых определений в том, что в  $\mathcal{F}_n$  можно добавить что-то ещё помимо  $X_i$ .

**Пример.** Пусть  $\xi_i$  — независимые случайные величины с  $E|\xi_i| < +\infty \forall i$ , а  $S_n = \xi_1 + \dots + \xi_n$ . Так как  $E(S_{n+1} | S_n, \dots, S_1) = E(\xi_{n+1} + S_n | S_n, \dots, S_1) = E\xi_{n+1} + S_n$ , то

- $S_n$  — мартингал, если  $E\xi_i = 0$ ;
- $S_n$  — субмартингал, если  $E\xi_i \geq 0$ ;
- $S_n$  — супермартингал, если  $E\xi_i \leq 0$ .

**Пример (мартингал Леви).** Если  $X$  — случайная величина с  $E|X| < +\infty$ ,  $\mathcal{F}_n$  — фильтрация (поток). Рассмотрим  $X_n = E(X | \mathcal{F}_n)$ . Это мартингал относительно потока  $\mathcal{F}_n$ , так как  $E(X_{n+1} | \mathcal{F}_n) = E(E(X | \mathcal{F}_{n+1}) | \mathcal{F}_n) = E(X | \mathcal{F}_n) = X_n$ .

**Свойства (суб/супер)мартингалов:**

1. Пусть  $X_m$  — (суб/супер)мартингал,  $m > n$ , тогда

$$E(X_m | \mathcal{F}_m) = E(E(X_m | \mathcal{F}_{m-1}) | \mathcal{F}_n) = E(X_{m-1} | \mathcal{F}_n) = \dots = X_n.$$

2.  $EX_m = EX_0$  — на уровне первого момента все (суб/супер)мартингалы одинаковые.

## Остановленный мартингал

Мы доказали, что у игрока в казино в среднем всегда меньше или равно денег, чем в начальный момент:  $EX_n \leq EX_0$ . Но мы можем уйти из казино

не в  $n$ -ый момент, а каждый раз выбирать  $n$ , зависящий от  $\omega$ . То есть найдём оптимальный момент ухода из казино  $T(\omega)$  и уйдём с выигрышем  $X_T$ .

**Пример (Петербургская игра).** Подбрасываем монетку. Если выпадает на орла, получаем в два раза больше денег, если на решку — теряем всю сумму. Игрок может увеличивать каждый раз, когда выпала решка, ставку в 2 раза, и когда орёл всё-таки выпадет, он вернёт все свои деньги и дополнительно заработает исходную ставку.

**Определение 5.3. Марковский момент** — такая случайная величина  $\tau$ , что событие  $\{\tau \leq n\} \in \mathcal{F}_n \forall n$ . **Момент остановки** — марковский момент с  $E T < +\infty$ .

Иногда марковский момент называют моментом остановки, но мы будем использовать такую терминологию.

В наших терминах условие  $\{\tau \leq n\}$  означает, что  $\tau$  не может забегать вперёд во времени. То есть решение игрока, ушёл ли он до момента  $n$  определяется исходя из того, что происходило в игре до момента  $n$ .

**Теорема 5.1 (об остановленных мартингалах).** Пусть  $T$  — марковский момент,  $E \sum_{i=1}^T |X_{i+1} - X_i| < +\infty$ . Тогда  $E X_T = E X_0$ .

Для субмартингалов и супермартингалов равенство меняется на соответствующее неравенство.

Так как в казино капитал игрока изначально ограничен, условия теоремы выполнены и игрок даже в момент остановки уйдёт с меньшим в среднем капиталом, чем пришёл. Не работает с таким лимитом и петербургская игра: когда выигрыш будет приближаться к размеру изначальному капиталу, будет реальный шанс всё проиграть.

## Теорема о сходимости субмартингалов

**Теорема 5.2 (о сходимости).** Пусть  $X_n$  — субмартингал и  $\sup_n E |X_n| < +\infty$ , тогда  $X_n$  сходится почти наверное.

Из этой теоремы получаем, что  $\frac{Z_n}{m^n}$  сходится почти наверное.

**Теорема 5.3.** Если  $X_n$  — равномерно интегрируема ( $E |X_n| I_{|X_n| > M} \xrightarrow{\text{по } n} 0$  при  $M \rightarrow +\infty$ ), то  $X_n$  сходится в  $L_1$  к какой-то случайной величине  $X_\infty$ , причём  $X_n$  — мартингал Леви по  $X_\infty$ :  $X_n = E(X_\infty | \mathcal{F}_1)$ .

Сходимость в  $L_1$  позволяет нам что-то понять про предел, потому что как минимум математическое ожидание  $X_n$  сходится к математическому ожиданию предела.

Для мартингалов выполнены ряд неравенств, которые похожи на неравенства для случайных блужданий.

**Теорема 5.4 (Дуба).**

1. Если  $X_n$  — субмартингал, то

$$P\left(\sup_{n \leq m} |X_n| > x\right) \leq \frac{EX_m^+}{x}, \quad x > 0, \quad X_m^+ = \max(X_m, 0);$$

2. Если  $X_n$  — мартингал,  $E|X_n|^p < +\infty$ ,  $p > 1$ , то  $P\left(\sup_{n \leq m} |X_n| > x\right) \leq \frac{E|X_m|^p}{x^p}$ .

## Семинар 6. Случайные процессы в общем виде

### Сигма-алгебры

Вспомним технологию построения меры и сигма-алгебр, начиная с  $\mathbb{R}$ . Возьмём отрезки  $[a, b]$  и рассмотрим минимальную сигма-алгебру, содержащую все отрезки  $\sigma([a, b], a, b)$ . Она совпадает с сигма-алгеброй, порождённой всеми открытыми множествами и называется *борелевской на прямой*:  $\mathcal{B}(\mathbb{R})$ .

В  $\mathbb{R}^n$  мы взяли параллелепипеды или декартово произведение любых борелевских множеств и породили сигма-алгебру:

$$\sigma([a_1, b_1] \times \dots \times [a_n, b_n]) = \sigma(B_1 \times \dots \times B_n, B_i \in \mathcal{B}(\mathbb{R}))$$

Она совпадает с сигма-алгеброй, порождённой открытыми множествами и называется *борелевской в пространстве*:  $\mathcal{B}(\mathbb{R}^n)$ .

В  $\mathbb{R}^\infty$  сигма-алгебру порождают цилиндры или декартово произведение любых борелевских множеств на  $\mathbb{R}^\infty$ :

$$\sigma([a_1, b_1] \times \dots \times [a_n, b_n] \times \mathbb{R}^\infty) = \sigma(B_1 \times \dots \times B_n \times \mathbb{R}^\infty) = \sigma(B \times \mathbb{R}^\infty)$$

Она совпадает с сигма-алгеброй, порождённой открытыми множествами с правильно введённой метрикой и называется *борелевской в пространстве последовательностей*:  $\mathcal{B}(\mathbb{R}^\infty)$ .

Теперь рассмотрим пространство функций  $\mathbb{R}^T$  — это множество отображений  $f: T \rightarrow \mathbb{R}$ ,  $T \subseteq \mathbb{R}^+$ . В качестве порождающего возьмём цилиндр  $I_{B_1, B_2, \dots, B_n}^{t_1, t_2, \dots, t_n} = \{f: f(t_1) \in B_1, \dots, f(t_n) \in B_n\}$ . Здесь  $B_i \in \mathcal{B}(\mathbb{R})$ ,  $t_i \in T$ .

То есть один цилиндр — множество функций, принимающих в заданных точках  $t_i$  значения из определенного множества  $B_i$  (рис. 6.1). Так как у нас нет никаких ограничений на функции, это довольно слабое условие на множество функций.

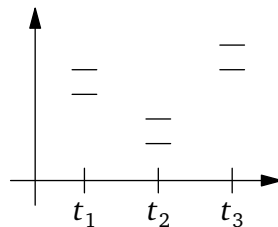


Рис. 6.1: Области, через которые должна пройти функция, лежащая в цилиндре

Хотим натянуть на эти цилиндры минимальную сигма–алгебру и назвать это **цилиндрической сигма–алгеброй в пространстве функций**:  $\mathcal{I}_T$ .

Мы задаём такие порождающие, потому что на них мы будем задавать меру. Для таких порождающих достаточно задать вероятность попадания в соответствующие множества конечного числа точек.

Эта сигма–алгебра будет совпадать с сигма–алгеброй

$$\sigma\left(\left\{f : (f(t_1), \dots, f(t_n)) \in B, B \in \mathcal{B}(\mathbb{R}^n)\right\}\right),$$

она же будет совпадать с сигма–алгеброй

$$\sigma\left(\left\{f : (f(t_1), \dots, f(t_n), \dots) \in B, B \in \mathcal{B}(\mathbb{R}^\infty)\right\}\right),$$

а эта сигма–алгебра в свою очередь равна множеству

$$\left\{\left\{f : (f(t_1), \dots) \in B\right\}\right\},$$

так как минимальная сигма–алгебра, содержащая такие множества — это просто сами эти множества, потому что они уже образуют сигма–алгебру.

Ни до какой борелевской сигма–алгебры, порождённой какими-то открытыми множествами, мы не дошли даже близко. Мы дошли только до множества функций, которые определяются конечным или счётным числом точек. Любое цилиндрическое множество в  $\mathbb{R}^T$  — это множество, для которого есть такое  $B \in \mathcal{B}(\mathbb{R}^\infty)$  и есть такая последовательность  $\{t_i\}_{i=1}^\infty$ .

**Пример.** Рассмотрим единичный шар в  $\mathbb{R}^T$  по равномерной норме:

$$A = \{f : |f(t)| \leq 1 \forall t\}.$$

Если  $A \in \mathcal{I}$ , то  $\exists B \in \mathcal{B}(\mathbb{R}^\infty)$  и  $\{t_i\} : A = \{f : (f(t_1), \dots, f(t_n)) \in B\}$ . Возьмём любую функцию, которая лежит в  $A$ , например,  $f_1 = 0$ . Положим  $f_2 = \begin{cases} 0, & t \neq s \\ 2, & t = s \end{cases}$ , где  $s \notin \{t_i\}$ .  $f_2$  в  $A$  не лежит, а  $f_1$  лежала. При этом значения у них в  $t_i$  одинаковые, поэтому раз одна лежала в  $B$ , то и вторая будет лежать в  $B$ . Мы получили противоречие. Значит,  $A \notin \mathcal{I}$ .

Сигма–алгебру в  $\mathbb{R}^T$  мы смогли построить, но она получилась очень бедная, в ней отсутствует даже единичный шар.

## Случайный процесс

**Определение 6.1.** Случайным процессом  $X_t$  называют отображение  $X_t: \Omega \rightarrow \mathbb{R}^T$ , такое, что прообраз любого элемента из цилиндрической сигма-алгебры является элементом  $\mathcal{F}: X^{-1}(B) \in \mathcal{F} \forall B \in \mathcal{I}_T$ .

Как всегда, достаточно проверить это только на порождающих: если

$$X^{-1}(\{f: f(t_1) \in B_1, \dots, f(t_n) \in B_n\}) \in \mathcal{F},$$

то  $X$  — случайный процесс. Можно смотреть на  $X$  как на конструкцию.  $X: \Omega \times T \rightarrow \mathbb{R}$ , тогда условие на случайный процесс запишется в виде

$$\{\omega: X_{t_1}(\omega) \in B_1, \dots, X_{t_n}(\omega) \in B_n\} \in \mathcal{F}.$$

Под  $X_{t_i}(\omega)$  подразумеваем значение функции  $X(\omega)$  в точке  $t_i$ .

Если взять  $n = 1$ , то получим, что  $X_{t_i}$  — случайные величины. И наоборот, если  $X_{t_i}$  — случайные величины, то при любом  $n$  такая конструкция есть пересечение элементов из  $\mathcal{F}$  и набор случайных величин будет случайным процессом.

То есть *случайный процесс* — просто набор случайных величин, индексированных индексом  $T$ , где их несчётное количество.

**Пример.** Рассмотрим на  $T = [0, 1]$   $\sup_{t \in [0,1]} |X_t|$ . Вообще говоря, это не случайная величина, потому что  $\{\omega: \sup_{t \in [0,1]} |X_t| \in [0, 1]\} = \{\omega: |X_t(\omega)| \leq 1 \forall t\}$ . Поэтому хоть случайный процесс и легко построить, даже максимум на отрезке случайной величиной не является. Построим пример, когда это прямо не будет случайной величиной. Возьмём  $\Omega = [0, 1]$ ,  $T = [0, 1]$ ,  $\mathcal{F}$  — борелевская сигма-алгебра на  $[0, 1]$ , случайный процесс

$$X_t(\omega) = \begin{cases} 0, & t \neq \omega \text{ или } \omega \notin A \text{ — неизмеримое множество на } [0, 1]; \\ 1, & t = \omega, \omega \in A \end{cases}$$

Зафиксируем  $\omega$  и будем менять  $t$ , чтобы понять как выглядит траектория процесса. Она либо всегда 0, если  $\omega \notin A$ , либо всегда 0, а в точке  $\omega$  равна 1, если  $\omega \in A$ . Тогда  $\sup_t |X_t| = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$  — не случайная величина, потому что прообраз множества, состоящего только из единицы, не является борелевским.

Теперь зафиксируем  $t$  и посмотрим как выглядит  $X_t(\omega)$ . Если  $t \notin A$ , то  $X_t(\omega) = 0$ ; если  $t \in A$ , то  $X_t(\omega) = I_{t=\omega}$ . Получается, что при фиксированном  $t$   $X_t(\omega)$  — случайная величина, так как прообраз — либо весь отрезок  $[0, 1]$ , либо этот отрезок без точки. Оба они являются борелевскими.

Чтобы задать распределение случайного процесса, нужно задать его *конечномерное распределение* — набор мер

$$p_{t_1, \dots, t_n}(B_1, \dots, B_n) := P(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n).$$

Однако не всегда эти меры определяют конкретный процесс.

**Пример.** Пусть мы задали случайным величинам  $\frac{1}{2}$  и  $\frac{1}{3}$  конечномерные распределения

$$p_{\frac{1}{2}}([0, x]) = x, \quad x \in [0, 1] \text{ и } p_{\frac{1}{3}, \frac{1}{2}}([0, x] \times [0, y]) = (1 - e^{-x})(1 - e^{-y})I_{x, y > 0}.$$

Такого не бывает, потому что  $\frac{1}{2}$  распределена равномерно, а пара  $(\frac{1}{3}, \frac{1}{2})$  независимо экспоненциально, поэтому  $X = \frac{1}{2}$  одновременно распределена равномерно и экспоненциально. Распределения не согласовались друг с другом.

То же самое может произойти при несимметричном распределении, если не указать, что  $t_i$  — монотонно возрастающие. Поэтому положим  $t_1 < \dots < t_n$ .

Будем требовать от  $p_{t_1, \dots, t_n}$  следующие условия согласованности:

- 1)  $p_{t_1, \dots, t_n}$  — мера, заданная на декартовом произведении;
- 2)  $p_{t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_n}(B_1, \dots, B_{i-1}, B_{i+1}, \dots, B_n) = p_{t_1, \dots, t_n}(B_1, \dots, B_{i-1}, \mathbb{R}, B_{i+1}, \dots, B_n)$ .

**Теорема 6.1 (Колмогорова).** Если выполнены условия 1) и 2) (меры согласованы), то существует пространство  $\Omega$  и случайный процесс  $X: \Omega \rightarrow \mathbb{R}^T$  с таким конечномерным распределением.

**Пример.** Хотим построить набор независимых  $X_t$  таких, что  $X_t \sim F_t$ ,  $F_t$  — заданный набор функций распределения,  $t \in [0, 1]$ , то есть хотим задать континуальный набор независимых величин на одном вероятностном пространстве. Для этого зададим конечномерные распределения

$$p_{t_1, \dots, t_n}(B_1, \dots, B_n) = P_{t_1}(B_1) \dots P_{t_n}(B_n),$$

где  $P_{t_i}$  — мера, соответствующая  $F_{t_i}$ . Условия 1) и 2) выполнены, значит, существует процесс  $X_t$  с такими конечномерными распределениями.

**Определение 6.2.**  $X_t$  — модификация  $Y_t$ , если  $P(X_t = Y_t) = 1 \quad \forall t \in T$ .

**Пример.** Возьмём  $\Omega = [0, 1]$ ,  $\mathcal{F} = \mathcal{B}([0, 1])$ ,  $X_t(\omega) = 0$ ,  $Y_t(\omega) = I_{t=\omega}$ . Траектории этих процессов выглядят по-разному, у них разные супремумы и другие характеристики, ни при каком  $\omega$  эти функции не совпадают, но  $X_t$  — модификация  $Y_t$ .

**Теорема 6.2 (о непрерывной модификации).** Пусть  $X$  — случайный процесс, такой, что выполнено условие  $E|X_t - X_s|^{a+1} \leq C|t - s|^{1+b}$ ,  $a, b, C > 0 \quad \forall t, s \in T$ .

Тогда  $\exists Y_t$  — модификация  $X_t$ , такая, что  $Y_t(\omega) \in C(T) \forall \omega$ .  $Y_t$  называется **непрерывной модификацией**  $X_t$ .

Если мы будем рассматривать непрерывную модификацию случайного процесса вместо самого процесса, то бедная цилиндрическая сигма-алгебра превратится в борелевскую на пространстве функций:  $\mathcal{B}(\mathbb{R}^T)$ . Например, единичный шар для непрерывных траекторий — цилиндрическое множество, потому что если в рациональных точках функция не больше 1, то во всех точках она не больше 1.

## Виды процессов

Рассмотрим некоторые виды процессов, с которыми в дальнейшем будем работать:

1) Процессы с независимыми приращениями

**Определение 6.3.**  $X_t$  имеет **независимые приращения**, если  $X_{t_n} - X_{t_{n-1}}$ ,  $X_{t_{n-1}} - X_{t_{n-2}}, \dots, X_{t_0}$  — независимые величины  $\forall t_0 < t_1 < \dots < t_n$ .

Если мы зададим распределение  $X_0$  и  $X_t - X_s$ , получим конечномерное распределение. Если они согласованы, тогда такой случайный процесс существует.

1a) **Винеровский процесс:**  $X_0 = 0, X_t - X_s \sim \mathcal{N}(0, t - s)$ ;

1b) **Пуассоновский процесс:**  $X_0 = 0, X_t - X_s \sim \text{Pois}(\lambda(t - s))$ .

Можно также рассмотреть процессы, у которых будет гамма-распределение или распределение Коши, но не равномерное, так как распределение должно быть безгранично делимым.

2) Марковские процессы

**Определение 6.4.** Случайный процесс  $X_t$  называется **марковским**, если

$$P(X_{t_n} \in A_n | X_{t_{n-1}}, \dots, X_0) = P(X_{t_n} \in A_n | X_{t_{n-1}}) \forall t_1 < t_2 < \dots < t_n.$$

3) Гауссовский процесс

**Определение 6.5.** Случайный процесс  $X_t$  называется **гауссовским**, если

$$(X_{t_1}, \dots, X_{t_n}) \sim \mathcal{N}(\vec{\mu}_{t_1, \dots, t_n}, \Sigma_{t_1, \dots, t_n}).$$

4) Стационарные процессы



**Определение 6.6.** Случайный процесс  $X_t$  называется **стационарным в узком смысле**, если  $(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \dots, X_{t_n+h})$ .

**Определение 6.7.** Случайный процесс  $X_t$  называется **стационарным в широком смысле**, если  $EX_t = \text{const}$ ,  $\text{cov}(X_t, X_s) = \text{cov}(X_{t+h}, X_{s+h})$ , в том числе  $DX_t = DX_{t+h}$ .

## Семинар 7. Марковские процессы

### Представление марковского процесса

Пусть есть  $S$  — конечное или счётное множество состояний, есть случайный процесс  $X_t: \Omega \rightarrow S$ . Он называется *марковским процессом с дискретным множеством состояний* (или *марковской цепью с непрерывным временем*), если  $\forall t_0 < t_1 < \dots < t_n$ :

$$\mathbb{P}(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1}, \dots, X_{t_0} = i_0) = \mathbb{P}(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1})$$

Мы будем рассматривать только *однородные* цепи Маркова, у которых эта функция равна  $p_{i_{n-1}, i_n}(t_n - t_{n-1})$

**Теорема 7.1** (уравнение Колмогорова–Чепмена).

$$\mathbb{P}(t+s) = \mathbb{P}(t)\mathbb{P}(s), \text{ где } \mathbb{P}(t) = (p_{i,j}(t)).$$

Это уравнение выполнено по формуле полной вероятности:

$$\mathbb{P}(X_{t+s} = j | X_0 = i) = \sum_k \mathbb{P}(X_t = k | X_0 = i) \mathbb{P}(X_{t+s} = j | X_t = k)$$

Введём дополнительные условия на нашу цепь:

1. (*сепарабельность*). Для любого момента времени  $t$  существует случайная величина  $\tau: \tau > 0$  почти наверное — время, которое я пробуду в состоянии  $X_t$ , прежде чем выйду.
2. (*стохастическая непрерывность*). Матрица  $\mathbb{P}(t) \rightarrow E$  при  $t \rightarrow 0+0$  равномерно по элементам, то есть  $p_{i,j}(s)$  стремится к 1, если  $i = j$ ; к 0, если  $i \neq j$ , причём сходимость равномерная по  $i$  и  $j$ . Считаем, что  $T = \mathbb{R}^+$ .

Представим, что в момент  $t$  я нахожусь в состоянии  $i$  цепи Маркова. Возьмём маленький шаг ширины  $\frac{s}{n}$  и будем проверять, поменялось ли состояние, если сделать несколько таких шагов (рис. 7.1).

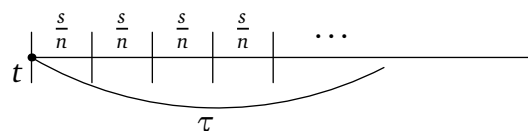


Рис. 7.1: Устройство сепарабельной цепи Маркова

$\tau$  — время, которое я просижу, прежде чем покину состояние.

$$\begin{aligned} P(\tau > s \mid X_t = i) &= P(X_r = i, r \in [t, t+s] \mid X_t = i) \stackrel{(a)}{\approx} \\ &\stackrel{(a)}{\approx} P\left(X_{t+\frac{s}{n}} = i, X_{t+\frac{2s}{n}} = i, \dots, X_{t+\frac{ns}{n}} = i \mid X_t = i\right) \stackrel{(b)}{=} \\ &\stackrel{(b)}{=} p_{i,i}^n\left(\frac{s}{n}\right) = e^{n \ln p_{i,i}\left(\frac{s}{n}\right)} \stackrel{(c)}{=} e^{n\left(1-p_{i,i}\left(\frac{s}{n}\right)\right) + o\left(n p_{i,i}\left(\frac{s}{n}\right)\right)} \stackrel{(d)}{\rightarrow} e^{-q_i(s)}. \end{aligned}$$

При переходе (a) мы потеряли события, когда в промежутке длины  $\frac{s}{n}$  цепь успела отскочить куда-то и вернуться назад. Можно показать, что вероятность такого перехода  $o\left(\frac{1}{n}\right)$ , поэтому это событие произойдёт хоть раз с вероятностью  $o(1)$ .

Мы получили обычную дискретную цепь Маркова, для которой можем честно выписать вероятность — это переход (b).

Переход (c) выполнен, так как  $\ln p_{i,i} = \ln(1 - (p_{i,i} - 1))$ .

Переход (d) возможен, если потребуем, чтобы  $n\left(1 - p_{i,i}\left(\frac{s}{n}\right)\right)$  стремилась к произведению  $s$  на некоторую константу  $q_i$ . Если  $p_{i,i}$  дифференцируема, то  $p'_{i,i}(0) = -q_i$ , так как

$$n\left(1 - p_{i,i}\left(\frac{s}{n}\right)\right) = \frac{\left(p_{i,i}(0) - p_{i,i}\left(\frac{s}{n}\right)\right)}{s/n} s \rightarrow -p'_{i,i}(0)s = q_i s.$$

Этим нестрогим рассуждением мы показали, что  $\tau$  имеет экспоненциальное распределение, поэтому цепь Маркова обязана сидеть в любом состоянии экспоненциальное время, а потом из него выходить.

## Матрица переходных интенсивностей

**Определение 7.1.** Рассмотрим  $q_{i,j} = \lim_{t \rightarrow 0} \frac{p_{i,j}(t)}{t}$ , если  $i \neq j$ ;  $q_i = \lim_{t \rightarrow 0} \frac{1 - p_{i,i}(t)}{t}$ ;  $q_{i,i} = -q_i$ . Иначе говоря,  $q_{i,j} = p'_{i,j}(0)$ . Матрица  $Q = (q_{i,j})$  называется **матрицей переходных интенсивностей**.

Посмотрим на конструктивное представление цепи Маркова в терминах матрицы  $Q$ . Оказывается, что справедливо такое представление:

**Утверждение 7.1.** Пусть  $Q$  — матрица переходных интенсивностей. Тогда можно рассмотреть следующий процесс:  $X_0 = i$  — разыгрывается с некоторым распределением. Тогда сгенерируем  $T_{0,j} \sim \exp(q_{i,j})$ ,  $j \neq i$  — независимые экспоненциальные случайные величины. Возьмём  $T_0 = \min(T_{0,j})$ ,  $i_1 = \operatorname{argmin} T_{0,j}$ . Теперь сгенерируем  $T_{1,j} \sim \exp(q_{i_1,j})$ ,  $j \neq i_1$  и посчитаем  $T_1 = \min T_{1,j}$ ,  $i_2 = \operatorname{argmin} T_{1,j}$ . Будем повторять эту операцию много раз.

Рассмотрим процесс

$$\tilde{X}_t = \begin{cases} i, & t < T_0; \\ i_1, & T_0 < t < T_0 + T_1; \\ i_2, & T_0 + T_1 < t < T_0 + T_1 + T_2; \\ \vdots & \end{cases}$$

Процесс устроен следующим образом: взяли начальное состояние  $i$  и сию же минуту в нём время  $T_0$ . По сути мы разыграли много экспоненциальных времён с параметрами  $q_{i,j}$  для всех  $j \neq i$ , и которое из них раньше наступило, то и победило — цепь переходит в это состояние. В этом состоянии запускаем новые экспоненциальные времена, уже с  $q$ , взятой из строки, в которой мы оказались. Кто из них победил, туда и переходим. И так далее (см. рис. 7.2).

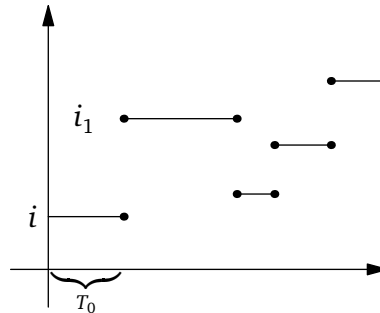


Рис. 7.2: Устройство процесса  $\tilde{X}_t$

Тезис в том, что  $\tilde{X}_t \stackrel{d}{=} X_n$  как процесс. То есть на каком-то вероятностном пространстве можно задать цепь Маркова таким конструктивным образом. По сути у нас идёт постоянное соревнование: мы из состояния разыграли кучу экспоненциальных величин, выбрали среди них победителя и перешли в него и так каждый раз.

Минимум экспоненциальных распределений — экспоненциальное распределение:

$$P(\min T_{0,j} > x) = P(T_{0,1} > x) \dots P(T_{0,n} > x) \dots = e^{-q_{i,1}x} \dots e^{-q_{i,n}x} \dots = e^{-\sum_{j \neq i} q_{i,j}x}$$

Но  $q_{i,i} = -\sum q_{i,j}$ , так как сумма в строке матрицы  $Q$  равна нулю, потому что это производная матрицы  $\mathbb{P}$ , а в строке матрицы  $\mathbb{P}$  сумма была равна 1. Тем самым мы получили, что минимум экспоненциальных величин экспоненциален с параметром  $q_i$ .

Нужно научиться из задачи получать матрицу  $Q$ .

**Пример** (Простое случайное блуждание с непрерывным временем). В момент  $t$  за время  $\Delta t \rightarrow 0$  мы выйдем из  $i$  с вероятностью  $\lambda \Delta t + o(\Delta t)$ . Если выходим, то с вероятностью  $p$  идём в  $i + 1$ ; с вероятностью  $1 - p$  идём в  $i - 1$ . Распишем сначала  $p_{i,i+2}(\Delta t)$ . Пусть  $\tau_1$  — время, прошедшее до первого перехода,  $\tau_2$  — после первого до второго перехода. Тогда

$$\begin{aligned} p_{i,i+2}(\Delta t) &\leq P(\tau_1 + \tau_2 \leq \Delta t) = \int_0^{\Delta t} P(\tau_2 \leq \Delta t - s) f_{\tau_1}(s) ds = \\ &= \int_0^{\Delta t} (1 - \lambda(\Delta t - s) + o(\Delta t - s)) \lambda e^{-\lambda s} ds = \int_0^{\Delta t} (1 - e^{-\lambda(\Delta t - s)}) \lambda e^{-\lambda s} ds = \\ &= (1 - e^{-\lambda \Delta t}) - \lambda \Delta t e^{-\lambda \Delta t} = \lambda \Delta t + o(\Delta t) - \lambda \Delta t (1 + o(1)) = o(\Delta t). \end{aligned}$$

Отсюда  $q_{i,i+2} = 0$ . Аналогично  $q_{i,j} = 0$ , если  $|i - j| > 1$ . Теперь будем искать  $p_{i,i+1}(\Delta t)$ :

$$\begin{aligned} p_{i,i+1}(\Delta t) &= P(\tau_1 < \Delta t, X_{\tau_1} = i + 1, \tau_1 + \tau_2 > \Delta t | X_0 = i) + \\ &+ \underbrace{P(\tau_1 + \tau_2 < \Delta t, X_{\Delta t} = i + 1 | X_0 = i)}_{=o(\Delta t)} = o(\Delta t) + p \int_0^{\Delta t} \underbrace{P(\tau_2 > \Delta t - s)}_{=e^{-\lambda(\Delta t - s)}} \underbrace{f_{\tau_1}(s)}_{=\lambda e^{-\lambda s}} ds = \\ &= o(\Delta t) + p \lambda \Delta t e^{-\lambda \Delta t}. \end{aligned}$$

Значит,  $q_{i,i+1} = p\lambda$ . Аналогично,  $q_{i,i-1} = (1 - p)\lambda$ . Тогда  $q_{i,i} = -\sum_{j \neq i} q_{i,j} = -\lambda$ . Мы получили общий вид матрицы  $Q$ :

$$Q = \begin{pmatrix} \ddots & & & & & & \\ & \ddots & & & & & \\ & & \lambda(1-p) & -\lambda & & & \\ & & & \lambda(1-p) & -\lambda & & \\ & & & & \lambda p & & \\ & & & & & \ddots & \\ 0 & & & & & & \ddots \end{pmatrix}$$

**Пример** (Ветвящиеся процессы с непрерывным временем). Есть  $Z_t$  — количество частиц в данный момент;  $\lambda > 0$  — параметр;  $\{p_i\}$  — набор вероятностей, описывающий сколько потомков даёт частица:  $p_i$  — вероятность дать  $i$  потомков,  $p_1 \neq 0$ . Каждая частица размножается спустя экспоненциальное  $\lambda$  время. Будем считать, что частица уходит из процесса после размножения (если хотим, чтобы частица оставалась, можем увеличить количество

потомков каждой частицы на 1). По аналогии с предыдущими выкладками:

$$P(\tau_1 + \tau_2 < \Delta t | X_0 = 1) = \sum_{j=i-1}^{\infty} \int_0^{\Delta t} \underbrace{f_{\tau_1}(s)}_{=\lambda i e^{-\lambda i s}} p_{j-i+1} \underbrace{P(\tau_2 < \Delta t - s | X_0 = j)}_{=1 - e^{-\lambda j(\Delta t - s)}} ds = o(\Delta t)$$

Последний переход не расписан, необходимо подставить выражения и проинтегрировать. У нас есть  $i$  частиц, у каждой своё экспоненциальное  $\lambda$ . Какое из них раньше всех наступит и соответствует вероятности перехода. Интенсивность перехода (показатель экспоненциального времени, спустя которое я выйду из состояния  $i$ ):  $q_i = \lambda i$ .

Теперь выпишем  $p_{i,j}(\Delta t)$ . Если  $j < i - 1$ , то это  $o(\Delta t)$ , поэтому  $q_{i,j} = 0$  если  $j < i - 1$ . Если же  $j \geq i - 1$ ,  $j \neq i$ , то

$$\begin{aligned} p_{i,j}(\Delta t) &= P(\tau_1 < \Delta t, \tau_2 + \tau_1 > \Delta t, X_{\tau_1} = j | X_0 = i) + o(\Delta t) = \\ &= \int_0^{\Delta t} \underbrace{f_{\tau_1}(s)}_{=i\lambda e^{-\lambda i s}} \underbrace{P(\tau_2 > \Delta t - s)}_{=e^{-\lambda j(\Delta t - s)}} p_{j-i+1} ds = p_{j-i+1} \int_0^{\Delta t} i\lambda e^{-\lambda i s - \lambda j(\Delta t - s)} ds = \\ &= p_{j-i+1} \underbrace{e^{-\lambda j \Delta t}}_{=1+o(1)} i\lambda \int_0^{\Delta t} e^{-\lambda(i-j)s} ds. \end{aligned}$$

Учитывая, что

$$\int_0^{\Delta t} e^{-\lambda(i-j)s} ds = \frac{1 - e^{-\lambda(i-j)\Delta t}}{\lambda(i-j)} = \frac{\lambda(i-j)\Delta t + o(\Delta t)}{\lambda(i-j)},$$

найдём

$$q_{i,j} = \lim_{\Delta t \rightarrow 0} \frac{p_{i,j}}{\Delta t} = i\lambda p_{j-i+1}, \quad j \geq i - 1, \quad j \neq i.$$

И, наконец,  $q_{i,i} = -\sum_{j \neq i} q_{ij} = -i\lambda$ . Итого матрица  $Q$  выглядит следующим образом:

$$Q = \begin{pmatrix} 0 & 0 & \dots & \dots & \dots & 0 & \dots \\ \lambda p_0 & -\lambda & \lambda p_2 & \dots & \dots & \lambda p_n & \dots \\ 0 & 2\lambda p_0 & -2\lambda & 2\lambda p_2 & \dots & \vdots & \\ & & 3\lambda p_0 & -3\lambda & 3\lambda p_2 & \dots & \\ 0 & & & \ddots & \ddots & \ddots & \end{pmatrix}$$

В примере возникла специфика, когда интенсивность зависит от состояния — чем больше частиц, тем быстрее наступают новые события.

**Утверждение 7.2.** Введём матрицу  $\tilde{P} = (\tilde{p}_{i,j})$ , где  $\tilde{p}_{i,j} = \frac{q_{i,j}}{q_i}$ ,  $j \neq i$ ;  $\tilde{p}_{i,i} = 0$ .

Получается, что матрица  $\tilde{P}$  устроена следующим образом: на диагонали находятся нули, вне диагонали какие-то неотрицательные числа, причём сумма в каждой строчке равна 1. Значит, это матрица вероятностей перехода какой-то цепи Маркова. Такая цепь Маркова называется **вложенной цепью Маркова**. Оказывается, что эта цепь — цепь Маркова с непрерывным временем. То есть цепь Маркова с непрерывным временем — какая-то цепь Маркова с дискретным временем, только теперь мы совершаем переходы через экспоненциальное, а не константное время. В сущности в этих цепях мы видим одно и то же, только в разные моменты времени.

Как найти  $P(X_t = i)$ ? Как найти  $\lim_{t \rightarrow \infty} P(X_t = i)$ ?

**Теорема 7.2 (уравнения Колмогорова).**  $P'(t) = P(t)Q = QP(t)$  — **прямое и обратное уравнения Колмогорова.**

Они выводятся из уравнений Колмогорова–Чепмена. Из этих дифференциальных уравнений можем найти  $P(t)$ , зная  $Q$ :

$$P(t) = e^{Qt}.$$

**Теорема 7.3 (эргодическая для марковских процессов).** Пусть  $X(t)$  — неразложимая цепь и существует решение уравнения  $\pi Q = 0$ , где  $\pi$  — вероятностный вектор, то есть  $\pi: \sum \pi_i = 1$ ,  $\pi_i \geq 0$ . Тогда  $p_{i,j}(t) \rightarrow \pi_j$  при  $t \rightarrow \infty$ .

Уравнение  $0 = Q\pi$  выполнено всегда.

Обратите внимание, что вложенная цепь описывает как цепь переходит, но не запоминает времена, в которые переход происходит. У них разные эргодические распределения. Если цепь в каком-то состоянии находится долго, шанс, что в момент времени  $t$  мы застанем её там будет большой, даже если цепь туда редко заходит с точки зрения вложенной цепи, но зато долго там сижу. Возьмём, например, цепь с тремя состояниями, которая равновероятно переходит в два остальных или остаётся. Тогда вне зависимости от интенсивности, стационарное распределение сложенной цепи Маркова будет  $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ . А у настоящей цепи распределение будет совсем другое, будет зависеть от того, сколько времени именно цепь находится в каждом состоянии.

Единственная принципиальная разница вложенной цепи и настоящей цепи

в том, что  $\lambda$  окажутся очень быстрорастущими, может оказаться так, что цепь взорвётся.

**Пример.** Возьмём цепь с

$$Q = \begin{pmatrix} -\lambda & \lambda & & & & & 0 \\ 0 & -2\lambda & 2\lambda & & & & \\ 0 & 0 & -4\lambda & 4\lambda & & & \\ & & & \ddots & \ddots & & \\ & & & & -2^n\lambda & 2^n\lambda & \\ 0 & & & & & \ddots & \ddots \end{pmatrix}.$$

Рассмотрим  $\tau_1 + \tau_2 + \dots + \tau_n$  — время до  $n$ -ого скачка.

$$\tau_1 \sim \exp(\lambda), \tau_2 \sim \frac{\exp(\lambda)}{2}, \dots, \tau_n \sim \frac{\exp(\lambda)}{2^n}$$

Из мартингальных соображений можно показать, что этот ряд сходится почти наверное. Поэтому с вероятностью 1 в какой-то момент наша цепь убежит в никуда.

Из-за того, что интенсивность сокращается, цепь за конечное время убегает в бесконечность, причём это время при каждом  $\omega$  наступит в своё случайное время, к которому сойдётся ряд из непрерывных случайных величин. Важно, что этот ряд почти наверное сходится, поэтому цепь будет взрываться. Что происходит после момента взрыва, не описывается цепью, потому что непонятно где она.



## Семинар 8. Пуассоновский процесс

### Определения

**Определение 8.1.** Пуассоновский процесс — процесс  $N_t$  со следующими свойствами:

- 1)  $N_0 = 0$  почти наверное;
- 2)  $N_t - N_s \sim \text{Poiss}(\lambda(t-s))$ ,  $\lambda > 0$ ,  $t > s$ ;
- 3) процесс с независимыми приращениями;
- 4) траектории непрерывны справа.

Поскольку приращения пуассоновские, процесс непрерывный справа, то, очевидно, траектория кусочно постоянна, монотонно неубывает (рис. 8.1). Если бы мы не потребовали условие 4, то можно было бы к процессу добавлять шумы вверх-вниз, которые выбрасывали бы его из этих точек вверх, а затем возвращали.

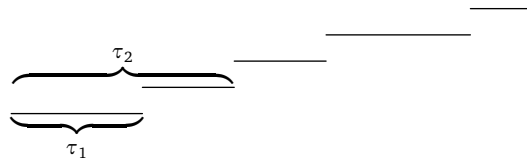


Рис. 8.1: Общий вид пуассоновского процесса

Построим явно такую модификацию, чтобы показать, что она существует.

Сначала определим, какова вероятность того, что  $\tau_1 > t$ . Первый скачок случится после момента  $t$  тогда и только тогда, когда в момент  $t$  процесс всё ещё в 0, поскольку траектория монотонна. Когда  $\tau_1$  наступит,  $N_t$  будет как минимум 1.

$$P(\tau_1 > t) = P(N_t = 0) = e^{-\lambda t}, \quad t > 0.$$

Поэтому у  $\tau_1$  экспоненциальное распределение.

Точно также можно посмотреть на вероятность

$$\begin{aligned} P(\tau_1 \in [t, t+d), \tau_2 \in [t+s, t+s+d)) &= \\ &= P(N_t = 0, N_{t+d} - N_t = 1, N_{t+s} - N_{t+d} = 0, N_{t+s+d} - N_{t+s} \geq 1) \stackrel{(a)}{=} \\ &\stackrel{(a)}{=} e^{-\lambda t} \lambda d e^{-\lambda d} e^{-\lambda(s-d)} (1 - e^{-\lambda d}). \end{aligned}$$

При переходе (а) мы пользовались независимостью приращений.

Отсюда можем найти плотность  $\tau_1$  и  $\tau_2$  (это предел при  $d \rightarrow 0$  выражения выше, делённого на  $d^2$ ):

$$f_{\tau_1, \tau_2}(t, t+s) = e^{-\lambda t} e^{-\lambda s} \lambda^2, \quad t, s > 0$$

Нетрудно видеть, что совместная плотность двух случайных величин: первая — экспоненциальная с параметром  $\lambda$ , а вторая — сумма её и еще одной экспоненциальной с параметром  $\lambda$ :

$$f_{\tau_1, \tau_2 - \tau_1}(t, s) = \lambda e^{-\lambda t} \lambda e^{-\lambda s}$$

Поэтому вместе  $\tau_1$  и  $\tau_2$  — две независимые экспоненциальные величины. Это верно для любого количества различных  $\tau$ . Промежутки между скачками пуассоновского процесса оказались экспоненциальными, а сами скачки при этом оказались ровно на 1. Если бы они были больше, плотности бы не получились. Раз процесс через экспоненциальные времена переходит на 1 вперед, можем дать новое определение:

**Определение 8.2.**  $N_t$  — пуассоновский процесс, если  $N_t = \sum_{k=1}^{\infty} I_{S_k \leq t}$ , где  $S_k = \sum_{i=1}^k T_i$ , а  $T_i \sim \exp(\lambda)$  — независимые.

Процесс из изначального определения говорил нам, сколько скачков случится к моменту  $t$ , теперь процесс показывает как устроены промежутки между точками скачков.

**Определение 8.3.**  $N_t$  — пуассоновский процесс, если  $N_t$  — марковский процесс с матрицей переходных интенсивностей  $Q = \begin{pmatrix} -\lambda & \lambda & 0 & \\ & -\lambda & \lambda & \\ 0 & & \ddots & \ddots \end{pmatrix}$ .

Заметим, что чтобы определить пуассоновский процесс, достаточно определить пуассоновский поток:

**Определение 8.4.** Пуассоновский (простейший) поток — набор точек  $\tau_1, \tau_2, \dots, \tau_k, \dots$ , в которых происходит подпрыгивание пуассоновского процесса.

**Определение 8.5.**  $\tau_1, \dots, \tau_k, \dots$  — пуассоновский поток, если

- 1) (ординарность).  $P(\exists i, j: i \neq j, \tau_i \in [t, t+s), \tau_j \in [t, t+s)) = o(s), s \rightarrow 0$ .
- 2) (стационарность).  $P(\exists k \text{ точек } \tau_i \text{ в } [t, t+s)) = P(\exists k \text{ точек } \tau_i \text{ в } [0, s))$ .

3) (отсутствие последствия). Число точек потока в полуинтервалах  $[0, t_1), [t_2, t_3), \dots, [t_{2k}, t_{2k+1})$  независимы при  $0 < t_1 < \dots < t_{2k+1}$ .

В последнем определении, в отличие от всех предыдущих, нет никаких распределений. Тут есть три простых свойства, которые могут возникнуть в реальных задачах из общих соображений.

**Пример.** Если мы стоим на берегу реки, где идёт золотоносная порода, можем считать, что частички плохо слипаются, их размывает по-отдельности (ординарность); река течёт довольно давно, вероятность не зависит от того, когда мы пришли к реке (стационарность); золото распределено по потоку воды и вероятности встретить его в разные промежутки времени независимы (отсутствие последствия). Поэтому сразу можем сделать вывод, что поток золотых песчинок будет пуассоновским. Эксперимент подтверждает, что число частиц, попавшихся в выделенной области реки, распределено экспоненциально с некоторым параметром  $\lambda$  (рис. 8.2).

Число частиц	Число наблюдавшихся случаев	Частота $\frac{m}{518}$	$\frac{\lambda^n e^{-\lambda}}{n!}$	Вычисленное число случаев
0	112	0,216	0,213	110
1	168	0,325	0,328	173
2	130	0,251	0,253	131
3	69	0,133	0,130	67
4	32	0,062	0,050	26
5	5	0,010	0,016	8
6	1	0,002	0,004	2
7	1	0,002	0,001	1

Рис. 8.2: Теоритическая и практическая вероятности попадания в область реки золотоносных частиц ( $\lambda$  взята равной выборочному среднему)

## Обобщенный пуассоновский поток

Заметим, что совершенно не обязательно пуассоновский поток рассматривать на прямой, его можно рассматривать где угодно. Пусть у нас есть  $\mathbb{R}^d$ . Тогда мы можем определить облако точек, обладающих теми же свойствами:

- 1) (*ординарность*). Вероятность того, что в области  $A \subset \mathbb{R}^n$  больше одной точки — это  $o(V(A))$  при  $V(A) \rightarrow 0$ .
- 2) (*стационарность*). Вероятность, что в области объёма  $V$   $k$  точек не зависит от формы.
- 3) (*отсутствие последствия*). Количество точек в непересекающихся областях независимы.

**Определение 8.6.** Поток, обладающий свойствами 1-3, называется **простейшим потоком в пространстве** или **обобщённым пуассоновским потоком**.

Можно показать, что для обобщённого потока

$$P(\text{в измеримой области } A \text{ } n \text{ точек}) = \frac{(\lambda V(A))^n e^{-\lambda V(A)}}{n!}, \quad V(A) \text{ — объём } A.$$

**Определение 8.7.** Если отказаться от условия 2, получим **неоднородный поток**. Определим для него **интенсивность** в каждой точке:

$$\lambda(x) = \lim \frac{P(\text{в окрестности } x \text{ есть точка})}{V(\text{окрестности})}.$$

Тогда получится похожая формула

$$P(\text{в } A \text{ } n \text{ точек потока}) = \frac{\Lambda(A)^n e^{-\Lambda(A)}}{n!}, \quad \text{где } \Lambda(A) = \int_A \lambda(x) dx.$$

Иначе говоря, число точек неоднородного потока в области пуассоновское с параметром  $\Lambda(A)$ , где  $\Lambda(A)$  — некоторая мера.

## Теоремы о пуассоновском потоке

**Теорема 8.1 (сложения).** Если есть  $k$  независимых однородных потоков с интенсивностями  $\lambda_i$ , то их объединение является потоком с интенсивностью  $\lambda_1 + \dots + \lambda_k$ .

**Теорема 8.2 (о раскраске).** Если есть пуассоновский поток с параметром  $\lambda$ . Каждая точка потока с вероятностью  $p_1$  красится в первый цвет, с вероятностью  $p_2$  — во второй цвет, ..., с вероятностью  $p_k$  — в  $k$ -ый цвет. Тогда

- 1) Каждый одноцветный набор точек — поток с параметром  $\lambda p_i$ .

2) Эти потоки независимы.

В частности, простым частным случаем этого утверждения является то, что если взять полиномиальную схему из пуассоновского количества величин, то полученное количество шаров будет независимыми и пуассоновскими.

Эти теоремы удобно применять для разных фактов, чтобы изучать несколько потоков вместе или чтобы делить части потока на разные куски.

**Теорема 8.3 (условное свойство).** Пусть есть пуассоновский поток и известно, что в множестве  $A$  положительной меры  $k$  точек потока. Тогда точки потока независимы и равномерно распределены по  $A$ .

В частности, на прямой возьмём отрезок  $[0, t]$  и посчитаем сколько точек туда попало:  $N_t = k$ . Тогда окажется, что моменты  $\tau_1, \tau_2, \dots, \tau_k$  распределены так же, как вариационный ряд равномерного распределения на  $[0, t]$ :

$$(\tau_1, \dots, \tau_k \mid N_t = k) \stackrel{d}{=} (X_{(1)}, \dots, X_{(k)}), X_{(i)} \sim R[0, t].$$

Это свойство позволяет утверждать, что в каком-то смысле пуассоновский поток — очень удобный механизм, чтобы приходить куда-то с инспекцией. Допустим, мы приходим с инспекцией через экспоненциальные времена с параметром  $\lambda$ , тогда сколько раз за год мы приходим с инспекцией случайное, а моменты прихода равномерно распределены на отрезке  $[0, t]$ . При этом за полгода мы всё равно приходили равномерно. Случайное число раз, но моменты прихода были равномерные. Получается, что это такое распределение, которое одновременно равномерно на всех отрезках.

## Эксцесс пуассоновского потока

**Определение 8.8.** Рассмотрим пуассоновский процесс. Зафиксируем точку  $t$  и посмотрим, через какое время случится очередной скачок. Если  $S_{N_t}$  — последняя точка потока, которая была до момента  $t$ , то мы смотрим на случайную величину  $\gamma_t = S_{N_{t+1}} - t$  — **эксцесс** процесса.

В силу того, что процесс марковский, ожидать скачка мы будем всё тоже экспоненциальное с параметром  $\lambda$  время. Это называется **парадоксом времени ожидания**.

**Пример.** Возьмём поток автобусов. Он неплохо аппроксимируется пуассоновским потоком, если мы находимся далеко от конечной остановки, где автобусы ходят по графику. Зафиксируем точку  $t$  прихода на остановку и будем считать время до прихода ближайшего автобуса — эксцесс потока.

В силу вышесказанного, он будет иметь тоже распределение, что и интервал между автобусами, поэтому ждать автобус с интервалом движения 10 минут мы будем в среднем 10 минут при приходе в случайное время. Это объясняется тем, что вероятность попасть в большие промежутки больше, чем в маленькие при равномерном времени прихода. Это можно увидеть, например, когда приходит несколько автобусов подряд, а потом большой интервал до следующего автобуса. Вероятность прийти, чтобы следующий автобус был последним в «паровозике» очень мала, так как с ухода предыдущего прошло мало времени, гораздо больше шансов прийти на удлинённый интервал и ждать автобус большее время. Для больших  $t$  время от последнего ушедшего автобуса до нашего прихода на остановку  $\eta_t = t - S_{N_t}$  также распределено экспоненциально с параметром  $\lambda$ , причём оно не зависит от временем ожидания следующего автобуса (большие  $t$  нужны, так как  $\eta_t$  ограничена сверху  $t$ , чтобы распределение не обрезалось).

## Семинар 9. Процессы восстановления

### Процесс восстановления

Представим себе следующую задачу: у нас есть лампочка, она работает в случайные времена  $T_i \geq 0$ , а затем сгорает. Иногда бывает удобно считать, что первую лампочку я купил новую, а остальные достаю из запасов, поэтому первое распределение может отличаться от остальных.

**Определение 9.1.**  $T_i$  — независимые одинаково распределённые (допускается, чтобы  $T_1$  был распределён по-другому, но также независимо). Рассмотрим момент  $t$ , тогда **процесс восстановления**  $N_t$ :

$$N_t = \max\{k : S_k = T_1 + \dots + T_k \leq t\}.$$

Процесс восстановления, у которого все величины распределены одинаково, называется **чистым процессом восстановления**.

Физически процесс показывает сколько раз сгорела лампочка к моменту  $t$ .

**Пример.** Частный случай процесса восстановления, построенный по экспоненциальным временам  $T_i \sim \exp(\lambda)$  — однородный пуассоновский процесс. Отметим, что единственный марковский процесс восстановления — пуассоновский процесс.

### Функция восстановления

Введём важную характеристику процесса восстановления — функцию восстановления:

**Определение 9.2.** **Функция восстановления** —  $H(t) = \mathbb{E}N_t$ .

Если расписать математическое ожидание, то

$$H(t) = \sum_{k=1}^{\infty} k \mathbb{P}(N_t = k) = \sum_{k=1}^{\infty} \mathbb{P}(N_t \geq k).$$

В терминах  $S_t$  функция восстановления представима в виде

$$H(t) = \sum_{k=1}^{\infty} \mathbb{P}(S_k \leq t) = \sum_{k=1}^{\infty} F^{*k}(t),$$

где  $F^{*k}$  —  $k$ -ая свёрточная степень функции распределения  $F$ , то есть свёртка  $F$  сама с собой  $k$  раз:  $F^{*k} = \underbrace{F * F * \dots * F}_{k \text{ раз}}$ .

**Пример.** Для пуассоновского процесса  $N_t \sim Poiss(\lambda t)$ , тогда функция восстановления  $H(t) = EN_t = \lambda t$  растёт линейно. Это соответствует интуитивному здравому смыслу: если восстановления происходят с математическим ожиданием  $a$ , то до момента  $t$  должно случиться примерно  $\frac{t}{a}$  восстановлений, что соответствует нашему  $H(t)$ . К сожалению, других чистых процессов с  $EN_t = \frac{t}{a}$ , где  $a = ET_1$ , кроме пуассоновского, нет. Если процесс не является чистым, всегда можно подобрать первый шаг по остальным, чтобы эта формула выполнялась.

## Теорема восстановления

**Определение 9.3.**  $X$  — решётчатая случайная величина, если

$$\exists a, d: P(X \in a + d\mathbb{Z}) = 1.$$

Максимальное такое  $d$  называется *шагом решётки*.

**Теорема 9.1 (восстановления).** Пусть  $ET_1 = a$  — конечно. Тогда

- 1)  $\frac{H(t)}{t} \rightarrow \frac{1}{a}, t \rightarrow +\infty$ ;
- 2) если  $T_i$  — нерешётчатые, то  $H(t) - H(t-u) \rightarrow \frac{u}{a}, t \rightarrow +\infty, u$  фиксировано;
- 3) если  $T_i$  — решётчатые с шагом  $d$ , то  $H(t) - H(t-u) \rightarrow \frac{u}{a}, \frac{u}{d} \in \mathbb{N}, t \rightarrow +\infty$ .

Грубо говоря, с течением времени первый шаг забывается и поэтому можно считать, что процесс становится похож на однородный, у которого первый шаг подобран так, чтобы было  $EN_t = \frac{t}{a}$  даже если первый шаг не подобран специальным образом.

Можно убрать математическое ожидание и сформулировать закон больших чисел и центральную предельную теорему для процесса восстановления.

**Теорема 9.2 (ЗБЧ для процесса восстановления).**

Пусть  $ET_1 = a$ . Тогда  $\frac{N_t}{t} \rightarrow \frac{1}{a}$  почти наверное.

**Теорема 9.3 (ЦПТ для процесса восстановления).**

Пусть  $ET_1 = a, DT_1 = \sigma^2 > 0$ . Тогда

$$\frac{N_t - \frac{t}{a}}{\sqrt{t}} \xrightarrow{d} Z \sim \mathcal{N}\left(0, \frac{\sigma^2}{a^3}\right).$$



Эти теоремы берутся из простого соотношения  $\{N_t \geq k\} = \{S_k \leq t\}$ .

## Уравнение восстановления

**Определение 9.4.** Уравнением восстановления называют конструкцию  $Z(t) = z_0(t) + Z * F(t)$ , где  $z_0$  и  $F$  — заданные функции, а  $Z$  неизвестна.

**Пример.**  $X_t$  — дискретная цепь Маркова. Хотим найти  $P(X_t = k)$ . Для удобства будем считать, что  $X_0 = k$ . Воспользуемся марковским свойством:

$$P(X_t = k) = I_{t=0} + \sum_{j=1}^t \underbrace{P(X_i \neq k, X_j = k, i \leq j)}_{\text{вышли не возвращаясь в } k} \underbrace{P(X_{t-j} = k)}_{\text{могли возвращаться}}$$

Так как в дискретном случае формула свёртки выглядит так:

$$a * u(k) = \sum a(k-j)(u(j) - u(j-1)),$$

то наше уравнение — это уравнение восстановления с  $z_0(t) = I_{t=0}$ ,  $Z(t) = P(X_t = k)$ ,  $F(x) = P(\tau \leq x)$ , где  $\tau$  — время первого прихода в  $k$ .

**Пример.** Рассмотрим  $\gamma_t = S_{N_t+1} - t$  — эксцесс процесса восстановления — время до ближайшей сгоревшей лампочки, если мы начали наблюдение в момент  $t$ . На функцию распределения эксцесса можно также составить уравнение восстановления:

$$P(\gamma_t > x) = P(T_1 > t + x) + \int_0^t f_{T_1}(u) P(\gamma_{t-u} > x).$$

Здесь  $z_0(t) = P(T_1 > t + x)$ ,  $F = F_{T_1}$ ,  $Z(t) = P(\gamma_t > x)$

Чтобы решить уравнение восстановления, нужно свернуть его с  $F$   $k$  раз:

$$Z * F^{*k} = z_0 * F^{*k} + Z * F^{*(k+1)}$$

и сложим по всем  $k$ :

$$\begin{aligned} Z * \sum_{k=0}^{\infty} F^{*k} &= z_0 * \sum_{k=0}^{\infty} F^{*k} + Z * \sum_{k=1}^{\infty} F^{*k} \implies \\ \implies Z &= z_0 * (I_{x>0} + H(x)) = z_0 + z_0 * H \end{aligned}$$

**Следствие 9.1** (к теореме восстановления). Пусть  $g$  — функция ограниченной вариации, тогда

$$\int_0^t g(t-u) dH(u) \xrightarrow{t \rightarrow \infty} \frac{1}{a} \int_0^{\infty} g(u) du.$$

Это следствие справедливо в нерешётчатом случае и в решётчатом, если  $t$  кратна шагу решётки  $\delta$ . Используя это свойство, можем расписать свёртку, фигурирующую в решении уравнения восстановления:

$$z_0 * H(x) = \int_0^x z_0(x-u) dH(u) \xrightarrow{x \rightarrow \infty} \frac{1}{a} \int_0^{\infty} z_0(u) du$$

**Пример.** Вернёмся к примеру с эксцессом.

$$Z(t) = P(\gamma_t > x) = z_0 + F * Z,$$

где  $z_0 = P(T_1 > t + x)$ . В силу решения уравнения восстановления

$$Z(t) = z_0(t) + z_0 * H(t).$$

При переходе к пределу в уравнении, получим

$$Z(t) \rightarrow \frac{1}{E T_1} \int_0^{\infty} P(T_1 > u + x) du = \frac{1}{E T_1} \int_x^{\infty} P(T_1 > u) du, \quad t \rightarrow \infty.$$

Вот такое получилось предельное распределение эксцесса. Можно увидеть, что для пуассоновского процесса распределение будет просто экспоненциальным.

**Пример.** В случае с цепью Маркова, ранее мы получали

$$Z(t) = P(X_t = k); \quad F(t) = P(\tau_1 < t); \quad z_0(t) = I_{t=0};$$

$$Z(t) = z_0(t) + Z * F(t).$$

В силу решения уравнения восстановления

$$Z(t) = I_{t=0} + I * H(t) = H(t) - H(t-1) \rightarrow \frac{1}{a}, \quad t > 0.$$

Получили эргодическую теорему для цепей Маркова в одной из её формулировок. Это работает, если время возвращения в себя конечно (возвратность цепи) и шаг решётки 1 (непериодичность цепи). Существенность в этой формулировке теоремы не важна.

Получается очень удобная конструкция, которая позволяет решать множество разных задач, где возникает регенерационная структура при возвращении в какое-то состояние спустя случайные промежутки времени. Теорема помогает учитывать, что эти промежутки случайные, что часто встречается в реальных задачах.

## Семинар 10. Гауссовские процессы

### Броуновское движение и броуновский мост

Начнём с истории этого процесса. В 1827 году Броун открыл *броуновское движение* — процесс хаотического движения молекулы под действием постоянных соударений с более мелкими частицами. Мы будем смотреть на движение вдоль прямой, хотя можно определить броуновское движение и в пространстве. Это движение является *марковским процессом* — если частица в какой-то момент где-то находится, то где она окажется в следующий момент не зависит от предыстории. Движение частицы происходит только под воздействием случайных соударений в каждый момент времени. Также процесс должен быть *однородным*, потому что мы предполагаем, что ничего во времени не меняется. Переходная плотность процесса за время  $t$  в точке  $x_0$  —  $p(x, t|x_0)$ . Из физических соображений получается *уравнение Фика* (*уравнение диффузии*):

$$\frac{\partial p}{\partial t} = D \frac{\partial^2 p}{\partial x^2},$$

которому подчиняется броуновское движение. Если взять  $D = \frac{1}{2}$  (это вопрос того, в каких единицах измеряется время и расстояние), то единственное решение этого уравнения — плотность нормального распределения:

$$p = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-x_0)^2}{2t}}.$$

**Определение 10.1.** Процесс  $W_t$  называется **броуновским движением** или **винеровским процессом**, если

- 1)  $W_0 = 0$ ;
- 2)  $W_t - W_s \sim \mathcal{N}(0, t-s)$ ,  $t > s$ ;
- 3)  $W_t$  — процесс с независимыми приращениями;
- 4)  $W_t$  имеет непрерывные траектории.

С другой стороны, процесс можно задать непосредственно его конечномерными распределениями.

**Определение 10.2.** Процесс  $W_t$  называется **броуновским движением** или **винеровским процессом**, если

- 1)  $W_0 = 0$ ;

- 2)  $W_t$  — гауссовский процесс;
- 3)  $EW_t = 0$ ,  $\text{cov}(W_t, W_s) = \text{cov}(W_s + W_t - W_s, W_s) = \min(t, s)$ ;
- 4)  $W_t$  имеет непрерывные траектории.

Рассмотрим ещё одно представление броуновского движения, связанное с Винером: пусть  $\xi_i \sim \mathcal{N}(0, 1)$  — независимые одинаково распределённые случайные величины. Рассмотрим ряд  $\sum_{n=1}^{\infty} \xi_n \frac{\sin(\pi n t)}{n}$  на отрезке  $t \in [0, 1]$ . Он сходится, то есть его частичные суммы  $S_n = \sum_{n=1}^m \xi_n \frac{\sin(\pi n t)}{n} \sim \mathcal{N}(0, \sigma_n^2)$  сходятся. Это сумма независимых величин с нулевым средним, то есть мартингал. Чтобы применить теорему о сходимости, нужно оценить

$$\sup E \left| \sum_{n=1}^m \xi_n \frac{\sin(\pi n t)}{n} \right| \leq \sup D \sum_{n=1}^m \xi_n \frac{\sin(\pi n t)}{n} = \sup \sum_{n=1}^m \underbrace{D \xi_n}_{=1} \frac{\sin(\pi n t)}{n} < +\infty.$$

$ES_m = 0 \rightarrow 0$ , значит, у предельной величины математическое ожидание нулевое.  $ES_m(t)S_m(s) = \sum_{n=1}^m D \xi_n \frac{\sin(\pi n t) \sin(\pi n s)}{n^2} = \min(t, s)$ . Часть преобразования этого выражения опущены. Получили, что  $S_n$  действительно является винеровским процессом. Это определение бывает удобным, потому что оно даёт явно как процесс задан.

Есть теорема Карунена–Лоэва, которая говорит, что любой процесс на компакте в довольно широких условиях можно представить такого рода конструкцией, что бывает очень удобно.

**Определение 10.3.** Процесс  $W_t^0 = W_t - tW_1$ ,  $t \in [0, 1]$  называют **броуновским мостом**. Это определение эквивалентно следующему:

- 1)  $W_0^0 = 0$ ;
- 2)  $W_t^0$  — гауссовский процесс;
- 3)  $\text{cov}(W_t^0, W_s^0) = \min(t, s) - ts$ ,  $t, s \in [0, 1]$ ;
- 4)  $W_t^0$  имеет непрерывные траектории.

## Свойства броуновского движения

Рассмотрим некоторые преобразования процесса, оставляющие его броуновским движением.

- 1) Если  $W_t$  — броуновское движение, то  $W_{t+s} - W_s$  — тоже, где  $s$  — фиксированный параметр,  $t > 0$ . Иначе говоря, если возьмём траекторию броуновского движения и начнём рассматривать её с момента  $s$  заново, то это будет выглядеть также, как броуновское движение. Доказывается по любому из определений напрямую. Например,

$$\text{cov}(W_{t+s} - W_s, W_{r+s} - W_s) = t + s - s - s + s = t, \quad t < r.$$

- 2) (*автомодельность*). Если  $W_t$  — броуновское движение, то  $\sqrt{c}W_{\frac{t}{c}}$  — тоже, где  $c > 0$ . Иначе говоря, процесс можно растянуть по оси времени в  $c$  раз, а по высоте в  $\sqrt{c}$  раз. Опять же, свойство проверяется непосредственно подстановкой в одно из определений.
- 3) Если  $W_t^{(1)}, W_t^{(2)}$  — два независимых броуновских движения, то процесс  $\frac{W_t^{(1)} + W_t^{(2)}}{\sqrt{2}}$  — тоже броуновское движение. Свойство просто проверяется через первое определение.
- 4) (*обратимость*).  $W_t$  — броуновское, значит,  $tW_{\frac{1}{t}}$ ,  $t > 0$ ;  $0, t = 0$  — тоже. Свойство очевидно для второго определения для всех пунктов, кроме непрерывности, которая следует из закона повторного логарифма, приведённого ниже.

Теперь рассмотрим свойства траектории броуновского движения.

- 1)  $W_t$  — мартингал.  $W_t^2 - t$ ,  $e^{W_t - \frac{t}{2}}$  — тоже.
- 2) Траектория броуновского движения почти наверное нигде не дифференцируема.
- 3) (*закон повторного логарифма*).

$$\overline{\lim}_{t \rightarrow \infty} \frac{W_t}{\sqrt{2t \ln \ln t}} = 1;$$

$$\underline{\lim}_{t \rightarrow \infty} \frac{W_t}{\sqrt{2t \ln \ln t}} = -1.$$

Иначе говоря, процесс  $\frac{W_t}{\sqrt{2t \ln \ln t}}$  болтается, но никогда не выходит из полосы  $[-1 - \varepsilon, 1 + \varepsilon]$  начиная с какого-то момента, при этом время от времени процесс возвращается как в  $[1 - \varepsilon, 1 + \varepsilon]$ , так и в  $[-1 - \varepsilon, -1 + \varepsilon]$ . Сам броуновский процесс будет ограничен графиком  $\sqrt{2t \ln \ln t}$  (см. рис. 10.1).

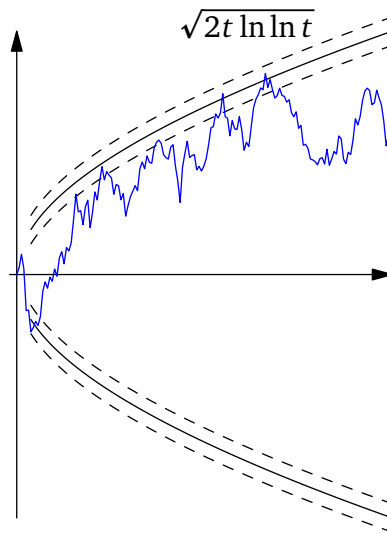


Рис. 10.1: Порядок роста траектории броуновского движения  $W_t$ . Пунктирной линией показана  $\varepsilon$ -окрестность графика  $\sqrt{2t \ln \ln t}$ , в которую траектория регулярно попадает.

- 4) Если подставить в закон повторного логарифма  $tW_{\frac{1}{t}}$  вместо  $W_t$  и ввести новую переменную  $s = \frac{1}{t}$ , то получим

$$\overline{\lim}_{s \rightarrow 0} \frac{W_s}{\sqrt{2s \ln \ln \frac{1}{s}}} = 1;$$

$$\underline{\lim}_{s \rightarrow 0} \frac{W_s}{\sqrt{2s \ln \ln \frac{1}{s}}} = -1.$$

В частности, это означает, что траектория броуновского движения бесконечно много раз пересекает ось  $x$  в окрестности нуля.

## Распределение максимума

Найдём  $P(\max_{t \in [0, s]} W_t > x)$ . Для этого будем использовать принцип отражения для простого симметричного случайного блуждания.

$$P\left(\max_{t \in [0, s]} W_t > x\right) = \underbrace{P\left(\max_{t \in [0, s]} W_t > x, W_s > x\right)}_{=1 - \Phi\left(\frac{x}{\sqrt{s}}\right) = P_1} + \underbrace{P\left(\max_{t \in [0, s]} W_t > x, W_s \leq x\right)}_{=P_2}$$

Найдём второе слагаемое (по условию  $\tau_x \in [0, s]$ ):

$$\begin{aligned}
 P_2 &= P(W_s \leq x, \tau_x \in dt) = \\
 &= \int_0^s P(W_r < x, r \leq t, \tau_x \in dt) \cdot \overbrace{P(W_s \leq x \mid W_t = x)}^{=E(I_{W_t \leq x} \mid \tau_x = t, \tau \leq t)} = \\
 &= P(W_s - W_t \leq 0) = P(W_s - W_t \geq 0) \\
 &= P(W_s \geq x, \tau_x \in dt) = P_1
 \end{aligned}$$

Мы взяли траекторию процесса, дождалась, пока она дошла до  $\tau_x$ . В силу марковости то, как она туда дошла, не зависит от того, что будет потом. Тогда траектории, которая ушла в итоге в точку, меньшую  $x$ , можно сопоставить траекторию, которая ушла в точку, большую  $x$ , причём вероятность этих траекторий одна и та же. Можно сказать, что мы воспользовались автомодельностью с  $c = -1$ . Значит,  $P(\max W_t > x, t \in [0, s]) = 2 \cdot \left(1 - \Phi\left(\frac{x}{\sqrt{s}}\right)\right)$ .



## Семинар 11. Слабая сходимость. Принцип инвариантности

### Сходимость в смысле конечномерных распределений

Мы знаем центральную предельную теорему: если  $X_i$  — независимые одинаково распределённые,  $EX_i = 0$ ,  $DX_i = \sigma^2 > 0$ , то

$$\frac{S_n}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim \mathcal{N}(0, 1).$$

Проблема в том, что этот факт бессодержателен с точки зрения теории случайных процессов, потому что для случайной последовательности  $S_n$  мы знаем только про распределение отдельных величин, а не то, как они вместе устроены. Представьте, что  $Y$  — это:

$$Y = \sigma\sqrt{n} \cdot Z, \quad Z \sim \mathcal{N}(0, 1) \implies \frac{Y}{\sigma\sqrt{n}} \xrightarrow{d} Z.$$

Предел по распределению у  $Y$  и  $S_n$  один и тот же, однако траектории этих двух последовательностей выглядят совершенно по-разному (см. рис. 11.1).

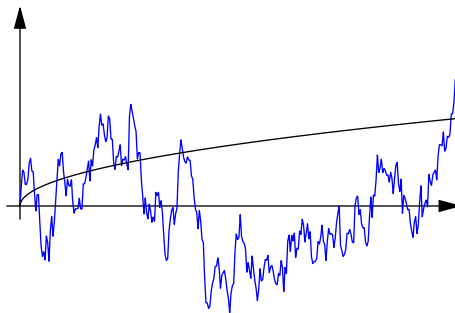


Рис. 11.1: Траектории двух случайных последовательностей с одинаковой центральной предельной теоремой. Синим цветом показана траектория  $S_n$ , чёрным цветом — траектория  $Y$ .

Проблема сходимости по распределению в том, что там идёт речь только про одномерные значения процесса. Чтобы получить какой-то содержательный факт про саму последовательность, нам нужно составить конечномерные распределения.

**Пример.** Рассмотрим вектор  $\left(\frac{S_{[nt_1]}}{\sigma\sqrt{n}}, \dots, \frac{S_{[nt_k]}}{\sigma\sqrt{n}}\right)$ , где  $t_1 < t_2 < \dots < t_k < 1$ . Чтобы посмотреть, куда такой вектор сходится, применим к нему линейное

преобразование:

$$\left( \frac{S_{[nt_1]}}{\sigma\sqrt{n}}, \frac{S_{[nt_2]} - S_{[nt_1]}}{\sigma\sqrt{n}}, \dots, \frac{S_{[nt_k]} - S_{[nt_{k-1}]}}{\sigma\sqrt{n}} \right) \rightarrow$$

$$\rightarrow (Z_1, \dots, Z_k) \sim \mathcal{N} \left( 0, \begin{pmatrix} t_1 & & & 0 \\ & t_2 - t_1 & & \\ & & \dots & \\ 0 & & & t_k - t_{k-1} \end{pmatrix} \right).$$

Тогда

$$\left( \frac{S_{[nt_1]}}{\sigma\sqrt{n}}, \dots, \frac{S_{[nt_k]}}{\sigma\sqrt{n}} \right) \xrightarrow{d} U = (W_{t_1}, \dots, W_{t_k}) \sim \mathcal{N} \left( 0, \begin{pmatrix} t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & \dots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \vdots & t_k \end{pmatrix} \right)$$

**Определение 11.1.** Процесс  $X_n(t)$  сходится в смысле конечномерных распределений к процессу  $X(t)$ , если любые конечномерные распределения последовательности этого процесса сходятся к конечномерным распределениям последовательностей другого процесса. Обозначение:  $X_n(t) \xrightarrow{\text{к.м.}} X(t)$ .

В примере выше мы ввели нормированное случайное блуждание  $Y_n(t) = \frac{S_{[nt]}}{\sigma\sqrt{n}}$  с нулевым средним и дисперсией  $\sigma^2$ , а затем показали, что этот процесс сходится к броуновскому движению  $W_t$  в смысле конечномерных распределений:  $Y_n(t) \xrightarrow{\text{к.м.}} W_t$ .

## Слабая сходимость

Вспомним эквивалентные определения слабой сходимости для случайных величин:

- 1)  $\forall g \in CB \ E g(X_n) = E g(X)$ ;
- 2)  $F_{X_n}(x) \rightarrow F_X(x) \ \forall x: P(X = x) = 0$ ;
- 3)  $\psi_{X_n}(t) \rightarrow \psi_X(t)$ ;
- 4)  $P(X_n \in A) \rightarrow P(X \in A) \ \forall A: P(X \in \partial A) = 0$ .

Условия 2 и 3 не переносятся на случайные процессы. 1 и 4 подходят для переноса, но в 1 возникает вопрос, что такое непрерывное отображение, а в 4 — что такое граница.

Введём общее определение слабой сходимости в пространстве, в котором есть метрика (или норма).

**Определение 11.2.** Рассмотрим пространство  $\mathcal{C}$  с метрикой  $\rho$ ;  $X_n: \Omega \rightarrow \mathcal{C}$  — измеримые относительно  $\mathcal{B}(\mathcal{C})$ . Будем говорить, что  $X_n$  **слабо сходится** к  $X$  (обозначение:  $X_n \xrightarrow{D} X$ ), если выполнено одно из эквивалентных условий:

$$\begin{aligned} \mathbb{P}(\omega: X_n(\omega) \in A \in \mathcal{B}(\mathcal{C})) \rightarrow \mathbb{P}(\omega: X(\omega) \in A) \quad \forall A: \mathbb{P}(X \in \partial A) = 0 &\iff \\ \iff \forall g: \mathcal{C} \rightarrow \mathbb{R} \mathbb{E}g(X_n) \rightarrow \mathbb{E}g(X), \end{aligned}$$

если  $g$  — непрерывная по метрике  $\rho$  и ограниченная.

С точки зрения понимания, что такое слабая сходимость, удобнее первое условие, а с точки зрения доказательства часто бывает удобно второе. В частности, из второго определения совсем просто следует факт, что если  $X_n \xrightarrow{D} X$ , то  $f(X_n) \xrightarrow{D} f(X)$ , где  $f$  — непрерывное отображение из  $\mathcal{C}$  в какое-то другое пространство.

## Связь слабой сходимости со сходимостью в конечномерных распределениях

Будем расстраивать процессы  $X_n \in C([0, 1])$ ;  $\|f\| = \sup_{[0,1]} f$ .

Из  $X_n \xrightarrow{D} X$  следует  $X_n \xrightarrow{\text{к.м.}} X$ , потому что конечномерное распределение — частный случай вероятности попадания процесса в множество: процесс попадает в множество функций, когда

$$\mathbb{P}(X_n \in B) = \mathbb{P}(X_n(t_1) \in A_1, \dots, X_n(t_k) \in A_k) \rightarrow \mathbb{P}(X(t_1) \in A_1, \dots, X(t_k) \in A_k),$$

где  $A_j: \mathbb{P}(X(t_j) \in \partial A_j) = 0$ ;  $B$  — множество функций, проходящих в момент  $t_1$  через  $A_1$ , и так далее, в момент  $t_k$  — через  $A_k$  (это просто частный случай принадлежности множеству). Это множество борелевское, и его граница состоит из функций, которые в один из моментов  $t_j$  проходят через  $\partial A_j$ . Поэтому сходимость в конечномерных распределениях — это просто частный случай слабой сходимости для таких «дырявых»  $B$ , а не всех возможных борелевских.

Обратное неверно. Из  $X_n \xrightarrow{\text{к.м.}} X$  не следует  $X_n \xrightarrow{D} X$ .

**Пример.** Рассмотрим процесс  $X_n(\omega) = \begin{cases} tn, & t \leq \frac{1}{n}; \\ 2 - tn, & \frac{1}{n} \leq t \leq \frac{2}{n}; \\ 0 & t \geq \frac{2}{n}. \end{cases}$

С течением времени «колышек» сужается, но не растёт вверх. С точки зрения конечномерных распределений для любых фиксированных  $t_1, \dots, t_k$  при достаточно большом  $n$  распределение  $X_n$  совпадает с распределением процесса, состоящего из одних нулей, потому что какие бы  $t_i$  мы не взяли, рано или поздно все  $t_i$  окажутся либо в нуле, либо правее этого скачка. Значит,  $X_n \xrightarrow{\text{к.м.}} 0$ . Возьмём теперь  $A = \left\{ f : \sup |f| \leq \frac{1}{2} \right\}$ . Для него  $P(X_n \in A) = 0$ ,  $P(X \in A) = 1$ . Поэтому  $X_n \not\xrightarrow{D} 0$ . Если говорить про непрерывно ограниченные  $g$ , можно взять равномерную норму. Это, очевидно, непрерывный функционал, раз это норма, но при этом он у исходного процесса при всех  $X_n$  равен 1, а у предельного процесса всегда 0.

Здесь видно в чём специфика сходимости в конечномерных распределениях — в том, что конечномерные распределения ставят фиксированную сетку, а процесс с ростом  $n$  может через неё «проскакать», что невозможно заметить ни на какой фиксированной решетке, только если смотреть на процесс целиком: у него будет отличаться супремум и другие характеристики, будет отсутствовать сходимость.

Сходимости в конечномерных распределениях не хватает для слабой сходимости. Это очень неудобно, потому что проверять слабую сходимость по определению не представляется возможным. Поэтому нам хочется добавить какое-нибудь условие, которое будет обеспечивать слабую сходимость. Такое условие есть и базируется оно на очень простой идее: если  $X_n$  сходится к  $X$  в смысле конечномерных распределений, то если оно и сходится куда-то слабо, то обязательно тоже к  $X$ , потому что если  $X_n$  сходится к  $X$  в смысле конечномерных распределений и  $X_n$  сходится к  $Y$  слабо, то конечномерные распределения  $X$  и  $Y$  одинаковы, тогда и распределения у процессов одинаковые. Это свойство называется *слабо относительно секвенциальной компактностью*.

**Определение 11.3.** Пусть  $\forall X_{n_k} \exists X_{n_{k_l}} \xrightarrow{D}$  куда-то, то есть из любой последовательности можно выделить слабо сходящуюся подпоследовательность. Тогда  $X_n$  — **слабо относительно секвенциально компактный процесс**.

Если добавить к сходимости в конечномерных распределениях слабо относительно секвенциальную компактность процесса, то это равносильно слабой сходимости.

**Определение 11.4.** Пространство называют **польским**, если оно полное метрическое и сепарабельное.

**Теорема 11.1 (Прохоров).** В польском пространстве  $X_n$  слабо относительно

секвенциально компактно тогда и только тогда, когда  $X_n$  плотна, то есть

$$\forall \varepsilon > 0 \exists K \text{ — компакт: } \forall n P(X_n \in K) \geq 1 - \varepsilon.$$

Иногда для доказательства слабой сходимости используют не относительно секвенциальную компактность, а вспомогательные факты про модуль и непрерывности. Если вы вспомните критерий компактности в  $C[0, 1]$ , то увидите, что он как раз в терминах модуля непрерывности и выписывается.

Итак, чтобы доказать слабую сходимоть, нам нужно показать, что последовательность процессов конечномерно сходится к какому-то процессу, а дальше доказать свойство плотности. Это определённые технические условия, которые нужно проверить. В случае  $C[0, 1]$ , например, это равномерная непрерывность и равномерная ограниченность функции.

## Принцип инвариантности и аналогичные теоремы

Условия для слабой сходимости наострились проверять довольно успешно, и с помощью этого доказывают разные теоремы. Одна из основных теорем такого рода называется принципом инвариантности Донскера–Прохорова:

**Теорема 11.2** (*принцип инвариантности Донскера–Прохорова*). Рассмотрим процесс  $Y_n(t)$ :

$$Y_n(t) = \frac{S_{[nt]} + X_{[nt]+1} \{nt\}}{\sigma \sqrt{n}},$$

где  $X_i$  — независимые одинаково распределённые с  $EX_i = 0$ ,  $DX_i = \sigma^2$ . Тогда  $Y_n \xrightarrow{D} W$  в  $C[0, 1]$ .

Выражение  $Y_t$  — непрерывная линейная интерполяция  $S_{nt}$ : линейно соединяем точки блуждания, чтобы в момент  $[nt] + 1$  процесс оказался в точке  $S_{[nt]+1}$  непрерывно. Это нужно, чтобы дискретное блуждание оказалось определённым в каждой точке процессом на  $[0, 1]$ .

Во-первых, это значит, что вероятности сходятся. Отчасти это вообще оправдывает нам понимание того, что такое броуновское движение — это случайное блуждание, которое сжали в отрезок  $[0, 1]$  и утрамбовали в  $\sqrt{n}$  раз. Во-вторых, это даёт множество полезных результатов, например,

$$f(Y_n) \xrightarrow{d} f(W) \forall f: C[0, 1] \rightarrow \mathbb{R}, f \text{ — непрерывное отображение.}$$

Любой такой функционал — это уже просто случайная величина, и такая последовательность случайных величин всегда будет сходиться. Значит, из

сходимости  $Y_n \xrightarrow{D} W$  следует, например,  $\sup Y_n \rightarrow \sup W$ ,  $\inf Y_n \rightarrow \inf W$ . Это позволяет нам исследовать функционал для броуновского движения с помощью функционала для блуждания, и наоборот.

Заметим, что от блуждания здесь ничего не зависит: какое бы блуждание не было, сходимость всё равно верна. Поэтому можно взять простое симметричное блуждание, для которого всё хорошо считается. Используя его специфику, сможем явно посчитать  $f(Y_n)$ , где  $f$  — интересующий функционал. Перейдя к пределу, найдём распределение  $f(W)$ . Так как теорема верна для всех блужданий, можем взять любое другое блуждание с нулевым средним и дисперсией  $\sigma^2$ , и его  $f(Y_n)$  в пределе будет вести себя также, как  $f(W)$ , которую мы нашли.

**Пример.** Рассмотрим непрерывный функционал  $f(g) = \max_{t \in [0,1]} g(t)$ .

Очевидно, что  $f$  — это непрерывный функционал по равномерной норме: если функцию изменить не больше, чем на  $\varepsilon$  в каждой точке, максимум тоже изменится не больше чем  $\varepsilon$ . Значит,  $f(Y_n) \xrightarrow{d} f(W)$ , где  $Y_n$  — линейно интерполированное случайное блуждание, а  $W$  — броуновское движение. Значит, функция распределения тоже сходится. Так как максимум линейной функции достигается на границе диапазона, нигде, кроме как в целой точке, не можем достигнуть максимума у линейно интерполированного блуждания:

$$P(f(Y_n) > x) = P(\max_{k \leq n} S_k > x\sigma\sqrt{n}).$$

Теперь возьмём  $X_i$  — простое симметричное случайное блуждание (у него  $\sigma = 1$ ), потому что для всех ответ один и тот же, а для него с максимумами мы умеем работать:

$$P(\max_{k \leq n} S_k > m, m \in \mathbb{Z}) = P(\max_{k \leq n} S_k \geq m, S_n > m) + P(\max_{k \leq n} S_k > m, S_n = m) + P(\max_{k \leq n} S_k > m, S_n < m).$$

Первая вероятность в правой части равна последней, потому что отражение: если возьмём момент последнего прохода через  $m$ , то каждой траектории, заканчивающейся в точке ниже  $m$ , можно сопоставить траекторию, которая заканчивается выше  $m$ . Заметим, что если  $m$  не целое, а какое-то  $x\sqrt{n}$ , то мы должны вместо  $> x\sqrt{n}$  поставить  $\geq [x\sqrt{n}]$ . Вероятность попасть в отдельную точку (второе слагаемое) стремится к нулю с ростом  $m$ .

В итоге получаем, что

$$P(f(Y_n) > x) = P(\max_{k \leq n} S_k > x \underbrace{\sigma}_{=1} \sqrt{n}) = 2P(S_n > x\sqrt{n}) + o(1) \rightarrow 2(1 - \Phi(x)),$$

где  $o(1)$  — это вероятность попасть в точку, даже если она целая. Переход к пределу выполнен с использованием центральной предельной теоремы. По принципу инвариантности Донскера–Прохорова мы нашли распределение максимума броуновского движения:

$$1 - P(f(Y_n) > x) = P(f(Y_n) \leq x) \rightarrow P(f(W) \leq x) = \boxed{f(W) = 2\Phi(x) - 1}$$

Кроме того, мы теперь доказали, что у всех блужданий такая же асимптотика, потому что по принципу инвариантности у всех блужданий предел один и тот же.

Принцип инвариантности стал предлогом множества всяких полезных открытий, в особенности в статистике. Давайте сформулируем еще три похожие теоремы.

**Теорема 11.3.** Пусть  $X_i$  — независимые одинаково распределённые решётчатые с шагом 1 и сдвигом 0, то есть они принимают значения из  $\mathbb{Z}$ , причем нельзя их вложить, например, в  $2\mathbb{Z}+3$  или в  $3\mathbb{Z}+5$ . Рассмотрим тот же самый процесс  $Y_n(t)$ . Тогда

$$P(Y_{2n}(t) \in A \mid S_{2n} = 0) \rightarrow P(W^0 \in A) \quad \forall A: P(W^0 \in \partial A) = 0.$$

Таким образом, процесс, рассмотренный при условии возвращения в 0, будет сходиться к броуновскому мосту. Это соотносится с тем, что броуновский мост — винеровский процесс, который заставили вернуться в 0.

**Теорема 11.4.** Пусть  $X_i$  — независимые одинаково распределённые с математическим ожиданием 0 и дисперсией  $\sigma^2$ , тогда  $\forall 0 < i \leq n$ :

$$P(Y_n(t) \in A \mid S_i > 0) \rightarrow P(W^+ \in A),$$

где  $W^+$  — броуновская извилина или броуновский меандр.

В статистике тоже верна аналогичная теорема, её обычно формулируют в терминах непрерывных справа функций, но мы сформулируем её в терминах непрерывных функций.

Пусть  $X_i \sim R[0, 1]$  — независимые одинаково распределённые величины.  $F_n(x)$  — эмпирическая функция распределения дискретной случайной величины. Введём непрерывную случайную величину с функцией распределения  $\tilde{F}_n(x)$  следующим образом: если исходная функция распределения имела разрывы в точках  $x_{(1)} \leq \dots \leq x_{(n)}$ , добавим к ним точку  $x_{(n+1)} = 1$  и на каждом из отрезков  $[x_{(i)}, x_{(i+1)}]$ ,  $1 \leq i \leq n$  заменим константу линейной функцией (см. рис. 11.2).

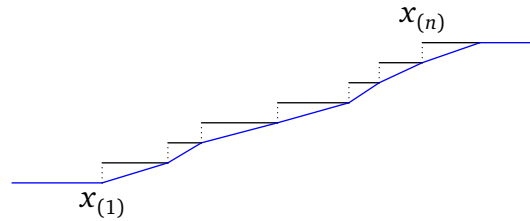


Рис. 11.2: Эмпирическая функция распределения  $F_n(x)$  показана чёрным цветом, построенная  $\tilde{F}_n(x)$  — синим.

**Теорема 11.5** (основная теорема статистики).

$$\sqrt{n}(\tilde{F}_n(x) - x) \xrightarrow{D} W^0$$

в пространстве непрерывных функций  $C[0, 1]$ .

Если мы дополнительно докажем, что броуновский мост имеет распределение Колмогорова, мы получим критерий Колмогорова из курса статистики.

К сожалению, непрерывный функционал в  $C[0, 1]$  — это довольно узкий класс.

**Пример.** Рассмотрим  $f(g) = \sup \{t : g(t) = 0, t \in [0, 1]\}$ . Возьмём функцию, показанную на рис. 11.3. У неё последний 0 в точке  $\frac{1}{3}$ , при этом в любой её окрестности есть функция, показанная на рисунке пунктирной линией, у которой последний 0 находится в точке 0. Значит, функционал на этой функции разрывен.

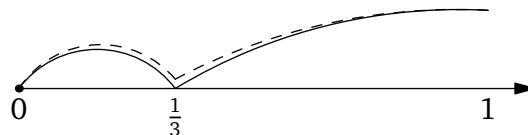


Рис. 11.3: Пример функции, в окрестности которой функционал будет разрывен.

При этом, чтобы функционал был разрывен, нужно брать своеобразные функции, особым образом касающиеся нуля. Если нуля не касаться, то тогда такой проблемы не возникнет. Расширим наше утверждение на подобного рода функционалы.

**Лемма 11.1.** Если  $f$  — функционал,  $C_f$  — множество точек (функций) его разрыва и  $P(W \in C_f) = 0$ , то  $f(Y_n) \xrightarrow{d} f(W)$ .



**Пример.** Вернёмся к  $f(g) = \sup \{t : g(t) = 0, t \in [0, 1]\}$  и докажем, что броуновское движение подаёт в множество функций разрыва с вероятностью 0. В  $C_f$  попадают функции, для которых

$$\exists \delta > 0 : \forall \varepsilon > 0 \exists g_n \in U_\varepsilon(g) : |f(g_n) - f(g)| > \delta.$$

Пусть  $f(g) = a$ . Тогда либо  $f(g_n) > a + \delta$ , либо  $f(g_n) < a - \delta$ .

Случай  $f(g_n) > a + \delta$  невозможен, потому что  $g$  — непрерывная функция, которая не обращается в 0 дальше точки  $a$ , значит её инфимум на отрезке  $[a + \delta, 1]$  положительный, а значит можно подобрать такую дельта-окрестность, что её инфимум на этом отрезке также будет положительным и новых нулей, больших  $a + \delta$  возникнуть не может.

При  $f(g_n) < a - \delta$  у нас пропадает 0. Рассмотрим отрезок  $[a - \delta, 1]$ . На него попадает точка  $a$ , в которой  $g = 0$ , значит, супремум не может быть меньше 0, а инфимум не может быть больше 0. Предположим, что инфимум отрицательный, а супремум положительный. Тогда у всех точек из некоторой окрестности это условие сохраняется, но тогда в этой окрестности есть 0 между инфимумом и супремумом, что противоречит тому, что на  $[a - \delta, 1]$  нет нулей. Значит, 0 может пропасть только если  $\inf_{[a-\delta, 1]} g(x) = 0$  или  $\sup_{[a-\delta, 1]} g(x) = 0$ .

Точка  $a$  сама по себе может быть иррациональной, но уж точно у неё в левой  $\delta$ -окрестности есть хотя бы одна рациональную точку  $q$ . Тогда

$$P(W \in C_f) \leq P(\exists q \in \mathbb{Q} : \sup_{[q, 1]} W = 0) + P(\exists q \in \mathbb{Q} : \inf_{[q, 1]} W = 0).$$

Посчитаем выражение для супремума, для инфимума всё делается аналогично. Зафиксируем  $q$ , в ней  $W(q) = x$ . Помним, что супремум изначального процесса равен нулю, тогда для процесса, стартующего из точки  $q$ , супремум будет равен  $-x$ . Рассмотрим следующее событие:

$$P(\sup_{[q, 1]} W = 0) = \int_{-\infty}^0 f_{W_q}(x) P(\sup_{[0, 1-q]} W = -x) dx \ominus$$

Воспользуемся свойством броуновского движения, которое говорит, что  $\frac{1}{\sqrt{c}} W_{ct} \stackrel{d}{=} W$ , соответственно можем отмасштабировать наш процесс:

$$\ominus \int_{-\infty}^0 f_{W_q}(x) \underbrace{P\left(\sup_{[0, 1]} W = -\frac{x}{\sqrt{1-q}}\right)}_{=0} dx = 0.$$

Получаем, что  $P(W \in C_f) = 0$  — счётное объединение таких вероятностей.

Значит,

$$P(\tau_0(Y_n) \leq x) \rightarrow P(\tau_0(W) \leq x).$$

Рассмотрим это выражение для простого симметричного блуждания. У него нулей, кроме как в целых точках, возникнуть не может, так что последний 0 у  $S_n$  не больше  $xn$ , а такую вероятность мы знаем:

$$P(\tau_0(Y_n) \leq x) = P(\text{последний } 0 \text{ у } S_n \leq xn) \rightarrow \frac{2}{\pi} \arcsin \sqrt{x}.$$

С помощью свойств случайного блуждания предельный переход можно не очень сложно доказать.

Тем самым мы получили два важных результата:

- 1) последний 0 броуновского движения на отрезке  $[0, 1]$  распределён как  $\frac{2}{\pi} \arcsin \sqrt{x}$ ;
- 2) этот результат верен для любого блуждания, если мы говорим не про последний приход блуждания в 0, а про пересечение линейно интерполированной траекторией 0.

Последний результат можно представить в виде общего закона арксинуса:

**Теорема 11.6 (общий закон арксинуса).** Для любого блуждания  $Y_n$  с нулевым средним и дисперсией  $\sigma^2$  момент перехода процесса из отрицательной точки в положительную  $\tau_0$  удовлетворяет соотношению

$$P(\tau_0(Y_n) \leq xn) \rightarrow \frac{2}{\pi} \arcsin \sqrt{x}.$$

Именно так и работает принцип инвариантности: мы доказываем что-то очень просто, комбинаторно, с явным заданием величин, а потом совершаем предельный переход и получаем сразу целую серию теорем. Каждый раз, когда вы про какой-то функционал доказали, что он непрерывный и нашли его распределение, вы вместе с этим дополнительно доказали теорему для всех блужданий.



МЕХАНИКО-  
МАТЕМАТИЧЕСКИЙ  
ФАКУЛЬТЕТ  
МГУ ИМЕНИ  
М.В. ЛОМОНОСОВА

*teach-in*  
ЛЕКЦИИ УЧЕНЫХ МГУ