

Московский государственный университет  
имени М.В. Ломоносова  
факультет Вычислительной математики и кибернетики  
кафедра системного анализа

Лекции по курсу  
**“Стохастический анализ и моделирование”**

Преподаватель:  
Смирнов Сергей Николаевич

Москва  
2008

# Содержание

<b>1</b>	<b>Глава I</b>	<b>2</b>
1.1	Математическая модель явлений, в которых случайность выступает существенным фактором. . . . .	2
1.2	Аксиоматика Колмогорова и непротиворечивость модели; терминология, содержание и интерпретация объектов. . . . .	2
1.3	Объективная (объективно-частотная) и субъективная интерпретации вероятности. . . . .	6
<b>2</b>	<b>Глава II</b>	<b>8</b>
2.1	Условная вероятность как модель с дополнительным ограничением. . . . .	8
2.2	Независимость событий. . . . .	9
2.3	Схема Бернулли. . . . .	14
2.4	Парадоксы фон Мизеса и д'Аламбера. . . . .	15
<b>3</b>	<b>Глава III</b>	<b>18</b>
3.1	Пренебрежимые модификации модели. . . . .	18
3.2	Непрерывные и дискретные модели. Теорема Улама (частный случай). . . .	20
3.3	Борелевская $\sigma$ -алгебра. . . . .	21
3.4	Нуль множества и пополнение вероятностного пространства. . . . .	23
3.5	Аксиома выбора. . . . .	24
3.6	Теорема об изоморфизме. . . . .	26
<b>4</b>	<b>Глава IV</b>	<b>27</b>
4.1	Распределение вероятностей как частный вид модели. Случайные величины, вектора, функции. Интерпретация условия измеримости. Распределение как индуцированная мера. . . . .	27
4.2	Функция распределения как характеристика распределения и ее свойства. Моделирование случайных величин методом обращения функций распределения. . . . .	32
<b>5</b>	<b>Глава V</b>	<b>40</b>
5.1	Альтернативный подход к аксиоматике на основе понятия среднего, эквивалентность подходу на основе понятия вероятности. Построение интеграла Лебега. . . . .	40
5.2	Свойства математического ожидания (замена переменных, неравенство Йенсена и т.д.). . . . .	45
<b>6</b>	<b>Глава VI</b>	<b>52</b>
6.1	Субъективная интерпретация среднего. . . . .	52
6.2	Формулировка закона больших чисел и объективная интерпретация среднего. .	53
6.3	Отличие интеграла Лебега от интеграла Римана. . . . .	54
6.3.1	Интеграл Римана. . . . .	54
6.3.2	Интеграл Лебега. . . . .	56
6.4	Метод Монте-Карло как численный метод нахождения интеграла Лебега. . .	57
<b>7</b>	<b>Глава VII</b>	<b>59</b>
7.1	Ковариация, дисперсия и их свойства. Ковариационные матрицы. . . . .	59
7.2	Линейные модели. Задача регрессии. . . . .	65
7.2.1	Линейная задача регрессии. . . . .	66

7.2.2	Другие задачи регрессии. . . . .	68
7.3	Свойства многомерного нормального распределения. Моделирование гауссовского случайного вектора. . . . .	70
<b>8</b>	<b>Глава VIII</b>	<b>72</b>
8.1	Теорема Радона-Никодима. . . . .	72
8.2	Абсолютно-непрерывные распределения, свойства плотности. . . . .	74
8.3	Моделирование методом элиминации фон Неймана. . . . .	76
<b>9</b>	<b>Глава IX. Дискретные распределения. Смеси распределений. Сингулярные распределения. Полная классификация распределений. Примеры наиболее часто употребляемых на практике распределений.</b>	<b>78</b>
9.1	Дискретные распределения. . . . .	78
9.2	Смеси распределений. . . . .	78
9.3	Сингулярные распределения. . . . .	79
9.4	Полная классификация распределений. . . . .	80
9.5	Примеры наиболее часто употребляемых на практике распределений. . . . .	81
<b>10</b>	<b>Глава X</b>	<b>83</b>
10.1	Совместные и маргинальные распределения. . . . .	83
10.2	Согласованность семейства распределений. . . . .	85
10.3	Теорема Колмогорова. . . . .	86
<b>11</b>	<b>Глава XI</b>	<b>90</b>
11.1	Условные распределения (существование регулярного варианта). Интегральные формулы типа полной вероятности. . . . .	90
11.2	Метод последовательного моделирования случайного вектора. . . . .	94
11.3	Независимость случайных величин. . . . .	95
<b>12</b>	<b>Глава XII</b>	<b>98</b>
<b>13</b>	<b>Глава XIII</b>	<b>102</b>
13.1	Марковское свойство. Уравнение Колмогорова-Чепмена. . . . .	102
13.2	Однородность по времени. . . . .	108
13.3	Представление для цепей Маркова (стохастическая динамическая система). . . . .	109
13.4	Стационарность. . . . .	110
13.5	Эргодичность. . . . .	114
13.6	Винеровский процесс. . . . .	115
13.7	Процесс Орнштейна-Уленбека. . . . .	119
<b>14</b>	<b>Виды сходимости случайных величин: по вероятности, в среднем, почти наверное, по распределению - слабая и по вариации.</b>	<b>120</b>

# 1 Математическая модель явлений, в которых случайность выступает существенным фактором. Аксиоматика Колмогорова и непротиворечивость модели; терминология, содержание и интерпретация объектов. Объективно-частотная и субъективная интерпретации вероятности.

## 1.1 Математическая модель явлений, в которых случайность выступает существенным фактором.

Теория вероятностей не является чисто математической дисциплиной. Например, Гильберт назвал ее физической дисциплиной.

**Определение 1.** *Теория вероятностей* занимается построением и анализом (+ иногда интерпретацией) математических моделей, в которых случайность выступает существенным фактором.

**Определение 2.** Математическая модель — формальное описание явления из реального мира.

Заметим, что определение теории вероятности не полно, но дать полное определение очень тяжело, запутано и не нужно;)

Построение математической модели не является чисто математической задачей, т.к. для ее построения нужно, в первую очередь, разобраться в предметной области. Последней стадией построения является формализация, т.е. перевод с естественного языка на математический или какой-нибудь другой из точных языков.

Анализ математических моделей — чисто математическая задача (анализ может проводиться как аналитически, так и численно). После анализа необходим опять перевод на естественный язык, т.е. интерпретация полученных результатов.

## 1.2 Аксиоматика Колмогорова и непротиворечивость модели; терминология, содержание и интерпретация объектов.

В 1933 году вышла книга А.Н.Колмогорова “Основные понятия теории вероятностей”, где давалась первая аксиоматика теории вероятностей. “Целью предлагаемой работы является аксиоматическое обоснование теории вероятностей. Ведущей мыслью автора было при этом естественное включение основ теории вероятностей, считавшихся еще недавно совершенно своеобразными, в ряд общих понятий современной математики. До возникновения лебеговой теории меры и интеграла эта задача была почти безнадежна. После исследований Лебега стала ясной аналогия между мерой множества и вероятностью события, а также между интегралом от функции и математическим ожиданием случайной величины. Эта аналогия допускает и дальнейшее продолжение: так, например, многие свойства независимых случайных величин вполне аналогичны соответствующим свойствам ортогональных функций. Для того чтобы, исходя из этой аналогии, обосновать теорию вероятностей, следовало еще освободить теорию меры и теорию интегрирования от геометрических элементов, которые еще имелись у Лебега. <...> Теория вероятностей как математическая дисциплина может и должна быть аксиоматизирована совершенно в том же смысле, как геометрия или алгебра. Это означает, что, после того как даны названия

изучаемым объектам и их основным отношениям, а также аксиомы, которым эти отношения должны подчиняться, все дальнейшее изложение должно основываться исключительно лишь на этих аксиомах, не опираясь на обычное конкретное значение этих объектов и их отношений.”<sup>1</sup>

Аксиоматика Колмогорова не первая из предлагавшихся (например, аксиоматика Бернштейна, где понятие вероятности не относилось к числу основных понятий, а само выражалось через другие понятия), однако ни одна из предшествующих не давала полной картины. Поэтому рассмотрим аксиоматику Колмогорова.

**Теория меры — это язык теории вероятности.** В этом и заключается основной постулат Колмогорова.

Теория меры возникла из геометрии как обобщение: множеству можно поставить в соответствие число, измеряющее это множество.

Пусть  $\Omega$  — некоторое непустое множество;  $\mathcal{F}$  — некий класс подмножеств  $\Omega$ .

**Определение 3.** Функция множества  $\mu : \mathcal{F} \rightarrow \mathbb{R}$  называется **мерой**, если выполнены следующие свойства:

аддитивность<sup>2</sup>:

$$\begin{aligned} \mu(A_1 + A_2) &= \mu(A_1) + \mu(A_2), \quad \text{если } A_1, A_2, A_1 + A_2 \in \mathcal{F}, \\ \text{или, другими словами,} \\ \mu\left(\sum_{i=1}^n A_i\right) &= \sum_{i=1}^n \mu(A_i), \end{aligned} \quad (1)$$

неотрицательность:

$$\mu(A) \geq 0, \quad (2)$$

непрерывность:

$$\begin{aligned} \forall A_i \in \mathcal{F}, A_i \subset A_{i+1}, \bigcup_{i=1}^{\infty} A_i = A \in \mathcal{F} &\implies \mu(A) = \lim_{i \rightarrow \infty} \mu(A_i), \\ \text{или, другими словами,} \\ A_i \nearrow A &\implies \mu(A_i) \rightarrow \mu(A)^3. \end{aligned} \quad (3)$$

Если выполнены свойства 1 и 2, то мера называется конечно-аддитивной. Мера называется конечной<sup>4</sup>, если  $\mu(\Omega) < \infty$ <sup>5</sup>. Если положить  $\mu(\Omega) = 1$ , то мера становится нормированной.

Заметим, что из свойства 1 следует  $\mu(\emptyset) = 0$  (так как  $A + \emptyset = A \implies \mu(A) + \mu(\emptyset) = \mu(A)$ ). А также из свойств 1 и 2 следует, что  $A \subseteq B \implies \mu(A) \leq \mu(B)$  ( $B = A + (B \setminus A) \implies \mu(B) = \mu(A) + \mu(B \setminus A)$ ,  $\mu(B \setminus A) \geq 0$ <sup>6</sup>).

В теории вероятности можно предложить **универсальный вид моделей. Любое непротиворечивое описание модели можно представить в виде вероятностного пространства, представляющего собой тройку объектов  $(\Omega, \mathcal{F}, \mathbb{P})$ .**

<sup>1</sup>А.Н.Колмогоров. “Основные понятия теории вероятностей.” (перевод с немецкого Г.М.Бавли) 1936г.

<sup>2</sup>запись  $A_1 + A_2$  означает объединение непересекающихся множеств ( $A_1 \cup A_2, A_1 \cap A_2 = \emptyset$ )

<sup>3</sup>монотонность стремления  $\mu(A_i) \nearrow \mu(A)$  следует из свойств меры 1 и 2

<sup>4</sup>для конечных мер в свойстве меры 3 можно взять и монотонно убывающую последовательность множеств  $A_i \supset A_{i+1}, \bigcap_{i=1}^{\infty} A_i = A$

<sup>5</sup>подразумевается, что  $\Omega \in \mathcal{F}$

<sup>6</sup>если, конечно,  $B \setminus A \in \mathcal{F}$

Обозначение	Терминология	Математическое содержание	Интерпретация (на метаязыке)
$\Omega$	Пространство элементарных событий	Непустое множество (произвольной природы)	Совокупность возможных исходов (эксперимента)
$\omega \in \Omega$	Элементарное событие		Исход
$\mathcal{F}$	"Класс событий" (нетрадиционное название)	Непустое множество подмножеств $\Omega$ , удовлетворяющее определенным свойствам замкнутости относительно теоретико-множественных операций (например, алгебра, полуалгебра, $\sigma$ -алгебра)	Описание информации, заложенной в модель
$A \in \mathcal{F}$	Событие		
$\mathbb{P}$	Вероятность	Нормированная мера: $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R}$	<смотри ниже>
$\mathbb{P}(A)$	Вероятность события $A$	(числовая функция множеств, являющаяся мерой)	$\mathbb{P}(A)$ — вероятность того, что " $\omega \in A$ " <sup>7</sup>

**Пример** (о непротиворечивости модели). Пусть  $A \subseteq B$ ; вероятность попадания во множество  $A$  равна 0.6, а во множество  $B$  — 0.2. *Построение вероятностного пространства в явном виде — это проверка на непротиворечивость модели.* То есть показать, что модель непротиворечива  $\iff$  построить вероятностное пространство. Пусть нам удалось построить модель в форме  $(\Omega, \mathcal{F}, \mathbb{P})$ ,  $A, B \in \mathcal{F}$ . Из свойств меры следует, что  $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$ , что противоречит исходным данным. Значит, описание противоречиво.<sup>8</sup>

Основным объектом в этом универсальном виде модели является вероятность  $\mathbb{P}$ , а  $(\Omega, \mathcal{F})$  — атрибуты вероятности<sup>10</sup>. Вероятность приписывается высказываниям.

Аксиома Колмогорова состоит в том, что  $\mathbb{P}$  — мера.

Получили абстрактную теорию, имеющую многочисленные интерпретации.

**Определение 4.**  $\Omega$  — достоверное событие (высказывание " $\omega \in \Omega$ " всегда верно).  $\emptyset$  — невозможное событие (высказывание " $\omega \in \emptyset$ " всегда ложно).

Обозначим через  $A^c$  дополнение множества  $A$ :  $A^c = \Omega \setminus A$ .

**Определение 5.**  $A^c$  — противоположное событие (высказывание " $\omega \in A^c$ " верно тогда и только тогда, когда высказывание " $\omega \in A$ " ложно).

<sup>7</sup>" $\omega \in A$ " является примером высказывания. Между теоретико-множественными и логическими операциями существует взаимно-однозначное соответствие (например, " $\omega \in A_1$ " & " $\omega \in A_2$ "  $\iff \omega \in A_1 \cap A_2$ ); поэтому вместо множеств можно оперировать и высказываниями.

<sup>8</sup>Заметим, что не обязательно каждый раз при построении модели строить  $(\Omega, \mathcal{F}, \mathbb{P})$  в явном виде, если нет сомнений в непротиворечивости.

<sup>9</sup> $\mathcal{F}$  — field — поле

<sup>10</sup>Аналогично тому, как для задания функции нужно знать ее область определения и область значений. Для чего указывают эти атрибуты  $(\Omega, \mathcal{F})$ , почему удобно говорить именно о тройке? Первая причина: при построении модели без этих атрибутов все равно не обойтись, о  $\Omega$  и  $\mathcal{F}$  все равно придется подумать (при построении мат. модели логично (и проще всего) построить пространство элементарных исходов). Вторая причина: можно рассматривать класс моделей с одинаковыми атрибутами  $(\Omega, \mathcal{F})$ , но разными вероятностями  $\mathbb{P}$ . Третья причина: бывают случаи, когда вероятность не важна и рассматривают только  $(\Omega, \mathcal{F})$  (например, свойство измеримости формулируется в терминах измеримых пространств).

**Определение 6.** Если  $A \cap B = \emptyset$ , то  $A$  и  $B$  называются несовместными событиями.

В общем случае класс событий  $\mathcal{F}$  произволен. Однако обычно требуют замкнутости  $\mathcal{F}$  относительно некоторого набора теоретико-множественных операций. Зачем это нужно? Чтобы имело смысл говорить о противоположных событиях, об объединении событий и т.д.

**Определение 7.** Множество подмножеств  $\Omega$  называется полукольцом  $S$ , если

- 1)  $A \cap B \in S, \forall A, B \in S$ ;
- 2)  $A \setminus B = \sum_{k=1}^n A_k, \forall A, B \in S, B \subset A \text{ и } A_k \in S$ .

**Определение 8.** Множество подмножеств  $\Omega$  называется кольцом  $K$ , если

- 1)  $A \cap B \in K, \forall A, B \in K$ ;
- 2)  $A \nabla B \in K, \forall A, B \in K$ .

Заметим, что  $A \setminus B = (A \nabla B) \cap A$  и  $A \cup B = (A \nabla B) \nabla (A \cap B)$ . Таким образом, кольцо замкнуто относительно операций объединения и разности.

**Определение 9.** Кольцо  $K$  называется  $\sigma$ -кольцом, если  $\bigcup_{k=1}^{\infty} A_k \in K, \forall A_k \in K$ .

**Определение 10.** Кольцо  $K$  называется алгеброй, если  $\Omega \in K$ .

Можно также дать и другое определение алгебры.

**Определение 11.** Алгебра — непустой класс подмножеств  $\Omega$ , замкнутый относительно операций дополнения и объединения.

Определение корректно, так как  $A \cap B = (A^c \cup B^c)^c$  и  $A \nabla B = (A^c \cup B) \cup (A \cup B^c)$ .

**Определение 12.** Алгебра  $K$  называется  $\sigma$ -алгеброй, если  $\bigcup_{k=1}^{\infty} A_k \in K, \forall A_k \in K$ .

Будем рассматривать нормированную меру ( $\mu(\Omega) = 1^{11}$ ) и  $\sigma$ -алгебру событий. Если не требовать непрерывности, то достаточно просто алгебры.

Мы уже говорили, что аксиома Колмогорова состоит в том, что вероятность есть мера. Таким образом, аксиомы меры становятся аксиомами вероятности. Рассмотрим их с этой точки зрения.

*Свойство аддитивности* основное.  $\mathbb{P}(A_1 + A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2)$ . Основываясь на геометрической интуиции ( $\mathbb{P}$  аналог площади), можно предложить еще одну формулу, эквивалентную свойству аддитивности:  $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$ .

Заметим, что из свойств аддитивности и неотрицательности следует, что  $A \subseteq B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$ .

Обозначим  $B_n = \sum_{i=1}^n A_i, B = \sum_{i=1}^{\infty} A_i$ . Так как  $B_n \subseteq B$ , то  $\mathbb{P}(B_n) \leq \mathbb{P}(B)$ . Из свойства непрерывности следует, что  $\mathbb{P}(B_n) \rightarrow \mathbb{P}(B)$ , т.е.  $\lim_{n \rightarrow \infty} \mathbb{P}(\sum_{i=1}^n A_i) = \{\text{св-во конечной аддитивности}\} = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{P}(A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) = \mathbb{P}(\sum_{i=1}^{\infty} A_i)$ . Получаем, что аксиомы аддитивности и непрерывности эквивалентны *свойству счетной аддитивности*:  $\mathbb{P}\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$  (напомним, что  $A_i \cap A_j = \emptyset, \forall i \neq j$ ).

<sup>11</sup>мы полагаем  $\mu(\Omega) = 1$ , однако это не обязательно; можно вместо единицы взять любое другое положительное число

**Пример.** Игральный кубик имеет 6 пронумерованных граней. Его бросают 1 раз. Мы знаем (неважно откуда), что вероятность события, при котором выпадает 2 или 3, равна 0.2. Построим универсальную модель (способов может быть много, приведем один из них):  $\Omega = \{1, 2, 3, 4, 5, 6\}$ ;  $\mathcal{F} = \{(2, 3), (1, 4, 5, 6), \Omega, \emptyset\}$ ;  $\mathbb{P}((2, 3)) = 0.2$ ,  $\mathbb{P}((1, 4, 5, 6)) = 0.8$ ,  $\mathbb{P}(\Omega) = 1$ ,  $\mathbb{P}(\emptyset) = 0$ . Заметим, что нельзя положить  $\mathcal{F} = 2^\Omega$ , т.к. мы не знаем, например, вероятности  $\mathbb{P}((1))$  и  $\mathbb{P}((5, 6))$ . Поэтому естественно взять в качестве  $\mathcal{F}$  минимальную  $\sigma$ -алгебру, содержащую события, о которых у нас имеется информация. Этот пример иллюстрирует, что смысловая интерпретация  $\mathcal{F}$  — это описание информации, заложенной в модели.

### 1.3 Объективная (объективно-частотная) и субъективная интерпретации вероятности.

Субъективная интерпретация основана на мнении субъекта, а объективная интерпретация инвариантна относительно субъектов и основана на эмпирических факторах (критерий истины — опыт). Объективная интерпретация вероятности в частотной форме выражается законом больших чисел.

**Закон больших чисел (ЗБЧ) — закон природы.** Его нельзя доказать, можно только пояснить. ЗБЧ — фундамент, на котором покоится объективная интерпретация теории вероятностей. Будем формулировать его не в общем виде (это очень трудно сделать), на метаязыке. Закон больших чисел основан на схеме Бернулли.

**Определение 13.** Эксперимент (или явление) — это проведение серии однородных<sup>12</sup> испытаний (временной фактор мы не затрагиваем, так как испытания могут проводиться одновременно).

Предположения:

1. Отсутствие взаимного влияния в испытаниях (одно на другое не оказывает влияния).
2. Воспроизводимость. Однородные испытания проводятся в сходных, аналогичных условиях (не одинаковых, иначе результат был бы один и тот же)<sup>13</sup>.
3. Существует признак, который реализуется (успех) или не реализуется (неуспех) в испытании<sup>14</sup>. Признак может быть отнесен к любому из испытаний (в силу их однородности).

Описанная модель называется *схемой Бернулли*.

Частотой признака называется доля реализации признака в эксперименте. *При неограниченном увеличении числа испытаний частота признака приближается<sup>15</sup> к некоторому числу, которое называется вероятностью появления признака в одном отдельном испытании (любом, т.к. испытания однородны). (То есть вероятность, по сути, является “абстрактным двойником частоты” (Г.Краммер).) В этом и состоит закон больших чисел, основанный на эмпирической реальности и сформулированный в частотной форме.*

Характерно, что закон больших чисел позволяет измерять вероятность практически, т.е. утверждает принципиальную возможность измерения вероятности. Однако заметим,

<sup>12</sup>т.е. одного типа

<sup>13</sup>например, бросание кубика много раз

<sup>14</sup>примером признака может служить выпадение тройки при бросании кубика

<sup>15</sup>не употребляем слова “независимый” и “стремится”, т.к. даем формулировку на метаязыке (если употребить слово “стремится”, то нужно обязательно указать как (по вероятности, в среднем и т.д.); говорим “приближается” (качественно)



что полученная по этому “рецепту” вероятность приближенная даже при очень большом числе испытаний. В теории мы получаем точное значение вероятности, но при этом делаются некоторые дополнительные предположения (например, о симметричности монеты). Если мы сомневаемся в корректности данных предположений, то нужно их проверять (теория проверки гипотез).

Вся теория вероятности и математическая статистика основываются на законе больших чисел, который является законом природы.

Теперь поговорим о *случайности*. Рассмотрим две последовательности из 0 и 1:

10101010...

10011000111...

В обеих последовательностях частота появления 1 равна 0.5. Но случайность ли это? Очевидно, что в данных последовательностях присутствует закономерность, есть некий алгоритм построения последовательностей. Как в общем случае определить, является ли заданная последовательность случайной?

Подход Мизеса к этому вопросу следующий: если не существует выигрышной стратегии игры при бросании монетки, то получившаяся последовательность случайна; т.е. залогом случайности полагается нерегулярность, отсутствие системной игры.

Подход Колмогорова: если не существует достаточно “легкого” алгоритма построения последовательности, то она случайна. Здесь мы приходим к вопросу о сложности алгоритмов. Тезис Чёрча: все определения алгоритма эквивалентны. Таким образом, машина Тьюринга, машина Клини, оценка длины программы — все это асимптотически одно и то же. Поэтому подход Колмогорова можно сформулировать так: если длина программы сопоставима с длиной генерируемого ею слова, то последовательность случайна.

Подход Мартин-Лёфа: если последовательность нельзя отличить никакими алгоритмическими фильтрами от “действительно” случайной, то она случайна.

Получили три эквивалентных подхода к случайности.

Пуанкаре утверждал, что случайность появляется тогда, когда появляется неустойчивость.

Теперь остановимся на *субъективной интерпретации вероятности*.

Подход Колмогорова к пониманию вероятности не единственный, но наиболее распространенный. Далее в курсе будет убедительно показано, почему должна использоваться именно аксиоматика Колмогорова.

Субъективное мнение может в какой-то мере отражать действительность; вопрос в том, в какой мере и от чего это зависит. Это зависит, в первую очередь, от осведомленности субъекта, является ли он экспертом в данной области. По экспертным мнениям можно построить оценку субъективной вероятности (экспертная система); аппарат здесь такой же — аксиоматика Колмогорова.

Байесовский подход состоит в комбинации объективной и субъективной вероятности при помощи общей аксиоматики Колмогорова (так происходит, когда данных недостаточно для объективной оценки вероятности и приходится прибегать к помощи экспертов).

## 2 Условная вероятность как модель с дополнительным ограничением. Независимость событий. Схема Бернулли. Парадоксы фон Мизеса и д'Аламбера.

### 2.1 Условная вероятность как модель с дополнительным ограничением.

Допустим, нам дана модель  $(\Omega, \mathcal{F}^{16}, \mathbb{P})$  (что автоматически означает ее непротиворечивость). Далее появляется дополнительная информация: произошло  $\omega \in B \subset \Omega$ . Получаем новую модель. Постулат Колмогорова к ней применим, т.е. она может быть приведена к виду  $(\Omega_B, \mathcal{F}_B, \mathbb{P}_B)$ . Аксиоматика теории вероятности в некотором смысле неполна: по модели вид  $(\Omega, \mathcal{F}, \mathbb{P})$  определен неоднозначно. При построении  $(\Omega_B, \mathcal{F}_B, \mathbb{P}_B)$  будем исходить из здравого смысла.

Интерпретация  $\Omega$  — совокупность возможных исходов.

Интерпретация  $(\Omega_B, \mathcal{F}_B, \mathbb{P}_B)^{17}$  (два равновозможных варианта):

1.  $\Omega_B := \Omega^{18};$

$$\mathcal{F}_B := \mathcal{F};$$

*аксиома связи моделей  $(\Omega, \mathcal{F}, \mathbb{P})$  и  $(\Omega_B, \mathcal{F}_B, \mathbb{P}_B)$ :*

$$\mathbb{P}_B(A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (4)$$

2.  $\Omega_B := B;$

$$\mathcal{F}_B := \{A \in \mathcal{F} : A \subseteq B\};$$

*аксиома связи моделей  $(\Omega, \mathcal{F}, \mathbb{P})$  и  $(\Omega_B, \mathcal{F}_B, \mathbb{P}_B)$ :*

$$\mathbb{P}_B(A) := \frac{\mathbb{P}(A)}{\mathbb{P}(B)}. \quad (5)$$

Легко проверить, что  $\mathbb{P}_B$  — нормированная мера.

Приведенные аксиомы связи моделей нельзя доказать. Однако вот аргументы в пользу этих формул. Используем объективно-частотную интерпретацию. Рассмотрим схему Бернулли с двумя признаками типа “А” и “В”.

номер испытания	1	2	3	4	5	6	7	8	9	10
“А”	1	0	0	1	1	1	1	0	0	1
“В”	0	1	1	0	0	0	1	0	1	1

Всего  $N = 10$  испытаний. Признак “В” реализовался в  $N_B = 5$  испытаниях. Признаки “А” и “В” вместе реализовались в  $N_{AB} = 2$  испытаниях. Пусть  $f_B^N(A)$  — условная частота признака “А” (при условии, что “В” произошло). Тогда  $f_B^N(A) = \frac{N_{AB}}{N_B} = \frac{2}{5}$ . Используем закон больших чисел. Предположим, что вероятность появления “В” больше 0:  $f^N(B) = \frac{N_B}{N}$  при неограниченном увеличении числа испытаний  $N$  приближается (в вероятностном смысле) к  $p > 0$  (при  $N \rightarrow \infty$   $N_B$  приближается (в вероятностном смысле) к  $\infty$ ). Перепишем выражение для  $f_B^N(A)$  в виде  $f_B^N(A) = \frac{N_{AB}}{N_B} = \frac{\frac{N_{AB}}{N}}{\frac{N_B}{N}}$ . При неограниченном увеличении числа

<sup>16</sup>по умолчанию  $\mathcal{F}$  —  $\sigma$ -алгебра

<sup>17</sup>при предположении, что  $B \in \mathcal{F}$  и  $\mathbb{P}(B) > 0$

<sup>18</sup>значок “:=” означает “равен по определению”

испытаний  $N$  числитель приближается (по закону больших чисел) к вероятности такого события, при котором реализуются признаки “А” и “В” вместе, а знаменатель — к вероятности появления признака “В”. Подчеркнем еще раз, что это не доказательство, а правдоподобные рассуждения, основанные на интерпретации вероятности и законе больших чисел.

Далее будем считать, что записи  $\mathbb{P}_B(A)$  и  $\mathbb{P}(A|B)$  эквивалентны.

Формулу (4) можно переписать в более общем виде ( $\mathbb{P}(B)$  может быть равен и нулю:  $\mathbb{P}(B) = 0 \implies 0 \leq \mathbb{P}(A \cap B) \leq \mathbb{P}(B) = 0 \implies \mathbb{P}(A \cap B) = 0$ , если из каких-либо соображений определим  $\mathbb{P}_B(A)$ , то следующая формула верна и при  $\mathbb{P}(B) = 0$ ):

$$\mathbb{P}(A \cap B) = \mathbb{P}_B(A) \cdot \mathbb{P}(B), \quad \forall \mathbb{P}(B). \quad (6)$$

## 2.2 Независимость событий.

**Определение 14.** Событие  $A$  не зависит от события  $B$ , если

$$\begin{cases} \mathbb{P}_B(A) = \mathbb{P}(A), \\ \mathbb{P}_{B^c}(A) = \mathbb{P}(A), \end{cases} \quad (7)$$

т.е. частота появления события  $A$  не изменится, если произойдет событие  $B$ .

Замечания:

1. Условия в (7) равносильны:  $\mathbb{P}_B(A) = \mathbb{P}(A) \iff \mathbb{P}_{B^c}(A) = \mathbb{P}(A)$ .<sup>19</sup>
2. На 2 курсе давалось следующее определение: события  $A$  и  $B$  называются независимыми, если  $\mathbb{P}(AB) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ .
3. Если  $A$  не зависит от  $B$ , то и  $B$  не зависит от  $A \implies$  говорят о независимости событий  $A$  и  $B$ .

Пусть события  $A$  и  $B$  независимы. Применим формулу (6) и, учитывая  $\mathbb{P}_B(A) = \mathbb{P}(A)$ , получим:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B). \quad (8)$$

Заметим, что формула (8) симметрична относительно событий  $A$  и  $B$ .

**Определение 15.** Событие  $A$  называется почти достоверным, если  $\mathbb{P}(A) = 1$ .

**Определение 16.** Событие  $A$  называется почти невозможным, если  $\mathbb{P}(A) = 0$ .

**Определение 17.** Событие  $A$  называется тривиальным, если оно либо почти достоверно, либо почти невозможно. Событие  $A$  называется нетривиальным, если  $0 < \mathbb{P}(A) < 1$ .

Пусть событие  $B$  не является почти невозможным. События  $A$  и  $B$  называются независимыми, если  $\mathbb{P}_B(A) = \mathbb{P}(A)$ . Так как  $\mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ , то  $\mathbb{P}_B(A) = \mathbb{P}(A) \iff \mathbb{P}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \iff \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ . Пусть событие  $B$  почти невозможно. Тогда формула (8) выполнена автоматически. Получаем второе определение двух независимых событий.

**Определение 18.** События  $A$  и  $B$  называются **независимыми**, если выполнено (8).

**Утверждение 1.** События  $A$  и  $B$  независимы  $\iff$  события  $A$  и  $B^c$  независимы.

<sup>19</sup> Действительно, пусть  $\mathbb{P}(A) = \mathbb{P}_B(A) \implies \{\text{определение условной вероятности}\} \implies \mathbb{P}(A) = \mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \implies \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B) \implies \{\Omega = B + B^c; A = A \cap B + A \cap B^c\} \implies \mathbb{P}(A)(1 - \mathbb{P}(B^c)) = \mathbb{P}(A) - \mathbb{P}(A \cap B^c) \implies \mathbb{P}(A) = \frac{\mathbb{P}(A \cap B^c)}{\mathbb{P}(B^c)} = \mathbb{P}_{B^c}(A)$ .

Доказательство. События  $A$  и  $B$  независимы  $\iff \mathbb{P}_{B^c}(A) = \mathbb{P}(A) \iff$  события  $A$  и  $B^c$  независимы.

**Утверждение 2.** Событие  $B$  независимо с любым событием  $A \iff B$  тривиально.

Доказательство. Если  $B$  тривиально, то очевидно, что  $B$  независимо с любым событием  $A$ . Докажем в обратную сторону. Пусть событие  $B$  независимо с любым событием  $A \implies$  событие  $B$  независимо с событием  $B^c \implies \mathbb{P}(B^c \cap B) = \mathbb{P}(B^c) \cdot \mathbb{P}(B) \implies \mathbb{P}(\emptyset) = (1 - \mathbb{P}(B)) \cdot \mathbb{P}(B) \implies 0 = (1 - \mathbb{P}(B)) \cdot \mathbb{P}(B) \implies 1 - \mathbb{P}(B) = 0 \quad \cup \quad \mathbb{P}(B) = 0$ .

**Определение 19.** События  $A$ ,  $B$  и  $C$  называются **попарно независимыми**, если

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B), \quad (9)$$

$$\mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(C), \quad (10)$$

$$\mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C). \quad (11)$$

**Определение 20.** События  $A$ ,  $B$  и  $C$  называются **независимыми в совокупности**, если

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B), \quad (12)$$

$$\mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(C), \quad (13)$$

$$\mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C), \quad (14)$$

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C). \quad (15)$$

**Пример** (Бернштейн). Рассмотрим правильный тетраэдр с разноцветными гранями: К (красная грань), С (синяя грань), Б (белая грань), Т (триколор — красно-сине-белая грань). Этот тетраэдр кидают на прозрачный стол. Из симметрии (физического свойства системы) вероятность выпадения отдельной грани равна 0.25. Покажем, что события  $A_1$  (выпадение синего цвета),  $A_2$  (выпадение красного) и  $A_3$  (выпадение белого) попарно независимы, но не независимы в совокупности. Построим вероятностную модель. Естественно полагать, что  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ , где  $\omega_1$  — появление грани С,  $\omega_2$  — появление грани К,  $\omega_3$  — появление грани Б,  $\omega_4$  — появление грани Т.  $\mathbb{P}(\{\omega_i\}) = 0.25$ ,  $A_1 = \{\omega_1, \omega_4\}$ ,  $A_2 = \{\omega_2, \omega_4\}$ ,  $A_3 = \{\omega_3, \omega_4\}$ . Очевидно, что  $\mathbb{P}(A_i) = \{\text{в силу независимости элементарных событий } \omega_i\} = \mathbb{P}(\omega_i) \cdot \mathbb{P}(\omega_4) = 0.25 \cdot 0.25 = 0.5$ ,  $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(\omega_4) = 0.25$  при  $i \neq j$ ,  $\mathbb{P}(A_1 \cap A_2 \cap A_3) = \mathbb{P}(\omega_4) = 0.25$ . Отсюда легко видеть, что события  $A_1$ ,  $A_2$  и  $A_3$  попарно независимы, но не независимы в совокупности (по определению).

Этому примеру можно дать и другую, физическую интерпретацию. Рассмотрим систему из трех частиц, каждая из которых может находиться в одном из двух возможных состояний: '0' и '1'. Данная система может находиться в четырех различных состояниях:  $\omega_1$  (в состоянии '1' находится только первая частица),  $\omega_2$  (в состоянии '1' находится только вторая частица),  $\omega_3$  (в состоянии '1' находится только третья частица),  $\omega_4$  (все три частицы находятся в состоянии '1'). По условию,  $\mathbb{P}(\{\omega_i\}) = 0.25$ ,  $i = 1, 2, 3, 4$ . Рассмотрим события  $A_1$  (первая частица находится в состоянии '1'),  $A_2$  (вторая частица находится в состоянии '1') и  $A_3$  (третья частица находится в состоянии '1'). Аналогично тому, как это было показано выше, можно доказать, что события  $A_1$ ,  $A_2$  и  $A_3$  попарно независимы, но не независимы в совокупности<sup>20</sup>. Этот пример показывает, что, в отличие от классической физики, где поведение системы полностью описывается попарными взаимодействиями частиц, входящих в эту систему, в квантовой механике нужно знать характер взаимодействия частиц системы в совокупности.

<sup>20</sup>например, если первая частица находится в состоянии '1', а вторая — в состоянии '0', то третья точно будет находиться в состоянии '0'

**Пример.** Имеется 12 карточек с написанными на них числами от 1 до 12. Испытание состоит в том, что выбирается наугад одна карточка.

Событие А: число, написанное на карточке, четное.

Событие В: число, написанное на карточке, делится на 3.

Событие С: число, написанное на карточке, содержится среди чисел 2, 3, 4, 10, 11, 12.

$$\mathbb{P}(A) = \frac{1}{2}, \mathbb{P}(B) = \frac{1}{3}, \mathbb{P}(C) = \frac{1}{2}.$$

$$1. A \cap B \cap C = \{12\} \Rightarrow \frac{1}{12} = \mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{2}.$$

$$2. A \cap B = \{6, 12\} \Rightarrow \frac{1}{6} = \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) = \frac{1}{2} \cdot \frac{1}{3}.$$

$$3. B \cap C = \{3, 12\} \Rightarrow \frac{1}{6} = \mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C) = \frac{1}{3} \cdot \frac{1}{2}.$$

$$4. A \cap C = \{2, 4, 10, 12\} \Rightarrow \frac{1}{3} = \mathbb{P}(A \cap C) \neq \mathbb{P}(A)\mathbb{P}(C) = \frac{1}{2} \cdot \frac{1}{2}.$$

Получаем, что из свойств (13)-(15) не следует свойство (12).

Дадим еще одно определение независимости (в совокупности) 3 событий по аналогии с определением 14.

**Определение 21.** События  $A$ ,  $B$  и  $C$  называются **независимыми в совокупности**, если

$$\begin{cases} \mathbb{P}(A|B) = \mathbb{P}(A), \\ \mathbb{P}(A|B^c) = \mathbb{P}(A), \end{cases} \quad (A \text{ не зависит от } B) \quad (16)$$

$$\begin{cases} \mathbb{P}(C|AB) = \mathbb{P}(C), \\ \mathbb{P}(C|AB^c) = \mathbb{P}(C), \\ \mathbb{P}(C|A^cB) = \mathbb{P}(C), \\ \mathbb{P}(C|A^cB^c) = \mathbb{P}(C). \end{cases} \quad (C \text{ не зависит от } A \text{ и } B) \quad (17)$$

**Утверждение 3.** Определения 20 и 21 эквивалентны.

Доказательство.

1.  $20 \Rightarrow 21$ :

$$\begin{aligned} & \cdot \mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B) \Rightarrow \{(6)\} \Rightarrow \mathbb{P}(A|B) = \mathbb{P}(A) \Rightarrow \mathbb{P}(A|B^c) = \mathbb{P}(A), \\ & \cdot \mathbb{P}(ABC) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \Rightarrow \{\mathbb{P}(C|AB) = \frac{\mathbb{P}(ABC)}{\mathbb{P}(AB)}, \mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)\} \Rightarrow \\ & \quad \mathbb{P}(C|AB)\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \Rightarrow \mathbb{P}(C|AB) = \mathbb{P}(C), \\ & \cdot \mathbb{P}(AC) = \mathbb{P}(A)\mathbb{P}(C) \Rightarrow \{AC = ABC + AB^cC, A = AB + AB^c\} \Rightarrow \mathbb{P}(ABC) + \\ & \quad \mathbb{P}(AB^cC) = (\mathbb{P}(AB) + \mathbb{P}(AB^c))\mathbb{P}(C) \Rightarrow \{\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B), \mathbb{P}(A|B^c) = \mathbb{P}(A) \Rightarrow \\ & \quad \mathbb{P}(AB^c) = \mathbb{P}(A)\mathbb{P}(B^c)\} \Rightarrow \mathbb{P}(ABC) + \mathbb{P}(AB^cC) = (\mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(A)\mathbb{P}(B^c))\mathbb{P}(C) \\ & \quad \Rightarrow \{\mathbb{P}(ABC) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)\} \Rightarrow \mathbb{P}(AB^cC) = \mathbb{P}(A)\mathbb{P}(B^c)\mathbb{P}(C) \Rightarrow \{\mathbb{P}(C|AB^c) = \\ & \quad \frac{\mathbb{P}(AB^cC)}{\mathbb{P}(AB^c)}, \mathbb{P}(AB^c) = \mathbb{P}(A)\mathbb{P}(B^c)\} \Rightarrow \mathbb{P}(C|AB^c)\mathbb{P}(A)\mathbb{P}(B^c) = \mathbb{P}(A)\mathbb{P}(B^c)\mathbb{P}(C) \Rightarrow \\ & \quad \mathbb{P}(C|AB^c) = \mathbb{P}(C), \\ & \cdot \mathbb{P}(BC) = \mathbb{P}(B)\mathbb{P}(C) \Rightarrow \{BC = A^cBC + ABC, B = AB + A^cB\} \Rightarrow \mathbb{P}(A^cBC) + \\ & \quad \mathbb{P}(ABC) = (\mathbb{P}(AB) + \mathbb{P}(A^cB))\mathbb{P}(C) \Rightarrow \{\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B), \mathbb{P}(A|B) = \mathbb{P}(A) \Rightarrow \\ & \quad \mathbb{P}(B|A) = \mathbb{P}(B) \Rightarrow \mathbb{P}(B|A^c) = \mathbb{P}(B) \Rightarrow \mathbb{P}(A^cB) = \mathbb{P}(A^c)\mathbb{P}(B)\} \Rightarrow \mathbb{P}(ABC) + \\ & \quad \mathbb{P}(A^cBC) = (\mathbb{P}(A)\mathbb{P}(B) + \mathbb{P}(A^c)\mathbb{P}(B))\mathbb{P}(C) \Rightarrow \{\mathbb{P}(ABC) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)\} \Rightarrow \\ & \quad \mathbb{P}(A^cBC) = \mathbb{P}(A^c)\mathbb{P}(B)\mathbb{P}(C) \Rightarrow \{\mathbb{P}(C|A^cB) = \frac{\mathbb{P}(A^cBC)}{\mathbb{P}(A^cB)}, \mathbb{P}(A^cB) = \mathbb{P}(A^c)\mathbb{P}(B)\} \Rightarrow \\ & \quad \mathbb{P}(C|A^cB)\mathbb{P}(A^c)\mathbb{P}(B) = \mathbb{P}(A^c)\mathbb{P}(B)\mathbb{P}(C) \Rightarrow \mathbb{P}(C|A^cB) = \mathbb{P}(C), \end{aligned}$$

$$\begin{aligned}
& \cdot \mathbb{P}(ABC) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C), \mathbb{P}(A^cBC) = \mathbb{P}(A^c)\mathbb{P}(B)\mathbb{P}(C), \mathbb{P}(AB^cC) = \mathbb{P}(A)\mathbb{P}(B^c)\mathbb{P}(C), \\
& C = ABC + A^cBC + AB^cC + A^cB^cC \implies \mathbb{P}(C) = \mathbb{P}(ABC) + \mathbb{P}(A^cBC) + \mathbb{P}(AB^cC) + \\
& \mathbb{P}(A^cB^cC) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) + \mathbb{P}(A^c)\mathbb{P}(B)\mathbb{P}(C) + \mathbb{P}(A)\mathbb{P}(B^c)\mathbb{P}(C) + \mathbb{P}(A^cB^cC) \implies \\
& \mathbb{P}(C)(1 - \mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(A^c)\mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B^c)) = \mathbb{P}(A^cB^cC) \implies \{\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B), \\
& \mathbb{P}(A^cB) = \mathbb{P}(A^c)\mathbb{P}(B), \mathbb{P}(AB^c) = \mathbb{P}(A)\mathbb{P}(B^c)\} \implies \mathbb{P}(C)(1 - \mathbb{P}(AB) - \mathbb{P}(A^cB) - \\
& \mathbb{P}(AB^c)) = \mathbb{P}(A^cB^cC) \implies \{\Omega = AB + A^cB + AB^c + A^cB^c\} \implies \mathbb{P}(C)\mathbb{P}(A^cB^c) = \\
& \mathbb{P}(A^cB^cC) \implies \{\mathbb{P}(A^cB) = \mathbb{P}(A^c)\mathbb{P}(B) \implies \mathbb{P}(A^c|B) = \mathbb{P}(A^c) \implies \mathbb{P}(A^c|B^c) = \\
& \mathbb{P}(A^c) \implies \mathbb{P}(A^cB^c) = \mathbb{P}(A^c)\mathbb{P}(B^c)\} \implies \mathbb{P}(C)\mathbb{P}(A^c)\mathbb{P}(B^c) = \mathbb{P}(A^cB^cC) \implies \{\mathbb{P}(C|A^cB^c) = \\
& \frac{\mathbb{P}(A^cB^cC)}{\mathbb{P}(A^cB^c)}, \mathbb{P}(A^cB^c) = \mathbb{P}(A^c)\mathbb{P}(B^c)\} \implies \mathbb{P}(C|A^cB^c)\mathbb{P}(A^c)\mathbb{P}(B^c) = \mathbb{P}(A^c)\mathbb{P}(B^c)\mathbb{P}(C) \\
& \implies \mathbb{P}(C|A^cB^c) = \mathbb{P}(C);
\end{aligned}$$

2. 21  $\implies$  20: аналогично;

Заметим, что в определении 21 два из шести соотношений зависят от других четырех (одно из (16) и одно из (17)).

Дадим определение независимости  $n$  событий.

**Определение 22.** События  $A_1, A_2, \dots, A_n$  называются **независимыми в совокупности**, если

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i), \quad \forall \text{ множества индексов } I \subseteq \{1, \dots, n\}, |I| \geq 2. \quad (18)$$

Посчитаем, сколько независимых соотношений написано в (18):  $C_n^2 + C_n^3 + \dots + C_n^n = (1+1)^n - C_n^1 - C_n^0 = 2^n - n - 1$ .

Дадим теперь еще одно эквивалентное определение независимости  $n$  событий.

Введем обозначение

$$A^\sigma = \begin{cases} A^c, & \sigma = 0, \\ A, & \sigma = 1, \end{cases} \quad \sigma \in \{0, 1\}.$$

**Определение 23.** События  $A_1, A_2, \dots, A_n$  называются **независимыми в совокупности**, если

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i^{\sigma_i}\right) = \prod_{i=1}^n \mathbb{P}(A_i^{\sigma_i}), \quad \forall \sigma_i \in \{0, 1\}, \quad i = 1, \dots, n. \quad (19)$$

Заметим, что в (19) написано  $2^n$  соотношений, т.е. на  $n+1$  больше, чем в (18).

**Утверждение 4.** Определения 22 и 23 эквивалентны.

Доказательство.

1. 22  $\implies$  23: Будем доказывать

$$\begin{aligned}
& \mathbb{P}\left(\bigcap_{i \in I} A_i^{\sigma_i}\right) = \prod_{i \in I} \mathbb{P}(A_i^{\sigma_i}), \\
& \forall \sigma_i \in \{0, 1\}, \quad i = 1, \dots, n, \quad \forall \text{ множества индексов } I \subseteq \{1, \dots, n\}, |I| \geq 2.
\end{aligned} \quad (20)$$

Индукция по  $k = \sum_{i=1}^n \sigma_i$ :

$k = n$  : очевидно;

$k + 1 \rightarrow k \geq 0$  : пусть  $\sigma_j = 0$ ; заметим, что выполнение (20) для  $\forall$  множества индексов  $I \subseteq \{1, \dots, n\}$ ,  $j \notin I$ ,  $|I| \geq 2$  очевидно (достаточно положить  $\sigma_j = 1$  и применить индуктивное предположение); докажем  $\mathbb{P}\left(\bigcap_{i=1}^n A_i^{\sigma_i}\right) = \prod_{i=1}^n \mathbb{P}(A_i^{\sigma_i})$ :  $\mathbb{P}\left(\bigcap_{i=1}^n A_i^{\sigma_i}\right) = \mathbb{P}\left(\bigcap_{\substack{i=1 \\ i \neq j}}^n A_i^{\sigma_i} \cap A_j^c\right) = \{\text{аддитивность меры } \mathbb{P}\} = \mathbb{P}\left(\bigcap_{\substack{i=1 \\ i \neq j}}^n A_i^{\sigma_i}\right) - \mathbb{P}\left(\bigcap_{\substack{i=1 \\ i \neq j}}^n A_i^{\sigma_i} \cap A_j\right) = \{\text{индуктивное предположение}\} = \prod_{\substack{i=1 \\ i \neq j}}^n \mathbb{P}(A_i^{\sigma_i}) - \prod_{\substack{i=1 \\ i \neq j}}^n \mathbb{P}(A_i^{\sigma_i}) \cdot \mathbb{P}(A_j) = \prod_{\substack{i=1 \\ i \neq j}}^n \mathbb{P}(A_i^{\sigma_i}) \cdot (1 - \mathbb{P}(A_j)) = \prod_{\substack{i=1 \\ i \neq j}}^n \mathbb{P}(A_i^{\sigma_i}) \cdot \mathbb{P}(A_j^c) = \prod_{i=1}^n \mathbb{P}(A_i^{\sigma_i})$ .

2. 23  $\implies$  22: Будем доказывать опять (20), откуда очевидным образом следует (18). Индукция по  $k = |I|$ :

$k = n$  : очевидно;

$k + 1 \rightarrow k \geq 2$  : пусть  $j \notin I$ ; докажем  $\mathbb{P}\left(\bigcap_{i \in I} A_i^{\sigma_i}\right) = \prod_{i \in I} \mathbb{P}(A_i^{\sigma_i})$ :  $\mathbb{P}\left(\bigcap_{i \in I} A_i^{\sigma_i}\right) = \{\text{аддитивность меры } \mathbb{P}\} = \mathbb{P}\left(\bigcap_{i \in I} A_i^{\sigma_i} \cap A_j\right) + \mathbb{P}\left(\bigcap_{i \in I} A_i^{\sigma_i} \cap A_j^c\right) = \{\text{индуктивное предположение}\} = \prod_{i \in I} \mathbb{P}(A_i^{\sigma_i}) \cdot \mathbb{P}(A_j) + \prod_{i \in I} \mathbb{P}(A_i^{\sigma_i}) \cdot \mathbb{P}(A_j^c) = \prod_{i \in I} \mathbb{P}(A_i^{\sigma_i}) \cdot (\mathbb{P}(A_j) + \mathbb{P}(A_j^c)) = \prod_{i \in I} \mathbb{P}(A_i^{\sigma_i})$ .

Приведем еще одно определение независимости событий в совокупности, эквивалентное первым двум (это легко показать).

**Определение 24.** События  $A_1, A_2, \dots, A_n$  называются **независимыми в совокупности**, если

$$\mathbb{P}\left(\bigcap_{i=1}^n B_i\right) = \prod_{i=1}^n \mathbb{P}(B_i), \quad \forall B_i \in \mathcal{A}_i, \quad i = 1, \dots, n, \quad (21)$$

где  $\mathcal{A}_i$  —  $\sigma$ -алгебра, порожденная событием  $A_i$ , т.е. минимальная  $\sigma$ -алгебра, содержащая событие  $A_i$ :  $\mathcal{A}_i = \{A_i, A_i^c, \emptyset, \Omega\}$ .

**Определение 25.** Выражение

$$C_{\sigma_1 \dots \sigma_n} = \bigcap_{i=1}^n A_i^{\sigma_i}$$

называется **конституэнтной**.

Легко понять, что всего  $2^n$  конституэнт. Заметим, что все конституэнты всегда попарно не пересекаются (действительно, возьмем два различных набора  $(\sigma_1 \dots \sigma_n) \neq (\sigma'_1 \dots \sigma'_n)$ ; тогда  $C_{\sigma_1 \dots \sigma_n} \cap C_{\sigma'_1 \dots \sigma'_n} = \emptyset$ ). Однако ничто не мешает какой-нибудь конституэнтке быть пустой.

**Утверждение 5.** Пусть есть  $n$  независимых в совокупности нетривиальных событий. Тогда пространство элементарных событий содержит не менее  $2^n$  точек, т.е.  $|\Omega| \geq 2^n$ .

Доказательство. Пусть события  $A_1, A_2, \dots, A_n$  независимы в совокупности и  $0 < \mathbb{P}(A_i) < 1 \quad \forall i$ . Тогда, по определению 23,

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i^{\sigma_i}\right) = \prod_{i=1}^n \mathbb{P}(A_i^{\sigma_i}), \quad \forall \sigma_i \in \{0, 1\}, \quad i = 1, \dots, n.$$



В силу нетривиальности событий  $A_i$ ,  $\prod_{i=1}^n \mathbb{P}(A_i^{\sigma_i}) > 0$ ,  $\forall \sigma_i \in \{0, 1\}$ ,  $i = 1, \dots, n$ . Таким образом, все конституэнты имеют положительную вероятность, а значит, непустые. Так как конституэнты попарно не пересекаются, то получаем, что они составляют множество попарно не пересекающихся  $2^n$  событий, что и доказывает утверждение.

Это утверждение помогает при построении примеров.

## 2.3 Схема Бернулли.

Рассмотрим пространство элементарных событий, состоящее из  $2^n$  элементов:  $|\Omega| = 2^n$ . По утверждению 5, на нем можно задать  $n$  независимых событий. Сделаем это явно, построив вероятностное пространство, используя схему Бернулли из  $n$  испытаний (см. первую главу).

1.  $\Omega := \left\{ \omega = (\alpha_1, \dots, \alpha_n) \mid \alpha_i = \begin{cases} 1, & \text{если успех в } i \text{ испытании} \\ 0, & \text{если неуспех в } i \text{ испытании} \end{cases} \right\} \Rightarrow |\Omega| = 2^n$ .
2.  $\mathcal{F} := 2^\Omega$  — множество всех подмножеств  $\Omega \Rightarrow |\mathcal{F}| = 2^{2^n}$ .
3.  $\mathbb{P}(\omega) = p_\omega \quad \forall \omega$ , причем  $p_\omega \geq 0$ ,  $\sum_{\omega \in \Omega} p_\omega = 1$ ; положим  $\mathbb{P}(A) := \sum_{\omega \in A} p_\omega$ . Легко видеть, что это нормированная мера.

Итак, осталось задать вероятность одноточечных множеств  $p_\omega$ . Для этого формализуем ту модель схемы Бернулли, которую ранее определили на метаязыке (см. первую главу).

**Определение 26.** Схемой Бернулли называется модель, состоящая из  $n$  событий  $A_1, A_2, \dots, A_n$  таких, что:

1. События  $A_i$  независимы в совокупности.
2. Вероятности всех событий одинаковы:  $\mathbb{P}(A_i) = p \quad i = 1, \dots, n$ .

Заметим, что событие  $A_i$  отвечает успеху в  $i$  испытании. Тогда независимость событий означает отсутствие взаимного влияния в испытаниях, а равновероятность событий — воспроизводимость однородных испытаний.

Приведем еще одну эквивалентную формализацию схемы Бернулли.

**Определение 27.** Схемой Бернулли называется модель, состоящая из последовательности  $n$  случайных величин<sup>21</sup>  $\xi_1, \xi_2, \dots, \xi_n$ , принимающих значения 0 и 1 и таких, что:

1. Случайные величины  $\xi_i$  независимы.
2. Случайные величины  $\xi_i$  одинаково распределены.

Заметим, что

$$\xi_i = \begin{cases} 1, & \text{если успех в } i \text{ испытании,} \\ 0, & \text{если неуспех в } i \text{ испытании.} \end{cases}$$

Тогда независимость случайных величин означает отсутствие взаимного влияния в испытаниях, а одинаковая распределенность — воспроизводимость однородных испытаний.

Заметим, что вторая модель сводится к первой, если положить  $\xi_i = I_{A_i}$ , где  $I_{A_i}$  — индикатор события  $A_i$ .

<sup>21</sup>определение случайной величины будет дано позже



Вернемся к построению вероятностного пространства. Вероятность одноточечных множеств  $p_\omega$  задается следующим образом:

$$\mathbb{P}(\{(\alpha_1, \alpha_2, \dots, \alpha_n)\}) = p_{(\alpha_1, \alpha_2, \dots, \alpha_n)} := p^{\sum_{i=1}^n \alpha_i} \cdot (1-p)^{n - \sum_{i=1}^n \alpha_i},$$

где  $p$  — вероятность успеха в  $i$  испытании.

Очевидно, что  $p_\omega \geq 0$ . Проверим условие нормировки:

$$\sum_{\omega \in \Omega} p_\omega = \sum_{(\alpha_1, \alpha_2, \dots, \alpha_n)} p^{\sum_{i=1}^n \alpha_i} \cdot (1-p)^{n - \sum_{i=1}^n \alpha_i} = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} = \{\text{бином Ньютона}\} = (p + (1-p))^n = 1.$$

Таким образом, мы построили вероятностное пространство.

Заметим, что события  $A_i = \{\omega = (\alpha_1, \dots, \alpha_n) : \alpha_i = 1\}$ , соответствующие успеху в  $i$  испытании, при таком построении являются независимыми в совокупности:

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^n A_i^{\sigma_i}\right) &= \mathbb{P}(\{(\sigma_1, \dots, \sigma_n)\}) = p^{\sum_{i=1}^n \sigma_i} \cdot (1-p)^{n - \sum_{i=1}^n \sigma_i} = \\ &= \{\mathbb{P}(A_i) = p\} = \prod_{i=1}^n \mathbb{P}(A_i^{\sigma_i}), \quad \forall \sigma_i \in \{0, 1\}, \quad i = 1, \dots, n. \end{aligned}$$

Итак, мы доказали, что в пространстве элементарных событий, состоящем из  $2^n$  элементов, можно задать  $n$  независимых событий.

Заметим, что построенное вероятностное пространство является, с одной стороны, доказательством непротиворечивости, корректности описанной выше модели схемы Бернулли, а с другой — еще одним способом формализации схемы Бернулли, эквивалентным первым двум.

## 2.4 Парадоксы фон Мизеса и д’Аламбера.

Рассмотрим *парадокс д’Аламбера*, называемый “Орел и решка”. Бросаем (честно) одновременно две идентичные симметричные монеты. Классическое определение вероятности таково:

$$\text{вероятность} = \frac{\text{число благоприятных событий}}{\text{число возможных событий}}.$$

Полагается, что в эксперименте возможно всего три исхода:

1. выпадение двух орлов;
2. выпадение одного орла и одной решки;
3. выпадение двух решек.

Тогда, по классическому определению вероятности, вероятность каждого исхода равна  $\frac{1}{3}$ . В рамках классической модели все верно, однако вывод противоречит здравому смыслу (опыту). Получается, что классический подход может приводить к неадекватным моделям.

---


$$\begin{aligned} {}^{22}\text{действительно, } \mathbb{P}(A_i) &= \sum_{(\alpha_1, \dots, \alpha_n): \alpha_i=1} p_{(\alpha_1, \alpha_2, \dots, \alpha_n)} = \sum_{(\alpha_1, \dots, \alpha_n): \alpha_i=1} p^{\sum_{j=1}^n \alpha_j} \cdot (1-p)^{n - \sum_{j=1}^n \alpha_j} = \\ &= \sum_{(\alpha_1, \dots, \alpha_n): \alpha_n=1} p^{\sum_{j=1}^n \alpha_j} \cdot (1-p)^{n - \sum_{j=1}^n \alpha_j} = p \cdot \sum_{(\alpha_1, \dots, \alpha_{n-1})} p^{\sum_{j=1}^{n-1} \alpha_j} \cdot (1-p)^{n-1 - \sum_{j=1}^{n-1} \alpha_j} = p \cdot \sum_{k=0}^{n-1} C_{n-1}^k p^k (1-p)^{n-1-k} = \\ &= \{\text{бином Ньютона}\} = p(p + (1-p))^{n-1} = p \end{aligned}$$

Проблема состоит в том, что теория вероятности не является чисто математической дисциплиной. Поэтому спор о построении адекватных моделей должен вестись на метаязыке. Д’Аламбер “не замечает”, что указанные три исхода не являются равновероятными. Ведь классическое определение вероятности исходит из равновероятности исходов, чего нет в этой модели. С точки зрения логики построения модели к адекватной модели приводят следующие рассуждения. Утверждается, что в рамках данного эксперимента имеется схема Бернулли с двумя независимыми однородными испытаниями. В силу симметрии монеты вероятность выпадения орла равна  $\frac{1}{2}$ . Из независимости получаем, что вероятность выпадения двух орлов равна  $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ .

Рассмотрим разбиение  $\Omega$  на конечное или счетное число попарно непересекающихся множеств  $H_i$ :  $\Omega = \sum_i H_i$ . Традиционно, множества  $H_i$  называют *гипотезами*. Рассмотрим две задачи. Пусть  $A$  — некоторое событие ( $A \in \mathcal{F}$ ). В обеих задачах нам даны  $\mathbb{P}(A|H_i)$ <sup>23</sup> и  $\mathbb{P}(H_i)$ .

1. Найти  $\mathbb{P}(A)$ .
2. Найти  $\mathbb{P}(H_i|A)$ .

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(A \cap \Omega) = \mathbb{P}(A \cap \sum_i H_i) = \mathbb{P}(\sum_i A \cap H_i) = \{\text{свойство счетной аддитивности}\} = \\ &= \sum_i \mathbb{P}(A \cap H_i) = \{\text{определение условной вероятности, а точнее аксиома связи моделей, а если еще точнее, то формула (6), т.к. } \mathbb{P}(H_i) \text{ может равняться нулю}\} = \sum_i \mathbb{P}(A|H_i) \cdot \mathbb{P}(H_i). \end{aligned}$$

Таким образом, получили **формулу полной вероятности**:

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A|H_i) \cdot \mathbb{P}(H_i). \quad (22)$$

Заметим, что фактически мы показали эквивалентность формул (6) и (22).<sup>24</sup>

Для решения второй задачи предположим, что  $\mathbb{P}(A) > 0$ .

$$\begin{aligned} \mathbb{P}(H_i|A) &= \{\text{аксиома связи моделей (4) в определении условной вероятности}\} = \\ &= \frac{\mathbb{P}(A \cap H_i)}{\mathbb{P}(A)} = \{\text{формула полной вероятности}\} = \frac{\mathbb{P}(A \cap H_i)}{\sum_k \mathbb{P}(A|H_k) \cdot \mathbb{P}(H_k)} = \\ &= \{\text{формула (6)}\} = \frac{\mathbb{P}(H_i) \cdot \mathbb{P}(A|H_i)}{\sum_k \mathbb{P}(A|H_k) \cdot \mathbb{P}(H_k)}. \end{aligned}$$

Таким образом, получили **формулу Байеса**:

$$\mathbb{P}(H_i|A) = \frac{\mathbb{P}(H_i) \cdot \mathbb{P}(A|H_i)}{\sum_k \mathbb{P}(A|H_k) \cdot \mathbb{P}(H_k)}. \quad (23)$$

<sup>23</sup>напомним, что записи  $\mathbb{P}_B(A)$  и  $\mathbb{P}(A|B)$  эквивалентны

<sup>24</sup>формулу полной вероятности мы вывели из формулы (6); обратно, формула (6) выводится из формулы (22), если заменить  $A \rightarrow A \cap B$ ,  $H_1 \rightarrow B$  и учесть, что  $\mathbb{P}(A \cap B|H_i) = 0$  при  $i \neq 1$

**Пример.** Пусть в группе 10% хороших студентов, 70% средних студентов, 20% плохих студентов. Вероятность получить плохую оценку на экзамене для хороших студентов равна 0.1, для средних студентов — 0.5, для плохих студентов — 0.9. Какова вероятность того, что вышедший с двойкой студент хороший/средний/плохой? Формализуем задачу:  $H_1$  — хороший студент,  $H_2$  — средний студент,  $H_3$  — плохой студент;  $A$  — получение неудовлетворительной оценки. Из условий задачи получаем, что  $\mathbb{P}(H_1) = 0.1$ ,  $\mathbb{P}(H_2) = 0.7$ ,  $\mathbb{P}(H_3) = 0.2$ ;  $\mathbb{P}(A|H_1) = 0.1$ ,  $\mathbb{P}(A|H_2) = 0.5$ ,  $\mathbb{P}(A|H_3) = 0.9$ . По формуле Байеса легко найти искомые вероятности  $\mathbb{P}(H_1|A)$ ,  $\mathbb{P}(H_2|A)$ ,  $\mathbb{P}(H_3|A)$ .

**Парадокс фон Мизеса.** Теннисист собирается участвовать в одном из двух турниров, один из которых проходит в Лондоне, а другой — в Париже (турниры происходят одновременно). Вероятность выигрыша в Лондоне равна 0.8, а в Париже — 0.7. Пусть событие  $A_1$  соответствует выигрышу турнира в Лондоне, а событие  $A_2$  — в Париже,  $A$  — просто факту выигрыша. Очевидно, что события  $A_1$  и  $A_2$  несовместны. Значит, по аксиоме конечной аддитивности получаем, что  $\mathbb{P}(A) = \mathbb{P}(A_1) + \mathbb{P}(A_2) = 0.8 + 0.7 = 1.5 > 1$ , что противоречит определению вероятности.

Парадокс объясняется неверно построенной моделью. Построим адекватную модель. Пусть событие  $P$  соответствует тому, что теннисист играл в Париже, а событие  $L$  — в Лондоне. Тогда вероятность выиграть при условии, что игра происходила в Париже, равна  $\mathbb{P}_P(A) = 0.7$ , а в Лондоне  $\mathbb{P}_L(A) = 0.8$ . Однако складывать эти величины нельзя, т.к. они относятся к разным моделям. По формуле полной вероятности получаем, что вероятность выигрыша составляет  $\mathbb{P}(A) = \mathbb{P}(P)\mathbb{P}_P(A) + \mathbb{P}(L)\mathbb{P}_L(A)$ . Отсюда легко видеть, что  $\mathbb{P}_P(A) = 0.7 \leq \mathbb{P}(A) \leq 0.8 = \mathbb{P}_L(A)$ .

Проведем три эксперимента.

**Эксперимент 1. Задание.** Рассмотрим схему Бернулли. Монету бросили 9 раз. Все 9 раз получили герб. Что будет следующим? *Ответ группы испытуемых.* Большинство думает, что в следующий раз скорее всего будет решка. *Ответ, основанный на математической статистике.* В следующий раз скорее всего будет герб. Основание: в схеме Бернулли испытания независимы и проводятся в сходных, аналогичных условиях  $\implies$  монета с высокой долей вероятности несимметричная.

**Эксперимент 2. Задание.** Сымитировать генератор случайных чисел, т.е. написать случайную последовательность из нулей и единиц длиной в 30 цифр. *Ответ группы испытуемых.* Будем считать длины подпоследовательностей, состоящих только из 0 или только из 1, например: 1100101000  $\rightarrow$  221113. Получаем:

длина подпоследовательности	1	2	3	4	5	6	7	8	9
сколько раз встретилась	141	86	48	15	5	5	0	1	1

Опираясь на закон больших чисел, получим из опыта:

длина п/п	1	2	3	4	5	6	7	8	9
вероятность	$\frac{141}{302} \approx 0.4669$	$\frac{43}{151} \approx 0.2848$	$\frac{24}{151} \approx 0.1589$	$\frac{15}{302} \approx 0.0497$	$\frac{5}{302} \approx 0.0166$	$\frac{5}{302} \approx 0.0166$	0	$\frac{1}{302} \approx 0.0033$	$\frac{1}{302} \approx 0.0033$

*Ответ, основанный на математической статистике.* По схеме Бернулли получаем

длина п/п	1	2	3	4	5	6	7	8	9
вероятность	$\frac{1}{2} = 0.5$	$\frac{1}{2^2} = 0.25$	$\frac{1}{2^3} = 0.125$	$\frac{1}{2^4} = 0.0625$	$\frac{1}{2^5} \approx 0.0313$	$\frac{1}{2^6} \approx 0.0156$	$\frac{1}{2^7} \approx 0.0078$	$\frac{1}{2^8} \approx 0.0039$	$\frac{1}{2^9} \approx 0.002$

**Вывод.** “Человеческий” генератор случайных чисел делает перекося в сторону коротких серий.

**Эксперимент 3.** (эксперимент Эдвардса) *Задание.* Даны два непрозрачных мешка, в которых находятся фишки двух цветов, зеленого и белого. В одном мешке 70% белых фишек (назовем такой мешок “беловатым”), а в другом — 70% зеленых (“зеленоватый” мешок). Предъявляется какой-то из этих мешков и осуществляется выбор с возвращением ( $\Rightarrow$  рассматривается схема Бернулли). Проводится 12 испытаний Бернулли, т.е. 12 раз вытаскивают фишку из данного мешка и кладут ее обратно. Событие  $A$  состоит в том, что вытащили 8 зеленых и 4 белых фишки. Спрашивается, какова вероятность того, что мешок “зеленоватый”? *Ответ группы испытуемых.* Большинство думает, что примерно 70%. *Ответ, основанный на математической статистике.* Воспользуемся формулой Байеса. Пусть событие  $G$  соответствует выбору “зеленоватого” мешка, а событие  $W$  — “беловатого”. Тогда

$$\mathbb{P}(A|G) = C_{12}^8 \cdot 0.7^8 \cdot 0.3^4, \quad \mathbb{P}(A|W) = C_{12}^4 \cdot 0.3^8 \cdot 0.7^4 \quad (C_{12}^4 = C_{12}^8);$$

$$\begin{aligned} \mathbb{P}(G|A) &= \frac{\mathbb{P}(A|G) \cdot \mathbb{P}(G)}{\mathbb{P}(G) \cdot \mathbb{P}(A|G) + \mathbb{P}(W) \cdot \mathbb{P}(A|W)} = \{\mathbb{P}(G) = \mathbb{P}(W) = 0.5\} = \\ &= \frac{C_{12}^8 \cdot 0.7^8 \cdot 0.3^4 \cdot 0.5}{0.5 \cdot C_{12}^8 \cdot 0.7^8 \cdot 0.3^4 + 0.5 \cdot C_{12}^4 \cdot 0.3^8 \cdot 0.7^4} = \frac{0.7^4}{0.7^4 + 0.3^4} \approx 97\% \end{aligned}$$

*Вывод.* Большинство людей сильно недооценивают небольшую выборку.

### 3 Пренебрежимые модификации модели. Нуль множества и пополнение вероятностного пространства. Непрерывные и дискретные модели. Теорема Улама (частный случай). Роль аксиомы выбора. Борелевская $\sigma$ -алгебра. Теорема об изоморфизме.

#### 3.1 Пренебрежимые модификации модели.

Одно и то же явление может быть описано разными способами, т.е. для одного явления можно построить 2 вероятностных пространства. Как определить, описывают ли 2 разных вероятностных пространства одно явление или нет?

Для этого рассмотрим преобразования вероятностного пространства.

##### 1. Переименование переменных.

Пусть задано вероятностное пространство  $(\Omega, \mathcal{F}, \mathbb{P})$ .

Пусть  $\varphi: \Omega \rightarrow \Omega'$  — биективное отображение.

$\mathcal{F}' = \{B \subseteq \Omega' : \varphi^{-1}(B) \in \mathcal{F}\}$  —  $\sigma$ -алгебра  $\Omega'$ , т.к.  $\mathcal{F}$  —  $\sigma$ -алгебра.

$\mathbb{P}'(B) := \mathbb{P}(\varphi^{-1}(B))$  — нормированная мера.

Получаем следующее преобразование:

$$\varphi: (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\Omega', \mathcal{F}', \mathbb{P}')$$

$$\varphi': \Omega' \rightarrow \Omega$$

Фактически  $\varphi$  и  $\varphi'$  задают переименование.

**Пример.** Переименование переменных.

$$\omega = 01100001$$

$$\omega' = \text{НУУННННУ}$$

При моделировании вероятностные пространства, получающиеся путем переименования, эквивалентны.

2. В теории вероятностей все изучается с точностью до "почти наверное" (т.е. мы пренебрегаем теми вещами, которые имеют вероятность нуль).

Формализуем этот основополагающий принцип (формализация может быть разная).

Рассмотрим вероятностное пространство  $(\Omega', \mathcal{F}', \mathbb{P}')$ .

$$\Omega' = \Omega + \Delta$$

Под "+" понимаем объединение непересекающихся множеств.

$$\mathbb{P}'(\Delta) = 0, (\Delta \in \mathcal{F}')$$

Рассмотрим вероятностное пространство  $(\Omega, \mathcal{F}, \mathbb{P})$ .

$$\mathcal{F} = \{B \in \mathcal{F}' | B \subseteq \Omega\} \Rightarrow \mathcal{F} \subseteq \mathcal{F}'$$

$\mathbb{P} = \mathbb{P}'|_{\mathcal{F}}$  – сужение  $\mathbb{P}'$  на  $\mathcal{F}$ .

Таким образом, вероятностное пространство  $(\Omega, \mathcal{F}, \mathbb{P})$  получается выкидыванием множества нулевой вероятности, а вероятностное пространство  $(\Omega', \mathcal{F}', \mathbb{P}')$  получается добавлением множества нулевой вероятности  $\Rightarrow$  эти 2 разных вероятностных пространства описывают одно и то же явление.

**Определение 28.** Два вероятностных пространства будем называть **изоморфными**, если они могут быть получены одно из другого с помощью двух описанных выше преобразований (переименование переменных; добавление или выкидывание множества нулевой вероятности).

Пространства изоморфны  $\Rightarrow$  пространства описывают одно и то же явление. Обратное не всегда верно.

**Пример.** Рассмотрим следующий эксперимент. Производится бросание симметричной игральной кости до тех пор, пока не выпадет первая шестерка. Введем пространство элементарных событий  $\Omega$ .

Переименование переменных.

$$\Omega = \{K_1, K_2, K_3, K_4, K_5, K_6\}$$

$$\Omega' = \{1, 2, 3, 4, 5, 6\}$$

Пусть при моделировании получились следующие элементарные исходы:

$$\Omega : 23114256; 44516; 6; \dots$$

Опишем эту модель:  $\mathbb{P}(6) = \frac{1}{6}$ .  $\mathbb{P}(44516) = \frac{1}{6^5}$  – всего может быть  $5^4$  исходов такого типа (это следует из симметрии и отсутствия взаимного влияния).

Условие нормировки проверим, просуммировав соответствующий ряд:

$$\frac{1}{6} + \frac{5}{6^2} + \frac{5^2}{6^3} + \dots = \sum_{i=1}^{\infty} \frac{1}{6^i} 5^{i-1} = 1$$

$\Rightarrow$  нормированная мера.

В результате мы получили дискретное вероятностное пространство, такое что мощность пространства элементарных исходов  $\Omega$  не более чем счетно.  $\sigma$ -алгебра событий может быть задана как множество всех подмножеств  $\Omega$ . Вероятность события задается как сумма вероятностей элементарных исходов, вошедших в событие.

Зададимся вопросом, возможна ли ситуация, когда 6 никогда не выпадет? В принципе такая ситуация возможна. В этом случае мы имеем исход бесконечной длины, например: 23352.... Рассмотрение такого исхода приводит нас к другой вероятностной модели. В данном случае  $\Omega'$  состоит из континуума событий, т.е.  $\Omega'$  имеет мощность континуума.

Заметим, что  $\Omega \subseteq \Omega'$ , и по условию нормировки естественно доопределить  $\mathbb{P}(\Omega' \setminus \Omega) = 0$

**Пример.** Рассмотрим некий прямоугольник и будем случайным образом выбирать точку внутри этого прямоугольника. Возьмем в качестве  $\Omega$  весь прямоугольник.  $\mathbb{P}(\text{попадания точки в прямоугольник}) = 0 \Rightarrow$  казалось бы, можно выкинуть из прямоугольника все точки. Однако же, если мы выкинем из прямоугольника все точки, то мы получим  $\Omega' = \emptyset \neq \Omega \Rightarrow$  использовать второй принцип, т.е. выкидывать множества нулевой вероятности надо очень осторожно.

### 3.2 Непрерывные и дискретные модели. Теорема Улама (частный случай).

**Определение 29.** Пусть мощность пространства элементарных событий  $\Omega$  не более, чем счетно  $\Rightarrow$  задано **дискретное вероятностное пространство**.

**Определение 30.** Пространство элементарных событий  $\Omega$  называется **дискретным**, если оно изоморфно пространству с не более, чем счетным числом элементов.

Если пространство не дискретно, то оно **непрерывно**.

$\sigma$ -алгебра событий отражает информацию о модели, а ее выбор зависит от информации, доступной в рамках задачи.

**Определение 31.** При фиксированном дискретном пространстве  $\Omega$  вероятностная модель, в которой  $\mathcal{F} = 2^\Omega$ , называется **максимально информативной моделью**, а модель, в которой  $\mathcal{F} = \{\emptyset, \Omega\}$ , называется **минимально информативной**.

*Рассмотрим дискретное вероятностное пространство.*

Пусть  $\Omega$  — дискретное пространство элементарных событий.

$$\exists \Omega' : \Omega' \subseteq \Omega, \mathbb{P}(\Omega') = 1, \Omega' = \{\omega_1, \omega_2, \dots\}$$

Рассмотрим множество  $\Omega \setminus \Omega'$  и исключим его из исходного пространства элементарных событий  $\Omega$ . Остальные элементы в  $\Omega$  оставим без изменений. Тогда  $\Omega$  изоморфно  $\Omega'$ .

Нас будут интересовать вероятности наступления всевозможных событий. В случае дискретного вероятностного пространства можно задать вероятность на всех подмножествах  $\Omega$ , т.е. можно взять  $\mathcal{F} = 2^\Omega$ . Для этого достаточно задать вероятности всех одноэлементных множеств  $\omega \in \Omega'$ , т.е. элементарных исходов, имеющих ненулевую вероятность (т.к. элементарные исходы, имеющие нулевую вероятность, можно исключить из вероятностного пространства):

$$\mathbb{P}(\{\omega_i\}) = p_{\omega_i} > 0$$

с учетом условия нормировки:

$$\sum_{\omega_i \in \Omega} p_{\omega_i} = 1$$

по аксиоме аддитивности вероятности вероятность любого события  $A \in \mathcal{F}$  равна:

$$\mathbb{P}(A) := \sum_{\omega_i \in A \cap \Omega'} p_{\omega_i}$$

*Рассмотрим непрерывное вероятностное пространство.*

Пусть пространство элементарных событий  $\Omega$  континуально, т.е. имеет мощность континуума (на практике чаще всего можно привести к такому виду). Будет ли в этом случае максимально информативной такая модель, для которой  $\mathcal{F} = 2^\Omega$ ? Другими словами, можно ли непротиворечиво задать вероятностную меру на всех подмножествах?

Сделаем дополнительные предположения. Будем рассматривать континуум гипотезу в следующем виде:

**Аксиома 1** (Континуум гипотеза). *Между множеством счетной мощности и множеством мощности континуум не существует множеств промежуточной мощности.*

Континуум гипотеза является аксиомой, т.к. ее нельзя доказать.

В 1930г. польский математик С.М.Улам доказал теорему, устанавливающую, что конечная мера, заданная на множестве всех подмножеств некоторого множества мощности континуума и приписывающая нулевую меру каждому множеству, содержащему ровно один элемент (одну точку), обязана приписывать нулевую меру любым другим множествам, т.е. является тривиальной.

**Теорема 1** (Улама). *Пусть пространство элементарных событий  $\Omega$  континуально. Пусть верны континуум гипотеза и аксиома выбора. Тогда вероятность на всех подмножествах  $\Omega$   $2^\Omega$  можно задать  $\iff$ , когда  $\Omega$  – дискретное пространство. Другими словами: мера, заданная на множестве всех подмножеств  $\Omega$ , дискретна (что равносильно дискретности пространства  $\Omega$ ), т.е. не существует абсолютно непрерывной меры, заданной на всех подмножествах континуального пространства.*

Проблемы возникают с непрерывностью.

### 3.3 Борелевская $\sigma$ -алгебра.

**Определение 32.** Борелевская  $\sigma$ -алгебра – это минимальная  $\sigma$ -алгебра, содержащая все открытые множества.

**Определение 33** (Эквивалентное определение). Борелевская  $\sigma$ -алгебра – это  $\sigma$ -алгебра, порожденная топологией.

Под топологией понимаем введение класса открытых множеств.

**Определение 34.** Непустая система множеств  $X$  называется **кольцом**, если она обладает тем свойством, что из  $A \in X$  и  $B \in X$  следует, что  $A \Delta B \in X$  и  $A \cap B \in X$  (из этого вытекает, что  $A \cup B \in X$  и  $A \setminus B \in X$ ).

**Определение 35.** Кольцо множеств называется  **$\sigma$ -кольцом**, если оно вместе с каждой последовательностью множеств  $A_1, A_2, \dots, A_n, \dots$  содержит сумму:

$$S = \cup_n A_n$$



**Определение 36.** Множество  $E$  называется **единицей** системы множеств  $X$ , если  $E \in X$  и если для любого множества  $A \in X$  имеет место равенство:

$$A \cap E = A$$

**Определение 37.**  $\sigma$ -алгебра – это  $\sigma$ -кольцо с единицей.

Как построить максимально информативную модель в  $\Omega$ -континуально? В случае  $\Omega$ -континуально в качестве  $\mathcal{F}$  возьмем борелевскую  $\sigma$ -алгебру (это хорошо при дополнительных предположениях на топологию, т.е. при условии, что  $\sigma$ -алгебра событий обладает некоторыми дополнительными свойствами).

**Определение 38.** Польское пространство – это полное сепарабельное метрическое пространство.

**Определение 39.** Метрическим пространством называется пара  $(X, \rho)$ , состоящая из некоторого множества (пространства)  $X$  элементов и расстояния, т.е. однозначной неотрицательной действительной функции  $\rho(x, y)$ , определенной для  $\forall x \in X, \forall y \in X$  и удовлетворяющей трем условиям:

1.  $\rho(x, y) = 0 \Leftrightarrow x = y$
2.  $\rho(x, y) = \rho(y, x)$  (аксиома симметрии)
3.  $\rho(x, z) \leq \rho(x, y) + \rho(y, z)$  (аксиома треугольника)

**Определение 40.** Пусть  $A$  и  $B$  – два множества в метрическом пространстве  $R$ . Множество  $A$  называется **плотным** в  $B$ , если  $[A] \supset B$ , где  $[A]$  – замыкание  $A$ . Множество  $A$  называется **всюду плотным** (в пространстве  $R$ ), если его замыкание  $[A]$  совпадает со всем пространством  $R$ .

**Определение 41.** Пространства, в которых имеется счетное всюду плотное множество, называются **сепарабельными**.

**Определение 42.** Если в пространстве  $R$  любая фундаментальная последовательность сходится, то это пространство называется **полным**.

**Определение 43.** Два пространства  $X$  и  $Y$  **гомеоморфны**, если существует отображение  $\varphi$ :

$$\varphi: X \rightarrow Y$$

причем  $\varphi$  – биекция (взаимно однозначное отображение) и  $\varphi$  – взаимно непрерывно, т.е.  $\varphi$  непрерывно и  $\varphi^{-1}$  непрерывно.

Гомеоморфные между собой пространства обладают одними и теми же топологическими свойствами, и с топологической точки зрения их можно рассматривать как два экземпляра одного и того же пространства. Топологии (открытые множества) в двух гомеоморфных пространствах служат образами и прообразами друг друга.

Полнота пространства не является топологическим свойством (не сохраняется при гомеоморфизме).



### 3.4 Нуль множества и пополнение вероятностного пространства.

**Определение 44.** Множество  $N$  называется **нуль-множеством** (по отношению к мере  $\mathbb{P}$ ), если существует событие  $A$  нулевой меры, такое что:

$$\exists A \in \mathcal{F} : \mathbb{P}(A) = 0, N \subseteq A$$

Если  $N \in \mathcal{F}$ , то  $0 \leq \mathbb{P}(N) \leq \mathbb{P}(A) = 0$ , т.е.  $\mathbb{P}(N) = 0$

Если  $N \notin \mathcal{F}$ , то  $\mathbb{P}(N)$  не определено.

**Определение 45.**  $\sigma$ -алгебра называется **полной** относительно меры  $\mathbb{P}$ , если она содержит все нуль-множества

Добавим все нуль-множества в  $\sigma$ -алгебру  $\mathcal{F} \Rightarrow$  пополнение  $\sigma$ -алгебры  $\mathcal{F}_{\mathbb{P}}^*$ :

$$(\Omega, \mathcal{F}_{\mathbb{P}}^*, \mathbb{P}^*), \mathcal{F} \in \mathcal{F}_{\mathbb{P}}^*$$

$$\mathbb{P}^*|_{\mathcal{F}} = \mathbb{P}$$

$\Rightarrow \mathbb{P}^*$  – продолжение меры  $\mathbb{P}$ .

**Определение 46.** Обозначим через  $\mathcal{N}$  – класс всех нуль множеств:

$$\mathcal{N} = \{N : \exists A \in \mathcal{F} : \mathbb{P}(A) = 0, N \subseteq A\}$$

$$\mathcal{N} \subseteq \mathcal{F}_{\mathbb{P}}^*$$

Определим  $\mathcal{F}_{\mathbb{P}}^*$  как минимальную  $\sigma$ -алгебру, содержащую  $\mathcal{F} \cup \mathcal{N}$ .

Покажем, что такая минимальная  $\sigma$ -алгебра существует. Пусть  $\mathcal{F}$  – некоторый класс множеств. Обозначим  $\sigma(\mathcal{K})$  – минимальную  $\sigma$ -алгебру, содержащую  $\mathcal{K}$ .

Заметим, что для  $\forall \sigma_1, \sigma_2$  –  $\sigma$ -алгебр,  $\sigma = \sigma_1 \cap \sigma_2$  –  $\sigma$ -алгебра.

$$\sigma(\mathcal{K}) = \cap_{A \supseteq \mathcal{K}} A,$$

где  $A$  –  $\sigma$ -алгебра.

$2^{\Omega}$  –  $\sigma$ -алгебра, содержащая  $\mathcal{K}$ .

Заметим, что такое построение  $\mathcal{F}_{\mathbb{P}}^*$  крайне неконструктивное.

Опишем конструктивное построение  $\mathcal{F}_{\mathbb{P}}^*$ :

$$\mathcal{F}_{\mathbb{P}}^* = \sigma(\mathcal{F} \cup \mathcal{N})$$

$$B \in \sigma(\mathcal{F} \cup \mathcal{N}) \iff B = B' + N, B' \in \mathcal{F}, N \in \mathcal{N}$$

Под "+" понимаем объединение непересекающихся множеств.

$$\mathbb{P}^*(B) := \mathbb{P}^*(B') + \mathbb{P}^*(N) = \mathbb{P}(B') + 0$$

Так как  $\mathbb{P}^*|_{\mathcal{F}} = \mathbb{P}$ .

$$\Rightarrow \mathbb{P}^*(B) = \mathbb{P}(B')$$

**Упражнение.** Доказать, что  $\mathcal{B} = \{B = B' + N, B' \in \mathcal{F}, N \in \mathcal{N}\}$  совпадает с  $\sigma(\mathcal{F} \cup \mathcal{N})$

**Упражнение.** Показать, что определение  $\mathbb{P}^*(B) := \mathbb{P}^*(B') + \mathbb{P}^*(N)$  корректно, т.е., если  $B = B' + N, B = B'' + M, N, M \in \mathcal{N}$ , то вне зависимости от разложения получаем одно и то же значение  $\mathbb{P}^*(B)$ . Показать, что  $\mathbb{P}^*$  – нормированная мера.

**Замечание.** Пополнение  $\sigma$ -алгебры  $\mathcal{F}_{\mathbb{P}}^*$  сильно зависит от меры  $\mathbb{P}$ .

Множество измеримо по Лебегу, если оно входит в пополнение  $\sigma$ -алгебры  $\mathcal{F}_{\mathbb{R}}^*$ . Мера Лебега — полная мера, т.е. она определена на всех нуль множествах.

**Определение 47.** Множество  $A$  называется **измеримым по Лебегу**, если для  $\forall \varepsilon > 0$  существует такая конечная совокупность непересекающихся прямоугольников  $A_\varepsilon$ , что:

$$\mu^*(A \Delta A_\varepsilon) < \varepsilon$$

**Меру Лебега**  $\mu(A)$  измеримого множества  $A$  полагают равной  $\mu^*(A)$ .

Рассмотрим в качестве  $\mathcal{F}$  борелескую  $\sigma$ -алгебру. Пополним  $\mathcal{F} \Rightarrow$  получим  $\mathcal{F}_{\mathbb{R}}^*$  (по теореме Улама в  $\Omega$  останутся неизмеримые по Лебегу множества).

**Определение 48. Особо выдающиеся меры в  $\mathbb{R}^n$**  — это меры, которые инвариантны относительно группы движения (вращения и сдвигов), т.е. при преобразовании движения мера не изменяется, например, это мера Лебега.

**Пример.** Особо выдающаяся мера на окружности — это мера, инвариантная относительно вращения, т.е. такая мера, которая углы переводит в углы. Такая мера называется **равномерной мерой на окружности**.

Единичная окружность изоморфна полуинтервалу  $[0, 2\pi)$ .

**Определение 49. Меры Хаара** — это меры на топологических пространствах, инвариантные относительно соответствующих преобразований, например, это мера Лебега, равномерная мера на окружности.

### 3.5 Аксиома выбора.

**Аксиома 2** (Аксиома выбора). Пусть  $I$  — некоторое множество индексов  $\alpha$ ,  $\alpha \in I$ . Пусть для  $\forall \alpha$  задано произвольное множество  $A_\alpha: \alpha \rightarrow A_\alpha \neq \emptyset$ . Тогда существует функция выбора  $\varphi$ , такая что :

$$\varphi: \alpha \rightarrow x_\alpha \in A_\alpha$$

**Пример** (неизмеримого по Лебегу множества). Рассмотрим в качестве  $\Omega$  полуинтервал  $[0, 1) = \Omega$ .

Рассмотрим сложение по  $\text{mod } 1: x + y \pmod{1}$ , причем если  $x \in [0, 1), y \in [0, 1)$ , то  $x + y \pmod{1} \in [0, 1)$ .

Рассмотрим  $\mathcal{Q}'$  — множество рациональных чисел на  $[0, 1)$ .

Множество  $\mathcal{Q}'$  с заданной на нем операцией сложение по  $\text{mod } 1: x + y \pmod{1}$  является подгруппой, т.к. есть замкнутость относительно этой операции и выполняется свойство ассоциативности.

Введем отношение *эквивалентности* (рефлексивность + симметричность + транзитивность) на  $[0, 1)$ :

$$x \sim y \iff x - y \pmod{1} \in \mathcal{Q}'$$

Введенное таким образом отношение эквивалентности разбивает  $[0, 1) = \Omega$  на непересекающиеся классы  $K_\alpha$ :

$$\Omega = \sum_{\alpha \in I} K_\alpha,$$

где  $I$  — множество индексов.

Выберем из каждого из непересекающихся классов  $K_\alpha$  по одному элементу, здесь используем аксиому выбора.

Рассмотрим множество  $D$ :

$$D = \{x_\alpha, x_\alpha \in K_\alpha, \alpha \in I\}$$

Множество  $D$  построено не конструктивно.

Рассмотрим множество  $r + D \pmod{1}$ , где  $r \in \mathcal{Q}'$ :

1. если  $r_1 \neq r_2$ , то

$$(r_1 + D) \cap (r_2 + D) = \emptyset, x \in (r_1 + D) \cap (r_2 + D) \Rightarrow$$

$$\Rightarrow x - r_1 \in D, x - r_2 \in D \Rightarrow$$

$$\Rightarrow x - r_1 \sim x - r_2 \Rightarrow r_1 = r_2 \Rightarrow$$

пришли к противоречию.

2.

$$\sum_{r \in \mathcal{Q}'} (r + D) = \Omega,$$

здесь под суммой понимаем объединение попарно непересекающихся множеств.

Рассмотрим в качестве  $\sigma$ -алгебры  $\mathcal{F}$  – множество рациональных чисел  $\mathcal{Q}'$  на  $[0, 1)$  с операцией сложение по  $\pmod{1}$ :  $x + y \pmod{1}$ , которое образует счетное всюду плотное множество на множестве действительных чисел на  $[0, 1)$ .

Обозначим эту  $\sigma$ -алгебру через  $\mathcal{B}$ . На  $\mathcal{B}$  можно ввести топологию окружности, в качестве открытых множеств рассматривать окрестности точек.

Введем также для  $\mathcal{B}$   $\pi$  – равномерную меру на окружности.

Тогда  $\pi(r + D) = \pi(r' + D)$ , т.к. мера  $\pi$  инвариантна относительно сдвига.

$$1 = \pi(\Omega) = \pi\left(\sum_{r \in \mathcal{Q}'} (r + D)\right) = \sum_{r \in \mathcal{Q}'} \pi(r + D)$$

Так как мера  $\pi$  инвариантна относительно сдвига, то если:

$$\pi(r + D) = 0 \Rightarrow 0 = 1$$

$$\pi(r + D) > 1 \Rightarrow \infty = 1$$

$\Rightarrow$  множество  $r + D$  – не измеримо по Лебегу.

Аксиому выбора можно заменить на *аксиому детерминированности* (из теории игр). Относится к аксиоме детерминированности надо так же, как и к аксиоме выбора.

Аксиома детерминированности не противоречит счетной аксиоме выбора.

*Рассмотрим игровую интерпретацию:* 2 игрока по очереди выбирают натуральное число, например, 5, 38, 7, 1, ... Первый игрок выбирает первое число, второй игрок – второе число и т.д.

$\mathbb{N}$  – множество всех натуральных чисел  $\Rightarrow$  получаем  $\mathbb{N}^{\mathbb{N}}$  – множество всех отображений из  $\mathbb{N}$  в  $\mathbb{N}$ , мощность  $\mathbb{N}^{\mathbb{N}}$  – континуум.

$A \subseteq \mathbb{N}^{\mathbb{N}}$ , т.е.  $A$  – некоторое множество последовательностей натуральных чисел.  $x$  – некоторая последовательность натуральных чисел.

Пусть  $x \in A$ , если выиграл первый игрок,  $x \notin A$ , если выиграл второй игрок.

**Определение 50.** Множество  $A$  называется **детерминированным**, если либо у первого игрока, либо у второго игрока существует выигрышная стратегия (т.е. всегда какой-то игрок побеждает, как бы ни ходил другой игрок)

**Аксиома 3** (Аксиома детерминированности). *Любое множество  $A$  является детерминированным.*

Аксиома детерминированности противоречит несчетной аксиоме выбора.

Если по аксиоме детерминированности производить пополнение борелевской  $\sigma$ -алгебры на прямой  $\Rightarrow$  все множества будут измеримы по Лебегу.

Если есть континуальное пространство  $\Omega$ , то можно задавать  $\sigma$ -алгебру на всех подмножествах.

Таким образом, применяя аксиому выбора, получаем неизмеримые множества. Если же применять аксиому детерминированности, то получаем, что все множества измеримы по Лебегу.

Договоримся, что по умолчанию мы принимаем аксиому выбора.

### 3.6 Теорема об изоморфизме.

Теорема об изоморфизме нужна для обоснования метода Монте-Карло (метода статистических испытаний).

Будем рассматривать пространства  $\Omega$ , изоморфные польскому пространству (польское пространство – метрическое, сепарабельное, полное).

Предположим, что рассматриваемая модель является *неатомической*.

Пусть  $\Omega$  – польское пространство,  $\mathcal{F}$  – борелевская  $\sigma$ -алгебра,  $\{x\} \in \mathcal{F}$ .

*Неатомическая мера*: любое одноточечное множество имеет нулевую вероятность:  $\mathbb{P}(x) = 0$ .

**Теорема 2** (об изоморфизме). *Пусть в вероятностном пространстве  $(\Omega, \mathcal{B}, \mathbb{P})$   $\Omega$  – польское пространство,  $\mathbb{P}$  – неатомическая мера. Тогда это вероятностное пространство изоморфно (с точностью до почти наверное) вероятностному пространству  $([0, 1], \mathcal{B}_{[0,1]}, U[0, 1])$ .*

**Пример.** Рассмотрим пространство  $\Omega = C[a, b]$ , т.е. элементарными исходами являются непрерывные функции. Пространство  $C[a, b]$  – полное нормированное сепарабельное пространство (в качестве нормы в  $C[a, b]$  можно взять:  $\|f\| = \max_{x \in [a, b]} |f(x)|$ , а множество всех полиномов с рациональными коэффициентами образует счетное всюду плотное множество в  $C[a, b]$ )  $\Rightarrow C[a, b]$  – польское пространство.

Рассмотрим неатомическую меру.

Пусть случайный вектор  $X \sim U[0, 1]$  имеет равномерное распределение на  $[0, 1] \Rightarrow$

$$\exists \varphi: [0, 1] \rightarrow \Omega' \simeq \Omega, \varphi(x) \sim \mathbb{P}$$

**Замечание.** Отображение  $\varphi$  – биекция, если  $\mathbb{P}$  – неатомическая мера.

Если  $\mathbb{P}$  – атомическая мера, то отображение  $\varphi$  – не биекция.

**Замечание.** Датчик равномерно распределенных случайных чисел может смоделировать почти любой процесс

## 4 Распределение вероятностей как частный вид модели. Случайные величины, вектора, функции. Интерпретация условия измеримости. Распределение как индуцированная мера. Функция распределения как характеристика распределения и ее свойства. Моделирование случайных величин методом обращения функций распределения.

### 4.1 Распределение вероятностей как частный вид модели. Случайные величины, вектора, функции. Интерпретация условия измеримости. Распределение как индуцированная мера.

Рассмотрим некоторый *случайный вектор* из пространства  $\mathbb{R}^n$ . Основной постулат Колмогорова гласит: любое непротиворечивое описание модели можно представить в виде вероятностного пространства. Построим для нашей модели вероятностное пространство в явном виде.

Положим  $\Omega = \mathbb{R}^n$ ,  $\mathcal{F} = \mathcal{B}^n$ <sup>25</sup>, где  $\mathcal{B}^n$  — борелевская  $\sigma$ -алгебра<sup>26</sup> на множестве  $\mathbb{R}^n$ . Вероятность  $\mathbb{P}$  в нашей модели, которую мы обозначим как  $Q$ , назовем распределением случайного вектора.

**Определение 51.** *Распределением случайного вектора* называется нормированная мера  $Q$  на пространстве  $(\mathbb{R}^n, \mathcal{B}^n, Q)$ .

При  $n = 1$  получаем определение *распределения случайной величины*.

Пусть теперь  $E$  — произвольное пространство,  $\mathcal{E}$  —  $\sigma$ -алгебра подмножеств  $E$ ,  $(E, \mathcal{E}, Q)$  — вероятностное пространство. Тогда  $Q$  можно рассматривать как *распределение случайного элемента со значениями из пространства  $E$* .

**Пример.** Рассмотрим вероятностное пространство  $(E, \mathcal{E}, Q)$ , где

$E = C[a, b]$  — пространство непрерывных на  $[a, b]$  функций,

$\mathcal{E} = \mathcal{B}_{C[a, b]}$  — борелевская  $\sigma$ -алгебра,

тогда  $Q$  — распределение случайной непрерывной на отрезке  $[a, b]$  функции.

Назовем модель  $(E, \mathcal{E}, Q)$  *канонической*<sup>27</sup>. Заметим, что она описывает поведение случайного элемента и *ничего больше*.

Однако в рамках одной и той же модели может быть интересно рассматривать что-то еще, так как чаще всего мы имеем дело не просто с каким-то случайным вектором, а случайный вектор возникает в контексте другой модели. Получаем, что канонической модели недостаточно. Рассмотрим более общий подход к определению случайного вектора.

Пусть случайный элемент  $X \in E$  является характеристикой исхода некоторого эксперимента. Эксперимент опишем универсальной моделью — тройкой  $(\Omega, \mathcal{F}, \mathbb{P})$ . Пусть нам известен исход эксперимента  $\omega \in \Omega$ . Очевидно, что тогда мы имеем полную информацию об эксперименте и можем узнать значение  $X$ :

$$\omega \mapsto X(\omega).$$

<sup>25</sup>по умолчанию мы принимаем аксиому выбора, а потому не можем взять  $\mathcal{F} = 2^\Omega$

<sup>26</sup>борелевская  $\sigma$ -алгебра —  $\sigma$ -алгебра, порожденная всеми открытыми множествами

<sup>27</sup>не стандартное определение

Таким образом, можно формально определить **случайный элемент как отображение**:

$$X: \Omega \rightarrow E.$$

В данном случае пространство  $(\Omega, \mathcal{F}, \mathbb{P})$  описывает поведение  $X$  и чего-то еще.

Построим теперь каноническую модель для случайного элемента  $X$ . Обозначим через  $\mathbb{P}_X$  распределение  $X$ . Каноническая модель примет вид  $(E, \mathcal{E}, \mathbb{P}_X)$ , где  $\mathcal{E}$  —  $\sigma$ -алгебра подмножеств  $E$ . Из соображений здравого смысла, следует положить

$$\mathbb{P}_X(B) := \mathbb{P}(\{\omega: X(\omega) \in B\}), \quad \forall B \in \mathcal{E}. \quad (24)$$

Это **аксиома, связывающая общую и каноническую модели**.

$\mathbb{P}_X(B)$  — вероятность того, что значение  $X$  попадает во множество  $B$ . Будем для краткости обозначать

$$\mathbb{P}(\{\omega: X(\omega) \in B\}) = \mathbb{P}(X^{-1}(B)).$$

Из формулы (24) получаем, что распределение  $\mathbb{P}_X$  строится по мере  $\mathbb{P}$ .

**Определение 52.** *Индукцированной отображением  $X$  мерой* называется мера  $\mathbb{P}_X$  такая, что

$$\mathbb{P}_X(B) \stackrel{\text{обозн.}}{=} X \circ \mathbb{P}(B) \stackrel{\text{def}}{=} \mathbb{P}(X^{-1}(B)).$$

Таким образом, используются следующие обозначения<sup>28</sup>:

$$\mathbb{P}_X = X \circ \mathbb{P} = \mathbb{P} \circ X^{-1}.$$

Получаем, что можно рассматривать *распределение случайного элемента  $X$  как индуцированную этим отображением  $X$  меру*.

Рассмотрим множество

$$\mathcal{F}_X = \{X^{-1}(B), B \in \mathcal{E}\}.$$

Это математический формализм, описывающий поведение  $X$ .  $\mathcal{F}_X$  образует  $\sigma$ -алгебру<sup>29</sup>. Очевидно, что для того, чтобы формула (24) была корректна, нужно, чтобы  $X^{-1}(B) \in \mathcal{F}$   $\forall B \in \mathcal{E}$ , или иначе

$$\mathcal{F}_X \subseteq \mathcal{F}.$$

В противном случае вероятность  $\mathbb{P}_X(B)$  может быть неопределена для какого-нибудь  $B$ .

**Пример.** Рассмотрим следующую модель: пусть  $\Omega$  — прямоугольник на плоскости, множества  $A$  и  $B$  заданы, как показано на рисунке 1. Пусть  $\mathcal{F} = \{A, A^c, \emptyset, \Omega\}$  ( $\sigma$ -алгебра, порожденная множеством  $A$ ). Положим  $\mathbb{P}(A) = 0.3$ , тогда можем однозначно определить вероятностную меру в нашей модели  $\mathbb{P}: \mathcal{F} \rightarrow \mathbb{R}$ :  $\mathbb{P}(A^c) = 0.7$ ,  $\mathbb{P}(\emptyset) = 0$ ,  $\mathbb{P}(\Omega) = 1$ <sup>30</sup>.

Теперь рассмотрим случайную величину  $X$ , определенную следующим образом:

$$X(\omega) = \begin{cases} 5, & \omega \in B, \\ -2, & \omega \in B^c. \end{cases}$$

<sup>28</sup>В первом случае значок 'o' — обозначение, а во втором — суперпозиция

<sup>29</sup>т.к. операция взятия прообраза сохраняет все теоретико-множественные операции, а  $\mathcal{E}$  есть  $\sigma$ -алгебра по построению

<sup>30</sup>использовали нормированность меры

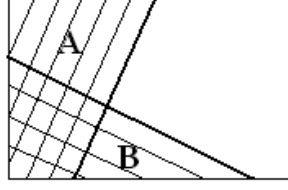


Рис. 1:

Очевидно,  $\mathcal{F}_X = \{B, B^c, \emptyset, \Omega\}$ <sup>31</sup>. Как теперь нам определить, например, вероятность того, что  $X = -2$ :  $\mathbb{P}(\{\omega: X(\omega) = -2\}) = \mathbb{P}(B^c) = ?$  Видно, что исходя из нашей модели ничего нельзя сказать о распределении  $X$ , т.е. случайная величина  $X$  не является характеристикой исхода эксперимента, описываемого вероятностным пространством  $(\Omega, \mathcal{F}, \mathbb{P})$ . А все дело в том, что  $\mathcal{F}_X \not\subseteq \mathcal{F}$ . Этот пример помогает определить и проинтерпретировать так называемое условие измеримости.

**Определение 53.** *Измеримым пространством* называют пару  $(E, \mathcal{E})$ , где  $E$  — некоторое пространство,  $\mathcal{E}$  — заданная на нем  $\sigma$ -алгебра подмножеств.

Заметим, что измеримое пространство  $(\Omega, \mathcal{F})$  выступает в роли атрибута вероятности в тройке вероятностного пространства  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Определение 54.** Пусть заданы два измеримых пространства:  $(\Omega, \mathcal{F})$  и  $(E, \mathcal{E})$ . отображение  $X: \Omega \rightarrow E$  называется  $(\mathcal{F}, \mathcal{E})$ -*измеримым отображением*, если

$$\forall B \in \mathcal{E} \quad X^{-1}(B) \in \mathcal{F}.$$

Итак,  $\mathcal{F}$  — это та информация, которая доступна в нашей модели,  $\mathcal{F}_X$  — формализм, описывающий информацию о поведении  $X$ .

**Условие измеримости:**

$$\mathcal{F}_X \subseteq \mathcal{F}. \quad (25)$$

Ясно, что  $(\mathcal{F}, \mathcal{E})$ -измеримость отображения эквивалентно условию измеримости.

Условие измеримости можно проинтерпретировать следующим образом: *информации, заложенной в модели, достаточно для описания поведения случайного элемента  $X$* .<sup>32</sup>

Заметим, что условие измеримости естественным образом вытекает из формулы (24), которая является аксиомой и обосновывается интерпретацией.

Таким образом, мы приходим к следующему определению случайного элемента:

<sup>31</sup>В роли  $E$  можно взять, например, просто  $\{-2, 5\}$  (это, безусловно, не единственно возможный вариант), а в качестве  $\mathcal{E}$  — множество всех подмножеств  $\{\emptyset, \{-2\}, \{5\}, \{-2, 5\}\}$ ; тогда

$$X^{-1}(C) = \{\omega: X(\omega) \in C\} = \begin{cases} \emptyset, & \text{если } C = \emptyset, \\ B, & \text{если } C = \{5\}, \\ B^c, & \text{если } C = \{-2\}, \\ \Omega, & \text{если } C = \{5, -2\}, \end{cases}$$

откуда следует, что  $\mathcal{F}_X$  определяется именно таким образом (причем оно не зависит от способа определения пары  $(E, \mathcal{E})$ )

<sup>32</sup>заметим, что т.к. измеримое пространство является атрибутом вероятностного, то вопрос об измеримости может быть решен лишь на основе атрибутов, для этого не нужно рассматривать вероятность



**Определение 55.** Пусть заданы два измеримых пространства:  $(\Omega, \mathcal{F})$  и  $(E, \mathcal{E})$ . *Случайным элементом* со значениями из множества  $E$ , определенным на пространстве  $\Omega$ , называется отображение

$$X: \Omega \rightarrow E,$$

которое является  $(\mathcal{F}, \mathcal{E})$ -измеримым<sup>33</sup>.

Это универсальное определение случайного элемента. Покажем, что данное в начале главы определение случайного элемента через каноническую модель  $(E, \mathcal{E}, Q)$  можно свести к данному. Действительно, для этого достаточно рассмотреть отображение<sup>34</sup>

$$X: E \rightarrow E,$$

которое, очевидно, является  $(\mathcal{E}, \mathcal{E})$ -измеримым. Индуцированная мера в данном случае:  $\mathbb{P}_X = Q$ .

Заметим, что при переходе от канонической формы к универсальной отображение  $X: E \rightarrow E$  можно выбрать разными способами. Рассмотрим, например, такую модель:

$$E = [0, 1], \quad \mathcal{E} = \mathcal{B}_{[0,1]}, \quad Q = U[0, 1].$$

Тогда<sup>35</sup>

$$\begin{aligned} X(x) = x &\Rightarrow \mathbb{P}_X = Q, \\ X'(x) = 1 - x &\Rightarrow \mathbb{P}_{X'} = Q. \end{aligned}$$

Такую неоднозначность можно объяснить тем, что в канонической модели есть информация только о пространстве  $E$ , в котором находятся значения случайного элемента, и о распределении этого случайного элемента, но нет информации о его связи с общей моделью  $(\Omega, \mathcal{F}, \mathbb{P})$ .

Еще раз подчеркнем, что *модель  $(E, \mathcal{E}, Q)$  описывает поведение только случайного элемента, а модель  $(\Omega, \mathcal{F}, \mathbb{P})$  может описывать что-то еще, т.е. содержать дополнительную информацию.*

**Договоренность.** Случайные величины будем обозначать либо большими латинскими буквами, либо греческими, а остальное (не случайные величины) — маленькими латинскими буквами.

В ходе предыдущих рассуждений мы построили следующую цепочку моделей:

$$(\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\Omega, \mathcal{F}_X, \mathbb{P}|_{\mathcal{F}_X}) \rightarrow (E, \mathcal{E}, \mathbb{P}_X).$$

По смыслу модели  $(\Omega, \mathcal{F}_X, \mathbb{P}|_{\mathcal{F}_X})$  и  $(E, \mathcal{E}, \mathbb{P}_X)$  описывают одно и то же (поведение  $X$  и больше ничего), но они не являются изоморфными<sup>36</sup>. В этом заключается недостаток данного определения через каноническую модель.

Заметим также, что информации из второй модели  $(\Omega, \mathcal{F}_X, \mathbb{P}|_{\mathcal{F}_X})$  достаточно для построения канонической модели  $(E, \mathcal{E}, \mathbb{P}_X)$ .

В заключении данного раздела рассмотрим вопрос о композиции измеримых отображений.

<sup>33</sup>таким образом, говоря о случайном элементе (в частности, о случайном векторе), мы будем подразумевать, что для данного отображения выполнено условие измеримости

<sup>34</sup>измеримые пространства  $(\Omega, \mathcal{F})$  и  $(E, \mathcal{E})$  из определения 55 совпадают с измеримым пространством канонической модели, а мера  $\mathbb{P}$  из общей модели  $(\Omega, \mathcal{F}, \mathbb{P})$  совпадает с  $Q$

<sup>35</sup>используется известный из второго курса факт о том, что случайные величины  $X(x) = x$  и  $X'(x) = 1 - x$  имеют одинаковое равномерное распределение на отрезке  $[0, 1]$

<sup>36</sup>см. третью главу, определение изоморфных пространств



**Утверждение 6.** Пусть заданы измеримые пространства  $(\Omega, \mathcal{F})$ ,  $(E, \mathcal{E})$ ,  $(D, \mathcal{D})$  и отображения  $\xi: \Omega \rightarrow E$  —  $(\mathcal{F}, \mathcal{E})$ -измеримое,  $\eta: E \rightarrow D$  —  $(\mathcal{E}, \mathcal{D})$ -измеримое. Тогда отображение  $\eta \circ \xi: \Omega \rightarrow D$  будет  $(\mathcal{F}, \mathcal{D})$ -измеримым.

Доказательство. Надо показать, что

$$\forall C \in \mathcal{D} \quad (\eta \circ \xi)^{-1}(C) \in \mathcal{F}.$$

Из измеримости отображения  $\eta$  следует, что  $\eta^{-1}(C) \in \mathcal{E}$ . Из измеримости отображения  $\xi$  следует, что

$$\forall B \in \mathcal{E} \quad \xi^{-1}(B) \in \mathcal{F},$$

а значит,  $\xi^{-1}(\eta^{-1}(C)) \in \mathcal{F}$ . Но  $(\eta \circ \xi)^{-1}(C) = \xi^{-1}(\eta^{-1}(C))$  по определению композиции отображений. Следовательно,  $(\eta \circ \xi)^{-1}(C) \in \mathcal{F}$ , ч.т.д.

**Замечание 1.** Дополним измеримые пространства в утверждении 6 до вероятностных:

$$\begin{array}{c} (\Omega, \mathcal{F}, \mathbb{P}) \\ \xi \downarrow \\ (E, \mathcal{E}, \mathbb{P}_\xi)^{37} \\ \eta \downarrow \\ (D, \mathcal{D}, \eta \circ \mathbb{P}_\xi) \end{array}$$

При этом  $\mathbb{P}_\xi = \xi \circ \mathbb{P}$ ,  $\eta \circ \mathbb{P}_\xi = \eta \circ (\xi \circ \mathbb{P})$ . С другой стороны

$$\begin{array}{c} (\Omega, \mathcal{F}, \mathbb{P}) \\ \eta \circ \xi \downarrow \\ (D, \mathcal{D}, (\eta \circ \xi) \circ \mathbb{P}) \end{array}$$

Получаем, что

$$\eta \circ (\xi \circ \mathbb{P}) = (\eta \circ \xi) \circ \mathbb{P}.$$

**Замечание 2.** Смысл утверждения 6 (интерпретируя условие измеримости): если информации из первой модели достаточно для описания второй модели, а информации из второй модели достаточно для описания третьей модели, то информации из первой модели хватит для описания третьей модели<sup>38</sup>.

Верно и, в некотором смысле, обратное утверждение.

**Утверждение 7.** Пусть заданы измеримые пространства  $(\Omega, \mathcal{F})$ ,  $(E, \mathcal{E})$ ,  $(D, \mathcal{D})$  и отображения  $\xi: \Omega \rightarrow E$  —  $(\mathcal{F}, \mathcal{E})$ -измеримое,  $\zeta: \Omega \rightarrow D$  —  $(\mathcal{F}, \mathcal{D})$ -измеримое. Пусть, кроме того, известно, что  $D$  — польское<sup>39</sup> пространство с борелевской  $\sigma$ -алгеброй  $\mathcal{D}$ . Тогда существует отображение  $\eta: E \rightarrow D$  —  $(\mathcal{E}, \mathcal{D})$ -измеримое, такое что

$$\zeta = \eta \circ \xi, \quad \text{т.е.} \quad \zeta(\omega) = \eta(\xi(\omega)).$$

Без доказательства.

<sup>37</sup> $(E, \mathcal{E}, \mathbb{P}_\xi)$  — каноническая модель, описывающая поведение  $\xi$  и только  $\xi$

<sup>38</sup>действительно, из соображений здравого смысла это и должны были получить, ведь по определению композиции отображений  $\eta \circ \xi(\omega) = \eta(\xi(\omega))$ , т.е. важна только информация о  $\xi(\omega)$ , а не само  $\omega$

<sup>39</sup>т.е. полное, сепарабельное, метрическое

## 4.2 Функция распределения как характеристика распределения и ее свойства. Моделирование случайных величин методом обращения функций распределения.

Рассмотрим вероятностное пространство  $(\Omega, \mathcal{F}, \mathbb{P})$ ,  $X$  — случайная величина, т.е.  $X: \Omega \rightarrow \mathbb{R}$  —  $(\mathcal{F}, \mathcal{B})$ -измеримое отображение,  $(\mathbb{R}, \mathcal{F}, \mathbb{P}_X)$  — каноническая модель для случайной величины  $X$ .

**Определение 56.** *Функцией распределения* случайной величины  $X$   $F_X(x)$  называют вероятность того, что  $X < x$ :

$$F_X(x) = \mathbb{P}(\{\omega: X(\omega) < x\})^{40} = \mathbb{P}(\{X < x\})^{41}.$$

Функция распределения случайного вектора  $X = (X_1, \dots, X_n)$  вводится аналогичным образом:

$$F_X(x_1, \dots, x_n) = \mathbb{P}(\{X_1 < x_1, \dots, X_n < x_n\}).$$

**Замечание.** Мы уверены, что вводимая функция распределения не вносит противоречия в модель, поэтому проверку на непротиворечивость, т.е. построение вероятностного пространства в явном виде, опустим.

**Свойства функции распределения.**

1. Функция распределения случайной величины является *характеристикой распределения* этой случайной величины, т.е. зная распределение случайной величины, мы можем построить для нее функцию распределения.

Действительно, для любого  $x$  имеем:

$$F_X(x) = \mathbb{P}(\{\omega: X(\omega) \in (-\infty, x)\}) = \mathbb{P}(\{X \in (-\infty, x)\}) \stackrel{\text{def}}{=} \mathbb{P}_X((-\infty, x)).$$

2. Функция распределения случайной величины является *исчерпывающей характеристикой распределения* случайной величины, т.е. зная функцию распределения случайной величины, мы можем построить ее распределение.

На доказательстве этого свойства остановимся более подробно.

Итак, пусть у нас задана функция распределения  $F_X(x)$ . Нам необходимо однозначно определить  $\mathbb{P}_X(B)$ , для любого  $B \in \mathcal{B}$ . Очевидно, мы сразу знаем  $\mathbb{P}_X(B)$ , где  $B = (-\infty, x)$ . Обозначим

$$\mathcal{K} = \{(-\infty, x), x \in \mathbb{R}\}.$$

Итак, нам известно  $\mathbb{P}_X|_{\mathcal{K}}$ .

**Утверждение 8.**

$$\sigma(\mathcal{K}) = \mathcal{B}.$$

Доказательство.

- $\subseteq$ : Обозначим за  $\tau$  класс всех открытых множеств прямой (топология). Напомним, что  $\mathcal{B}$  —  $\sigma$ -алгебра, порожденная всеми открытыми множествами на прямой, т.е.  $\sigma(\tau) = \mathcal{B}$ . Очевидно, что  $\mathcal{K} \subset \tau$  и, значит,  $\sigma(\mathcal{K}) \subseteq \sigma(\tau) = \mathcal{B}$ .

<sup>40</sup>предикат — высказывание, истинность которого зависит от переменной;  $\{\omega: X(\omega) < x\}$  — пример предиката ( $x$  — переменная)

<sup>41</sup>сокращенная форма записи

$\supseteq$ : Докажем вложение в обратную сторону. Известно, что любое открытое множество прямой можно представить в виде объединения не более чем счетного числа открытых интервалов<sup>42</sup>. С другой стороны, любой интервал представляется в виде

$$(a, b) = \bigcup_{n=1}^{\infty} [a + \frac{1}{n}, b),$$

а любой полуинтервал можно представить в виде разности множеств из  $\mathcal{K}$ :

$$[a, b) = (-\infty, b) \setminus (-\infty, a).$$

Таким образом,  $\tau \subset \sigma(\mathcal{K}) \Rightarrow \sigma(\tau) \subseteq \sigma(\mathcal{K})$ .

Итак, нам известно  $\mathbb{P}_X|_{\mathcal{K}}$  и известно, что  $\sigma(\mathcal{K}) = \mathcal{B}$ . Можно ли утверждать, что существует однозначное продолжение  $\mathbb{P}_X|_{\mathcal{B}}$ ? Рассмотрим пример.

**Пример.** Пусть  $\Omega$  — прямоугольник на плоскости, множества  $A$  и  $B$  заданы, как показано на рисунке 2.

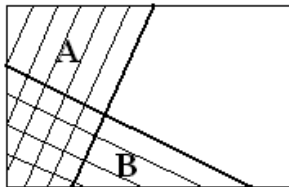


Рис. 2:

Обозначим  $\mathcal{M} = \{A, B\}$ . Рассмотрим, что представляет из себя  $\sigma(\mathcal{M})$ . Существует четыре конституэнты множеств  $A$  и  $B$ :  $A^\alpha \cap B^\beta$ ,  $\alpha, \beta \in \{0, 1\}$ <sup>43</sup>. Они попарно не пересекаются и образуют разбиение пространства  $\Omega$ . Очевидно, что  $\sigma(\mathcal{M})$  можно представить в виде всевозможных объединений конституэнт:

$$\sigma(\mathcal{M}) = \left\{ \emptyset, \bigcup_{i=1}^k C_i, \quad k = 1, 2, 3, 4 \mid C_i \text{ — одна из четырех конституэнт вида } A^\alpha \cap B^\beta \right\}.$$

Всего в  $\sigma(\mathcal{M})$   $2^4$  множеств ( $C_4^0 + C_4^1 + \dots + C_4^4 = (1 + 1)^4 = 2^4$ ).

Пусть теперь у нас заданы вероятности  $\mathbb{P}(A) = 0.4$ ,  $\mathbb{P}(B) = 0.7$ . Видно, что мы не можем однозначно определить вероятность на  $\sigma(\mathcal{M})$ , нам нужно знать вероятность хотя бы одной конституэнты.

Таким образом, в этом примере мера продолжается неоднозначно с  $\mathcal{M}$  на  $\sigma(\mathcal{M})$ .

В каком же случае можно однозначно продлить меру на  $\sigma$ -алгебру? Ответ на этот вопрос дает теорема Каратеодори.

**Теорема 3** (Каратеодори). *Если на алгебре задана непрерывная мера, то ее можно однозначно продлить до непрерывной меры на  $\sigma$ -алгебре, порожденной этой алгеброй.*

<sup>42</sup>см. курс функционального анализа

<sup>43</sup>напомним, что  $A^1 = A$ ,  $A^0 = A^c$

Однако пока мы не можем применить эту теорему к нашей ситуации, т.к.  $\mathcal{K}$  не является алгеброй. Мало того,  $\{\mathcal{K}, (-\infty; +\infty), \emptyset\}$  даже не является полуалгеброй<sup>44</sup>. Можно показать, что минимальная алгебра, содержащая  $\mathcal{K}$ , состоит из всевозможных объединений множеств вида  $(-\infty, b)$ ,  $[a, b)$ ,  $[a, +\infty)$ , а также всей прямой и пустого множества<sup>45</sup>:

$$\mathcal{A}(\mathcal{K}) = \left\{ (-\infty, b), \bigcup_{i=1}^n [a_i, b_i), [a, +\infty), \{(-\infty, c) \cup \bigcup_{i=1}^n [a_i, b_i)\}, \{(-\infty, a) \cup [b, +\infty)\}, \right. \\ \left. \{ \bigcup_{i=1}^n [a_i, b_i) \cup [c, +\infty) \}, \{(-\infty, c) \cup \bigcup_{i=1}^n [a_i, b_i) \cup [d, +\infty)\}, (-\infty, +\infty), \emptyset \mid a, b, a_i, b_i, c, d \in \mathbb{R}, n = 1, 2, \dots \right\}.$$

На эту алгебру мера продолжается однозначно. Действительно, в силу аддитивности меры  $((-\infty, a) + [a, b) = (-\infty, b), \forall a \in \mathbb{R}, b \in \mathbb{R} \cup \{+\infty\})$ , необходимо положить на полуинтервале<sup>46</sup>:

$$\begin{aligned} \mathbb{P}_X([a, b)) &:= \mathbb{P}_X((-\infty, b)) - \mathbb{P}_X((-\infty, a)) = F_X(b) - F_X(a), \\ \mathbb{P}_X([a, +\infty)) &:= 1 - \mathbb{P}_X((-\infty, a)) = 1 - F_X(a). \end{aligned}$$

А дальше любое множество из нашей алгебры представляем в виде конечного объединения множеств, на которых мера уже задана<sup>47</sup>, и суммируем соответствующие значения меры.

Здесь возникают два вопроса. Во-первых, однозначно ли такое продолжение? Ведь множество можно представить в виде конечного объединения “элементарных” несколькими способами. В данном случае однозначность доказывается при помощи тех же рассуждений, что используются при построении интеграла Римана, а именно, при помощи измельчения разбиений.

И во-вторых, будет ли функция множеств, определенная нами на алгебре, порожденной  $\mathcal{K}$ , непрерывной мерой, как того требуют условия теоремы Каратеодори. Здесь ответ такой: да, будет, в силу определенных свойств функции распределения, которые мы сейчас рассмотрим.

Укажем класс функций  $\mathfrak{M}$ , такой что  $F \in \mathfrak{M}$  тогда и только тогда, когда функция  $F$  является функцией распределения некоторой случайной величины.

$\mathfrak{M}$  — класс числовых функций  $F: \mathbb{R} \rightarrow \mathbb{R}$ , удовлетворяющих свойствам:

1.  $F$  монотонно неубывает.
2. а)  $F(x) \rightarrow 0$  при  $x \rightarrow -\infty$ ;  
б)  $F(x) \rightarrow 1$  при  $x \rightarrow +\infty$ .
3.  $F$  непрерывна слева

---

<sup>44</sup>Напомним, что множество подмножеств  $\mathbb{R}$  называется полуалгеброй  $S$ , если  $\mathbb{R} \in S$  и

1)  $A \cap B \in S, \forall A, B \in S$ ;

2)  $A \setminus B = \sum_{k=1}^n A_k, \forall A, B \in S, B \subset A \text{ и } A_k \in S$ .

Возьмем произвольные множества  $A = (-\infty, a)$  и  $B = (-\infty, b)$  из  $\mathcal{K}$ , такие что  $a > b$ . Ясно, что  $A \setminus B = [b, a)$  нельзя представить конечным объединением непересекающихся множеств из  $\{\mathcal{K}, (-\infty; +\infty), \emptyset\}$ . Следовательно,  $\{\mathcal{K}, (-\infty; +\infty), \emptyset\}$  не является полуалгеброй.

<sup>45</sup>Напомним, что множество подмножеств  $\mathbb{R}$  называется алгеброй  $A$ , если оно замкнуто относительно операций дополнения и объединения.

<sup>46</sup>далее будет показано, что в силу свойств функции распределения  $F_X(+\infty) = 1, F_X(-\infty) = 0$

<sup>47</sup>очевидно, что надо положить  $\mathbb{P}_X((-\infty, +\infty)) = 1$

Из первого свойства следует, что у функции  $F$  существует конечный или бесконечный предел в любой точке. Соответственно, второе свойство утверждает, что на бесконечности эти пределы конечны. Кроме того, из первых двух свойств следует, что  $F$  принимает значения между 0 и 1.

Известно, что у монотонной функции могут быть разрывы только первого рода (т.е. нет устранимых разрывов и разрывов второго рода). Функция имеет в точке разрыв первого рода (будем также говорить “скачок”), если в этой точке у нее существуют правый и левый пределы, но они не равны между собой.

Множество точек разрыва монотонной функции не более, чем счетно. Это можно доказать разными способами. Например, так: поставим каждой точке разрыва в соответствие интервал  $(a, b)$ , где  $a$  — нижний предел функции в точке,  $b$  — верхний предел. Получим множество непересекающихся интервалов. В каждом интервале выберем по рациональной точке (используя аксиому выбора). Получим, что множество таких интервалов, а следовательно, и точек разрыва монотонной функции, не более, чем счетно (т.к. множество всех рациональных чисел счетно). Для монотонной функции, удовлетворяющей свойству 2, справедливо и такое доказательство: число скачков такой функции величины  $\geq 1$  не превосходит 1, число скачков величины  $\geq \frac{1}{2}$  не превосходит 2, число скачков величины  $\geq \frac{1}{3}$  не превосходит 3, и т.д. В итоге получаем счетное число скачков.

Заметим также, что в силу того, что функция распределения  $F_X(x) \in \mathfrak{M}$  (что мы докажем ниже), функция множеств, определенная нами на алгебре, порожденной  $\mathcal{K}$ , действительно будет непрерывной мерой (для этого, на самом деле, достаточно первых двух свойств). Значит, по теореме Каратеодори существует и единственно непрерывное продолжение меры на  $\sigma(\mathcal{K}) = \mathcal{B}$ <sup>48</sup> (нормированность меры, очевидно, сохраняется)  $\implies$  функция распределения действительно является исчерпывающей характеристикой распределения случайной величины.

**Теорема 4.** Пусть  $X$  — случайная величина, а  $F_X(x)$  — ее функция распределения; тогда  $F_X(x) \in \mathfrak{M}$ . Обратно: пусть  $F$  — некоторая функция, принадлежащая классу  $\mathfrak{M}$ ; тогда существует случайная величина  $X$  такая, что  $F$  является функцией распределения  $X$ , т.е.  $F_X = F$ .

**Замечание.** Важность этой теоремы состоит в том, что  $X$  можно указать **конструктивно**. Если у нас имеется датчик равномерно распределенной случайной величины  $\xi \sim U[0, 1]$ , то мы можем моделировать любую случайную величину, зная ее функцию распределения.

**Доказательство.**

1. Пусть  $F_X(x)$  — функция распределения:

$$F_X(x) = \mathbb{P}(X < x) = \mathbb{P}_X((-\infty, x)).$$

(a)  $F_X$  монотонно неубывает в силу монотонности меры:

$$\begin{aligned} \forall x_1 < x_2 &\implies (-\infty, x_1) \subset (-\infty, x_2) \implies \\ \mathbb{P}_X((-\infty, x_1)) &\leq \mathbb{P}_X((-\infty, x_2)) \implies F_X(x_1) \leq F_X(x_2) \end{aligned}$$

(b) Напомним, что непрерывность меры  $\mu$  означает, что если  $A_n \nearrow A$ , то  $\mu(A_n) \nearrow \mu(A)$ . В случае, если  $\mu$  конечна, это эквивалентно тому, что если  $B_n \searrow B$ , то

---

<sup>48</sup>при этом, в силу непрерывности слева функции распределения, получим  $\mathbb{P}_X([a, b]) = \sum_{k=1}^{\infty} \mathbb{P}_X([a_k, a_{k+1})) = \sum_{k=1}^{\infty} F_X(a_{k+1}) - F_X(a_k) = \lim_{k \rightarrow \infty} F_X(a_k) - F_X(a) = F_X(b) - F_X(a)$  при  $\sum_{k=1}^{\infty} [a_k, a_{k+1}) = [a, b]$

$\mu(B_n) \searrow \mu(B)$ <sup>49</sup>. Для бесконечной меры это неверно. В качестве примера достаточно взять меру Лебега, а в качестве системы множеств — систему сужающихся полос с пустым пересечением (рисунок 3). Тогда  $B_n \searrow \emptyset$ , но  $\mu(B_n) = +\infty \not\rightarrow 0$ .

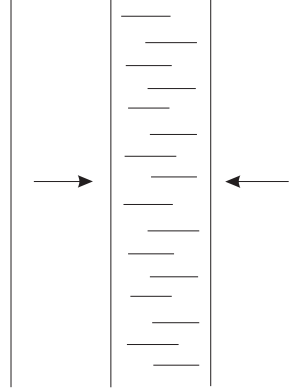


Рис. 3:

- i.  $F_X(x) \rightarrow 0$  при  $x \rightarrow -\infty$  (достаточно взять систему множеств  $\{B_n = (-\infty, x_n)\}$ ,  $x_n > x_{n+1}$ ,  $x_n \rightarrow -\infty$  при  $n \rightarrow \infty$ ,  $B_n \searrow \emptyset$ , в силу конечности, аддитивности и непрерывности вероятностной меры  $\lim_{n \rightarrow \infty} \mathbb{P}_X(B_n) = \mathbb{P}_X(\emptyset) = 0$ , а значит, по определению функции распределения,  $\lim_{n \rightarrow \infty} F_X(x_n) = 0$ );
  - ii.  $F_X(x) \rightarrow 1$  при  $x \rightarrow +\infty$  (достаточно взять систему множеств  $\{A_n = (-\infty, x_n)\}$ ,  $x_n < x_{n+1}$ ,  $x_n \rightarrow +\infty$  при  $n \rightarrow \infty$ ,  $A_n \nearrow (-\infty, +\infty)$ , в силу нормированности и непрерывности вероятностной меры  $\lim_{n \rightarrow \infty} \mathbb{P}_X(A_n) = \mathbb{P}_X((-\infty, +\infty)) = 1$ , а значит, по определению функции распределения,  $\lim_{n \rightarrow \infty} F_X(x_n) = 1$ ).
  - (с)  $F_X$  непрерывна слева (достаточно взять систему множеств  $\{A_n = (-\infty, x_n)\}$ ,  $x_n < x_{n+1}$ ,  $x_n \rightarrow x$  при  $n \rightarrow \infty$ ,  $A_n \nearrow (-\infty, x)$ , в силу непрерывности вероятностной меры  $\lim_{n \rightarrow \infty} \mathbb{P}_X(A_n) = \mathbb{P}_X((-\infty, x))$ , а значит, по определению функции распределения,  $\lim_{n \rightarrow \infty} F_X(x_n) = F_X(x)$ ).
2. Изложение **метода обращения функции распределения** (одного из методов построения одномерного датчика) составит *неформальное* доказательство существования случайной величины  $X$  в теореме. Попутно покажем, что случайная величина  $X$  определена неоднозначно по своей функции распределения.

Пусть  $F \in \mathfrak{M}$ .

Рассмотрим три случая. Сначала наложим на функцию  $F$  более жесткие требования, а потом будем постепенно от них отказываться.

I  $F$  строго монотонна и непрерывна (рисунок 4).

В этом случае у  $F$  существует обратная функция  $F^{-1}$  также строго монотонная и непрерывная. Мы предположили, что у нас имеется датчик равномерно распределенной случайной величины  $\xi \sim U[0, 1]$ . Положим

$$X := F^{-1}(\xi).$$

<sup>49</sup> действительно, возьмем  $A_n = B_1 \setminus B_n$ , тогда  $A = B_1 \setminus B \implies \{\lim_{n \rightarrow \infty} \mu(A_n) = \mu(A)\} \implies \lim_{n \rightarrow \infty} \mu(B_1 \setminus B_n) = \mu(B_1 \setminus B) \implies \{B_n \subseteq B_1, B \subseteq B_1\} \implies \lim_{n \rightarrow \infty} \mu(B_1) - \mu(B_n) = \mu(B_1) - \mu(B) \implies \lim_{n \rightarrow \infty} \mu(B_n) = \mu(B)$

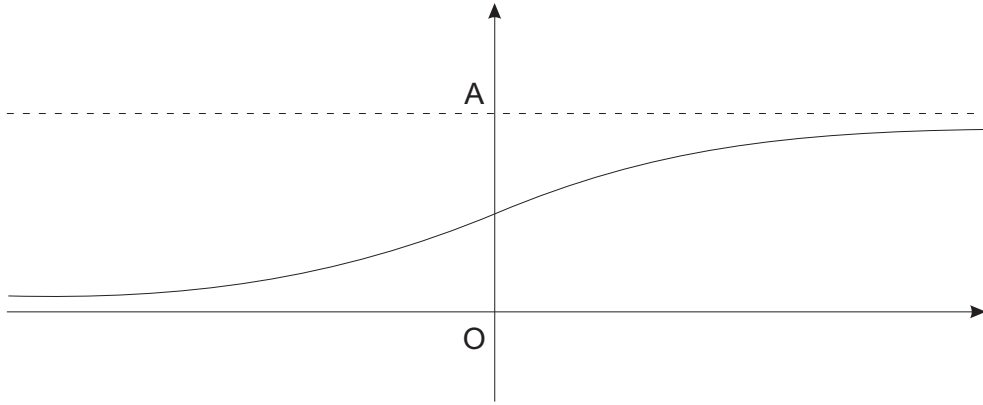


Рис. 4:

Покажем, что функцией распределения этой случайной величины является  $F$ .

$$\begin{aligned} F_X(x) &= \mathbb{P}(X < x) = \mathbb{P}(F^{-1}(\xi) < x) = \mathbb{P}(\xi < F(x)) = \mathbb{P}(\xi \in [0, F(x)]) = \\ &= \{\xi \sim U[0, 1]\} = F(x). \end{aligned}$$

Таким образом, в данном случае датчик для  $X$  построен.

## II Отказ от непрерывности.

В этом случае у функции  $F$  могут быть скачки (рисунок 5).

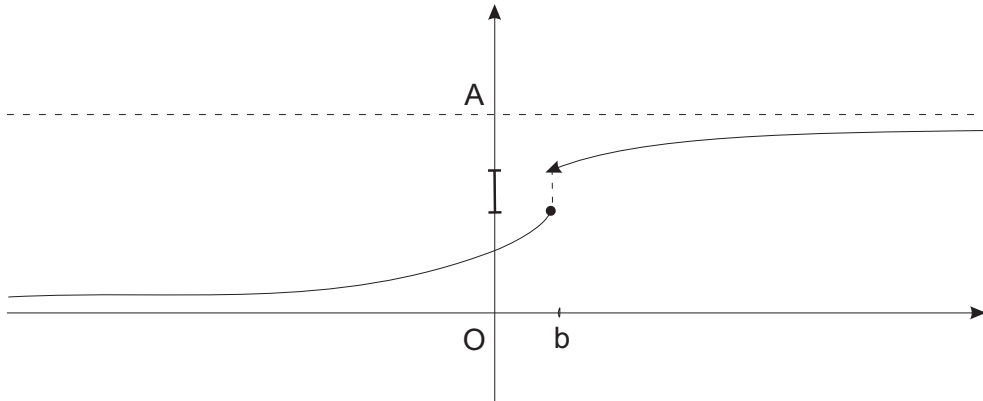


Рис. 5:

Заметим, что в этом случае обратной функции уже может и не быть (если скачки все-таки есть). При построении квазиобратной функции доопределим ее на интервалах, соответствующих скачкам, константой из соображений монотонности (на рисунке эта константа равна  $b$ ). Проверим, оправдано ли такое продолжение.

Из свойств функции распределения следует, что  $\mathbb{P}(X = b)$  (см. рисунок 5) равна величине скачка функции  $F_X$  в точке  $b$ . Посмотрим, чему получилась равна эта вероятность у нас.  $\xi$  разыгрывается на отрезке  $OA$ , т.е. на сегменте  $[0, 1]$ . Когда  $\xi$  находится в выделенном интервале,  $X$  по построению принимает одно и то же значение  $b$ . Но вероятность попадания  $\xi$  в выделенный интервал равна длине этого интервала, т.е. величине скачка, что и требовалось.

Итак, положим

$$X := \tilde{F}(\xi), \quad \xi \sim U[0, 1]$$

где  $\tilde{F}$  — квазиобратная к  $F$  функция. Покажем, что функцией распределения этой случайной величины является  $F$  (без ограничения общности положим, что  $F$  имеет единственный скачок в точке  $x = b$ ).

$$\begin{aligned}
 F_X(x) &= \mathbb{P}(X < x) = \mathbb{P}(\tilde{F}(\xi) < x) = \\
 &= \begin{cases} \mathbb{P}(\xi < F(x)), & \text{если } x \leq b \\
 \mathbb{P}(\tilde{F}(\xi) < b) + \mathbb{P}(\tilde{F}(\xi) = b) + \mathbb{P}(b < \tilde{F}(\xi) < x) = & \\
 \mathbb{P}(\xi < F(b)) + \mathbb{P}(\xi \in (F(b), F(b+0))) + \mathbb{P}(\xi \in (F(b+0), F(x))) = & \text{если } x > b \\
 \mathbb{P}(\xi < F(x)), & \\
 & = F(x).
 \end{cases}
 \end{aligned}$$

### III Отказ от строгой монотонности (наиболее общий случай).

В этой ситуации у функции могут быть “полочки” (рисунок 6).

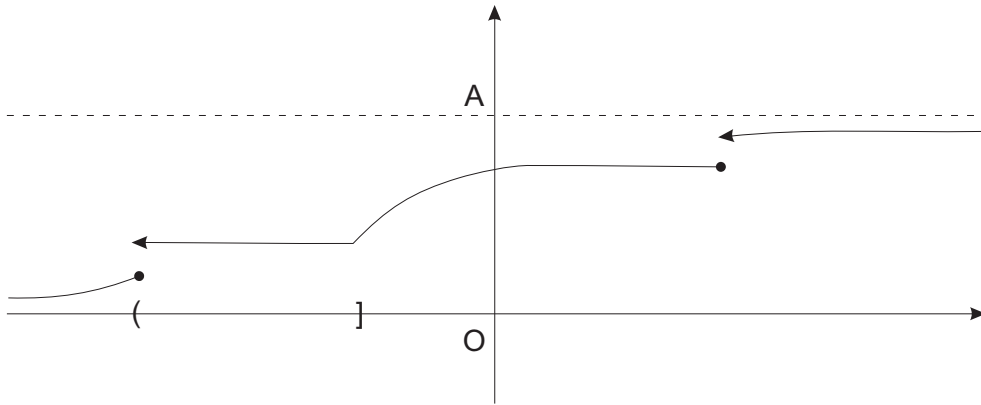


Рис. 6:

При построении квазиобратной функции выбрасываем “полочки” и доопределяем квазиобратную функцию в этих точках произвольным образом, например, можно положить  $\tilde{F}(y) = \sup\{x : F(x) = y\}$  (как для правой “полочки” на рисунке 7).

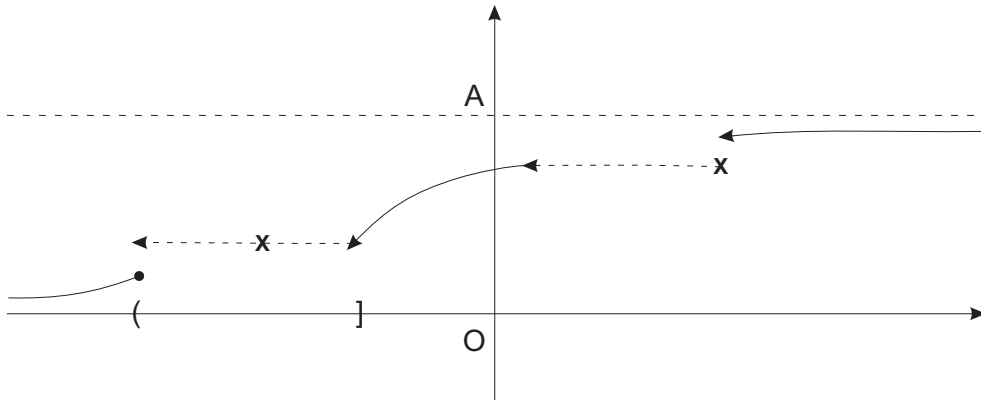


Рис. 7:

Из свойств функции распределения следует, что вероятность попадания  $X$  в выделенный полуинтервал равна нулю. Но аналогичный результат мы получаем и при моделировании  $X$  при помощи обращения датчика, т.к. для  $\xi$  вероятность попасть в точку равна нулю.



Итак, положим

$$X := \tilde{F}(\xi), \quad \xi \sim U[0, 1]$$

где  $\tilde{F}$  — квазиобратная к  $F$  функция. Покажем, что функцией распределения этой случайной величины является  $F$  (без ограничения общности положим, что  $F$  имеет единственную “полочку” на  $(a, b]$ ).

$$\begin{aligned} F_X(x) &= \mathbb{P}(X < x) = \mathbb{P}(\tilde{F}(\xi) < x) = \\ &= \begin{cases} \mathbb{P}(\xi < F(x)), & \text{если } x \leq a \\ \mathbb{P}(\tilde{F}(\xi) \leq a) + \mathbb{P}(a < \tilde{F}(\xi) < x) = \\ \mathbb{P}(\xi < F(a)) + 0 = \mathbb{P}(\xi < F(x)), & \text{если } x \in (a, b] \\ \mathbb{P}(\tilde{F}(\xi) \leq a) + \mathbb{P}(a < \tilde{F}(\xi) \leq b) + \mathbb{P}(b < \tilde{F}(\xi) < x) = \\ \mathbb{P}(\xi < F(a)) + 0 + \mathbb{P}(\xi \in (F(b), F(x))) = \\ \mathbb{P}(\xi < F(x)), & \text{если } x > b \end{cases} \\ &= F(x). \end{aligned}$$

**Замечание.** Способ обращения функции распределения является *универсальным методом* моделирования случайных величин. А это значит, что для некоторых конкретных случайных величин он может быть не самым эффективным.

**Пример.** I Рассмотрим экспоненциальное распределение:

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Разыграем  $\xi$  на  $[0, 1]$ . Если выпало  $\xi = 1$  или  $\xi = 0$ , то, согласно нашим договоренностям, все равно, какое мы при этом возьмем значение  $X$ , т.к.  $\mathbb{P}(\xi = 1) = \mathbb{P}(\xi = 0) = 0$ . Однако экспоненциальное распределение возникает в тех приложениях, где значения случайных величин неотрицательны. Поэтому естественно будет при определении  $X$  не рассматривать полупрямую  $(-\infty; 0)$  и полагать, например,  $X = 0$  при выпадении  $\xi = 0$ .

Обращаем функцию распределения:  $\tilde{F}(y) = -\frac{1}{\lambda} \ln(1 - y)$ . Хотелось бы сократить количество операций для вычисления  $X$ , поскольку при моделировании  $X$  число операций для вычисления одного значения умножается на число испытаний  $N$ . Т.е. для каждой операции получаем дополнительные  $N$  вычислений, а на практике  $N$  имеет порядок около 10000.

Итак, замечаем, что если  $\xi \sim U[0, 1]$ , то  $\eta = 1 - \xi$  имеет такое же распределение, что и  $\xi$ . Поэтому можем положить  $x = -\frac{1}{\lambda} \ln y$ , где  $y$  — значение равномерно распределенной на  $[0, 1]$  случайной величины.

II Рассмотрим стандартное нормальное распределение:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Для этой функции распределения нельзя аналитически построить обратную, а численно каждый раз это делать слишком долго. В таких случаях  $F(x)$  заранее рассчитывают в некоторых значениях и получают, таким образом, обратную функцию, заданную на неравномерной сетке. Однако тут есть одна тонкость с сеткой: для того чтобы полученная сеточная функция приближала искомую обратную функцию с сопоставимой точностью во всех точках, необходимо концы функции  $F(x)$  приблизить аналитически (иначе получим сильное сгущение сетки на концах интервала  $(0, 1)$ )<sup>50</sup>.

<sup>50</sup> см. Гнеденко

## 5 Альтернативный подход к аксиоматике на основе понятия среднего, эквивалентность подходу на основе понятия вероятности. Построение интеграла Лебега. Свойства математического ожидания (замена переменных, неравенство Йенсена и т.д.).

### 5.1 Альтернативный подход к аксиоматике на основе понятия среднего, эквивалентность подходу на основе понятия вероятности. Построение интеграла Лебега.

Аксиоматика Колмогорова брала за первичное понятие вероятности. Теперь рассмотрим альтернативный подход к аксиоматике на основе понятия среднего<sup>51</sup>. Заметим, что второй подход (когда за первичное берется понятие среднего) легче, т.к. в этом случае результаты получаются быстрее и проще, причем от понятия среднего легко перейти к понятию вероятности, обратное же требует построения интеграла Лебега. К тому же после дачи интерпретации среднего увидим, что вероятность значительно менее понятна с точки зрения того, что именно она формализует. Однако колмогоровский подход более традиционный (поэтому знать нужно и то, и другое;)).

Основной постулат Колмогорова гласит: любое непротиворечивое описание модели можно представить в виде универсальной формы — вероятностного пространства — тройки  $(\Omega, \mathcal{F}, \mathbb{P})$ , причем вероятность  $\mathbb{P}$ , являющаяся нормированной мерой, — основной объект, выступающий в роли первичного понятия, а измеримое пространство  $(\Omega, \mathcal{F})$  — его атрибут.

Другой подход заключается в том, чтобы в роли универсальной формы рассматривать тройку  $(\Omega, \mathcal{F}, \mathbb{E})$ , где в роли основного объекта<sup>52</sup> выступает  $\mathbb{E}$ <sup>53</sup>:

$$(\Omega, \mathcal{F}, \mathbb{P}) \longleftrightarrow (\Omega, \mathcal{F}, \mathbb{E})$$

Что же  $\mathbb{E}$  за объект? Так же, как и в свое время вероятности, дадим рассматриваемому объекту математическую и смысловую интерпретации (см. табличку в первой главе: обозначение/терминология/математическое содержание/интерпретация на метаязыке). В этой главе мы остановимся на математическом содержании, а про интерпретацию на метаязыке поговорим в следующей главе.

**Определение 57.**  $\mathbb{E}$  — *оператор математического ожидания (среднее)*.

Термин “оператор математического ожидания” больше отражает математическое содержание, а термин “среднее” — смысловую интерпретацию.

Рассмотрим пространство  $\mathcal{L}_+$  неотрицательных случайных величин<sup>54</sup>, т.е.

$$\mathcal{L}_+ = \{\xi: \Omega \rightarrow [0, +\infty), \xi \text{ — } (\mathcal{F}, \mathcal{B})\text{-измеримое отображение}\}^{55}.$$

<sup>51</sup>см. книгу Уиттла

<sup>52</sup>атрибуты  $(\Omega, \mathcal{F})$  удобно указывать по тем же причинам, что и в случае вероятностного пространства (см. первую главу)

<sup>53</sup> $\mathbb{P}$  от англ. probability (вероятность),  $\mathbb{E}$  от англ. expectation (математическое ожидание), однако в отечественной литературе встречается и другое обозначение —  $\mathbb{M}$  (мат. ожидание, middle (англ.), moyenne (фр.))

<sup>54</sup>заметим, что для корректного задания случайной величины достаточно иметь  $(\Omega, \mathcal{F})$ , т.е. вероятность, которая у нас пока не определена, нам и не нужна

Заметим, что  $\mathcal{L}_+$  — выпуклый конус<sup>56</sup>.

Итак, займемся определением  $\mathbb{E}$  как математического объекта.

$\mathbb{E}$  — это отображение

$$\mathbb{E}: \mathcal{L}_+ \rightarrow \mathbb{R}_+ \cup \{+\infty\}^{57},$$

обладающее следующими свойствами:

1) неотрицательность<sup>58</sup>:

$$\mathbb{E}: \mathcal{L}_+ \rightarrow [0, +\infty]^{59};$$

2) аддитивность:

$$\mathbb{E}(\xi_1 + \xi_2) = \mathbb{E}(\xi_1) + \mathbb{E}(\xi_2)^{60};$$

3) однородность (положительная)<sup>61</sup>:

$$\mathbb{E}(a\xi) = a\mathbb{E}\xi \quad \forall a > 0;$$

4) нормированность:

$$\mathbb{E}\mathbf{I}_\Omega = 1,$$

где  $\mathbf{I}_\Omega$  — случайная величина, тождественно равная единице<sup>62</sup>;

5) непрерывность:

$$\xi_n \nearrow \xi \quad (\text{т.е. } \xi_n(\omega) \geq 0 \text{ сходится монотонно поточечно к } \xi(\omega)) \implies \mathbb{E}\xi_n \nearrow \mathbb{E}\xi^{63}.$$

<sup>55</sup>можно рассматривать  $\xi: \Omega \rightarrow [0, +\infty]$ , но в этом нет необходимости

<sup>56</sup>Напомним, что множество  $K$  называется конусом, если  $\lambda K \equiv K \quad \forall \lambda > 0$ . Конус  $K$  называется выпуклым, если  $K$  — выпуклое множество. Можно также дать такое определение: конус  $K$  называется выпуклым, если  $x + y \in K, \forall x, y \in K$ .

Доказательство того, что  $\mathcal{L}_+$  является выпуклым конусом, опирается на следующие рассуждения. Понятие измеримости случайных величин, определенных в пространстве  $(\mathbb{R}, \mathcal{B})$ , эквивалентно понятию измеримости функций, введенному в курсе функционального анализа. Там же доказывалось, что линейная комбинация измеримых функций есть измеримая функция.

<sup>57</sup>Пространство  $\mathbb{R}_+ = [0, +\infty)$  локально компактно (пространство локально компактно, если для любой точки существует окрестность, замыкание которой компактно; примером пространства, не являющегося локально компактным, является пространство  $l_1$  (бесконечномерное банахово), в котором любой шар не является компактом). Есть такое утверждение: всегда можно к локально компактному пространству прибавить одну точку и получить компактное пространство. При добавлении к  $\mathbb{R}_+$  одной точки  $+\infty$  получим компакт, гомеоморфный отрезку.

<sup>58</sup>выделяем в силу важности свойства, хотя это уже записано

<sup>59</sup>Заметим, что возможен случай  $\mathbb{E}(\xi) = +\infty$  (включать точку  $+\infty$  не совсем типично, но мы делаем так). В связи с этим введем некоторые **соглашения**:

(а)  $+\infty + a = +\infty \quad \forall a \in [0, +\infty]$  (из соображений непрерывности),

(б)  $a \cdot (+\infty) = +\infty \quad \forall a \in (0, +\infty]$  (из соображений непрерывности),

(с)  $0 \cdot (+\infty)$  — не определено.

<sup>60</sup>заметим, что если  $\xi_1, \xi_2 \in \mathcal{L}_+$ , то и  $\xi_1 + \xi_2 \in \mathcal{L}_+$ , т.к.  $\mathcal{L}_+$  — выпуклый конус

<sup>61</sup>не можем взять  $a = 0$  только потому, если  $\mathbb{E}\xi = +\infty$ , то получаем  $a\mathbb{E}\xi = 0 \cdot (+\infty)$  — не определено; если же  $\mathbb{E}\xi \neq +\infty$ , то при  $a = 0$  получим  $\mathbb{E}\mathbf{I}_\emptyset = 0$  (действительно,  $\Omega + \emptyset = \Omega \implies \mathbf{I}_\Omega + \mathbf{I}_\emptyset = \mathbf{I}_\Omega \implies$  по свойству аддитивности  $\mathbb{E}\mathbf{I}_\Omega + \mathbb{E}\mathbf{I}_\emptyset = \mathbb{E}\mathbf{I}_\Omega \implies \mathbb{E}\mathbf{I}_\emptyset = 0$ )

<sup>62</sup>вообще говоря,  $\mathbf{I}_A$  — индикатор события  $A$ , т.е.  $\mathbf{I}_A(\omega) = \begin{cases} 1, & \text{если } \omega \in A, \\ 0, & \text{если } \omega \notin A. \end{cases}$

<sup>63</sup>не исключая ситуации, когда  $\mathbb{E}\xi = +\infty$

Итак, мы определили  $\mathbb{E}$  как математический объект. Про интерпретацию (объективную и субъективную) поговорим в следующей главе.

Вопрос: не является ли пустым класс объектов вида  $(\Omega, \mathcal{F}, \mathbb{E})$ ? Ответ: нет; обоснование: можно привести примеры с конечным  $\Omega$  (построить в явном виде  $\mathbb{E}$ ).

Покажем эквивалентность подходов через понятия вероятности и среднего.

Заметим, что основной постулат Колмогорова с учетом этой эквивалентности можно сформулировать следующим образом. *Любое непротиворечивое описание можно представить в виде пространства со средним  $(\Omega, \mathcal{F}, \mathbb{E})$*

Рассмотрим связь двух моделей. Покажем, что из одной модели можно всегда получить другую, что и доказывает эквивалентность двух подходов.

1.  $(\Omega, \mathcal{F}, \mathbb{P}) \longleftarrow (\Omega, \mathcal{F}, \mathbb{E})$ :

Пусть первично математическое ожидание. Тогда имеет место **формула связи (аксиома)**:

$$\mathbb{P}(A) := \mathbb{E}\mathbf{I}_A \quad \forall A \in \mathcal{F}, \quad (26)$$

где

$$\mathbf{I}_A = \begin{cases} 1, & \text{с вероятностью } \mathbb{P}(A), \\ 0, & \text{с вероятностью } 1 - \mathbb{P}(A). \end{cases}$$

Проверим, что так определенное отображение  $\mathbb{P}: \mathcal{F} \rightarrow \mathbb{R}$  действительно будет нормированной мерой.

(a) аддитивность: из определения индикатора и свойства аддитивности для  $\mathbb{E}$

$$\mathbf{I}_{A_1+A_2} = \mathbf{I}_{A_1} + \mathbf{I}_{A_2} \implies \mathbb{E}\mathbf{I}_{A_1+A_2} = \mathbb{E}\mathbf{I}_{A_1} + \mathbb{E}\mathbf{I}_{A_2} \implies \mathbb{P}(A_1+A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2);$$

(b) неотрицательность следует из свойства неотрицательности для  $\mathbb{E}$ ;

(c) непрерывность: из определения индикатора и свойства непрерывности для  $\mathbb{E}$

$$A_n \nearrow A \iff \mathbf{I}_{A_n} \nearrow \mathbf{I}_A \quad (\text{поточечно}) \implies \mathbb{E}\mathbf{I}_{A_n} \nearrow \mathbb{E}\mathbf{I}_A \implies \mathbb{P}(A_n) \nearrow \mathbb{P}(A);$$

(d) нормированность следует из нормированности для  $\mathbb{E}$ .

Вывод: если взять  $\mathbb{E}$  в качестве первичного объекта и постулировать аксиому связи (26), то  $\mathbb{P}$  будет нормированной мерой.

2.  $(\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\Omega, \mathcal{F}, \mathbb{E})$ :

Пусть теперь первична вероятность. Тогда формула связи (26) должна оставаться такой же. Но можно ли из нее однозначно определить  $\mathbb{E}$ ? На этот вопрос отвечает следующая теорема.

**Теорема 5.**  $\exists! \mathbb{E}$ , удовлетворяющее формуле связи (26) (по формуле связи можно однозначно задать  $\mathbb{E}$ ).

Доказательство (**построение интеграла Лебега**). Итак, нам дана  $\mathbb{P}$ , надо построить  $\mathbb{E}$ . Будем применять рассуждения вида: если мы хотим построить среднее, то неизбежно должны быть выполнены следующие свойства.

Из построения  $\mathbb{E}$  будет видно, что по сути это построение интеграла Лебега. Поэтому применяется следующее обозначение:

$$\mathbb{E}\xi = \int \xi d\mathbb{P} = \int_{\Omega} \xi(\omega) \mathbb{P}(d\omega)^{64}.$$

- 0)  $\mathbb{E}\mathbf{I}_A := \mathbb{P}(A) \quad \forall A \in \mathcal{F}$  (из аксиомы связи).
- 1)  $\mathbb{E}(a\mathbf{I}_A) := a\mathbb{E}\mathbf{I}_A = a\mathbb{P}(A) \quad \forall a \geq 0 \quad \forall A \in \mathcal{F}$  (из свойства однородности математического ожидания<sup>65</sup>).
- 2)  $\mathbb{E}\left(\sum_{i=1}^n a_i \mathbf{I}_{A_i}\right) := \sum_{i=1}^n a_i \mathbb{P}(A_i) \quad \forall a_i \geq 0 \quad \forall A_i \in \mathcal{F} \quad n = 1, 2, \dots$  (из свойства аддитивности математического ожидания).
- 3) Определим математическое ожидание от положительной дискретной случайной величины. Дискретная случайная величина — случайная величина, принимающая не более, чем счетное число значений, т.е. представимая в виде  $\xi = \sum_{i=1}^{\infty} a_i \mathbf{I}_{A_i}$ ,  $A_i \cap A_j = \emptyset$ <sup>66</sup>,  $\forall i \neq j$ ,  $\sum_{i=1}^{\infty} A_i = \Omega$ . Положим  $\mathbb{E}\left(\sum_{i=1}^{\infty} a_i \mathbf{I}_{A_i}\right) := \sum_{i=1}^{\infty} a_i \mathbb{P}(A_i) \quad \forall a_i \geq 0 \quad \forall A_i \in \mathcal{F}$  (из соображений непрерывности:  $\xi_n = \sum_{i=1}^n a_i \mathbf{I}_{A_i}$ ,  $\xi_n \nearrow \xi = \sum_{i=1}^{\infty} a_i \mathbf{I}_{A_i} \implies$  по свойству непрерывности математического ожидания  $\mathbb{E}\xi_n \nearrow \mathbb{E}\xi \implies$  из 2) получим  $\sum_{i=1}^n a_i \mathbb{P}(A_i) \nearrow \mathbb{E}\left(\sum_{i=1}^{\infty} a_i \mathbf{I}_{A_i}\right)$ )<sup>67</sup>.
- 4) Определим математическое ожидание от произвольной неотрицательной случайной величины путем ее равномерного приближения дискретной. Рассмотрим произвольную неотрицательную случайную величину  $\xi: \Omega \rightarrow [0, +\infty]$ . Разобьем  $[0, +\infty]$  на счетное число полуинтервалов длины меньше  $\varepsilon$ :  $[0, +\infty] = \bigcup_{k=1}^{\infty} [a_{k-1}, a_k)$ ,  $a_0 = 0$ ,  $a_k < a_{k+1}$ ,  $a_k - a_{k-1} < \varepsilon$ . Обозначим через  $I_k = \{\omega : \xi(\omega) \in [a_{k-1}, a_k)\} = \xi^{-1}([a_{k-1}, a_k))$ , в силу условия измеримости  $I_k \in \mathcal{F}$ . Рассмотрим неотрицательные дискретные случайные величины  $\xi'_\varepsilon = \sum_{k=1}^{\infty} a_{k-1} \mathbf{I}_{I_k}$ ,  $\xi''_\varepsilon = \sum_{k=1}^{\infty} a_k \mathbf{I}_{I_k}$ . Очевидно, что  $\xi'_\varepsilon \leq \xi \leq \xi''_\varepsilon$ ,  $0 \cdot \mathbf{I}_\Omega \leq \xi''_\varepsilon - \xi'_\varepsilon < \varepsilon \cdot \mathbf{I}_\Omega \implies \exists$  неотрицательная дискретная случайная величина  $\xi_\varepsilon$  такая, что  $\|\xi - \xi_\varepsilon\|_\infty < \varepsilon$ <sup>68</sup>. Таким образом, мы заменили произвольную неотрицательную измеримую функцию на кусочно-постоянную, отличающуюся по норме от исходной меньше, чем на  $\varepsilon$ . Положим  $\mathbb{E}\xi := \lim_{\varepsilon \rightarrow 0} \mathbb{E}\xi_\varepsilon$  (из соображений непрерывности,  $\xi_\varepsilon \nearrow \xi$ )<sup>69</sup>. Замечание: неотрицательную случайную величину всегда можно интегрировать, просто иногда может получиться  $\mathbb{E}\xi = \int \xi(\omega) \mathbb{P}(d\omega) = +\infty$ .

Итак, мы определили однозначно отображение  $\mathbb{E}: \mathcal{L}_+ \rightarrow [0, +\infty]$ . Очевидно, что оно удовлетворяет всем 5 свойствам (непрерывность доказывается просто, но довольно громоздко, если действовать формально).

<sup>64</sup>  $\Omega$  можно опустить, т.к. интегрирование всегда идет по всему пространству

<sup>65</sup> можем брать и  $a = 0$ , т.к.  $\mathbb{E}\mathbf{I}_A \leq 1$

<sup>66</sup> это обеспечивает  $\xi(\omega) < +\infty \quad \forall \omega \in \Omega$ , если же  $A_i$  пересекаются, то нужно требовать сходимости ряда  $\sum_{i=1}^{\infty} a_i \mathbf{I}_{A_i} < \infty$

<sup>67</sup> Заметим, что дискретную случайную величину можно представить разными способами. Однако измельчением можно показать, что определение корректно.

<sup>68</sup>  $\|\nu\|_\infty = \sup_{\omega \in \Omega} |\nu(\omega)|$

<sup>69</sup> корректность (независимость от разбиения) можно опять же показать измельчением разбиения

Итак, мы показали эквивалентность двух подходов. При этом попутно был построен интеграл Лебега для неотрицательной случайной величины (это неотрицательное число, возможно равное  $+\infty$ ). Хотелось бы иметь возможность интегрировать и знакопеременные случайные величины, т.е. определить  $\mathbb{E}\xi$ , где  $\xi: \Omega \rightarrow \mathbb{R}$ . Заметим, что в этом случае придется отказаться от неотрицательности математического ожидания.

Основная идея в интегрировании знакопеременных случайных величин состоит в представлении их в виде разности неотрицательных случайных величин:

$$\xi = \xi^1 - \xi^2, \quad \xi^1 \geq 0, \quad \xi^2 \geq 0.$$

Ясное дело, что такое представление далеко не единственно. Будем брать одно из возможных, например<sup>70</sup>:

$$\xi = \xi^+ - \xi^-, \quad \xi^+(\omega) = \max\{\xi(\omega), 0\} \geq 0, \quad \xi^-(\omega) = \max\{-\xi(\omega), 0\} \geq 0.$$

Из соображений линейности (для сохранения свойства аддитивности) положим

$$\mathbb{E}\xi := \mathbb{E}\xi^+ - \mathbb{E}\xi^-. \quad (27)$$

Легко можно проверить сохранение свойств аддитивности, однородности, нормированности и непрерывности.

**Замечание.** Когда хотят сказать, что существует конечное математическое ожидание (т.е. функция интегрируема (суммируема) по Лебегу), то пишут  $\mathbb{E}|\xi| < \infty$ . Поясним эту запись. Рассмотрим 4 случая:

1.  $\mathbb{E}\xi^+ < +\infty, \mathbb{E}\xi^- < +\infty \implies$  формула (27) применима и  $\mathbb{E}\xi$  конечно;
2.  $\mathbb{E}\xi^+ < \infty, \mathbb{E}\xi^- = +\infty \implies$  формула (27) применима и  $\mathbb{E}\xi = -\infty$ ;
3.  $\mathbb{E}\xi^+ = +\infty, \mathbb{E}\xi^- < +\infty \implies$  формула (27) применима и  $\mathbb{E}\xi = +\infty$ ;
4.  $\mathbb{E}\xi^+ = \mathbb{E}\xi^- = +\infty \implies$  формула (27) не применима и, значит,  $\mathbb{E}\xi$  не существует (не определено).

Получили, что  $\exists \mathbb{E}\xi$  в первых трех случаях и  $\nexists \mathbb{E}\xi$  при  $\mathbb{E}\xi^+ = \mathbb{E}\xi^- = +\infty$ . Конечное математическое ожидание существует только в первом случае, следовательно:

$$\mathbb{E}|\xi| < \infty \iff \begin{cases} \mathbb{E}\xi^+ < +\infty, \\ \mathbb{E}\xi^- < +\infty. \end{cases}$$

Это можно объяснить кратко так:  $|\xi| = \xi^+ + \xi^-$ .

Мы построили интеграл Лебега для конечной меры (вероятностная мера  $\mathbb{P}$  — нормированная мера). Теперь рассмотрим  $\sigma$ -конечную меру.

**Определение 58.** Мера  $\mu: X \rightarrow [0, +\infty]$  называется  **$\sigma$ -конечной мерой**, если существует такое разбиение пространства  $X$  на непересекающиеся множества  $X = \sum_{k=1}^{+\infty} X_k$ , что  $\mu(X_k) < \infty$ .

**Утверждение 9.** Для любой  $\sigma$ -конечной меры  $\mu$  можно найти конечную меру  $\nu$ , такую что  $\nu \sim \mu$  (меры эквивалентны, т.е. взаимно абсолютно непрерывны:  $\nu \sim \mu \iff \nu \ll \mu, \mu \ll \nu$ ), причем производная Радона-Никодима положительна почти всюду по мере  $\nu$ :  $g(x) = \frac{d\mu}{d\nu}(x) > 0$ .

---

<sup>70</sup>здесь используется известный прием:  $a \in \mathbb{R}, a^+ = \max\{a, 0\}, a^- = \max\{-a, 0\} \implies a = a^+ - a^-, |a| = a^+ + a^-$

Доказательство. По определению  $\nu \sim \mu$ :

$$\nu(A) = 0 \iff \mu(A) = 0.$$

По определению  $\sigma$ -конечной меры:

$$\exists X_k : X = \sum_{k=1}^{+\infty} X_k, \quad \mu(X_k) < \infty.$$

Возьмем

$$\nu(A) = \sum_{k=1}^{\infty} w_k \frac{\mu(A \cap X_k)}{\mu(X_k)}, \quad \sum_{k=1}^{\infty} w_k = 1, \quad w_k \geq 0.$$

Данная мера конечна:  $\nu(X) = 1$  (в том, что это мера, сомнений нет). Покажем эквивалентность мер.

Пусть  $\mu(A) = 0$ . Тогда  $\mu(A \cap X_k) = 0 \forall k$ , откуда  $\nu(A) = 0$ .

Пусть  $\nu(A) = 0$ . Тогда  $\frac{\mu(A \cap X_k)}{\mu(X_k)} = 0 \forall k$ , откуда  $\mu(A \cap X_k) = 0 \forall k$ . Получаем, что  $\mu(A) = \sum_{k=1}^{\infty} \mu(A \cap X_k) = 0$ .

Покажем, что производная Радона-Никодима положительна почти всюду по мере  $\nu$ :  $g(x) = \frac{d\mu}{d\nu}(x) > 0$ . Положим  $A = \{x : g(x) = 0\}$ . Тогда по определению производной Радона-Никодима  $\mu(A) = \int_A g(x) \nu(dx) = 0$ . Так как  $\mu(A) = 0$ , по в силу эквивалентности мер  $\nu(A) = 0$ .

Утверждение доказано.

Таким образом, получаем, что нет нужды отдельно рассматривать интеграл Лебега для  $\sigma$ -конечной меры, ведь

$$\int f d\mu = \int f g d\nu,$$

где  $\mu$  —  $\sigma$ -конечная мера, а  $\nu$  — конечная (вероятностная) мера. Другими словами, интеграл по  $\sigma$ -конечной мере можно свести к интегралу по конечной мере.

## 5.2 Свойства математического ожидания (замена переменных, неравенство Йенсена и т.д.).

Теперь рассмотрим свойства математического ожидания. Так как с математической точки зрения это интеграл Лебега, то ему присущи все свойства интеграла Лебега. Перечислим некоторые из них (свойств интеграла очень много, их можно найти любой книжке).

1.  $\mathbb{E}(a\xi_1 + b\xi_2) = a\mathbb{E}\xi_1 + b\mathbb{E}\xi_2$  (линейность);
2.  $\xi_1 \leq \xi_2 \implies \mathbb{E}\xi_1 \leq \mathbb{E}\xi_2$  (монотонность);
3.  $|\mathbb{E}\xi| \leq \mathbb{E}|\xi|$ ;

и т.д.

Перечислим теоремы о предельном переходе под знаком интеграла.

**Теорема 6** (Лебег). Пусть функции  $f_n$  и  $f$  измеримы на  $X$ ,  $f_n \rightarrow f$  по мере на  $X$  (т.е.  $\forall \varepsilon > 0 \mu(|f - f_n| > \varepsilon) \rightarrow 0$  при  $n \rightarrow \infty$ ), существует интегрируемая функция  $F(x)$  такая, что  $|f_n(x)| \stackrel{n.б.}{\leq} F(x) \forall n = 1, 2, \dots$ . Тогда  $f_n$  и  $f$  интегрируемы и  $\lim_{n \rightarrow \infty} \int_X f_n d\mu = \int_X f d\mu$ .



Заметим, что из сходимости почти всюду следует сходимость по мере.

Интерпретация в терминах теории вероятностей:  $\xi_n \rightarrow \xi$  по вероятности,  $\exists \nu: \mathbb{E}|\nu| < \infty$ ,  $|\xi_n(\omega)| \leq \nu(\omega) \forall n = 1, 2, \dots \implies \mathbb{E}|\xi_n| < \infty, \mathbb{E}|\xi| < \infty, \lim_{n \rightarrow \infty} \mathbb{E}\xi_n = \mathbb{E}\xi$ .

**Лемма 1** (Беппо Леви). Пусть функции  $f_n$  интегрируемы по  $X$ ,  $f_n(x) \overset{n.б.}{\leq} f_{n+1}(x) \forall n = 1, 2, \dots$ ,  $\exists C > 0: \int f_n d\mu \leq C \forall n = 1, 2, \dots$ . Тогда  $\lim_{n \rightarrow \infty} f_n = f$  — интегрируемая по  $X$  функция и  $\lim_{n \rightarrow \infty} \int_X f_n d\mu = \int_X f d\mu$ .

Свойство непрерывности математического ожидания  $\xi_n \nearrow \xi \implies \mathbb{E}\xi_n \nearrow \mathbb{E}\xi$  является расширенным вариантом леммы Беппо Леви, т.к. мы не исключаем случая  $\mathbb{E}\xi = +\infty$ .

**Теорема 7** (Фату). Пусть функции  $f_n(x) \geq 0$  интегрируемы по  $X$ ,  $\lim_{n \rightarrow \infty} f_n(x) \overset{n.б.}{=} f(x)$ ,  $\exists C > 0: \int f_n d\mu \leq C \forall n = 1, 2, \dots$ . Тогда  $f(x)$  — интегрируема и  $\int_X f d\mu \leq C$ .

Это следствие леммы Леви (достаточно положить  $F_n(x) = \inf_{k \geq n} \{f_k(x)\}$ ). Заметим, что в теореме Фату предельный переход невозможен (пример:  $f_n(x) = 0$  при  $x \in 0 \cup [\frac{1}{n}, 1]$  и  $f_n(x) = n$  при  $x \in (0, \frac{1}{n}) \implies \lim_{n \rightarrow \infty} f_n = 0$ ,  $\int_{[0,1]} f_n d\mu = 1 \leq C = 1$ ).

Интерпретация в терминах теории вероятностей:  $\xi_n \geq 0$ ,  $\xi_n \xrightarrow{n.б.} \xi$ ,  $\exists C > 0: \mathbb{E}\xi_n \leq C \forall n = 1, 2, \dots \implies \mathbb{E}\xi \leq C$ .

Теперь докажем *теорему о замене переменных в интеграле Лебега*, сформулированную в терминах теории вероятностей<sup>71</sup>.

**Теорема 8** (о замене переменных в интеграле Лебега). Рассмотрим два измеримых пространства  $(\Omega, \mathcal{F})$  и  $(E, \mathcal{E})$  и два отображения:

$$\begin{aligned} X: \Omega &\rightarrow E \text{ — } (\mathcal{F}, \mathcal{E})\text{-измеримое,} \\ f: E &\rightarrow \mathbb{R} \text{ — } (\mathcal{E}, \mathcal{B})\text{-измеримое.} \end{aligned}$$

Пусть  $\mathbb{P}$  — вероятностная мера в первом пространстве, т.е. получаем вероятностное пространство  $(\Omega, \mathcal{F}, \mathbb{P})$ . Тогда можем рассмотреть индуцированную меру  $\mathbb{P}_X = X \circ \mathbb{P} = \mathbb{P} \circ X^{-1}$ , являющуюся, по определению, распределением случайного элемента  $X$  (каноническая модель  $(E, \mathcal{E}, \mathbb{P}_X)$ ).

Утверждается, что

$$\int_{\Omega} f(X(\omega)) \mathbb{P}(d\omega) = \int_E f(x) \mathbb{P}_X(dx), \quad (28)$$

причем под равенством интегралов подразумевается, что

либо оба интеграла существуют и тогда они равны (включая случай, когда интегралы равны  $+\infty$ ),

либо оба интеграла не существуют.

Доказательство. Схема доказательства полностью совпадает со схемой построения интеграла Лебега (см. выше).

<sup>71</sup>эта вещь простая и часто используемая, мы останавливаемся на ней подробнее для того, чтобы лучше понять ее суть

0)  $f(x) = \mathbf{I}_A(x), \forall A \in \mathcal{E}$ .

Тогда, по определению математического ожидания от случайной величины  $\mathbf{I}_A: E \rightarrow \mathbb{R}$ , получим  $\mathbb{E}\mathbf{I}_A = \mathbb{P}_X(A)$  или, если расписать через интеграл,  $\int_E f(x)\mathbb{P}_X(dx) = \mathbb{P}_X(A)$ .

С другой стороны,  $\mathbf{I}_A(X(\omega)) \stackrel{\text{def}}{=} \begin{cases} 1, & \text{если } X(\omega) \in A, \\ 0, & \text{если } X(\omega) \notin A. \end{cases} = \begin{cases} 1, & \text{если } \omega \in X^{-1}(A), \\ 0, & \text{если } \omega \notin X^{-1}(A). \end{cases} \stackrel{\text{def}}{=} \mathbf{I}_{X^{-1}(A)}(\omega)$ .

Значит,  $f(X(\omega)) = \mathbf{I}_A(X(\omega)) = \mathbf{I}_{X^{-1}(A)}(\omega)$ . Заметим, что в силу измеримости  $X$   $X^{-1}(A) \in \mathcal{F}$ . Тогда, по определению математического ожидания от случайной величины  $\mathbf{I}_{X^{-1}(A)}: \Omega \rightarrow \mathbb{R}$ , получим  $\mathbb{E}\mathbf{I}_{X^{-1}(A)} = \mathbb{P}(X^{-1}(A))$  или, если расписать через интеграл,  $\int_{\Omega} f(X(\omega))\mathbb{P}(d\omega) = \mathbb{P}(X^{-1}(A))$ .

По определению  $\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A))$ . Следовательно, интегралы существуют и равны:

$$\int_{\Omega} f(X(\omega))\mathbb{P}(d\omega) = \int_E f(x)\mathbb{P}_X(dx)$$

$\implies$  формула (28) выполнена.

1)  $f(x) = a\mathbf{I}_A(x), \forall A \in \mathcal{E}, \forall a \geq 0$ .

Доказательство формулы (28) полностью аналогично случаю 0), за исключением того, что вместо  $\mathbb{E}\mathbf{I}_A = \mathbb{P}_X(A)$  используется формула  $\mathbb{E}(a\mathbf{I}_A) = a\mathbb{P}_X(A)$ .

2)  $f(x) = \sum_{i=1}^n a_i \mathbf{I}_{A_i}, \forall A_i \in \mathcal{E}, \forall a_i \geq 0, n = 1, 2, \dots$

Доказательство формулы (28) полностью аналогично случаю 0), за исключением того, что вместо  $\mathbb{E}\mathbf{I}_A = \mathbb{P}_X(A)$  используется формула  $\mathbb{E}\left(\sum_{i=1}^n a_i \mathbf{I}_{A_i}\right) = \sum_{i=1}^n a_i \mathbb{P}_X(A_i)$ .

3)  $f(x) = \sum_{i=1}^{\infty} a_i \mathbf{I}_{A_i}, A_i \cap A_j = \emptyset, \forall i \neq j, \sum_{i=1}^{\infty} A_i = E, A_i \in \mathcal{E}, \forall a_i \geq 0$ .

Доказательство формулы (28) полностью аналогично случаю 0), за исключением того, что вместо  $\mathbb{E}\mathbf{I}_A = \mathbb{P}_X(A)$  используется формула  $\mathbb{E}\left(\sum_{i=1}^{\infty} a_i \mathbf{I}_{A_i}\right) = \sum_{i=1}^{\infty} a_i \mathbb{P}_X(A_i)$ .

4)  $f(x) \geq 0$ .

Аналогично тому, как это делалось при построении интеграла Лебега, находим неотрицательную дискретную случайную величину  $f_{\varepsilon} = \sum_{i=1}^{\infty} a_i \mathbf{I}_{A_i}$  такую, что  $\|f - f_{\varepsilon}\|_{\infty} < \varepsilon$

( $\implies \|f \circ X - f_{\varepsilon} \circ X\|_{\infty} < \varepsilon, f_{\varepsilon} \circ X: \Omega \rightarrow \mathbb{R}, f_{\varepsilon} \circ X = \sum_{i=1}^{\infty} a_i \mathbf{I}_{X^{-1}(A_i)}$  — неотрицательная дискретная случайная величина).

Из предыдущего пункта  $\mathbb{E} \sum_{i=1}^{\infty} a_i \mathbf{I}_{A_i} = \mathbb{E} \sum_{i=1}^{\infty} a_i \mathbf{I}_{X^{-1}(A_i)}$ , т.е.  $\mathbb{E}f_{\varepsilon} = \mathbb{E}(f_{\varepsilon} \circ X)$ . Далее используем переход к пределу  $\mathbb{E}f = \lim_{\varepsilon \rightarrow 0} \mathbb{E}f_{\varepsilon}, \mathbb{E}f \circ X = \lim_{\varepsilon \rightarrow 0} \mathbb{E}f_{\varepsilon} \circ X$  и получаем  $\mathbb{E}f = \mathbb{E}(f \circ X)$ , что эквивалентно формуле (28).

5)  $f(x)$  — знакопеременная случайная величина.

Воспользуемся формулой  $\mathbb{E}f = \mathbb{E}f^{+} - \mathbb{E}f^{-}$ , где  $f^{+} \geq 0$  и  $f^{-} \geq 0$ . Ясно, что  $(f \circ X)^{+} = f^{+} \circ X, (f \circ X)^{-} = f^{-} \circ X$ . Значит,  $\mathbb{E}(f \circ X) = \mathbb{E}f^{+} \circ X - \mathbb{E}f^{-} \circ X$ . Из предыдущего пункта  $\mathbb{E}f^{+} = \mathbb{E}(f^{+} \circ X), \mathbb{E}f^{-} = \mathbb{E}(f^{-} \circ X)$ . Следовательно,  $\mathbb{E}f = \mathbb{E}(f \circ X)$ , что и т.д.

Заметим, что в пунктах 0)–4) интегралы всегда существуют (конечные или бесконечные). В случае 5) нужно рассмотреть 4 случая, как это делалось при построении интеграла Лебега (в случае 4 интегралы не существуют, т.к. возникает неопределенность вида  $(+\infty) - (+\infty)$ ).

Теорема доказана.

**Смысл теоремы.** Для того чтобы посчитать математическое ожидание от функции от случайного элемента  $\mathbb{E}f(x)$ , достаточно знать распределение этого случайного элемента  $\mathbb{P}_X$  (т.е. знать полностью вероятность  $\mathbb{P}$  не нужно, модель  $(\Omega, \mathcal{F}, \mathbb{P})$  можно без потери необходимой для подсчета  $\mathbb{E}f(x)$  информации заменить на  $(E, \mathcal{E}, \mathbb{P}_X)$ ):

$$\mathbb{E}f(x) = \int_E f(x) \mathbb{P}_X(dx) = \int f d\mathbb{P}_X.$$

Таким образом,  $\mathbb{E}f(x)$  — это характеристика распределения  $X$ <sup>72</sup>.

В заключении этой главы рассмотрим два неравенства (и их следствия), очень часто используемые в теории вероятностей.

Первое неравенство используется для оценки хвостов распределений.

**Теорема 9.** Для любой случайной величины  $Y \geq 0$  и любого числа  $b > 0$  справедливо следующее неравенство<sup>73</sup>:

$$\mathbb{P}(Y \geq b) \leq \frac{\mathbb{E}Y}{b}, \quad (29)$$

причем равенство достигается  $\iff$  когда носитель распределения  $Y$  сосредоточен на множестве  $\{0, b\}$  (т.е.  $Y$  с вероятностью 1 принимает лишь два значения — 0 и  $b$ ).

Доказательство. Приведем графическое доказательство, из которого легко понять, когда достигается равенство.

Рассмотрим две числовые функции  $f_1: [0, +\infty) \rightarrow [0, +\infty)$ ,  $f_2: [0, +\infty) \rightarrow [0, +\infty)$ :

$$f_1(y) = \frac{y}{b}, \quad f_2(y) = \mathbf{I}_{[b, +\infty)}(y) = \begin{cases} 1, & \text{если } y \in [b, +\infty), \\ 0, & \text{если } y \in [0, b). \end{cases}$$

Из рисунка ( ) видно, что

$$\frac{y}{b} \geq \mathbf{I}_{[b, +\infty)}(y), \quad \forall y \in [0, +\infty),$$

причем равенство достигается  $\iff y \in \{0, b\}$ .

Рассмотрим неотрицательные случайные величины  $f_1 \circ Y$  и  $f_2 \circ Y$  ( $f_1 \circ Y(\omega) = f_1(Y(\omega))$ ,  $f_2 \circ Y(\omega) = f_2(Y(\omega))$ ). В силу  $f_1(y) \geq f_2(y)$  получаем, что  $f_1 \circ Y \geq f_2 \circ Y$ . Заметим, что  $f_1 \circ Y \geq 0$  и  $f_2 \circ Y \geq 0$ , а значит, математическое ожидание от этих случайных величин существует всегда. В силу свойства монотонности  $\mathbb{E}(f_1 \circ Y) \geq \mathbb{E}(f_2 \circ Y)$ .

Так как  $f_1 \circ Y = \frac{Y}{b}$ ,  $f_2 \circ Y = \mathbf{I}_{[b, +\infty)}(Y) = \mathbf{I}_{Y^{-1}([b, +\infty))}$ , то получим

$$\begin{aligned} \mathbb{E}(f_1 \circ Y) &= \mathbb{E} \frac{Y}{b} = \{\text{свойство однородности мат. ожидания}\} = \frac{\mathbb{E}Y}{b}, \\ \mathbb{E}(f_2 \circ Y) &= \mathbb{E} \mathbf{I}_{Y^{-1}([b, +\infty))} = \{\mathbb{E} \mathbf{I}_A = \mathbb{P}(A)\} = \mathbb{P}(Y^{-1}([b, +\infty))) = \mathbb{P}(Y \geq b), \\ \mathbb{E}(f_1 \circ Y) &\geq \mathbb{E}(f_2 \circ Y) \implies \frac{\mathbb{E}Y}{b} \geq \mathbb{P}(Y \geq b), \end{aligned}$$

<sup>72</sup>В отличие от функции распределения,  $\mathbb{E}f(x)$  не является, очевидно, исчерпывающей характеристикой распределения

<sup>73</sup>включая случай  $\mathbb{E}Y = +\infty$  (тривиальный результат)

т.е. получили искомое неравенство (29).

Из свойств интеграла Лебега следует следующее свойство математического ожидания:

$$\xi \geq 0, \quad \mathbb{E}\xi = 0 \quad \implies \quad \xi = 0 \text{ почти наверное.}$$

Воспользуемся этим свойством. В неравенстве (29) достигается равенство  $\mathbb{E}(\frac{Y}{b} - \mathbf{I}_{Y^{-1}([b, +\infty))}) = 0 \iff \{\frac{Y}{b} - \mathbf{I}_{Y^{-1}([b, +\infty))} \geq 0\} \iff \frac{Y}{b} = \mathbf{I}_{Y^{-1}([b, +\infty))}$  почти наверное  $\iff \mathbb{P}(\frac{Y}{b} = \mathbf{I}_{Y^{-1}([b, +\infty))}) = 1 \iff \mathbb{P}(Y \in \{0, b\}) = 1$ .

Теорема доказана.

Заметим, что усилить неравенство, не накладывая дополнительных ограничений, не удастся.

**Следствие 1** (неравенство Маркова). *Для любой случайной величины  $X$  справедливо*

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}|X|^\kappa}{a^\kappa}, \quad \forall a > 0, \quad \forall \kappa \geq 0. \quad (30)$$

Доказательство. В неравенстве (29) в качестве  $Y$  возьмем  $|X|^\kappa \geq 0$ , а в качестве  $b$  —  $a^\kappa > 0$ :

$$\mathbb{P}(|X|^\kappa \geq a^\kappa) \leq \frac{\mathbb{E}|X|^\kappa}{a^\kappa},$$

откуда и получаем искомое неравенство.

Из неравенства Маркова получаем, что если  $\mathbb{E}|X|^\kappa < \infty$ , то хвосты распределения случайной величины  $X$  убывают со скоростью порядка  $\frac{C}{a^\kappa}$ .

**Следствие 2** (неравенство Чебышёва). *Для любой случайной величины  $X$  справедливо<sup>74</sup>*

$$\mathbb{P}(|X - \mathbb{E}X| \geq a) \leq \frac{\mathbb{V}ar X}{a^2}, \quad \forall a > 0. \quad (31)$$

Доказательство. В неравенстве (29) в качестве  $Y$  возьмем  $|X - \mathbb{E}X|^2 \geq 0$ , а в качестве  $b$  —  $a^2 > 0$ :

$$\mathbb{P}(|X - \mathbb{E}X|^2 \geq a^2) \leq \frac{\mathbb{E}|X - \mathbb{E}X|^2}{a^2},$$

откуда, учитывая выражение для дисперсии  $\mathbb{V}ar X = \mathbb{E}(X - \mathbb{E}X)^2$ , и получаем искомое неравенство.

Заметим, что равенство в (31) достигается  $\iff$  когда  $\mathbb{P}(|X - \mathbb{E}X|^2 \in \{0, a^2\}) = 1 \iff$  носитель распределения  $X$  сосредоточен не более, чем в трех точках  $\{c, c + a, c - a\}$  ( $\forall c$ ), причем  $\mathbb{P}(X = c + a) = \mathbb{P}(X = c - a)$ .

Неравенство Чебышёва достаточно грубое. Для его улучшения надо сужать класс распределений (например, для одновершинных распределений<sup>75</sup> неравенство Чебышёва можно уточнить).

**Следствие 3.** *Для любой случайной величины  $X$  справедливо*

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}e^{\lambda X}}{e^{\lambda a}}, \quad \forall \lambda > 0, \quad (32)$$

причем если  $\mathbb{E}e^{\lambda X} < \infty$ , то хвосты распределения случайной величины  $X$  убывают со скоростью порядка  $\frac{C}{e^{\lambda a}}$  (экспоненциальная скорость убывания).

<sup>74</sup>Международное обозначение дисперсии  $\mathbb{V}ar$  (или  $\mathbb{V}$ ), а не  $\mathbb{D}$ , от англ. variance (variation — это вариация, не путать!).

<sup>75</sup>плотность распределения имеет один локальный максимум, например, плотность нормального распределения

Для доказательства достаточно в неравенстве (29) в качестве  $Y$  взять  $e^{\lambda X} \geq 0$ , а в качестве  $b - e^{\lambda a} > 0$ .

Примером распределения, для которого имеет место быть экспоненциальная скорость убывания хвостов, может служить гауссовское распределение ( $\exists \lambda > 0: \mathbb{E}e^{\lambda X} < \infty$ ).

Теперь преступим к рассмотрению второго из двух основных неравенств. Но для начала напомним понятие *выпуклой функции*. Функция  $f: D \rightarrow \mathbb{R}$ ,  $D \subseteq \mathbb{R}^n$  — выпуклое множество, называется выпуклой, если ее надграфик  $epif = \{(x, \alpha) : \alpha \geq f(x)\}$  является выпуклым множеством. Другое определение: функция  $f: D \rightarrow \mathbb{R}$ ,  $D \subseteq \mathbb{R}^n$  — выпуклое множество, называется выпуклой, если

$$f(p_1x_1 + p_2x_2) \leq p_1f(x_1) + p_2f(x_2), \quad \forall p_1, p_2 \geq 0, \quad p_1 + p_2 = 1, \quad \forall x_1, x_2 \in D. \quad (33)$$

По индукции можно показать, что  $f: D \rightarrow \mathbb{R}$ ,  $D \subseteq \mathbb{R}^n$  — выпуклое множество, является выпуклой функцией  $\iff$

$$f\left(\sum_{i=1}^n p_i x_i\right) \leq \sum_{i=1}^n p_i f(x_i), \quad \forall p_i \geq 0, \quad \sum_{i=1}^n p_i = 1, \quad \forall x_i \in D. \quad (34)$$

Функция  $f: D \rightarrow \mathbb{R}$ ,  $D \subseteq \mathbb{R}^n$  — выпуклое множество, называется строго выпуклой, если

$$f(p_1x_1 + p_2x_2) < p_1f(x_1) + p_2f(x_2), \quad \forall p_1, p_2 > 0, \quad p_1 + p_2 = 1, \quad \forall x_1, x_2 \in D. \quad (35)$$

Для гладкой функции справедлив следующий критерий выпуклости: функция  $f: D \rightarrow \mathbb{R}$ ,  $D \subseteq \mathbb{R}^n$  — выпуклое множество, является выпуклой, если ее график лежит выше любой своей касательной гиперплоскости.

Однако касательная гиперплоскость существует не всегда. В случае отсутствия гладкости вводят понятие опорной гиперплоскости, обобщающее понятие касательной гиперплоскости (см. рисинки ( )). Обобщением градиента в этом случае будет субдифференциал — конус касательных направлений. Определим субдифференциал для выпуклой функции формально.

**Определение 59.** Пусть функция  $f: D \rightarrow \mathbb{R}$ ,  $D \subseteq \mathbb{R}^n$  — выпуклое множество, выпукла. Тогда вектор  $d$  называется субградиентом функции  $f$  в точке  $x^* \in D$ , если выполнено субградиентное неравенство:

$$f(x) \geq f(x^*) + \langle d, x - x^* \rangle, \quad \forall x \in D. \quad (36)$$

**Определение 60.** Пусть функция  $f: D \rightarrow \mathbb{R}$ ,  $D \subseteq \mathbb{R}^n$  — выпуклое множество, выпукла. Тогда множество всех ее субградиентов в точке  $x^* \in D$  называется ее субдифференциалом  $\partial f(x^*)$ .

Говорят, что функция выпукла, если в любой точке ее выпуклой области определения существует опорная гиперплоскость<sup>76</sup>.

(33) и (34) — это частные случаи неравенства Йенсона.

**Теорема 10** (неравенство Йенсена). Для любой выпуклой функции  $f: D \rightarrow \mathbb{R}$ ,  $D \subseteq \mathbb{R}^n$  — выпуклое замкнутое множество, и любого случайного вектора  $X$ , у которого конечно математическое ожидание  $\mathbb{E}|X| < \infty$  и для которого выполнено  $X(\omega) \in D$  почти наверное, справедливо следующее неравенство<sup>77</sup>:

$$f(\mathbb{E}X) \leq \mathbb{E}f(X). \quad (37)$$

<sup>76</sup>подразумевается, что график функции лежит выше любой своей опорной гиперплоскости

<sup>77</sup>математическое ожидание от случайного вектора  $X = (X_1, \dots, X_n)'$  определяется покомпонентно:  $\mathbb{E}X = (\mathbb{E}X_1, \dots, \mathbb{E}X_n)'$

Доказательство. Заметим, что получить неравенство Йенсена предельным переходом из (34) достаточно проблематично. Поэтому используем другой подход.

Во-первых, в силу  $\mathbb{E}|X| < \infty$  получаем, что  $|\mathbb{E}X| \stackrel{\text{def}}{=} |\mathbb{E}X^+ - \mathbb{E}X^-| < +\infty$ .

Во-вторых, из того, что  $D$  — выпукло и замкнуто, а  $X(\omega) \in D$  почти наверное, следует, что  $\mathbb{E}X \in D$ <sup>78</sup>.

Так как по условию функция  $f$  выпукла, то можем записать субградиентное неравенство (36) в любой точке  $x^* \in D$ :

$$\exists d \in \mathbb{R}^n : f(x) \geq f(x^*) + \langle d, x - x^* \rangle, \quad \forall x \in D.$$

Так как  $X(\omega) \stackrel{\text{п.н.}}{\in} D$ ,  $\mathbb{E}X \in D$ , то положим  $x^* = \mathbb{E}X$ ,  $x = X$ :

$$f(X) \geq f(\mathbb{E}X) + \langle d, X - \mathbb{E}X \rangle.$$

Справа и слева стоят случайные величины, причем

$$\begin{aligned} -[f(X)]^- &\geq -[f(\mathbb{E}X) + \langle d, X - \mathbb{E}X \rangle]^- \implies \\ [f(X)]^- &\leq [f(\mathbb{E}X) + \langle d, X - \mathbb{E}X \rangle]^- \implies \\ \mathbb{E}[f(X)]^- &\leq \mathbb{E}[f(\mathbb{E}X) + \langle d, X - \mathbb{E}X \rangle]^- \leq |f(\mathbb{E}X)| + \mathbb{E} \sum_{i=1}^n |d_i(X_i - \mathbb{E}X_i)| \leq \\ &\leq |f(\mathbb{E}X)| + \sum_{i=1}^n |d_i| \mathbb{E}|X_i - \mathbb{E}X_i| \leq |f(\mathbb{E}X)| + \sum_{i=1}^n |d_i| (\mathbb{E}|X_i| + |\mathbb{E}X_i|) \leq \{ \mathbb{E}|X| < \infty \} < +\infty \implies \\ &\exists \mathbb{E}(f(X)) \quad (\text{возможно } \mathbb{E}(f(X)) = +\infty). \end{aligned}$$

Значит, имеем право написать

$$\mathbb{E}(f(X)) \geq \mathbb{E}(f(\mathbb{E}X) + \langle d, X - \mathbb{E}X \rangle) = f(\mathbb{E}X).$$

Что и т.д.

**Замечание.** Когда достигается равенство в неравенстве Йенсена? Ясно, что если распределение  $X$  сосредоточено в одной точке (одноточечное распределение Дирака), то  $X \stackrel{\text{п.н.}}{=} \mathbb{E}X$  и неравенство Йенсена превращается в равенство. Предположим, что функция  $f$  строго выпукла. Тогда в субградиентном неравенстве (36) достигается равенство  $\iff x = x^*$ . Таким образом, если функция  $f$  строго выпукла, то равенство в неравенстве Йенсена достигается  $\iff X \stackrel{\text{п.н.}}{=} \mathbb{E}X$ , т.е. распределение  $X$  сосредоточено в одной точке.

<sup>78</sup>Схема доказательства аналогична схеме построения интеграла Лебега. Для простоты будем считать, что  $X$  скалярно.

Пусть  $X = \mathbf{I}_A$ , тогда, в силу того, что  $X(\omega) \stackrel{\text{п.н.}}{\in} D$  и  $D$  — выпукло,  $[0, 1] \in D$ , а значит  $\mathbb{E}X = \mathbb{P}(A) \in D$ .

Пусть  $X = \sum_{i=1}^n a_i \mathbf{I}_{A_i}$ ,  $a_i \geq 0$ ,  $A_i \cap A_j = \emptyset$ ,  $\sum_{i=1}^n A_i = \Omega$ . Из того, что  $X(\omega) \stackrel{\text{п.н.}}{\in} D$ , получаем  $a_i \in D$ . Из выпуклости  $D$  следует, что  $\mathbb{E}X = \sum_{i=1}^n a_i \mathbb{P}(A_i) \in D$ , т.к.  $\sum_{i=1}^n \mathbb{P}(A_i) = 1$ . Заметим, что условие  $\sum_{i=1}^n A_i = \Omega$  можно опустить.

Пусть  $X = \sum_{i=1}^{\infty} a_i \mathbf{I}_{A_i}$ ,  $a_i \geq 0$ ,  $A_i \cap A_j = \emptyset$ ,  $\sum_{i=1}^{\infty} A_i = \Omega$ . Так как  $\exists X_n = \sum_{i=1}^n a_i \mathbf{I}_{A_i} \stackrel{\text{п.н.}}{\in} D$ :  $X_n \nearrow X$ , и в силу непрерывности мат. ожидания  $\mathbb{E}X_n \nearrow \mathbb{E}X$ , то, используя уже доказанный факт  $\mathbb{E}X_n \in D$ , получаем из замкнутости  $D$ , что  $\mathbb{E}X \in D$ .

Пусть  $X \geq 0$ . Тогда  $X$  можно приблизить неотрицательной дискретной случайной величиной (такой, которую мы только что рассмотрели), которая также почти наверное принадлежит  $D$ . Проводя аналогичные рассуждения, получим, что  $\mathbb{E}X \in D$ .

Если же  $X$  — знакопеременная случайная величина, то, представляя  $X$  в виде  $X = X^+ - X^-$ , получим  $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$ . Заметим, что если  $X$  — не знакостоянна, то  $0 \in D$  в силу выпуклости  $D$ , а значит  $X^+(\omega) \stackrel{\text{п.н.}}{\in} D$ ,  $-X^-(\omega) \stackrel{\text{п.н.}}{\in} D$ . Из предыдущих рассуждений  $\mathbb{E}X^+ \in D$ ,  $-\mathbb{E}X^- \in D$ . Ясно, что  $-\mathbb{E}X^- \leq \mathbb{E}X \leq \mathbb{E}X^+$ , а значит, в силу выпуклости  $D$ ,  $\mathbb{E}X \in D$ .

**Пример.** Поезд ходит каждый день по 100 км (по одной и той же дороге). Средняя скорость поезда равна 50 км/ч:

$$\mathbb{E}V = 50 \text{ км/ч}, \quad S = 100 \text{ км},$$

где  $V$  — случайная величина, равная средней скорости прохождения пути в данный день. Среднее время нахождения в пути:

$$\mathbb{E}T = \mathbb{E}\frac{S}{V} = S \cdot \mathbb{E}\frac{1}{V}.$$

Функция  $f(x) = \frac{1}{x}$  строго выпукла, значит, по неравенству Йенсена, получим:

$$\mathbb{E}\frac{1}{V} \geq \frac{1}{\mathbb{E}V},$$

причем равенство достигается  $\iff V \stackrel{\text{п.н.}}{=} \mathbb{E}V$ . Следовательно,

$$\mathbb{E}T \geq \frac{S}{\mathbb{E}V} = 2 \text{ часа},$$

причем  $\mathbb{E}T = 2 \text{ часа} \iff$  когда поезд ходит точно по расписанию. Заметим, что, вообще говоря, может быть  $\mathbb{E}T = +\infty$ .

**Пример.** Пусть  $\mathbb{E}|X| < \infty$ . Тогда дисперсия

$$\text{Var}X \stackrel{\text{def}}{=} \mathbb{E}(X - \mathbb{E}X)^2$$

определена. Если  $\mathbb{E}|X| < \infty$ , то дисперсия определена, хотя может быть и  $\text{Var}X = +\infty$ . Так как  $(X - \mathbb{E}X)^2 \geq 0$ , то  $\text{Var}X \geq 0$ . С другой стороны

$$\text{Var}X = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2) = \mathbb{E}X^2 - (\mathbb{E}X)^2.$$

Значит,

$$\mathbb{E}X^2 \geq (\mathbb{E}X)^2,$$

причем равенство достигается  $\iff$  когда  $\text{Var}X = 0 \iff X \stackrel{\text{п.н.}}{=} \mathbb{E}X$ . Получили частный случай неравенства Йенсена при  $f(x) = x^2$ .

## 6 Субъективная интерпретация среднего. Формулировка закона больших чисел и объективная интерпретация среднего. Отличие интеграла Лебега от интеграла Римана. Метод Монте-Карло как численный метод нахождения интеграла Лебега.

### 6.1 Субъективная интерпретация среднего.

В предыдущей главе была рассмотрена универсальная модель вида  $(\Omega, \mathcal{F}, \mathbb{E})$ . Интерпретация  $\Omega, \mathcal{F}$  — та же, что и для вероятностного пространства (см. первую главу). Дадим интерпретацию среднего  $\mathbb{E}$ .

Сначала рассмотрим субъективную интерпретацию среднего. Покажем, что  $\mathbb{E}X$  как объект, описанный в предыдущей главе, — адекватный аналог субъективного среднего. Для этого рассмотрим несколько примеров.



Рассмотрим семью из двух человек. Муж получает в среднем  $X_1$  \$ в месяц, а жена —  $X_2$  \$ в месяц. Интуиция: в среднем семья получает в месяц  $(X_1 + X_2)$  \$. Таким образом, мы интуитивно применили свойство аддитивности математического ожидания:  $\mathbb{E}X_1 + \mathbb{E}X_2 = \mathbb{E}(X_1 + X_2)$ . Заметим, что при этом  $X_1$  и  $X_2$  могут быть зависимыми (например, если жена находится в подчинении у мужа).

Теперь пусть муж получает в среднем  $X$  \$ в месяц, а жена получает всегда в два раза больше, чем муж. Интуиция: в среднем жена получает в месяц  $2X$  \$. Таким образом, мы интуитивно применили свойство положительной однородности математического ожидания:  $\mathbb{E}(X) = \mathbb{E}X$ ,  $a > 0$ .

Пусть человек стабильно получаем одинаковую зарплату  $X$  \$. Интуиция: средняя зарплата равна  $X$  \$. Здесь неосознанно применили свойства нормированности  $\mathbb{E}\mathbf{1}_\Omega = 1$  и однородности математического ожидания.

Также естественно на интуитивном уровне свойство неотрицательности: если  $X \geq 0$ , то  $\mathbb{E}X \geq 0$ .

А вот непрерывность

$$0 \leq X_n \nearrow X \implies \mathbb{E}X_n \nearrow \mathbb{E}X$$

не вытекает из интуиции, однако это достаточно разумное, логичное свойство, оно не вызывает раздражения.

Итак, мы показали логичность всех 5 аксиом, определяющих среднее.

Заметим, что вероятность не имеет такой хорошей субъективной интерпретации; она интуитивно плохо воспринимается и оценивается, в отличие от среднего.

## 6.2 Формулировка закона больших чисел и объективная интерпретация среднего.

Объективная интерпретация вероятности давалась через частотную форму ЗБЧ. Для объективной интерпретации среднего нам также понадобится ЗБЧ, однако уже не в частотной форме. Как уже говорилось, ЗБЧ — это закон природы, это фундамент, на котором покоится объективная интерпретация теории вероятностей. Теорем о ЗБЧ много, но их смысл примерно похожий. В общем виде ЗБЧ очень тяжело сформулировать, поэтому мы рассматриваем частные случаи.

Рассмотрим эксперимент, состоящий из серии однородных испытаний.

Предположения:

1. Отсутствие взаимного влияния в испытаниях (одно на другое не оказывает влияния).
2. Воспроизводимость. Однородные испытания проводятся в сходных, аналогичных условиях (не одинаковых, иначе результат был бы один и тот же).

Однако теперь, в отличие от схемы Бернулли, где испытания характеризовались бинарным признаком (успех/неуспех), теперь мы будем наблюдать некоторую *числовую характеристику* испытания, имеющую одинаковый смысл во всех испытаниях. Заметим, что схема Бернулли является частным случаем описанной общей схемы: за числовую характеристику можно принять величину  $p = \begin{cases} 1, & \text{если произошел успех,} \\ 0, & \text{если произошел неуспех.} \end{cases}$

В качестве примера можно привести стендовые испытания лампочек. Числовая характеристика в общей схеме — максимальное время работы лампочек до перегорания; признак в схеме Бернулли — перегорит или не перегорит).

Примем дополнительное упрощающее предположение о *неотрицательности числовой характеристики* (без этого предположения можно обойтись, но это гораздо сложнее).

При проведении ряда испытаний можно посчитать *среднее арифметическое* значение характеристики (в теории вероятностей его называют выборочным средним, или эмпирическим средним).

**ЗБЧ (для средних)** гласит: при неограниченном увеличении числа испытаний *среднее арифметическое значение числовой характеристики приближается к некоторому числу (неотрицательному), которое мы интерпретируем как **среднее** значение рассматриваемой числовой характеристики в отдельном испытании.*

Замечания.

1. Мы не исключаем случая, когда полученное среднее значение равно  $+\infty$ .
2. Среднее значение числовой характеристики всегда определено. Для этого и вводилось дополнительное предположение о неотрицательности. Если отказаться от данного предположения, то возникают 4 случая (см. 5 главу), в последнем из которых математическое ожидание не существует; следовательно, пришлось бы формулировать ЗБЧ так, чтобы охватить и этот случай несуществования среднего.
3. ЗБЧ для частот — частный случай приведенного ЗБЧ для средних. Как уже говорилось, в роли неотрицательной числовой характеристики можно взять

$$p = \begin{cases} 1, & \text{если произошел успех,} \\ 0, & \text{если произошел неуспех.} \end{cases}$$

Тогда ее среднее арифметическое значение будет равно частоте успеха в эксперименте, а среднее значение при неограниченном увеличении числа испытаний — вероятности появления признака в одном отдельном испытании.

Так же, как мы в свое время формализовали схему Бернулли (см. вторую главу), формализуем и эту общую схему как модель, состоящую из последовательности независимых одинаково распределенных случайных величин  $X_1, \dots, X_n$ . По ЗБЧ выборочное среднее  $\frac{1}{n} \sum_{i=1}^n X_i$  приближается при  $n \rightarrow \infty$  к некоторому числу  $a \geq 0$  (возможно  $a = +\infty$ ).

## 6.3 Отличие интеграла Лебега от интеграла Римана.

### 6.3.1 Интеграл Римана.

Область интегрирования  $D \subset \mathbb{R}^n$ :

- $D$  — ограниченная область;
- $D$  измерима по Жордану.

**Определение 61.** Множество  $D$  называется измеримым по Жордану (квадрируемым множеством), если его граница является нуль-множеством, т.е. содержится во множестве нулевой меры (имеется в виду мера Лебега)<sup>79</sup>:  $\exists A: \partial D \in A, \mu_L(A) = 0$ .

Из измеримости по Жордану следует измеримость по Лебегу.

<sup>79</sup>В силу полноты меры Лебега можно утверждать, что  $\partial D$  измеримо и  $\mu_L(\partial D) = 0$  (примером меры, не являющейся полной, может служить борелевская мера, т.к. канторово множество не является борелевским)

Из функционального анализа известно, что

$$\mu_L(A) = \inf_{A \subset \bigcup_{i=1}^{\infty} B_i} \sum_{i=1}^{\infty} m(B_i), \quad \mu_L(A) + \mu_L(X \setminus A) = m(X),$$

$$\mu_J(A) = \inf_{A \subset \bigcup_{i=1}^n B_i} \sum_{i=1}^n m(B_i), \quad \mu_J(A) + \mu_J(X \setminus A) = m(X),$$

где  $\mu_L$  — мера Лебега,  $\mu_J$  — мера Жордана, множества  $B_i \in K$  ( $K$  — алгебра элементарных множеств из  $X$ ),  $m$  — счетно-аддитивная мера на  $K$ . Мы рассматриваем случай, когда  $X = \mathbb{R}^n$ , элементарные множества — это  $n$ -мерные прямоугольники, а  $m(B_i)$  — площадь прямоугольника  $B_i$ .

Интегрируемая функция  $f: D \rightarrow \mathbb{R}$ :

- $f$  — ограниченная функция;
- множество точек разрыва функции  $f$  есть нуль-множество.

**Замечание 1.** От условия квадратуемости множества  $D$  можно избавиться, используя индикатор этого множества. Другими словами, положив  $f(x) = 0$   $x \notin D$ , получим, что условие квадратуемости есть частный случай второго требования для функции (граница множества  $D$  принадлежит множеству точек разрыва функции  $f$ , которое по условию является нуль-множеством).

Фактически, мы сформулировали критерий Лебега интегрируемости по Риману.

**Теорема 11** (критерий Лебега интегрируемости по Риману). *Для того чтобы числовая функция, заданная на ограниченном множестве<sup>80</sup>  $\mathbb{R}^n$ , была интегрируемой по Риману необходимо и достаточно, чтобы функция  $f$  была ограниченной и непрерывной почти всюду.*

**Замечание 2.** Немного обобщим. Вместо  $\mathbb{R}^n$  рассмотрим абстрактное пространство с топологией. Тогда условие ограниченности лучше заменить на предкомпактность (из любой последовательности можно выбрать фундаментальную, из предкомпактности следует ограниченность). Известно, что если рассматривать польское пространство  $P$  (полное, сепарабельное, метрическое) с биевской  $\sigma$ -алгеброй ( $\sigma$ -алгеброй, порожденной топологией, т.е. системой открытых множеств), то любая ограниченная функция  $f: P \rightarrow \mathbb{R}$ , множество точек разрыва которой есть нуль-множество, интегрируема на любом предкомпактном множестве, вне которого  $f$  равна нулю.

**Замечание 3.** Можно также на прямой брать за основу не меру Лебега, а меру Лебега-Стилтьеса (но этот случай, также как и другие всевозможные обобщения, мы рассматривать не будем).

Заметим, однако, что от требований на функцию (ограниченность и относительная гладкость) отказаться нельзя (это будет видно из построения интеграла Римана).

Построение интеграла Римана. Область, по которой ведется интегрирование, разбивается на маленькие (не по мере, а по расстоянию, т.е. по диаметру<sup>81</sup>) квадратуемые (граница нулевой меры) куски (например, на прямоугольники). В каждом кусочке вычисляем наибольшее и наименьшее значения интегрируемой функции, составляем верхнюю

<sup>80</sup>множество достаточно простое, как минимум квадратуемое; например в  $\mathbb{R}$  можно взять отрезок  $[a, b]$ ; как уже говорилось, в нужных местах функцию всегда можно положить равной нулю; интегрирование по всему пространству мы здесь просто не рассматриваем

<sup>81</sup>на прямой мера равна расстоянию, диаметр множества — максимальное расстояние между точками этого множества

и нижнюю интегральные суммы. Устремляя диаметр разбиения (максимальный диаметр по всем кусочкам) к нулю<sup>82</sup>, получим верхний и нижний интегралы Дарбу. Если полученные интегралы равны, то они равны интегралу Римана от рассматриваемой функции по области, а сама эта функция называется интегрируемой по Риману по заданной области.

Ограниченность функции и нулевая мера множеств точек разрыва нужны для того, чтобы верхняя и нижняя суммы сходились (значения функции в одном кусочке были близки).

### 6.3.2 Интеграл Лебега.

Область интегрирования  $D$ :

·  $D$  — измеримое множество.

Интегрируемая функция  $f: D \rightarrow \mathbb{R}$ :

·  $f$  — измеримая функция.

Построение интеграла Лебега. Разбивается область значений функции. Основная идея построения интеграла Лебега состоит в том, что здесь, в отличие от интеграла Римана, точки группируются не по признаку их близости на оси аргумента, а по признаку близости значений функции в этих точках. Это сразу же позволяет распространить понятие интеграла на весьма широкий класс функций.

**Теорема 12.** Если функция  $f(x)$ , заданная на  $[a, b]$ , интегрируема по Риману, то она интегрируема по Лебегу и

$$R \int_a^b f(x) dx = L \int_{[a,b]} f(x) dx.$$

В обратную сторону неверно.

**Пример.** 1. Функция Дирихле

$$f(x) = \begin{cases} 1, & \text{если } x \in Q \cap [0, 1], \\ 0, & \text{если } x \in [0, 1] \setminus Q, \end{cases}$$

где  $Q$  — множество рациональных чисел, интегрируема по Лебегу (интеграл равен нулю, т.к.  $\mu_L(Q) = 0$ ), но не интегрируема по Риману. Однако  $f$  можно изменить на множестве нулевой меры так, чтобы она стала интегрируема по Риману (при этом заметим, что множество точек разрыва у функции Дирихле — весь отрезок  $[0, 1]$ ).

2. “Классическое” канторово множество получается путем выкидывания из отрезка  $[0, 1]$  интервалов длины  $\frac{1}{3}$ . Его мера равна нулю ( $1 - \left(\frac{1}{3} + \frac{2}{3^2} + \frac{2^2}{3^3} + \dots\right) = 1 - \left(\sum_{i=1}^{\infty} \frac{2^{i-1}}{3^i}\right) = 0$ ). Рассмотрим канторово множество положительной меры, получающееся путем выкидывания из отрезка  $[0, 1]$  интервалов длины  $< \frac{1}{3}$ . Оно является совершенным<sup>83</sup> нигде не плотным<sup>84</sup> множеством. Индикатор этого множества не эквивалентен ни одной функции, интегрируемой по Риману (т.к. все точки канторова

<sup>82</sup>стремление диаметра к нулю означает стремление к нулю разности аргументов функции, а не их значений

<sup>83</sup>Совершенное множество — замкнутое множество, не имеющее изолированных точек, т.е. совпадающее со множеством всех своих предельных точек.

<sup>84</sup>Множество называется нигде не плотным, если его замыкание не содержит ни одного шара.

множества являются точками разрыва соответствующего индикатора в силу того, что каждая точка множества является граничной, а рассматриваемое множество имеет ненулевую меру), но интегрируем по Лебегу.

Получаем, что интеграл Лебега даже на прямой — более общее понятие, чем интеграл Римана<sup>85</sup>.

**Замечание.** Связь между измеримыми и непрерывными функциями показана в теореме Лузина.

**Теорема 13** (Лузин). *Для того чтобы функция  $f(x)$ , заданная на отрезке  $[a, b]$ , была измерима, необходимо и достаточно, чтобы для любого  $\varepsilon > 0$  существовала такая непрерывная на  $[a, b]$  функция  $\phi(x)$ , что*

$$\mu\{x: f(x) \neq \phi(x)\} < \varepsilon.$$

Иначе говоря, измеримая функция может быть сделана непрерывной на  $[a, b]$  путем ее изменения на множестве сколь угодно малой меры. Подчеркнем, что в теореме говорится о множестве положительной меры, а значит, класс измеримых функций шире класса функций с нуль-множеством точек разрыва.

**Квадратурные формулы.** Приближенным методом вычисления интеграла Римана являются квадратурные формулы. Суть их в следующем: если функция достаточно гладкая, то мы можем на малом отрезке заменить ее более простой. Чем более гладкая функция, тем сложнее заменяющая функция и тем точнее получаются формулы. Примером квадратурной формулы является метод Симпсона.

## 6.4 Метод Монте-Карло как численный метод нахождения интеграла Лебега.

**Метод Монте-Карло** является численным методом интегрирования *по Лебегу*. Этот метод был придуман Джоном фон Нейманом, к числу достижений которого также принадлежат аксиоматика квантовой механики, изобретение вычислительной техники, развитие теории игр, аксиоматика теории ожидаемой полезности, теория автоматов.

Метод Монте-Карло появился одновременно с первыми ЭВМ, и одной из первых программ была реализация этого метода. Идея метода возникла при решении задачи расчета критической массы, которая сводится к вычислению кратных интегралов высокой кратности. Для их вычисления невозможно применять квадратурные формулы по нескольким причинам, одна из которых носит название «проклятие размерности».

Действительно, пусть нам необходимо посчитать интеграл от функции, заданной в пространстве  $\mathbb{R}^k$ . Тогда для достижения требуемой точности  $\varepsilon$  необходимо посчитать значения функции в порядка  $\frac{1}{\varepsilon^k}$  точках. Таким образом, методы вычисления интеграла Римана являются очень чувствительными к размерности задачи и к гладкости подынтегральной функции.

Заметим, что метод Монте-Карло является *единственным* численным методом вычисления интеграла Лебега. Название метода происходит от названия города в княжестве Монако, широко известного своими многочисленными казино, поскольку именно рулетка является одним из самых широко известных генераторов случайных чисел.

Итак, пусть у нас имеется программный датчик *псевдослучайных* чисел (генерируется программой — поэтому «псевдо»). Генерируется последовательность  $\xi_1, \xi_2, \dots$  независимых одинаково распределенных случайных величин с распределением  $U[0, 1]$ . (Заметим,

<sup>85</sup>Если рассматривать интегрирование по неограниченному множеству, то возможен случай, когда интеграл Римана существует, а Лебега — нет. Например, интегрирование знакопеременной функции  $f(x) = \frac{\sin \frac{1}{x}}{x}$  на  $(0, 1]$  (попадаем на 4 случай, неопределенность вида  $(+\infty) - (+\infty)$ ).

что особенностью вычислительной техники является ограниченное число разрядов для представления вещественного числа, поэтому, вообще говоря, мы не можем получить *любое* число из  $[0, 1]$ .)

Теорема об изоморфизме гласит, что польское пространство  $\Omega$  с борелевской  $\sigma$ -алгеброй и неатомической мерой  $\mathbb{P}$  изоморфно вероятностному пространству  $([0, 1], \mathcal{B}_{[0,1]}, U[0, 1])$ , поэтому мы можем считать, что умеем строить последовательности случайных элементов из некоторого абстрактного пространства. Итак, имеем последовательность случайных элементов  $\xi_1, \xi_2, \dots$  с распределением  $Q$  и со значениями из измеримого пространства  $(E, \mathcal{E})$ .

Предположим, что необходимо посчитать интеграл

$$I = \int_E f(x)Q(dx),$$

где  $f(x)$  — числовая измеримая функция,  $f(x) \geq 0$ . В качестве приближенного значения интеграла будем брать

$$\tilde{I} = \frac{1}{n} \sum_{k=1}^n f(\xi_k).$$

По закону больших чисел значение  $\tilde{I}$  стремится к  $\mathbb{E}f(\xi)$ , которое по определению равно

$$\mathbb{E}f(\xi) = \int_{\Omega} f(\xi(\omega))\mathbb{P}(d\omega),$$

а по теореме о замене переменных под знаком интеграла

$$\int_{\Omega} f(\xi(\omega))\mathbb{P}(d\omega) = \int_E f(x)Q(dx).$$

Метод Монте-Карло является универсальным методом, поэтому обладает низкой скоростью сходимости, однако важно то, что эту скорость сходимости можно оценить легче и точнее, чем в квадратурных формулах. Например, если  $\mathbb{E}f^2(\xi) < \infty$  (это довольно слабое требование, оно обычно выполняется на практике), то скорость сходимости имеет порядок  $\frac{\sigma(f(\xi))}{\sqrt{n}}$ . Докажем это, воспользовавшись центральной предельной теоремой.

Пусть  $\varepsilon$  — точность вычисления интеграла,  $\gamma$  — коэффициент доверия:

$$\mathbb{P}(|\tilde{I} - I| \leq \varepsilon) = 1 - \gamma.$$

Оценим точность вычислений при заданном числе генераций  $n$  и коэффициенте доверия  $\gamma$ . Обозначим  $\text{Var}f(\xi) = \sigma^2$ , тогда  $\text{Var}\tilde{I} = \frac{\sigma^2}{n}$ . Согласно центральной предельной теореме имеем

$$\mathbb{P}(\tilde{I} - I \leq \varepsilon) = \mathbb{P}\left(|\tilde{I} - I| \frac{\sqrt{n}}{\sigma} \leq \varepsilon \frac{\sqrt{n}}{\sigma}\right) \xrightarrow{n \rightarrow \infty} 2\Phi\left(\varepsilon \frac{\sqrt{n}}{\sigma}\right) - 1.$$

Если взять в качестве оценки дисперсии  $\text{Var}f(\xi)$  величину  $S$ :

$$S = \frac{1}{n-1} \sum_{i=1}^n (f(\xi_i) - \tilde{I})^2,$$



а в качестве оценки среднего квадратического  $\bar{\sigma} = \sqrt{S}$ , то оценку для точности можно найти из равенства

$$\Phi\left(\varepsilon \frac{\sqrt{n}}{\bar{\sigma}}\right) = 1 - \frac{\gamma}{2} \quad \text{или} \quad \varepsilon = \frac{\bar{\sigma}}{\sqrt{n}} p_{1-\frac{\gamma}{2}},$$

где  $p_{1-\frac{\gamma}{2}}$  — квантиль стандартного нормального распределения порядка  $1 - \frac{\gamma}{2}$ . Что и требовалось показать.

Итак, серьезно о методе Монте–Карло имеет смысл говорить, если  $n$  имеет порядок 10000 ( $n$  — число испытаний).

В заключение заметим, что иногда можно увеличить скорость сходимости метода Монте–Карло, несколько изменив функцию  $f(x)$  или наложив на нее некоторые условия, так чтобы уменьшить дисперсию. Поиск и применение таких приемов делают метод Монте–Карло искусством.

## 7 Ковариация, дисперсия и их свойства. Ковариационные матрицы. Линейные модели. Задача регрессии. Свойства многомерного нормального распределения. Моделирование гауссовского случайного вектора.

### 7.1 Ковариация, дисперсия и их свойства. Ковариационные матрицы.

Прежде, чем приступить к изложению основного материала, напомним *неравенство Коши–Буняковского–Шварца*<sup>86</sup>, которым мы в дальнейшем не раз будем пользоваться:

$$|\langle x, y \rangle| \leq \|x\| \cdot \|y\|, \quad (38)$$

где  $\|x\| = \sqrt{\langle x, x \rangle}$  — евклидова норма.

Пусть дано вероятностное пространство  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Определение 62.** *Ковариацией* двух случайных величин  $X$  и  $Y$  со значениями в  $\mathbb{R}$  называется среднее значение произведения соответствующих центрированных<sup>87</sup> случайных величин:

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)].$$

О ковариации имеет смысл говорить только тогда, когда  $\mathbb{E}|X| < \infty$ ,  $\mathbb{E}|Y| < \infty$ . Однако при этом не факт, что она определена.

Достаточное условие того, что ковариация определена:  $\mathbb{E}X^2 < \infty$ ,  $\mathbb{E}Y^2 < \infty$ . Во-первых, по неравенству Йенсона  $\mathbb{E}X^2 \geq (\mathbb{E}X)^2$ , а значит, если второй момент существует, то существует и первый, причем из  $\mathbb{E}X^2 < \infty$  следует, что  $\mathbb{E}|X| < \infty$ . Во-вторых, по определению ковариации  $|\text{cov}(X, Y)| = |\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]| = |\mathbb{E}XY - \mathbb{E}X\mathbb{E}Y| \leq \{\text{нер-во Буняковского (38)}\} \leq \sqrt{\mathbb{E}X^2} \sqrt{\mathbb{E}Y^2} + \mathbb{E}|X| \mathbb{E}|Y|$ . Следовательно, если  $\mathbb{E}X^2 < \infty$ ,  $\mathbb{E}Y^2 < \infty$ , то  $|\text{cov}(X, Y)| < \infty$ .

<sup>86</sup>Коши сформулировал это неравенство для сумм, Буняковский — для интегралов, а Шварц — в общем случае (для евклидовых пространств).

<sup>87</sup> $\xi \mapsto (\xi - \mathbb{E}\xi)$  — операция центрирования ( $\mathbb{E}(\xi - \mathbb{E}\xi) = 0$ )



**Замечание.** Ковариация может быть определена и для комплекснозначных случайных величин<sup>88</sup> (случайных элементов со значениями на комплексной плоскости):

$$\text{cov}(X, Y) = \mathbb{E} \left[ (X - \mathbb{E}X) \overline{(Y - \mathbb{E}Y)} \right].$$

Однако этот случай мы в дальнейшем рассматривать не будем.

### Свойства ковариации.

1. Билинейность<sup>89</sup>:

$$\begin{aligned} \text{cov}(a_1 X_1 + a_2 X_2, Y) &= a_1 \text{cov}(X_1, Y) + a_2 \text{cov}(X_2, Y), \\ \text{cov}(X, b_1 Y_1 + b_2 Y_2) &= b_1 \text{cov}(X, Y_1) + b_2 \text{cov}(X, Y_2), \end{aligned}$$

где маленькими буквами  $(a_i, b_i)$  обозначены неслучайные величины.

Это свойство следует из свойств однородности и аддитивности математического ожидания.

2. Симметричность:

$$\text{cov}(X, Y) = \text{cov}(Y, X).$$

3.

$$\text{cov}(X, X) \stackrel{\text{def}}{=} \text{Var}(X) \geq 0.$$

Получаем, что ковариация  $\text{cov}(X, Y)$  — билинейная, симметричная, неотрицательная квадратичная форма.

Рассмотрим пространство случайных величин с конечными вторыми моментами:

$$L_2(\mathbb{P}) = \{ \xi - \text{сл.в.} : \mathbb{E}\xi^2 < \infty \}.$$

Очевидно, что  $L_2(\mathbb{P})$  — линейное пространство<sup>90</sup>.

Положим:

$$\langle \xi, \eta \rangle = \mathbb{E}\xi\eta = \int \xi(\omega)\eta(\omega)\mathbb{P}(d\omega). \quad (39)$$

Это “почти” скалярное произведение, т.к. из  $\langle \xi, \xi \rangle = 0$ , вообще говоря, не следует, что  $\xi(\omega) = 0 \ \forall \omega \in \Omega$ . Соответственно получаем, что

$$\|\xi\|_2 = \sqrt{\langle \xi, \xi \rangle} \quad (40)$$

— это не норма, а полунорма (свойство  $\|\xi\|_2 = 0 \iff \xi = 0$  не выполнено). Заметим, однако, что  $\|\xi\|_2 = 0 \implies \xi \stackrel{\text{п.в.}}{=} 0$ .

<sup>88</sup> комплекснозначные случайные величины используются в статистической радиомеханике при описании колебаний

<sup>89</sup> в случае комплекснозначных случайных величин — полуторалинейность

<sup>90</sup> Действительно,  $\xi \in \mathcal{L}_2(\mathbb{P}), \eta \in \mathcal{L}_2(\mathbb{P}) \implies \mathbb{E}(a\xi + b\eta)^2 = a^2\mathbb{E}\xi^2 + 2ab\mathbb{E}\xi\eta + b^2\mathbb{E}\eta^2 \leq \{\text{нр-во Буняковского}\} \leq a^2\mathbb{E}\xi^2 + 2ab(\sqrt{\mathbb{E}\xi^2}\sqrt{\mathbb{E}\eta^2}) + b^2\mathbb{E}\eta^2 < \infty$ .

“Подправим” пространство  $L_2(\mathbb{P})$  так, чтобы для введенных операций  $\langle \xi, \eta \rangle$  по формуле (39) и  $\|\xi\|_2$  по формуле (40) выполнялись все аксиомы скалярного произведения и все аксиомы нормы соответственно.

Легко видеть, что отношение  $\xi \sim \eta \iff \xi \stackrel{\text{п.в.}}{=} \eta$  есть отношение эквивалентности<sup>91</sup>.

Рассмотрим пространство классов эквивалентности случайных величин с конечными вторыми моментами:

$$\mathcal{L}_2(\mathbb{P}) = \left\{ \xi - \text{класс эквивалентности сл.в. : } \xi \sim \eta \iff \xi \stackrel{\text{п.в.}}{=} \eta; \xi \in L_2(\mathbb{P}) \right\}.$$

Под операцией сложения классов подразумевается сложение представителей соответствующих классов, а под операцией умножения класса на число — умножение на число представителя этого класса. Очевидно, что эти операции корректны и  $\mathcal{L}_2(\mathbb{P})$  является линейным пространством. Формулы (39) и (40) задают скалярное произведение и норму в пространстве классов эквивалентности. Из функционального анализа известно, что описанное пространство с так введенной нормой полно, а значит,  $\mathcal{L}_2(\mathbb{P})$  — гильбертово пространство.

Обозначим через

$$\mathbb{M}X = \mathbf{I}_\Omega \cdot \mathbb{E}X.$$

Ясно, что  $\forall X \in \mathcal{L}_2(\mathbb{P})$  ( $\forall X \in L_2(\mathbb{P})$ )  $\mathbb{M}X \in \mathcal{L}_2(\mathbb{P})$  ( $\mathbb{M}X \in L_2(\mathbb{P})$ ), причем  $\mathbb{M}X$  — проектор на одномерное подпространство константных случайных величин.

Применяя данное обозначение, получим:

$$\text{cov}(X, Y) = \langle X - \mathbb{M}X, Y - \mathbb{M}Y \rangle, \quad \forall X, Y \in \mathcal{L}_2(\mathbb{P}).$$

Заметим, что мы берем разность элементов пространства (поэтому-то нам и понадобилось дополнительно вводить обозначение).

4.

$$\text{Var}X = 0 \iff X \stackrel{\text{п.в.}}{=} \mathbb{M}X \iff X \stackrel{\text{п.в.}}{=} \text{Const} \cdot \mathbf{I}_\Omega \quad (\mathbb{P}(X = \mathbb{E}X) = \mathbb{P}(X = \text{Const}) = 1),$$

т.е.  $X$  имеет вырожденное распределение.

Заметим, что данный факт был получен нами еще в 5 главе как следствие неравенства Йенсона.

5.  $\text{Var}X$  — характеристика распределения  $X$  (т.е. зная распределение  $\mathbb{P}_X$ , можем посчитать дисперсию как математическое ожидание от известной функции от  $X$ , используя теорему о замене переменных, при этом полностью зная  $\mathbb{P}$  на всем пространстве не нужно);

$\text{cov}(X, Y)$  — характеристика совместного распределения  $\mathbb{P}_{XY}$ .

6. Если  $\mathbb{E}|X| < \infty$ ,  $\mathbb{E}|Y| < \infty$ ,  $X$  и  $Y$  независимы, то  $\text{cov}(X, Y) = 0$ .

Доказательство. По определению ковариации получаем, что

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y.$$

Покажем, что если  $\mathbb{E}|X| < \infty$ ,  $\mathbb{E}|Y| < \infty$ ,  $X$  и  $Y$  независимы, то  $\mathbb{E}XY < \infty$  и  $\mathbb{E}XY = \mathbb{E}X \cdot \mathbb{E}Y$ .

<sup>91</sup>Рефлексивность:  $\xi \sim \xi$ ; симметричность:  $\xi \sim \eta \implies \eta \sim \xi$ ; транзитивность:  $\xi \sim \eta, \eta \sim \nu \implies \xi \sim \nu$ .

По определению независимости случайных величин (см. 11 главу):

$$\mathbb{P}_{XY}(A, B) = \mathbb{P}_X(A)\mathbb{P}_Y(B) \quad \forall A \in \mathcal{A}, \forall B \in \mathcal{B}.$$

Воспользуемся теоремой о замене переменных в интеграле Лебега.

$$\mathbb{E}XY = \int xy\mathbb{P}_{XY}(dx, dy) = \int xy\mathbb{P}_X(dx)\mathbb{P}_Y(dy) = \int x\mathbb{P}_X(dx) \cdot \int y\mathbb{P}_Y(dy) = \mathbb{E}X \cdot \mathbb{E}Y.$$

Более строгое доказательство можно провести так же, как и в теореме о замене переменных.

**Определение 63.** Случайные величины, ковариация которых равна нулю, называются *некоррелируемыми случайными величинами*.

**Пример** (из некоррелируемости не вытекает независимость).  $\xi \sim \mathcal{N}(0, 1)$  — стандартное нормальное распределение,  $\eta = \xi^2$ . Зависимость  $\xi$  и  $\eta$  очевидна. Однако  $\text{cov}(\xi, \eta) = \mathbb{E}(\xi\eta) - \mathbb{E}\xi\mathbb{E}\eta = \mathbb{E}(\xi^3) - \mathbb{E}\xi\mathbb{E}\xi^2 = \{\mathbb{E}\xi = 0, \mathbb{E}\xi^3 = 0 \text{ в силу симметричности распределения и } \exists \mathbb{E}\xi^3\} = 0$ .

Выше была получена формула

$$\text{cov}(X, Y) = \langle X - \mathbb{M}X, Y - \mathbb{M}Y \rangle, \quad \forall X, Y \in \mathcal{L}_2(\mathbb{P}).$$

Введем оператор центрирования

$$CX = X - \mathbb{M}X.$$

Ясно, что  $\forall X \in \mathcal{L}_2(\mathbb{P}) \quad CX \in \mathcal{L}_2(\mathbb{P})$ , причем  $CX$  — проектор<sup>92</sup> на подпространство случайных величин, имеющих нулевое математическое ожидание.

По неравенству Шварца получим, что

$$|\text{cov}(X, Y)| = |\langle CX, CY \rangle| \leq \|CX\|_2 \cdot \|CY\|_2 = \sqrt{\mathbb{E}(CX)^2} \cdot \sqrt{\mathbb{E}(CY)^2} = \sqrt{\text{Var}X} \cdot \sqrt{\text{Var}Y}.$$

**Определение 64.** Если  $\mathbb{E}X^2 < \infty$ , то для случайной величины  $X$  определено *среднеквадратичное отклонение*<sup>93</sup>

$$\sigma_X = \sqrt{\text{Var}X}.$$

**Замечание.** Часто говорят, что дисперсия показывает отклонение случайной величины от среднего значения. Но верно ли это? По определению дисперсии  $\text{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2$ . Значит, если математическое ожидание (среднее)  $\mathbb{E}X$  имеет размерность, например, кг или см, то дисперсия измеряется в кг<sup>2</sup> и см<sup>2</sup>. Таким образом, дисперсия не может измерять отклонение случайной величины от среднего значения, т.к. имеет другую размерность. Физический смысл имеет только  $\sigma_X$ , которое и показывает отклонение.

7.

$$|\text{cov}(X, Y)| \leq \sigma_X \cdot \sigma_Y.$$

<sup>92</sup> $C^2 = C$

<sup>93</sup>раньше в отечественной литературе этот термин переводили как “стандартное отклонение”

**Определение 65.** Если для случайных величин  $X$  и  $Y$  существуют конечные вторые моменты ( $\mathbb{E}X^2 < \infty$ ,  $\mathbb{E}Y^2 < \infty$ ) и они имеют невырожденные распределения (т.е. распределения  $X$  и  $Y$  не сосредоточены в одной точке), то для этих случайных величин определено понятие *коэффициента корреляции*

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

**Замечания.**

- (a) Существование конечных вторых моментов обеспечивает существование конечной ковариации.
  - (b) Невырожденность распределений эквивалентна условию  $\sigma_X \neq 0$ ,  $\sigma_Y \neq 0$  (см. свойство 4).
  - (c)  $\rho_{XY}$  — безразмерная величина.
8. Для коэффициента корреляции справедливы следующие утверждения:

- (a)  $|\rho_{XY}| \leq 1$  (из свойства 7).
- (b) Если  $\mathbb{E}X^2 < \infty$ ,  $\mathbb{E}Y^2 < \infty$ ,  $X$  и  $Y$  независимы и имеют невырожденные распределения, то  $\rho_{XY} = 0$  (из свойства 6).
- (c) Коэффициент корреляции достигает своего максимального или минимального значения тогда и только тогда, когда имеет место аффинная зависимость случайных величин (при этом неявно подразумевается, что коэффициент корреляции существует, т.е.  $\mathbb{E}X^2 < \infty$ ,  $\mathbb{E}Y^2 < \infty$  и распределения невырождены):

$$\rho_{XY} = \pm 1 \iff \exists \text{ константы } a \neq 0, b \neq 0, c \text{ такие, что } aX + bY = c \text{ почти наверное.}$$

**Замечания.**

- i. Если  $a = 0$  или  $b = 0$ , то одна из случайных величин имеет вырожденное распределение, что противоречит исходному предположению о существовании коэффициента корреляции.
  - ii. Так как  $a \neq 0$ ,  $b \neq 0$ , то  $aX + bY = c$  можно переписать в виде  $Y = a'X + b'$ .
- Доказательство этого свойства будет дано чуть позже, при рассмотрении задачи регрессии.

Получаем, что  $\rho_{XY}$  показывает степень линейной зависимости. При этом из  $\rho_{XY} = 0$  не следует независимость  $X$  и  $Y$  (см. пример ранее). При  $|\rho_{XY}|$  близким к 1 можно утверждать, что  $X$  и  $Y$ , скорее всего, сильно зависимы.

Вывод: ковариационный анализ хорош при исследовании зависимостей, близких к линейным.

9. Если  $\mathbb{E}|X| < \infty$ ,  $\mathbb{E}|Y| < \infty$ , распределения  $X$  и  $Y$  невырождены,  $Y \stackrel{\text{п.н.}}{=} f(X)$ ,  $f$  — строго монотонная функция, то ковариация  $\text{cov}(X, Y)$  определена (но может быть равной  $\infty$ ) и

$$\begin{aligned} \text{cov}(X, Y) &> 0, \text{ если } f \text{ монотонно возрастает,} \\ \text{cov}(X, Y) &< 0, \text{ если } f \text{ монотонно убывает.} \end{aligned}$$

Доказательство. Пусть  $f$  монотонно возрастает.

$$\begin{aligned}
\text{cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}X)f(X)] = \int_{\mathbb{R}^n} (x - \mathbb{E}X)f(x)\mathbb{P}_X(dx) = \\
&= \int_{x > \mathbb{E}X} (x - \mathbb{E}X)f(x)\mathbb{P}_X(dx) + \int_{x < \mathbb{E}X} (x - \mathbb{E}X)f(x)\mathbb{P}_X(dx) > \\
&> \int_{x > \mathbb{E}X} (x - \mathbb{E}X)f(\mathbb{E}X)\mathbb{P}_X(dx) + \int_{x < \mathbb{E}X} (x - \mathbb{E}X)f(\mathbb{E}X)\mathbb{P}_X(dx) = \\
&= f(\mathbb{E}X) \int_{\mathbb{R}^n} (x - \mathbb{E}X)\mathbb{P}_X(dx) = 0.
\end{aligned}$$

10. Рассмотрим теперь случай случайных векторов:

$$X \in \mathbb{R}^m, \quad Y \in \mathbb{R}^n.$$

Будем считать, что  $X$  и  $Y$  представлены в виде столбцов, т.е. матриц  $m \times 1$  и  $n \times 1$  соответственно.

Математическое ожидание вектора / матрицы определяется покомпонентно.

**Определение 66.** *Взаимной ковариационной матрицей* двух случайных векторов  $X$  и  $Y$  называется следующая матрица размера  $m \times n$ :

$$C^{X,Y} = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)^T].$$

Из определения следует, что элементы взаимной ковариационной матрицы определяются по правилу:

$$C_{i,j}^{X,Y} = \text{cov}(X_i, Y_j).$$

Частным случаем взаимной ковариационной матрицы является **ковариационная матрица** (многомерный аналог дисперсии):

$$B^X = C^{X,X}.$$

11. Рассмотрим преобразование вида:

$$X' = AX + a, \quad Y' = DY + d,$$

где  $A \in \mathbb{R}^{k \times n}$ ,  $D \in \mathbb{R}^{l \times m}$  — константные матрицы,  $a \in \mathbb{R}^{k \times 1}$ ,  $d \in \mathbb{R}^{l \times 1}$  — константные вектора.

Пользуясь определением взаимной ковариационной матрицы и свойствами математического ожидания, получим

$$\begin{aligned}
C^{X',Y'} &= \mathbb{E}[(AX + a - (A\mathbb{E}X + a))(DY + d - (D\mathbb{E}Y + d))^T] = \mathbb{E}[A(X - \mathbb{E}X)(Y - \mathbb{E}Y)^T D^T] \\
&\implies C^{X',Y'} = AC^{X,Y}D^T.
\end{aligned}$$

В частности, для ковариационной матрицы получим:

$$B^{AX+a} = AB^X A^T.$$

12. Рассмотрим произвольную матрицу  $B \in \mathbb{R}^{m \times m}$ . Для того чтобы существовал  $m$ -мерный вектор  $X$  такой, что  $B^X = B$ , необходимо и достаточно, чтобы

- (а) матрица  $B$  была симметрична:  $B^T = B$ ,
- (б) матрица  $B$  была неотрицательно определена:  $B \geq 0$  ( $\forall t \in \mathbb{R}^m < Bt, t > \geq 0$ ).

Доказательство.

Пусть  $\exists X \in \mathbb{R}^m$ :  $B^X = B$ . Симметричность очевидна  $B^X$  ( $\mathbb{B}_{i,j}^X = \text{cov}(X_i, X_j) = \mathbb{B}_{j,i}^X$ ). Покажем неотрицательную определенность:  $\forall t \in \mathbb{R}^m$

$$\begin{aligned} 0 \leq \text{Var}(< t, X >) &= \text{Var} \left( \sum_{i=1}^m t_i X_i \right) = \text{cov} \left( \sum_{i=1}^m t_i X_i, \sum_{j=1}^m t_j X_j \right) = \\ &= \{ \text{свойство билинейности ковариации} \} = \sum_{i=1}^m \sum_{j=1}^m t_i t_j \text{cov}(X_i, X_j) = < B^X t, t >. \end{aligned}$$

Заметим, что попутно показали  $< B^X t, t >$  — дисперсия проекции случайного вектора  $X$  на направление  $t$ .

В обратную сторону. Пусть матрица  $B$  симметрична и положительно определена. Тогда  $X \sim \mathcal{N}(0, B)$ . Если матрица  $B \geq 0$ , то рассмотрим возмущенную матрицу  $B^\varepsilon = B + \varepsilon E > 0$ . Есть теорема о характеристических функциях, говорящая о том, что если последовательность характеристических функций  $f_{X^\varepsilon}(x)$  сходится к некоторой функции  $f(x)$ , то существует такой случайный вектор  $X$ , что  $f_X(x) = f(x)$ .

Можно предложить другое доказательство. Известно, что матрица  $B$  является симметричной и неотрицательно определенной тогда и только тогда, когда она представима в виде  $B = AA^T$  ( $A$  — некоторая матрица). Положим  $Y \sim \mathcal{N}(0, E)$  ( $Y_k \sim \mathcal{N}(0, 1)$ ). Тогда  $X = AY$  — искомый случайный вектор ( $B^X = \mathbb{E}XX^T - (\mathbb{E}X)(\mathbb{E}X)^T = AA^T = B$ ).

## 7.2 Линейные модели. Задача регрессии.

Дадим неформальную постановку задачи регрессии (различных формализаций много).

Следует отличать *вероятностную* задачу регрессии от *статической*. Мы дадим постановку именно **вероятностной задачи регрессии**.

Будем считать, что в рамках нашей модели известно совместное распределение невырожденных<sup>94</sup> случайных величин  $X$  и  $Y$  —  $\mathbb{P}_{XY}$  (на практике иногда достаточно знать лишь некоторые характеристики совместного распределения). Вероятностная задача регрессии состоит в том, чтобы найти приближенную зависимость  $X$  и  $Y$ :

$$Y \approx f(X).$$

Функция  $f$  называется функцией регрессии.

Сразу возникают два вопроса:

1. В каком смысле понимается “приближенная зависимость”? Нужна формализация приближения.

<sup>94</sup>В задаче регрессии мы рассматриваем невырожденные случайные величины, т.к. если одна из них вырождается (константа п.н.), то задача регрессии не имеет смысла.

2. В каком классе ищется функция регрессии (рассматриваем ли мы линейную / квадратичную или какую-либо другую зависимости)? Обычно фиксируется класс зависимостей, в котором ищется функция  $f$ :

$$f \in \mathfrak{M},$$

где  $\mathfrak{M}$  — некоторый “хороший” класс функций, в котором ищется наилучшее приближение.

Для формализации вероятностной задачи регрессии необходимо и достаточно ответить на эти два вопроса, т.е. определить понятие приближения и указать класс зависимостей.

### 7.2.1 Линейная задача регрессии.

**Одномерный случай.** Пусть  $X$  и  $Y$  — невырожденные случайные величины с конечными вторыми моментами  $\mathbb{E}X^2 < \infty$ ,  $\mathbb{E}Y^2 < \infty$  и известным совместным распределением  $\mathbb{P}_{XY}$ .

1. Зададим метрику (в пространстве  $\mathcal{L}_2(\mathbb{P})$ ):

$$\|X\|_2 = \sqrt{\mathbb{E}X^2}.$$

Таким образом,  $X_1$  и  $X_2$  близки, если  $\|X_1 - X_2\|_2 = \sqrt{\mathbb{E}(X_1 - X_2)^2}$  мало (минимизируется среднеквадратическая невязка, характеризующая степень близости).

2. Класс зависимостей:

$$\mathfrak{M} = \{f : f(x) = ax + b\} \text{ — класс аффинных зависимостей.}$$

Получаем задачу линейной регрессии:

$$\|Y - f(X)\|_2 \longrightarrow \min_{f \in \mathfrak{M}}. \quad (41)$$

Ее можно переписать по-другому, в более удобном для решения виде:

$$\mathbb{E}(Y - aX - b)^2 \longrightarrow \min_{a, b \in \mathbb{R}}.$$

Решение.

Положим

$$\phi(a, b) = \mathbb{E}(Y - aX - b)^2.$$

Тогда необходимое условие экстремума запишется в виде

$$\begin{aligned} \begin{cases} \frac{\partial \phi}{\partial a} = -\mathbb{E}2X(Y - aX - b) = 0, \\ \frac{\partial \phi}{\partial b} = -\mathbb{E}2(Y - aX - b) = 0; \end{cases} &\implies \begin{cases} \mathbb{E}XY - a\mathbb{E}X^2 - b\mathbb{E}X = 0, \\ \mathbb{E}Y - a\mathbb{E}X - b = 0; \end{cases} \\ &\implies \begin{cases} a = \frac{\mathbb{E}XY - \mathbb{E}Y\mathbb{E}X}{\mathbb{E}X^2 - (\mathbb{E}X)^2}, \\ b = \mathbb{E}Y - \frac{\mathbb{E}XY - \mathbb{E}Y\mathbb{E}X}{\mathbb{E}X^2 - (\mathbb{E}X)^2} \mathbb{E}X. \end{cases} \end{aligned}$$

В силу того, что  $\phi(a, b)$  — неотрицательная квадратичная форма, получаем, что найденное значение есть единственная точка минимума функции  $\phi(a, b)$ .

Ответ.

$$\begin{cases} a^* = \frac{\sigma_Y}{\sigma_X} \cdot \rho_{XY}, \\ b^* = \mathbb{E}Y - a^* \mathbb{E}X. \end{cases}$$



**Замечание 1.** Достаточно знать  $\mathbb{E}X$ ,  $\mathbb{E}Y$ ,  $\sigma_X$ ,  $\sigma_Y$ ,  $\rho_{XY}$  вместо совместного распределения.

**Замечание 2.** Подставим найденные значения параметров  $a^*$  и  $b^*$  в функцию  $\phi(a, b)$ :

$$\begin{aligned}
\phi(a^*, b^*) &= \mathbb{E}(Y - a^*X - b^*)^2 = \\
&= \mathbb{E}Y^2 + a^{*2}\mathbb{E}X^2 + (\mathbb{E}Y - a^*\mathbb{E}X)^2 - 2a^*\mathbb{E}XY - 2(\mathbb{E}Y - a^*\mathbb{E}X)\mathbb{E}Y + 2a^*(\mathbb{E}Y - a^*\mathbb{E}X)\mathbb{E}X = \\
&= \mathbb{E}Y^2 + a^{*2}\mathbb{E}X^2 + (\mathbb{E}Y)^2 - 2a^*\mathbb{E}Y\mathbb{E}X + a^{*2}(\mathbb{E}X)^2 - 2a^*\mathbb{E}XY - \\
&\quad - 2(\mathbb{E}Y)^2 + 2a^*\mathbb{E}X\mathbb{E}Y + 2a^*\mathbb{E}Y\mathbb{E}X - 2a^{*2}(\mathbb{E}X)^2 = \\
&= \mathbb{E}Y^2 + a^{*2}\mathbb{E}X^2 - (\mathbb{E}Y)^2 - a^{*2}(\mathbb{E}X)^2 - 2a^*\mathbb{E}XY + 2a^*\mathbb{E}Y\mathbb{E}X = \\
&= a^{*2}\mathbb{V}\text{ar}X - 2a^*\text{cov}(X, Y) + \mathbb{V}\text{ar}Y = \frac{\mathbb{V}\text{ar}Y}{\mathbb{V}\text{ar}X} \cdot \rho_{XY}^2 \mathbb{V}\text{ar}X - 2\frac{\sigma_Y}{\sigma_X} \cdot \rho_{XY} \text{cov}(X, Y) + \mathbb{V}\text{ar}Y = \\
&= \sigma_Y^2 \rho_{XY}^2 - 2\sigma_Y^2 \rho_{XY}^2 + \sigma_Y^2 = \sigma_Y^2 (1 - \rho_{XY}^2).
\end{aligned}$$

Получаем, что если  $|\rho_{XY}| = 1$  для некоторых случайных величин  $X$  и  $Y$ , то  $\mathbb{E}(Y - a^*X - b^*)^2 = 0$ , а значит,  $Y \stackrel{\text{п.п.}}{=} a^*X + b^*$ . Чем ближе  $|\rho_{XY}|$  к 1, тем сильнее линейная зависимость и тем лучше приближение.

Итак, мы доказали 8 свойство предыдущего пункта главы:

$$\rho_{XY} = \pm 1 \iff \exists \text{ константы } a \neq 0, b \neq 0, c \text{ такие, что } aX + bY = c \text{ почти наверное.}$$

Заметим только, что если известно, что зависимость между  $X$  и  $Y$  линейная, то  $\mathbb{E}(Y - a^*X - b^*)^2 = 0$ , а значит,  $|\rho_{XY}| = 1$ , т.к.  $\sigma_Y^2 \neq 0$  в силу невырожденности.

**Замечание 3.** В результате решения линейной задачи регрессии найдены числа  $a^*$  и  $b^*$ , характеризующие наилучшее в среднеквадратическом смысле приближение  $Y$  линейной функцией от  $X$ :

$$Y = (a^*X + b^*) + Z.$$

Покажем, что  $(a^*X + b^*) - \mathbb{E}(a^*X + b^*)$  и  $Z - \mathbb{E}Z$  ортогональны как элементы пространства  $\mathcal{L}_2(\mathbb{P})$ , т.е. случайные величины  $(a^*X + b^*)$  и  $Z$  некоррелируемы:

$$\begin{aligned}
\text{cov}(a^*X + b^*, Z) &= \mathbb{E}[(a^*X + b^* - \mathbb{E}(a^*X + b^*))(Z - \mathbb{E}Z)] = \\
&= \mathbb{E}[a^*(X - \mathbb{E}X)(Y - a^*X - b^* - \mathbb{E}(Y - a^*X - b^*))] = \\
&= \mathbb{E}[a^*(X - \mathbb{E}X)((Y - \mathbb{E}Y) - a^*(X - \mathbb{E}X))] = \\
&= a^* \text{cov}(X, Y) - a^{*2} \mathbb{V}\text{ar}X = \frac{\sigma_Y}{\sigma_X} \cdot \rho_{XY} \text{cov}(X, Y) - \frac{\sigma_Y^2}{\sigma_X^2} \cdot \rho_{XY}^2 \mathbb{V}\text{ar}X = \sigma_Y^2 \rho_{XY}^2 - \sigma_Y^2 \rho_{XY}^2 = 0.
\end{aligned}$$

**Многомерный случай.** (кратко)  $X \in \mathbb{R}^n$ ,  $Y \in \mathbb{R}^m$  (невырожденные). Задача состоит в том, чтобы найти линейную оценку, оптимальную в среднеквадратическом смысле (ЛООСК):

$$\begin{aligned}
f(x) &= Ax + b, \quad x \in \mathbb{R}^n, \quad A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m; \\
\mathbb{E}\|Y - AX - b\|^2 &\longrightarrow \min_{A, b}.
\end{aligned}$$

Дифференцированием можно показать, что искомая функция регрессии примет вид (см. курс теории идентификации):

$$f(x) = \mathbb{E}Y + C^{Y, X} (B^X)^{-1} (x - \mathbb{E}X). \quad (42)$$

При  $n = 1$ ,  $m = 1$  получим рассмотренный выше одномерный случай.

Заметим, что если происходит вырождение по некоторым координатам  $Y$ , то запись (42) остается корректной; если же происходит вырождение по  $X$ , то

### 7.2.2 Другие задачи регрессии.

1. Рассмотрим теперь более широкий класс функций:

$$\mathfrak{M} = \{f : \mathbb{E}f^2(X) < \infty\}.$$

Предположим, что случайная величина  $Y$  имеет конечный второй момент  $\mathbb{E}Y^2 < \infty$ , а  $X$  — любой случайный элемент (без ограничений).

Степень близости будем по-прежнему характеризовать среднеквадратической невязкой:

$$\|Y - f(X)\|_2 = \sqrt{\mathbb{E}(Y - f(X))^2}.$$

Итак, рассмотрим задачу регрессии:

$$\|Y - f(X)\|_2^2 = \mathbb{E}(Y - f(X))^2 \longrightarrow \min_{f \in \mathfrak{M}}.$$

Эта задача имеет вполне определенное решение. Для того, чтобы его найти, рассмотрим вспомогательную задачу:

$$\begin{aligned} \mathfrak{M}_o &= \{f : f(x) = \text{const}\}, \\ \mathbb{E}(Y - f(X))^2 &\longrightarrow \min_{f \in \mathfrak{M}_o}. \end{aligned}$$

Запишем немного по-другому, в более удобной для решения форме:

$$\mathbb{E}(Y - a)^2 \longrightarrow \min_{a \in \mathbb{R}}.$$

Положим

$$\phi(a) = \mathbb{E}(Y - a)^2.$$

Используя необходимое условие экстремума, находим:

$$\frac{d\phi}{da} = -2\mathbb{E}(Y - a) = 0 \implies a = \mathbb{E}Y.$$

Очевидно, что найденное значение параметра является искомым:

$$Y \approx f(X) = \mathbb{E}Y \text{ — решение вспомогательной задачи.}$$

Вернемся к исходной задаче.

$$\begin{aligned} \mathbb{E}(Y - f(X))^2 &= \{\text{теорема о замене переменных в интеграле Лебега}\} = \\ &= \int (y - f(x))^2 \mathbb{P}_{XY}(dx, dy) = \\ &= \{\mathbb{P}_{XY}(A, B) = \int_A \mathbb{P}_{Y|X=x}(B) \mathbb{P}_X(dx) \text{ по опр. условного распределения (см. 11 главу)}\} = \\ &= \int \left[ \int (y - f(x))^2 \mathbb{P}_{Y|X=x}(dy) \right] \mathbb{P}_X(dx). \end{aligned}$$

Заметим, что интегрирование везде производится по всему пространству. По соглашению, в этом случае пределы интегрирования можно опустить.

Используя понятие условного математического ожидания, запишем:

$$\mathbb{E}(Y - f(X))^2 = \int \mathbb{E}[(Y - f(x))^2 | X = x] \mathbb{P}_X(dx) \longrightarrow \min_{f \in \mathfrak{M}},$$

т.е. при каждом фиксированном  $x$  нужно найти  $f(x)$ , минимизирующую  $\mathbb{E}[(Y - f(x))^2 | X = x]$ . Таким образом, мы пришли к уже рассмотренной выше вспомогательной задаче, используя которую, получаем ответ:

$$f(x) = \mathbb{E}(Y | X = x) \stackrel{\text{def}}{=} \int y \mathbb{P}_{Y|X=x}(dy). \quad (43)$$

**Замечание.** Вообще говоря, функция  $f(x)$  определена с точностью до почти всюду по мере  $\mathbb{P}_X$ , т.к. при ее на множестве меры нуль значение  $\mathbb{E}(Y - f(X))^2$  не изменится. Однако заметим, что выражение (43) корректно, т.к. условное распределение (переходное ядро)  $Q(x, B) = \mathbb{P}_{Y|X=x}(B)$  определено почти всюду по мере  $\mathbb{P}_X$  (см. 11 главу).

2. Возьмем теперь такое расстояние (среднеабсолютная невязка):

$$\|X_1 - X_2\|_1 = \mathbb{E}|X_1 - X_2|.$$

Соответственно изменим и класс зависимостей:

$$\mathfrak{M} = \{f : \mathbb{E}|f(X)| < \infty\}.$$

Получаем следующую задачу регрессии:

$$\|Y - f(X)\|_1 = \mathbb{E}|Y - f(X)| \longrightarrow \min_{f \in \mathfrak{M}}.$$

Аналогично предыдущему случаю, рассмотрим вспомогательную задачу:

$$\begin{aligned} \mathfrak{M}_0 &= \{f : f(x) = \text{const}\}, \\ \mathbb{E}|Y - f(X)| &\longrightarrow \min_{f \in \mathfrak{M}_0}. \end{aligned}$$

Запишем немного по-другому, в более удобной для решения форме:

$$\mathbb{E}|Y - a| \longrightarrow \min_{a \in \mathbb{R}}.$$

Положим

$$\phi(a) = \mathbb{E}|Y - a| = \int |y - a| \mathbb{P}_Y(dy) = \int_{-\infty}^a (a - y) \mathbb{P}_Y(dy) + \int_a^{+\infty} (y - a) \mathbb{P}_Y(dy).$$

Используя необходимое условие экстремума, находим:

$$\frac{d\phi}{da} = \int_{-\infty}^a \mathbb{P}_Y(dy) - \int_a^{+\infty} \mathbb{P}_Y(dy) = \mathbb{P}_Y((-\infty, a]) - \mathbb{P}_Y([a, +\infty)) = 0 \implies a = \text{med } Y.$$

Очевидно, что найденное значение параметра является искомым:

$$Y \approx f(X) = \text{med } Y \text{ (медиана) — решение вспомогательной задачи.}$$

Вернемся к исходной задаче.

$$\begin{aligned}\mathbb{E} |Y - f(X)| &= \int |y - f(x)| \mathbb{P}_{XY}(dx, dy) = \\ &= \int \left[ \int |y - f(x)| \mathbb{P}_{Y|X=x}(dy) \right] \mathbb{P}_X(dx) = \\ &= \int \mathbb{E} [|Y - f(x)| | X = x] \mathbb{P}_X(dx) \longrightarrow \min_{f \in \mathfrak{M}},\end{aligned}$$

т.е. при каждом фиксированном  $x$  нужно найти  $f(x)$ , минимизирующую  $\mathbb{E} [|Y - f(x)| | X = x]$ . Таким образом, мы пришли к уже рассмотренной выше вспомогательной задаче, используя которую, получаем ответ — условную медиану:

$$f(x) = \text{med}(Y | X = x) \stackrel{\text{def}}{=} \mathbb{P}_{Y|X=x}((-\infty, \text{med}(Y | X = x)]) = \mathbb{P}_{Y|X=x}([\text{med}(Y | X = x), +\infty)).$$

**Замечание.** Как и в предыдущей задаче, функция  $f(x)$  определена с точностью до почти всюду по мере  $\mathbb{P}_X$ .

Вывод: много постановок задачи регрессии  $\longrightarrow$  много решений.

### 7.3 Свойства многомерного нормального распределения. Моделирование гауссовского случайного вектора.

Рассмотрим несколько способов введения многомерного нормального распределения (гауссовского распределения).

1. Независимые случайные величины  $X$  и  $Y$  имеют гауссовское распределение, если случайные величины  $X + Y$  и  $X - Y$  также независимы. (Идея: независимость означает ортогональность  $\implies$  поворот на 45 градусов сохраняет ортогональность).
2. Рассмотрим модель идеального газа (совокупность молекул, взаимодействиями между которыми можно пренебречь). Проекция вектора импульса одной молекулы на оси  $p_X, p_Y, p_Z$  независимы в силу изотропности (симметричности) модели (нет предпочтительных направлений, при вращении осей распределение не меняется). Тогда распределение вектора импульса будет гауссовским (Максвелл).
3. Пусть имеется ряд независимых однородных (не сильно отличающихся, но и не обязательно одинаково распределенных) случайных величин, у которых существует достаточно много моментов (распределение без тяжелых хвостов). Тогда распределение их суммы стремится (при увеличении количества слагаемых) к гауссовскому распределению (ЦПТ).
4. Энтропийная характеристика и т.д. (статей по этому поводу очень много)

Дадим теперь определение нормального распределения через плотность (им мы и будем в дальнейшем пользоваться).

**Определение 67.** Случайная величина  $X$  имеет *стандартное нормальное распределение* (обозначается  $X \sim \mathcal{N}(0, 1)$ ), если ее распределение  $\mathbb{P}_X$  абсолютно непрерывно и непрерывный вариант плотности имеет вид<sup>95</sup>:

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

<sup>95</sup>напомним, что плотность определена почти всюду относительно меры Лебега

Заметим, что

$$\begin{aligned}\mathbb{E}X &= \{\text{т. замене переменных}\} = \int_{-\infty}^{+\infty} x \mathbb{P}_X(dx) = \\ &= \int_{-\infty}^{+\infty} x p_X(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{x^2}{2}} dx = 0; \\ \text{Var}X &= \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = 1.\end{aligned}$$

**Определение 68.** Случайная величина  $Y$  имеет **нормальное (гауссовское) распределение с параметрами  $\mu$  и  $\sigma \geq 0$**  (обозначается  $Y \sim \mathcal{N}(\mu, \sigma^2)$ ), если существует такая случайная величина  $X \sim \mathcal{N}(0, 1)$ , что:

$$Y = \mu + \sigma X.$$

Из определения видно, что  $\mu$  — параметр сдвига, а  $\sigma$  — параметр масштаба:

$$\begin{aligned}\mathbb{E}Y &= \mathbb{E}(\mu + \sigma X) = \mu + \sigma \mathbb{E}X = \mu; \\ \text{Var}Y &= \mathbb{E}(Y - \mathbb{E}Y)^2 = \mathbb{E}(\mu + \sigma X - \mu)^2 = \mathbb{E}(\sigma X)^2 = \sigma^2 \mathbb{E}X^2 = \sigma^2 \\ &\implies \text{среднеквадратичное отклонение } \sigma_Y = \sigma.\end{aligned}$$

Обычно рассматривают случай  $\sigma > 0$ . Тогда  $Y \sim \mathcal{N}(\mu, \sigma)$  имеет абсолютно непрерывное распределение с непрерывным вариантом плотности<sup>96</sup>

$$p_Y(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Если  $\sigma = 0$ , то  $Y \sim \mathcal{N}(\mu, \sigma)$  имеет вырожденное распределение ( $\mathbb{P}(Y = \mu) = 1$ ). Договоренность: вырожденное распределение является частным случаем гауссовского (так удобнее при рассмотрении многомерного случая).

**Определение 69.** Случайный вектор  $X$  имеет **многомерное нормальное (гауссовское) распределение** (обозначается  $Y \sim \mathcal{N}(\mu, \sigma^2)$ ), если при проектировании этого вектора на любое направление с любым центром получается одномерное гауссовское распределение (т.е. координата оси направления имеет одномерное гауссовское распределение).

Корректность определения (существование и единственность распределения) принимаем без доказательства.

Плотность многомерного нормального распределения (в случае отсутствия вырождения по всем направлениям) выражается в виде (имеется в виду непрерывный вариант плотности):

$$p_{(X_1, \dots, X_n)}(x) = \frac{1}{\sqrt{(2\pi)^n \det(B)}} e^{-\frac{(B^{-1}(x-\mu), x-\mu)}{2}}, \quad x \in \mathbb{R}^n.$$

Обозначение:  $X \sim \mathcal{N}(\mu, B)$ . Заметим, что  $\mathbb{E}X = \mu$ ,  $B^X = B$ .

**Свойства гауссовского распределения.**

---

<sup>96</sup>функция распределения  $F_Y(t) = \mathbb{P}(Y < t) = \mathbb{P}(X < \frac{t-\mu}{\sigma}) = F_X(\frac{t-\mu}{\sigma}) = \int_{-\infty}^{\frac{t-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx =$   
 $\{y = \mu + \sigma x\} = \int_{-\infty}^t \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$

1. Аффинное преобразование гауссовского случайного вектора будет гауссовским случайным вектором (возможно другой размерности).

Доказательство из определения. Пусть  $X$  имеет многомерное гауссовское распределение. Аффинные преобразования:

- (а) растяжение/сжатие  $\implies$  растяжение/сжатие любой проекции; если  $Y \mathbb{R}$  имеет одномерное гауссовское распределение, то и  $aY \mathbb{R}$ ,  $a \geq 0$ , будет иметь одномерное гауссовское распределение (по определению);
- (б) поворот  $\implies$  распределение останется многомерным гауссовским, т.к. в определении речь идет о *любом* направлении;
- (с) сдвиг  $\implies$  распределение останется многомерным гауссовским, т.к. в определении речь идет о *любом* центре.

**Следствие.** Маргинальные распределения гауссовского распределения будут гауссовскими (взятию маргинального распределения соответствует линейное преобразование вида  $X = (X_1, \dots, X_n) \rightarrow X_i$ ).

2. Условное маргинальное распределение части координат при условии, что известны другие координаты, также является гауссовским.

**Теорема 14.** Пусть  $X \in \mathbb{R}^n$ ,  $Y \in \mathbb{R}^m$ ,  $Z = (X, Y)$ . Известно, что их совместное распределение  $\mathbb{P}_{XY} = \mathbb{P}_Z$  гауссовское. Тогда условное распределение  $\mathbb{P}_{X|Y=y}$  является гауссовским, причем условные математическое ожидание и дисперсия имеют вид:

$$\begin{aligned}\mu(y) &= \mathbb{E}(X | Y = y) = \mathbb{E}X + C^{X,Y} (B^Y)^{-1} (Y - \mathbb{E}Y), \\ B(y) &= B^{X|Y=y} = B^X - C^{X,Y} (B^Y)^{-1} C^{Y,X},\end{aligned}$$

т.е.  $\mathbb{P}_{X|Y=y} \longrightarrow \mathcal{N}(\mu(y), B(y))$ .

Доказательство этой теоремы можно найти в курсе теории идентификации.

Заметим, что  $\mu(y)$  — аффинная функция от  $y$ ,  $B(y) = \text{const}$  как функция от  $y$ .

Ранее было получено, что

$$f(x) = \mathbb{E}(Y | X = x)$$

— решение общей задачи регрессии при выборе соответствующей метрики. Теперь мы знаем, что в случае гауссовского распределения функция  $\mu(x) = \mathbb{E}(Y | X = x)$  линейна по  $x$ , а значит решения линейной и общей задачи регрессии совпадают (минимум в классе всех функции достигается на линейной функции).

## 8 Теорема Радона-Никодима. Абсолютно-непрерывные распределения, свойства плотности. Моделирование методом элиминации фон Неймана.

### 8.1 Теорема Радона-Никодима.

**Определение 70.** Пусть меры  $\mu$  и  $\nu$  счетно-аддитивны ( $\sigma$ -аддитивны) и заданы на общей  $\sigma$ -алгебре  $\Sigma$  подмножеств из  $X$ . Тогда **мера**  $\nu$  называется **абсолютно непрерывной относительно меры**  $\mu$ , если

$$\mu(A) = 0 \implies \nu(A) = 0 \quad \forall A \in \Sigma.$$

Обозначение:

$$\nu \ll \mu.$$

**Определение 71.** Меры  $\mu$  и  $\nu$  называются *эквивалентными мерами*, если  $\mu \ll \nu$  и  $\nu \ll \mu$ . Обозначение:

$$\nu \sim \mu.$$

Очевидно, что отношение абсолютной непрерывности транзитивно:

$$\nu \ll \mu \ll \eta \implies \nu \ll \eta.$$

Теорема Радона-Никодима, вообще говоря, формулируется для зарядов (см. книжку Колмогорова, Фомина по функциональному анализу). Поэтому напомним понятие заряда и дадим классическую формулировку теоремы.

Рассмотрим пространство  $X$ ,  $\Sigma$  —  $\sigma$ -алгебра подмножеств  $X$ .

**Определение 72.** *Зарядом*  $\Phi$  называется функция  $\Phi: \Sigma \rightarrow \mathbb{R}$  такая, что

$$\Phi\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \Phi(A_k).$$

Заметим, что главное отличие заряда от меры заключается в том, что заряд может принимать отрицательные значения.

Справедливо следующее утверждение.

**Утверждение 10.** Если  $\Phi$  — заряд, то существуют две  $\sigma$ -аддитивные меры  $\mu$  и  $\nu$  такие, что  $\Phi = \mu - \nu$ .

**Определение 73.** *Заряд*  $\Phi$  называется *абсолютно непрерывным относительно меры  $\mu$* , если

$$\mu(A) = 0 \implies \Phi(A) = 0 \quad \forall A \in \Sigma.$$

**Теорема 15** (Радон, Никодим). Пусть  $\mu$  — некоторая конечная  $\sigma$ -аддитивная мера, определенная на  $\sigma$ -алгебре  $\Sigma$  подмножеств из  $X$ , а  $\Phi$  — заряд, определенный на той же  $\sigma$ -алгебре. Если заряд  $\Phi$  абсолютно непрерывен относительно меры  $\mu$ , то существует функция  $f$  такая, что

$$\Phi(A) = \int_A f d\mu, \quad \forall A \in \Sigma.$$

**Замечание.** Функция  $f$ , называемая производной заряда  $\Phi$  по мере  $\mu$ , определена однозначно, с точностью до  $\mu$ -эквивалентности (две функции называются  $\mu$ -эквивалентными, если они совпадают почти всюду относительно меры  $\mu$ ).

Теперь приведем ту формулировку теоремы, которой мы будем пользоваться.

Рассмотрим измеримое пространство  $(\Omega, \mathcal{F})$ .

**Теорема 16** (Радон, Никодим). Пусть  $\mu$  и  $\nu$  — некоторые  $\sigma$ -конечные меры, определенные на  $\sigma$ -алгебре  $\mathcal{F}$  подмножеств из  $\Omega$ . Мера  $\mu$  абсолютно непрерывна относительно меры  $\nu$  ( $\mu \ll \nu$ )  $\iff$  существует функция  $f$  такая, что

$$\mu(A) = \int_A f(\omega) \nu(d\omega) \equiv \int_A f d\nu, \quad \forall A \in \mathcal{F}.$$



**Замечание 1.** Очевидно, что если  $\mu(A) = \int_A f(\omega) \nu(d\omega)$ , то  $\mu \ll \nu$ . Поэтому мы и можем сформулировать теорему в обе стороны.

**Замечание 2.** (о  $\sigma$ -конечных мерах) ?????????

**Определение 74.** Пусть  $\mu \ll \nu$ . Тогда  $\mu(A) = \int_A f d\nu$ . Функция  $f$  называется *производной Радона-Никодима от меры  $\mu$  по мере  $\nu$*  и обозначается<sup>97</sup>

$$f = \frac{d\mu}{d\nu}.$$

Физический смысл  $f$  — плотность меры  $\mu$  относительно меры  $\nu$ . Эта функция определена однозначно, с точностью до  $\nu$ -эквивалентности.

## 8.2 Абсолютно-непрерывные распределения, свойства плотности.

Прежде чем перейти непосредственно к изложению материала, связанного с абсолютно непрерывным распределением, напомним понятие абсолютно непрерывной функции.

**Определение 75.** Функция  $f$ , заданная на некотором отрезке  $[a, b]$ , называется *абсолютно непрерывной функцией* на отрезке  $[a, b]$ , если для любого  $\varepsilon > 0$  найдется такое  $\delta > 0$ , что, какова бы ни была конечная система попарно непересекающихся интервалов

$$(a_k, b_k), \quad k = 1, \dots, n,$$

с суммой длин, меньшей  $\delta$ :

$$\sum_{k=1}^n (b_k - a_k) < \delta,$$

выполнено неравенство

$$\sum_{k=1}^n |f(b_k) - f(a_k)| < \varepsilon.$$

**Замечание 1.** В определении 75 можно вместо любой *конечной* системы интервалов с суммой длин  $< \delta$  рассматривать любую *конечную* или *счетную* систему интервалов, сумма длин которых  $< \delta$  (просто переходя в определении 75 к пределу при  $n \rightarrow \infty$ ).

**Замечание 2.** Всякая абсолютно непрерывная функция имеет ограниченное изменение (поскольку отрезок  $[a, b]$  можно разбить на конечное число отрезков длины  $< \delta$ ).

**Замечание 3.** Ясно, что всякая абсолютно непрерывная функция равномерно непрерывна. Обратное, вообще говоря, неверно, например: “канторова лестница” непрерывна (а значит, и равномерно непрерывна) на отрезке  $[0, 1]$ , однако она не абсолютно непрерывна, т.к. множество точек роста этой функции есть канторово множество меры нуль.

Следующие две теоремы устанавливают тесную связь между понятием абсолютной непрерывности и интегралом Лебега.

**Теорема 17.** Функция  $F(x) = \int_a^x f(t) dt$  (интеграл Лебега) абсолютно непрерывна.

Доказательство теоремы опирается на абсолютную непрерывность интеграла Лебега.

<sup>97</sup>Формальное обозначение можно пояснить так:  $\int_A f d\nu = \int_A \frac{d\mu}{d\nu} d\nu = \int_A d\mu = \mu(A)$ .

**Теорема 18** (Лебег). Производная  $f = F'$  абсолютно непрерывной функции<sup>98</sup>, заданной на отрезке  $[a, b]$ , суммируема на этом отрезке и для каждого  $x$  ( $a \leq x \leq b$ )

$$\int_a^x f(t)dt = F(x) - F(a).$$

Эти две теоремы показывают, что абсолютно непрерывные функции, и только они, восстанавливаются с точностью до постоянного слагаемого по своей производной с помощью операции интегрирования.

Заметим, что теорема Радона-Никодима представляет собой, очевидно, естественное обобщение теоремы Лебега о том, что абсолютно непрерывная функция есть интеграл от своей производной (мера Лебега-Стилтьеса  $\mu_F([a, b]) = F(b) - F(a)$ , где  $F(x)$  — абсолютно непрерывная функция, абсолютно непрерывна (относительно меры Лебега)).

Рассмотрим меру Лебега — единственную с точностью до постоянного множителя меру, обладающую свойством инвариантности относительно сдвигов ( $\mu(A + x) = \mu(A)$ ,  $A + x = \{z = y + x | y \in A\}$ )<sup>99</sup>.

**Определение 76.** Мера, абсолютно непрерывная относительно меры Лебега, называется просто **абсолютно непрерывной мерой**.

Вероятностная мера может быть абсолютно непрерывной только относительно тех мер, которые заданы на  $\mathcal{F}$ . Вероятностная мера  $\mathbb{P}: \mathcal{F} \rightarrow [0, 1]$  не может быть абсолютно непрерывной, если на  $\mathcal{F}$  не задана мера Лебега<sup>100</sup>. Зато абсолютно непрерывным может быть распределение некоторой случайной величины.

Напомним определение случайной величины.

**Определение 77.** Пусть заданы два измеримых пространства:  $(\Omega, \mathcal{F})$  и  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , где  $\mathcal{B}(\mathbb{R})$  — борелевская  $\sigma$ -алгебра множеств на прямой. **Случайной величиной** называется отображение

$$X: \Omega \rightarrow \mathbb{R},$$

которое является  $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$ -измеримым.

Канонической моделью будет тройка  $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ , где  $\mathbb{P}_X = X \circ \mathbb{P} = \mathbb{P} \circ X^{-1}: \mathcal{B}(\mathbb{R}) \rightarrow [0, 1]$  — индуцированная мера (распределение случайной величины  $X$ ). Ясно, что на  $\mathcal{B}(\mathbb{R})$  можно задать меру Лебега, а значит, индуцированная мера может быть абсолютно непрерывной.

**Определение 78.** Распределение  $\mathbb{P}_X$  случайной величины  $X$  называется **абсолютно непрерывным распределением**, если мера  $\mathbb{P}_X$  абсолютно непрерывна.

Если распределение случайной величины  $X$  абсолютно непрерывно, то по теореме Радона-Никодима существует функция  $p$  такая, что

$$\mathbb{P}_X(A) = \int_A p(t)dt, \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

<sup>98</sup>Функции, имеющие ограниченное изменение, дифференцируемы почти всюду, т.е. производная таких функций определена почти всюду.

<sup>99</sup>понятие меры Лебега вводится, исходя из обычного определения длины/площади/объема и их аналогов для многомерного случая

<sup>100</sup>мера Лебега может быть задана, если  $\Omega = \mathbb{R}^n$

Тогда соответствующая функция распределения представима в виде

$$F_X(x) = \mathbb{P}(X < x) = \mathbb{P}_X((-\infty, x)) = \int_{-\infty}^x p(t) dt.$$

По теореме 17 функция  $F_X(x)$  будет абсолютно непрерывной функцией.

**Определение 79.** Функция  $p(x)$  называется **плотностью** абсолютно непрерывного распределения или плотностью вероятностей или просто плотностью случайной величины  $X$ .

Плотность абсолютно непрерывного распределения является производной абсолютно непрерывной функции распределения и *определена почти всюду*.

**Замечание 1.** Используя теорему Радона-Никодима, можем дать другое определение абсолютно непрерывного распределения.

**Определение 80.** Распределение  $\mathbb{P}_X$  случайной величины  $X$  называется **абсолютно непрерывным распределением**, если существует функция  $p$  такая, что

$$\mathbb{P}_X(A) = \int_A p(t) dt, \quad \forall A \in \mathcal{B}(\mathbb{R}).$$

**Замечание 2.** Вообще говоря, можно дать определение абсолютно непрерывного распределения относительно любой меры.

**Определение 81.** Пусть  $\mu$  — мера, заданная на измеримом пространстве  $(\mathbb{R}, \mathcal{B})$ . Распределение  $\mathbb{P}_X$  случайной величины  $X$  называется **абсолютно непрерывным распределением относительно меры  $\mu$** , если мера  $\mathbb{P}_X$  абсолютно непрерывна относительно меры  $\mu$ .

В этом случае производная Радона-Никодима  $p = \frac{d\mathbb{P}_X}{d\mu}$  называется обобщенной плотностью распределения или просто плотностью.

### 8.3 Моделирование методом элиминации фон Неймана.

Метод Фон-Неймана моделирования случайной величины заключается в том, что если у нас есть датчик для одного распределения, то при некоторых условиях на его параметры на основе этого датчика можно построить датчик для другого распределения.

Будем рассматривать моделирование стандартного нормального распределения на основе распределения Коши с параметрами  $a$  и  $b$ .

Пусть имеется некоторое пространство  $(\mathcal{X}, \mathcal{A})$ , на котором заданы вероятностные меры  $\mathbf{P}$  и  $\mathbf{Q}$ , такие, что  $\mathbf{P} < k \cdot \mathbf{Q}$ , где  $k$  — некоторое вещественное число, т.е.:

$$\forall A \in \mathcal{A} \Rightarrow \mathbf{P}(A) < k \cdot \mathbf{Q}(A).$$

Это равносильно тому, что  $\frac{d\mathbf{P}}{d\mathbf{Q}} < k$ , где  $\frac{d\mathbf{P}}{d\mathbf{Q}}$  — производная Радона-Никодима.

В нашем случае распределения абсолютно непрерывны. Значит они абсолютно непрерывны относительно меры Лебега,  $\Rightarrow$  выполняются условия теоремы Радона-Никодима, т.е.  $\exists \frac{d\mathbf{P}}{d\mathbf{Q}}$  — производная Радона-Никодима, а, следовательно, выполняется соотношение  $\frac{d\mathbf{P}}{d\mathbf{Q}} < k$ , где  $d\mathbf{P}$  и  $d\mathbf{Q}$  соответственно плотности стандартного нормального распределения и распределения Коши с параметрами сдвига  $b$  и масштаба  $a$ .

Так как  $\frac{dP}{dQ} < k$ , то получаем, что чем меньше  $k$ , тем эффективнее будет алгоритм. Итак для реализации метода надо выбрать  $k^*$ :

$$k^* = \inf_k (f_n \leq k \cdot f_c),$$

где  $f_n$  и  $f_c$  — плотности стандартного нормального распределения и распределения Коши соответственно. Эта задача сводится к следующей задаче минимакса:

$$k^* = \min_{a,b} \max_{x \in \mathbb{R}} \left( \frac{f_n(x)}{f_c(x, a, b)} \right).$$

Сначала рассмотрим случай, когда  $b = 0$ , тогда имеем:

$$k_0^* = \min_a \max_{x \in \mathbb{R}} \left( \frac{f_n(x)}{f_c(x, a)} \right).$$

Эта задача решается методами математического анализа. Рассмотрим функцию  $g(x) = \frac{f_n}{f_c}$ . Имеем:

$$g(x) = \sqrt{\frac{\pi}{2}} \cdot \frac{1}{a} e^{-\frac{x^2}{2}} (x^2 + a^2) = C e^{-\frac{x^2}{2}} (x^2 + a^2)$$

Максимум функции  $g(x)$  достигается при  $x^* = \pm\sqrt{2-a^2}$ , если  $|a| \leq \sqrt{2}$ , и при  $x = 0$ , если  $|a| > \sqrt{2}$ . Тогда выражение для  $k_0^*$  перепишется в виде:

$$k_0^* = \min \left( \min_{|a| > \sqrt{2}} \left\{ a \sqrt{\frac{\pi}{2}} \right\}, \min_{|a| \leq \sqrt{2}} \left\{ \frac{\sqrt{2\pi}}{a} e^{\frac{a^2}{2}-1} \right\} \right)$$

Теперь рассмотрим случай, когда  $b \neq 0$ . Тогда:

$$\begin{aligned} k_0^* &= \min_a \max_{x \in \mathbb{R}} \left\{ \sqrt{\frac{\pi}{2}} \cdot \frac{1}{a} e^{-\frac{x^2}{2}} ((x-b)^2 + a^2) \right\} = \\ &= \min \left( \min_{|a| > \sqrt{2}} \max_{x \in \mathbb{R}} \left\{ \sqrt{\frac{\pi}{2}} \cdot \frac{1}{a} e^{-\frac{x^2}{2}} ((x-b)^2 + a^2) \right\}, \min_{|a| \leq \sqrt{2}} \max_{x \in \mathbb{R}} \left\{ \sqrt{\frac{\pi}{2}} \cdot \frac{1}{a} e^{-\frac{x^2}{2}} ((x-b)^2 + a^2) \right\} \right) \geq \\ &\geq \min \left( \min_{|a| > \sqrt{2}} \left\{ \sqrt{\frac{\pi}{2}} \cdot \frac{1}{a} e^{-\frac{x^2}{2}} ((x-b)^2 + a^2) \right\} \Big|_{x=0}, \min_{|a| \leq \sqrt{2}} \left\{ \sqrt{\frac{\pi}{2}} \cdot \frac{1}{a} e^{-\frac{x^2}{2}} ((x-b)^2 + a^2) \right\} \Big|_{x=\pm\sqrt{2-a^2}} \right) = \\ &= \min \left( \min_{|a| > \sqrt{2}} \left\{ \sqrt{\frac{\pi}{2}} \cdot \frac{1}{a} (b^2 + a^2) \right\}, \min_{|a| \leq \sqrt{2}} \left\{ \sqrt{\frac{\pi}{2}} \cdot \frac{1}{a} e^{\frac{a^2-2}{2}} ((|b| + \sqrt{2-a^2})^2 + a^2) \right\} \right) > \\ &> \min \left( \min_{|a| > \sqrt{2}} \left\{ \sqrt{\frac{\pi}{2}} \cdot \frac{1}{a} a^2 \right\}, \min_{|a| \leq \sqrt{2}} \left\{ \sqrt{\frac{\pi}{2}} \cdot \frac{1}{a} e^{\frac{a^2-2}{2}} ((\sqrt{2-a^2})^2 + a^2) \right\} \right) = k_0^* \end{aligned}$$

Отсюда получаем неравенство  $k^*(b) > k_0^*, \forall b \neq 0$ . Поэтому при поиске параметров распределения Коши, для того чтобы добиться максимальной эффективности метода Фон-Неймана, надо взять  $b = 0$ .

Итак,  $b = 0$ . Максимум функции  $g(x)$  достигается при  $x^* = \pm\sqrt{2-a^2}$ , если  $|a| \leq \sqrt{2}$ , и при  $x = 0$ , если  $|a| > \sqrt{2}$ . При  $x = 0$  получаем, что  $k^* = \sqrt{\pi}$ . Теперь подставим значение  $x^* = \pm\sqrt{2-a^2}$  в  $g(x)$ , получим  $g(x^*) = w(a)$ , где:

$$w(a) = \frac{\sqrt{2\pi}}{e} \frac{e^{\frac{a^2}{2}}}{a}$$

Минимум функции  $w(a)$  достигается при  $a = 1$ . Тогда получаем, что максимальная эффективность метода Фон-Неймана достигается при использовании распределения Коши с параметром сдвига 0 и параметром масштаба 1. В итоге коэффициент  $k^* = \min \left\{ \frac{\sqrt{2\pi}}{\sqrt{e}}, \sqrt{\pi} \right\} = \frac{\sqrt{2\pi}}{\sqrt{e}}$

Для моделирования стандартного нормального распределения методом Фон-Неймана используем следующий алгоритм:

1. Генерируем независимые пары  $(X_i, \nu_i)$ , где случайные величины  $X_i$  имеют распределение Коши с параметрами  $a = 1$ ,  $b = 0$ , а случайные величины  $\nu_i$  имеют распределение Бернулли с вероятностью успеха:

$$\mathbb{P}(\nu_i = 1, X_i = x) = \frac{1}{k} \sqrt{\frac{\pi}{2}} e^{-\frac{x^2}{2}} (x^2 + 1) = \frac{\sqrt{e}}{2} e^{-\frac{x^2}{2}} (x^2 + 1).$$

2. Выбираем только те пары, где  $\nu_i = 1$ . Соответствующая последовательность  $X_i$  будет стандартное нормальное распределение.

Действительно:

$$\begin{aligned} \mathbb{P}(X \in A | \nu = 1) &= \frac{\mathbb{P}(X \in A, \nu = 1)}{\mathbb{P}(\nu = 1)} = \frac{\int_A \mathbb{P}(\nu = 1, X = x) dQ(x)}{\int_X \mathbb{P}(\nu = 1, X = x) dQ(x)} = \\ &= \frac{\int_A \frac{1}{k} \frac{dP}{dQ} dQ(x)}{\int_X \frac{1}{k} \frac{dP}{dQ} dQ(x)} = P(A). \end{aligned}$$

## 9 Глава IX. Дискретные распределения. Смеси распределений. Сингулярные распределения. Полная классификация распределений. Примеры наиболее часто употребляемых на практике распределений.

### 9.1 Дискретные распределения.

**Определение 82.** Дискретным распределением называется мера, сосредоточенная не более чем на счетном множестве, т.е. не более чем на счетном числе точек.

Рассмотрим вероятностное пространство  $(\Omega, \mathcal{A}, \mathbb{P})$ . Вероятностная мера  $\mathbb{P}$  может быть дискретной, если она задана не более, чем на счетном  $\Omega$ , т.е. это означает, что должно существовать не более чем счетное множество  $\Omega' = \{\omega_i\}$ ,  $\omega_i \in \Omega$ , причем  $\forall A \in \mathcal{A} : \mathbb{P}(A) = \mathbb{P}(A \cap \Omega')$ , т.е. меру  $\mathbb{P}$  мы задаем только на множестве  $\Omega'$ , т.е. сужаем меру на множество  $\Omega'$ , таким образом получаем дискретную вероятностную меру  $\mathbb{P}$ .

Иначе говоря, если мы можем без потери сущности  $\mathbb{P}$  свести исходное вероятностное  $(\Omega, \mathcal{A}, \mathbb{P})$ , к дискретному вероятностному пространству, для которого по определению мощность  $\Omega$  не более чем счетна, то тогда  $\mathbb{P}$  – это дискретная вероятностная мера.

### 9.2 Смеси распределений.

Рассмотрим конечное или счетное число распределений  $Q_i$ , заданных на одном и том же пространстве  $(\Omega, \mathcal{A})$ , напомним, что распределение – это мера. Пусть, кроме того, задан набор неотрицательных чисел  $q_i \geq 0$ , таких что  $\sum_i q_i = 1$ .

**Определение 83.** Распределение  $Q = \sum_i q_i Q_i$ , определенное для  $\forall A \in \mathcal{A}$ ,  $Q(A) = \sum_i q_i Q_i(A)$ , называется **смесью распределений**  $Q_i$  с весами  $q_i$ .

**Пример** (Смесь дискретного и абсолютно непрерывного распределений.). Рассмотрим систему массового обслуживания. Очередь в этой системе будем характеризовать 2 параметрами:

1. есть очередь или нет
2. длина очереди

$\Rightarrow$  зная эти 2 параметра, можем определить время ожидания в очереди, это будет некая случайная величина.

Хотим задать распределение  $Q$  времени ожидания в очереди. В данном случае это распределение можно представить в виде смеси дискретного и абсолютно непрерывного распределений:  $Q = q\delta_a + (1 - q)Q'$ , где  $q$  – вероятность застать систему в стационарном состоянии (когда очереди нет),  $\Rightarrow 1 - q$  – вероятность того, что очередь есть. Если очереди нет, то не надо определять и длину очереди, если же очередь есть, то надо определить длину очереди для нахождения времени ожидания в очереди.  $\delta_a$  – это мера Дирака, сосредоточенная в точке, которая определяется как:

$$\delta_a = \begin{cases} 1, & a \in A \\ 0, & a \notin A \end{cases}$$

Пусть  $a \in A$  означает, что система находится в стационарном состоянии (т.е. очереди нет).  $Q'$  – это некоторое абсолютно непрерывное распределение, задающее длину очереди.

На практике часто используют смеси нормальных распределений. Практически любое абсолютно непрерывное распределение можно приблизить смесью нормальных распределений. Так часто поступают при оценке неизвестных распределений, приближают его смесью из 10 нормальных распределений.

### 9.3 Сингулярные распределения.

**Определение 84.** Распределение называется **сингулярным**, если оно сосредоточено на континуальном множестве с нулевой мерой Лебега.

Примером множества в пространстве  $\mathbb{R}^1$ , на котором сосредоточено сингулярное распределение, является канторово множество.

Напомним процесс построения канторова множества.

При построении Канторова множества  $F$  на отрезке  $[0, 1]$  мы выбрасываем из него интервалы  $(\frac{1}{3}, \frac{2}{3}), (\frac{1}{9}, \frac{2}{9}), (\frac{7}{9}, \frac{8}{9}), \dots$ . В итоге получаем замкнутое множество  $F$  (как пересечение замкнутых). Оно получается из отрезка  $[0, 1]$  выбрасыванием счетного числа интервалов. Рассмотрим структуру множества  $F$ . Ему принадлежат точки:

$$0, 1, \frac{1}{3}, \frac{2}{3}, \frac{1}{9}, \frac{2}{9}, \frac{7}{9}, \frac{8}{9}, \dots$$

- концы выбрасываемых интервалов. Однако множество  $F$  не исчерпывается этими точками. Те точки отрезка  $[0, 1]$ , которые входят в множество  $F$ , можно охарактеризовать следующим образом. Запишем каждое из чисел  $x$ ,  $0 \leq x \leq 1$  в троичной системе:

$$x = \frac{a_1}{3} + \frac{a_2}{3^2} + \dots + \frac{a_n}{3^n} + \dots$$

где числа  $a_n$  могут принимать значения 0, 1 и 2. При этом, как и в случае десятичных дробей, некоторые числа допускают двоякую запись. Например,

$$\frac{1}{3} = \frac{1}{3} + \frac{0}{3^2} + \dots + \frac{0}{3^n} + \dots = \frac{0}{3} + \frac{2}{3^2} + \frac{2}{3^3} + \dots + \frac{2}{3^n} + \dots$$

Множеству  $F$  принадлежат только те числа  $x$ ,  $0 \leq x \leq 1$ , которые могут быть записаны хотя бы одним способом в виде троичной дроби так, чтобы в последовательности  $a_1, a_2, \dots, a_n, \dots$  единица ни разу не встречалась. Таким образом, каждой точке  $x \in F$  можно поставить в соответствие последовательность

$$a_1, a_2, \dots, a_n, \dots,$$

где  $a_n$  равно 0 или 2. Совокупность таких последовательностей образует множество мощности континуума.

Таким образом, мы получили отображение множества  $F$  на весь отрезок  $[0,1] \Rightarrow F$  имеет мощность континуума, т.е. содержит столько же точек, сколько и весь отрезок  $[0,1]$ .

Лебеговская мера канторова множества равна 0, поскольку при построении канторова множества  $F$  на отрезке  $[0,1]$  мы выбрасываем из него интервалы  $(\frac{1}{3}, \frac{2}{3}), (\frac{1}{9}, \frac{2}{9}), (\frac{7}{9}, \frac{8}{9}), \dots$ , общая длина которых равна 1.

**Определение 85.** Меры  $\mu$  и  $\nu$  называются **взаимносингулярными**, если существуют два непересекающихся множества  $A_1, A_2$ ,  $A_1 \cap A_2 = \emptyset$ , такие что меры  $\mu$  и  $\nu$  сосредоточены на соответствующих множествах.

Сингулярность мер  $\mu$  и  $\nu$  будем обозначать символом:  $\mu \perp \nu$ .

Используя определение сингулярности мер, можно по-другому сформулировать определение сингулярного распределения.

**Определение 86.** Распределение называется **сингулярным**, если оно сингулярно к смесям дискретных и абсолютно непрерывных распределений.

**Пример.** Пусть  $\nu$  – дискретная мера,  $\mu$  – мера Лебега, тогда  $\mu \perp \nu$ , поскольку эти меры сосредоточены на непересекающихся множествах, дискретная мера сосредоточена не более чем на счетном множестве точек, т.е. на множестве нулевой меры Лебега, а мера Лебега (это абсолютно непрерывная мера) сосредоточена на множестве ненулевой меры Лебега.

**Упражнение.** Распределение  $\mu$  сингулярно распределению  $\nu$ :

$$\mu \perp \nu \iff \|\mu - \nu\| = 2$$

## 9.4 Полная классификация распределений.

1. Дискретное распределение
2. Абсолютно непрерывное распределение, определение абсолютно непрерывного распределения дано в главе 8.
3. Сингулярное распределение

Любое распределение  $Q$  представимо в виде смеси трех видов распределений:

$$Q = \alpha_1 Q_1 + \alpha_2 Q_2 + \alpha_3 Q_3$$

где  $Q_1$  – дискретное распределение,  $Q_2$  – абсолютно непрерывное распределение,  $Q_3$  – сингулярное распределение, все  $\alpha_i \geq 0$  и  $\alpha_1 + \alpha_2 + \alpha_3 = 1$ .

Так как распределение – это мера, то данное представление следует из того, что любая мера представима как сумма абсолютно непрерывной, дискретной и сингулярной мер.



## 9.5 Примеры наиболее часто употребляемых на практике распределений.

Рассмотрим сначала примеры *дискретных распределений*.

### 1. Распределение Бернулли.

Схема Бернулли описана в первой главе.

### 2. Биномиальное распределение.

Это распределение возникает при реализации схемы Бернулли из  $n$  испытаний с вероятностью успеха  $p$ . Биномиальное распределение – это распределение числа появления события, имеющего вероятность  $p$  в  $n$  независимых испытаниях (т.е. распределение количества успехов в схеме Бернулли). Тогда:

$$\mathbb{P}_X(A) = \sum_{k=0}^n I_A(k) C_n^k p^k (1-p)^{n-k}, \quad 0 < p < 1$$

$$I_A(k) = \begin{cases} 1, & k \in A \\ 0, & k \notin A \end{cases}$$

### 3. Отрицательное биномиальное распределение.

Это распределение также возникает при реализации схемы Бернулли из  $n$  испытаний с вероятностью успеха  $p$ . Отрицательное биномиальное распределение – это распределение числа неудач, произошедших до  $n$ -ого успеха. Тогда:

$$\mathbb{P}_X(A) = \sum_{k=0}^n I_A(k) C_{k+n-1}^k p^n (1-p)^k, \quad 0 < p < 1$$

### 4. Геометрическое распределение.

Это распределение также возникает при реализации схемы Бернулли из  $n$  испытаний с вероятностью успеха  $p$ . Предположим, что испытания продолжаются до тех пор, пока не произойдет событие  $A$ . Вероятность того, что событие  $A$  впервые произойдет во время  $n$ -ого испытания, равна  $(1-p)^{n-1}p$ . Тогда:

$$\mathbb{P}_X(A) = \sum_{k=1}^{\infty} I_A(k) p (1-p)^{k-1}, \quad 0 < p < 1$$

### 5. Распределение Пуассона.

Рассмотрим последовательность серий из  $1, 2, \dots, n$  испытаний Бернулли, причем вероятность успеха в каждой серии различна и равна  $p = p(n)$ , т.е. считаем, что  $p$  является функцией от  $n$ . Пусть  $P_n(k)$  – вероятность появления  $k$  успехов в серии из  $n$  испытаний,  $P_n(k) = C_n^k p^k q^{n-k}$ . Оказывается, что при малых значениях  $p$  хорошую аппроксимацию для  $\{P_n(k)\}$  дает пуассоновское распределение вероятностей. В этом случае верна теорема Пуассона.

**Теорема 19** (Пуассона). Пусть  $p(n) \rightarrow 0$ ,  $n \rightarrow \infty$ , причем так, что  $np(n) \rightarrow \lambda$ , где  $\lambda > 0$ . Тогда для любого  $k = 0, 1, \dots$

$$P_n(k) \rightarrow \pi_k, \quad n \rightarrow \infty, \tag{1}$$

где

$$\pi_k = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \dots \quad (2)$$

Запишем распределение вероятности в уже привычной форме:

$$\mathbb{P}_X(A) = \sum_{k=0}^{\infty} I_A(k) \frac{\lambda^k e^{-\lambda}}{k!}, \quad 0 < p < 1$$

Теперь рассмотрим примеры *непрерывных распределений*.

1. *Равномерное распределение.*

Равномерное распределение на множестве  $C \in \mathcal{B}_{\mathbb{R}}$ , где  $\mathcal{B}_{\mathbb{R}}$  – борелевская  $\sigma$ -алгебра, – это распределение с плотностью  $I_C(y)(m(C))^{-1}$ , где  $m$  – мера Лебега. Равномерное распределение имеет вид  $P_X(A) = (m(C))^{-1}m(C \cap A)$ , где  $m(C) < \infty$ . Если  $C = [a, b]$ , то получим равномерное распределение на  $[a, b]$  с плотностью  $I_{[a,b]}(x)(b-a)^{-1}$ .

2. *Нормальное распределение.*

Нормальное распределение очень хорошо описано в главе 7.

3. *Экспоненциальное распределение.*

Экспоненциальное распределение характеризуется свойством отсутствия памяти (единственное непрерывное распределение, обладающее этим свойством).

Пусть:

$$G(x) = \mathbb{P}(T \geq x).$$

Свойство отсутствия памяти можно записать следующим образом:

$$\begin{aligned} \mathbb{P}(T \geq x+y | T \geq x) &= \mathbb{P}(T \geq y) > 0 \quad \forall x, y > 0 \Rightarrow \\ \Rightarrow \frac{G(x+y)}{G(x)} &= G(y) \Rightarrow G(x+y) = G(x)G(y) \Rightarrow \\ \Rightarrow \ln G(x+y) &= \ln G(x) + \ln G(y) \end{aligned}$$

Рассмотрим функциональное уравнение:  $G(x+y) = G(x)G(y)$ , т.к.  $G(x) = \mathbb{P}(T \geq x)$ , функция  $G(x)$  – абсолютно непрерывная невозрастающая функция (т.к. в качестве  $T$  подразумевается время, а распределение  $T$  абсолютно непрерывное) и нас интересуют те решения данного уравнения, которые удовлетворяют условиям:

$$0 \leq G(x) \leq 1, \quad G(0) = 1$$

Тогда получим, что данное уравнение будет иметь единственное решение:

$$G(x) = e^{-\lambda x}, \quad \lambda > 0$$

Отсюда получаем, что функция экспоненциального распределения имеет вид (следует из определения функции распределения):

$$F(x) = 1 - e^{-\lambda x}, \quad \lambda > 0$$

Таким образом, получили функцию экспоненциального распределения.

Свойством отсутствия памяти обладает еще геометрическое распределение (единственное в классе дискретных распределений).

Геометрическое распределение в пределе дает экспоненциальное распределение.

## 10 Совместные и маргинальные распределения. Согласованность семейства распределений. Теорема Колмогорова.

### 10.1 Совместные и маргинальные распределения.

Пусть имеются два случайных элемента:  $X$ , принимающий значения в измеримом пространстве  $(\mathcal{X}, \mathcal{A})$ , и  $Y$ , принимающий значения в измеримом пространстве  $(\mathcal{Y}, \mathcal{B})$ <sup>101</sup>. Рассмотрим декартово произведение  $\mathcal{X} \times \mathcal{Y}$  и некоторую  $\sigma$ -алгебру  $\mathcal{C}$  на нем. Пусть на измеримом пространстве  $(\mathcal{X} \times \mathcal{Y}, \mathcal{C})$  задано распределение, т.е. мера  $\mathbb{P}_{XY}$ , которую мы будем интерпретировать как совместное распределение случайных элементов  $X$  и  $Y$ .

Мы будем рассматривать случайные элементы в рамках канонической модели, однако будем помнить, что каноническую модель можно свести к общей, т.е. указать вероятностное пространство  $(\Omega, \mathcal{F}, \mathbb{P})$ , такое что  $X$  и  $Y$  можно рассматривать как отображения

$$X: \Omega \rightarrow \mathcal{X}, \quad Y: \Omega \rightarrow \mathcal{Y}.$$

Тогда совместное распределение  $X$  и  $Y$  определяется следующим образом:

$$\mathbb{P}_{XY}(A, B) = \mathbb{P}(X \in A, Y \in B) \quad \forall A, B: A \times B \in \mathcal{C}.$$

Фактически получили, что совместное распределение случайных величин  $X$  и  $Y$  — это распределение случайного вектора  $(X, Y): \Omega \rightarrow \mathcal{X} \times \mathcal{Y}$ .

Итак, известно совместное распределение  $\mathbb{P}_{XY}$ , и мы хотим по нему узнать маргинальные распределения  $\mathbb{P}_X$  и  $\mathbb{P}_Y$ . Посмотрим, каким свойствам должны удовлетворять  $\sigma$ -алгебра  $\mathcal{C}$  и распределение  $\mathbb{P}_{XY}$ , чтобы можно было корректно определить маргинальные распределения.

Имея ввиду общую модель задания случайных элементов, запишем:

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X \in A, Y \in \mathcal{Y}) = \mathbb{P}_{XY}(A, \mathcal{Y}).$$

Итак, при наличии совместного распределения маргинальное распределение определяется по формуле

$$\mathbb{P}_X(A) = \mathbb{P}_{XY}(A \times \mathcal{Y}). \quad (44)$$

Таким образом, для того чтобы можно было определить маргинальные распределения, необходимо

$$A \times \mathcal{Y} \in \mathcal{C}, \quad \forall A \in \mathcal{A}. \quad (45)$$

Множества вида  $A \times \mathcal{Y}$  называются *цилиндрическими*.

Сформулируем свойство (45) в эквивалентной форме. Для этого рассмотрим функции — *проекторы*:

$$\begin{aligned} \pi_X: \mathcal{X} \times \mathcal{Y} &\rightarrow \mathcal{X}, & \pi_X(\omega, \theta) &= \omega \\ \pi_Y: \mathcal{X} \times \mathcal{Y} &\rightarrow \mathcal{Y}, & \pi_Y(\omega, \theta) &= \theta \end{aligned}$$

для любых  $\omega \in \mathcal{X}$ ,  $\theta \in \mathcal{Y}$ .

Поскольку

$$A \times \mathcal{Y} = \pi_X^{-1}(A),$$

<sup>101</sup>Напомним, что по определению измеримого пространства  $\mathcal{A}$  и  $\mathcal{B}$  —  $\sigma$ -алгебры соответствующих подмножеств.

то условие (45) можно переформулировать так: для того, чтобы можно было определить маргинальные распределения, проектор  $\pi_X$  должен быть  $(\mathcal{C}, \mathcal{A})$ -измеримым, а проектор  $\pi_Y$  —  $(\mathcal{C}, \mathcal{B})$ -измеримым.

**Определение 87.** Пусть заданы два измеримых пространства  $(\mathcal{X}, \mathcal{A})$ ,  $(\mathcal{Y}, \mathcal{B})$ . Минимальная  $\sigma$ -алгебра  $\mathcal{C}$  на декартовом произведении  $\mathcal{X} \times \mathcal{Y}$ , относительно которой измеримы проекторы на пространства-сомножители декартова произведения, т.е. проектор  $\pi_X$   $(\mathcal{C}, \mathcal{A})$ -измерим, а проектор  $\pi_Y$   $(\mathcal{C}, \mathcal{B})$ -измерим, называется **прямым произведением  $\sigma$ -алгебр  $\mathcal{A}$  и  $\mathcal{B}$**  и обозначается

$$\mathcal{C} = \mathcal{A} \otimes \mathcal{B}.$$

Легко видеть, что  $\mathcal{C} = \{(A, B) | A \in \mathcal{A}, B \in \mathcal{B}\}$ .

Итак, в дальнейшем, чтобы можно было определить маргинальные распределения, мы будем считать, что совместное распределение задано на прямом произведении  $\sigma$ -алгебр. (Прямое произведение  $\sigma$ -алгебр —  $\sigma$ -алгебра, обеспечивающая минимальную информацию, необходимую для работы с маргинальными распределениями.)

Заметим, что маргинальных распределений недостаточно для построения совместного распределения, необходимо знать еще корреляцию случайных величин.

**Пример.** Пусть  $X$  и  $Y$  имеют маргинальные равномерные распределения на отрезке  $[0, 1]$ :

$$\mathbb{P}_X \sim U[0, 1], \quad \mathbb{P}_Y \sim U[0, 1].$$

Тогда их совместное распределение может быть как равномерным на единичном квадрате, если случайные величины являются независимыми, так и равномерным на диагонали этого квадрата, если они равны (см. рис. 8).



Рис. 8: Распределение  $\mathbb{P}_{XY}$ .

**Замечание.** Мы рассмотрели случай совместного распределения двух случайных величин. Легко распространить эту теорию на случай  $n$  случайных элементов.

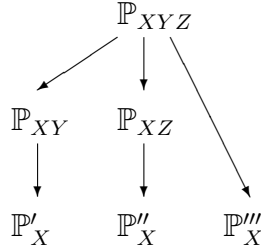
**Определение 88.** Пусть  $X = (X_1, X_2, \dots, X_n)$ ,  $X' = (X_{k_1}, X_{k_2}, \dots, X_{k_m})$ ,  $1 \leq m < n$ ,  $\{k_1, k_2, \dots, k_m\} \subset \{1, 2, \dots, n\}$ . Тогда распределение  $\mathbb{P}_{X'}$  называется **маргинальным** по отношению к  $\mathbb{P}_X$ , которое называется **совместным**.

Общее количество маргинальных распределений в этом случае равно  $C_n^1 + C_n^2 + \dots + C_n^{n-1} = 2^n - 2$ .

Условия, выполнение которых необходимо для построения маргинальных распределений по совместному, выписываются аналогичным образом.

## 10.2 Согласованность семейства распределений.

Итак, если имеется совместное распределение, то можно задать маргинальные распределения по формуле (44). Однако необходимо, чтобы такое задание было непротиворечивым. Поясним эту мысль на примере. Пусть задано совместное распределение  $\mathbb{P}_{XYZ}$ . Тогда маргинальное распределение  $\mathbb{P}_X$  можно найти несколькими способами:



Очевидно, что совместное распределение  $\mathbb{P}_{XYZ}$  должно быть таким, чтобы распределения  $\mathbb{P}'_X$ ,  $\mathbb{P}''_X$  и  $\mathbb{P}'''_X$ , полученные различными способами, совпадали. Таким образом, мы приходим к свойству совместного распределения, которое называется свойством согласованности. Более строго оно формулируется следующим образом.

Пусть  $\{X_j\}$  — семейство случайных элементов,  $j \in J$ .  $J$  — множество индексов произвольной мощности,  $P_J$  — совместное распределение случайных элементов.  $I_1, I_2, I_3$  — произвольные подмножества  $J$ , такие что

$$I_1 \subseteq I_2 \subseteq J, \quad I_1 \subseteq I_3 \subseteq J.$$

Обозначим  $\pi_{I_2}^{I_1} \mathbb{P}_{I_2}$  взятие маргинального распределения случайных элементов с индексами из  $I_1$  из совместного распределения случайных элементов с индексами из  $I_2$ . Тогда **условие согласованности маргинальных распределений** записывается следующим образом:

$$\pi_{I_3}^{I_1} \mathbb{P}_{I_3} = \pi_{I_2}^{I_1} \mathbb{P}_{I_2} = \mathbb{P}_{I_1}. \quad (46)$$

Заметим, что если  $\mathbb{P}_J$  в действительности является совместным распределением некоторых реальных случайных элементов, то свойство согласованности для его маргинальных распределений всегда выполнено.

Теперь зададимся обратным вопросом: если имеются всевозможные маргинальные распределения, удовлетворяющие свойству согласованности, можно ли по ним построить совместное распределение? Т.е. если, например, заданы  $\mathbb{P}_X, \mathbb{P}_Y, \mathbb{P}_Z, \mathbb{P}_{XY}, \mathbb{P}_{YZ}, \mathbb{P}_{XZ}$ , удовлетворяющие свойству согласованности, можно ли по ним построить  $\mathbb{P}_{XYZ}$ ? В общем случае ответ на этот вопрос отрицательный. Приведем пример.

**Пример.** Пусть  $\mathbb{P}_X = \mathbb{P}_Y = \mathbb{P}_Z \sim U[0, 1]$ , распределения  $\mathbb{P}_{XY}, \mathbb{P}_{YZ}, \mathbb{P}_{XZ}$  являются равномерными на диагоналях квадрата, как показано на рисунке 9.

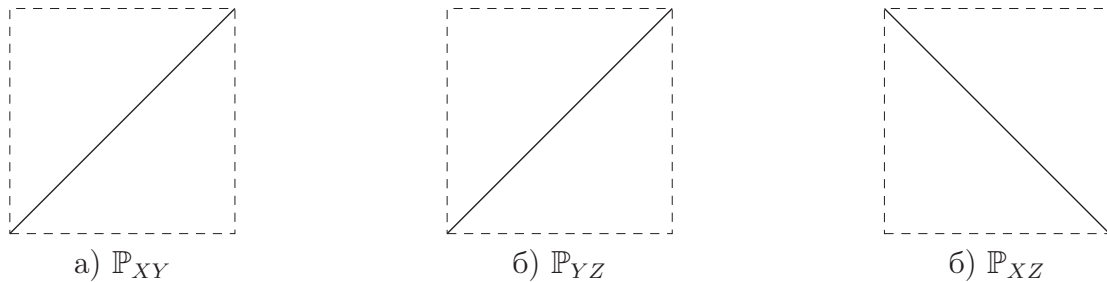


Рис. 9: Маргинальные распределения.

Очевидно, свойство согласованности выполнено. Из рисунка а) следует, что  $X = Y$ , из рисунка б) следует, что  $Y = Z$ , а из рисунка в) — что  $X = 1 - Z$ . Значит, совместного распределения  $\mathbb{P}_{XYZ}$ , обладающего такими маргинальными не существует.

Таким образом, свойство согласованности является необходимым, но не достаточным условием того, что семейство распределений, удовлетворяющее этому свойству, является маргинальными распределениями некоторого совместного распределения.

Однако в случае когда число случайных величин в рассматриваемом семействе бесконечно, условие является и достаточным, т.е. по набору маргинальных распределений, удовлетворяющих условию согласованности, можно восстановить совместное распределение, причем однозначно. Это утверждение составляет содержание теоремы Колмогорова. Оставшаяся часть главы будет посвящена формулировке этой теоремы в общем случае и изложению идеи доказательства.

### 10.3 Теорема Колмогорова.

Пусть  $\{X_j\}$  — семейство случайных элементов,  $j \in J$ .  $J$  — множество индексов произвольной мощности. Каждый случайный элемент  $X_j$  принимает значения из измеримого пространства  $(\mathcal{X}_j, \mathcal{A}_j)$  (будем рассматривать польские пространства).

Сначала определим пространство, на котором будет рассматриваться совместное распределение. Им будет декартово произведение

$$\bigotimes_{j \in J} \mathcal{X}_j. \quad (47)$$

В случае, когда множество  $J$  было конечно:  $|J| = n$ , декартово произведение (47) определялось как множество всевозможных наборов  $(x_1, x_2, \dots, x_n)$ ,  $x_j \in \mathcal{X}_j$ . В случае, когда множество  $J$  счетно, декартово произведению (47) можно поставить в соответствие множество всевозможных бесконечных последовательностей. В случае, когда  $J$  имеет мощность больше чем счетную, такое определение давать нельзя. Поэтому объект (47) мы будем интерпретировать как множество функций, которые каждому числу  $j$  из множества  $J$  ставят в соответствие элемент  $x_j$  из множества  $\mathcal{X}_j$  (множество значений функции зависит от аргумента!):

$$J \rightarrow x \in \bigcup_{j \in J} \mathcal{X}_j, \quad \text{причем } j \rightarrow x_j \in \mathcal{X}_j.$$

Таким образом,  $\bigotimes_{j \in J} \mathcal{X}_j$  — это множество всевозможных функций, отображающих индекс в компоненту, находящуюся по этому индексу.

Очевидно, утверждение о непустоте определяемого таким образом множества эквивалентно аксиоме выбора: принимаем аксиому выбора  $\implies$  если  $\mathcal{X}_j \neq \emptyset$ , то  $\bigotimes_{j \in J} \mathcal{X}_j \neq \emptyset$ , т.е.

$\forall j \in J$  существует функция выбора  $j \rightarrow x_j \in \mathcal{X}_j$ .

На множестве  $\bigotimes_{j \in J} \mathcal{X}_j$  рассмотрим  $\sigma$ -алгебру подмножеств  $\mathcal{A}$  — минимальную  $\sigma$ -алгебру подмножеств  $\bigotimes_{j \in J} \mathcal{X}_j$ , относительно которой измеримы проекторы  $\pi_J^{\{j\}}: \bigotimes_{j \in J} \mathcal{X}_j \rightarrow \mathcal{X}_j$ :

$$\sigma(\pi_J^{\{j\}}) = \bigotimes_{j \in J} \mathcal{A}_j = \mathcal{A}.$$

Такая  $\sigma$ -алгебра существует, т.к. существует, по крайней мере, одна  $\sigma$ -алгебра, относительно которой измеримы рассматриваемые проекторы: множество всех подмножеств  $\bigotimes_{j \in J} \mathcal{X}_j$ .

Итак, мы имеем измеримое пространство  $(\bigotimes_{j \in J} \mathcal{X}_j, \bigotimes_{j \in J} \mathcal{A}_j)$ , на котором мы будем строить меру.

**Утверждение 11.** *Из того, что относительно  $\mathcal{A}$  измеримы проекторы  $\pi_J^{\{j\}}$ , следует, что относительно  $\mathcal{A}$  измеримы проекторы  $\pi_J^I$ , где  $I \subset J$ ,  $I$  конечно или счетно.*

**Доказательство.**

1. Измеримость  $\pi_J^{\{j\}}$

$$\forall A \in \mathcal{A}_j \quad (\pi_J^{\{j\}})^{-1}(A) \in \mathcal{A}$$

следует из построения.

2. Рассмотрим случай конечного множества индексов  $I$ . Пусть без ограничения общности  $I = \{1, 2, \dots, n\}$ . Любое множество  $A \in \bigotimes_{i \in I} \mathcal{A}_i$  можно представить в виде  $A = A_1 \times A_2 \times \dots \times A_n$ , где  $A_i \in \mathcal{A}_i$ . Это следует из того, что прямое произведение  $\sigma$ -алгебр есть множество, полученное из семейства цилиндрических множеств путем замыкания относительно операций пересечения и симметрической разности.

$$\begin{aligned} (\pi_J^I)^{-1}(A) &= \{x \in \bigotimes_{j \in J} \mathcal{X}_j \mid \pi_J^I(x) \in A\} = \\ &= \{x \in \bigotimes_{j \in J} \mathcal{X}_j \mid \pi_I^{\{1\}} \circ \pi_J^I(x) \in A_1, \dots, \pi_I^{\{n\}} \circ \pi_J^I(x) \in A_n\} = \\ &= \{x \in \bigotimes_{j \in J} \mathcal{X}_j \mid \pi_J^{\{1\}}(x) \in A_1, \dots, \pi_J^{\{n\}}(x) \in A_n\} = \\ &= \bigcap_{i=1}^n \{x \in \bigotimes_{j \in J} \mathcal{X}_j \mid \pi_J^{\{i\}}(x) \in A_i\} = \bigcap_{i=1}^n (\pi_J^{\{i\}})^{-1}(A_i) \in \mathcal{A}. \end{aligned}$$

3. Аналогично проводится доказательство для случая, когда  $I$  счетное. (Учитываем, что  $\sigma$ -алгебра замкнута относительно счетного объединения и дополнения, а значит, замкнута относительно счетного пересечения.)  $\square$

Итак, мы задали измеримое пространство, на котором будем строить совместное распределение  $\mathbb{P}_J$ , имеющее заданные маргинальные. Задано семейство маргинальных распределений  $\{\mathbb{P}_I\}$ , где  $I$  — всевозможные конечные подмножества  $J$ :

$$I \subset J, \quad |I| < \infty.$$

Все распределения  $\mathbb{P}_I$  удовлетворяют свойству согласованности (46).

**Теорема 20** (Колмогоров). *Если выполнены сформулированные выше условия, то на  $(\bigotimes_{j \in J} \mathcal{X}_j, \bigotimes_{j \in J} \mathcal{A}_j)$ , где  $J$  — бесконечное множество индексов, можно задать меру  $\mathbb{P}_J$ , причем единственным образом, такую что ее маргинальные распределения будут совпадать с заданными.*

*Другими словами, бесконечномерное распределение можно однозначно восстановить по совокупности всевозможных конечномерных маргинальных распределений.*

Идея доказательства<sup>102</sup> теоремы совпадает с идеей построения распределения по функции распределения, проведенного в главе 4. Напомним, что там мы сначала естественным образом определяли меру на алгебре, порожденной элементарными множествами вида  $(-\infty, x)$ , а затем по теореме Каратеодори продолжали ее на  $\sigma$ -алгебру.

В данном случае в качестве элементарных множеств будем рассматривать цилиндрические.

<sup>102</sup>полное доказательство можно найти в книге Ширяева «Вероятность»



**Определение 89.** Пусть  $A \in \bigotimes_{j \in I} \mathcal{A}_j$ ,  $|I| < \infty$ . Тогда множество

$$(\pi_J^I)^{-1}(A) \in \bigotimes_{j \in J} \mathcal{A}_j$$

будем называть **цилиндрическим**.

На цилиндрических множествах естественно определить меру следующим образом:

$$C = (\pi_J^I)^{-1}(A) \Rightarrow \mathbb{P}_J(C) = \mathbb{P}_I(A).$$

Так же, как и при рассмотрении функции распределения, на этом этапе возникают два вопроса.

Во-первых, проблема неединственности представления множества  $C$  в виде цилиндрического множества. Корректность задания меры гарантируется условием согласованности. Действительно, пусть  $I \subset I'$ ,  $|I'| < \infty$ ,  $A \in \bigotimes_{j \in I} \mathcal{A}_j$ ,  $C = (\pi_J^I)^{-1}(A) = (\pi_J^{I'})^{-1}((\pi_{I'}^I)^{-1}A)$ . Получаем, с одной стороны  $\mathbb{P}_J(C) = \mathbb{P}_I(A)$ , а с другой —  $\mathbb{P}_J(C) = \mathbb{P}_{I'}((\pi_{I'}^I)^{-1}A)$  (заметим, что распределения  $\mathbb{P}_I$  и  $\mathbb{P}_{I'}$  заданы, т.к. по условию заданы *всевозможные* конечномерные распределения). В силу согласованности маргинальных распределений  $\pi_{I'}^I \mathbb{P}_{I'} = \mathbb{P}_I$ . Следовательно,  $\mathbb{P}_I(A) = \mathbb{P}_{I'}((\pi_{I'}^I)^{-1}A)$ .

Можно показать, что заданная таким образом мера является непрерывной.

Кроме того, т.к. семейство цилиндрических множеств является алгеброй<sup>103</sup> (операция взятия прообраза сохраняет теоретико-множественные операции), по теореме Каратеодори заданная мера однозначно продолжается на  $\sigma$ -алгебру, порожденную семейством цилиндрических множеств, которая по построению равна  $\bigotimes_{j \in J} \mathcal{A}_j$ .

Заметим, что построенная таким образом мера является вероятностной, т.к.  $\mathbb{P}_J \left( \bigotimes_{j \in J} \mathcal{X}_j \right) = \mathbb{P}_i(\mathcal{X}_i)$ .

**Замечание.** Бесконечность множества  $J$  существенным образом используется при доказательстве возможности корректного построения меры на семействе цилиндрических множеств. Поясним это на примере. Пусть  $A = A_1 \times \mathcal{X}_2 \times \mathcal{X}_3$ . Тогда цилиндр  $C$  с основанием  $A$  можно представить двумя способами:

$$(\pi_J^{\{1,2\}})^{-1}(A_1 \times \mathcal{X}_2) = (\pi_J^{\{1,3\}})^{-1}(A_1 \times \mathcal{X}_3).$$

Соответственно необходимо, чтобы

$$\mathbb{P}_{\{1,2\}}(A_1, \mathcal{X}_2) = \mathbb{P}_{\{1,3\}}(A_1, \mathcal{X}_3) \quad (48)$$

В случае бесконечного  $J$  распределения  $\mathbb{P}_{\{1,2\}}$  и  $\mathbb{P}_{\{1,3\}}$  в силу условия согласованности являлись маргинальными распределениями одного и того же заданного совместного распределения  $\mathbb{P}_{\{1,2,3\}}$ . Поскольку  $A_1 \times \mathcal{X}_2$  и  $A_1 \times \mathcal{X}_3$  являются проекциями одного и того же множества, то выполнение (48) тем самым гарантируется. В случае же если, например,  $|J| = 3$ , совместного распределения  $\mathbb{P}_{\{1,2,3\}}$  у нас нет (собственно, его-то мы и хотим построить), поэтому, вообще говоря равенство (48) может быть нарушено (что и продемонстрировано в примере, предшествующем теореме Колмогорова).

<sup>103</sup>но не  $\sigma$ -алгеброй, т.к. при пересечении счетного числа цилиндрических множеств может получиться бесконечномерное основание цилиндра

**Пример.** Доказать существование последовательности независимых случайных величин.

Напомним, что последовательность случайных величин  $X_1, X_2, \dots, X_n$  называется независимой, если для любого конечного набора индексов  $I$  случайные величины  $\{X_i\}, i \in I$  являются независимыми в совокупности.

Вообще говоря, существование бесконечного числа независимых случайных величин не является очевидным. Понятно, что можно построить конечное число независимых случайных величин, однако, возможно, среди бесконечного числа случайных величин всегда найдется набор зависимых?

Выберем некоторые одномерные маргинальные распределения  $\mathbb{P}_{X_i}$ . Многомерные маргинальные распределения определим по формуле

$$\mathbb{P}_{X_{i_1} X_{i_2} \dots X_{i_n}}(A_1, A_2, \dots, A_n) = \mathbb{P}_{X_{i_1}}(A_1) \cdot \mathbb{P}_{X_{i_2}}(A_2) \cdot \dots \cdot \mathbb{P}_{X_{i_n}}(A_n).$$

Очевидно, условие согласованности будет выполнено. Значит, по теореме Колмогорова существует совместное распределение  $\mathbb{P}_J$  (а значит, существует и соответствующая последовательность случайных величин), маргинальные распределения которого совпадают с заданными. По построению полученная последовательность является независимой.

**Пример** (Гауссовские системы).

**Определение 90.** Семейство случайных величин  $\{X_t\}, t \in T$  ( $T$  — множество индексов произвольной мощности) называется **гауссовской системой**, если для любого конечного набора индексов  $t_1, t_2, \dots, t_n$  случайный вектор

$$(X_{t_1}, X_{t_2}, \dots, X_{t_n})$$

является гауссовским.

Если существует гауссовская система, то для нее определены функции математического ожидания и ковариации

$$m(t) = \mathbb{E}(X_t), \quad \text{cov}(t, s) = \text{cov}(X_t, X_s).$$

Пусть теперь, наоборот, произвольным образом задана некоторая функция  $m(t)$  и функция  $\text{cov}(t, s)$ , удовлетворяющая определению ковариационной матрицы.

**Определение 91.** Функция  $\text{cov}(t, s)$  называется ковариационной матрицей, если для любого конечного набора чисел  $t_1, t_2, \dots, t_n$  матрица

$$R(t_1, t_2, \dots, t_n) = \begin{pmatrix} \text{cov}(t_1, t_1) & \text{cov}(t_1, t_2) & \dots & \text{cov}(t_1, t_n) \\ \text{cov}(t_2, t_1) & \text{cov}(t_2, t_2) & \dots & \text{cov}(t_2, t_n) \\ \dots & \dots & \dots & \dots \\ \text{cov}(t_n, t_1) & \text{cov}(t_n, t_2) & \dots & \text{cov}(t_n, t_n) \end{pmatrix}$$

является симметричной и положительно определенной.

Тогда существует гауссовская система с заданными математическим ожиданием и ковариацией. Действительно, построим маргинальные распределения с плотностями

$$p_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x) = \frac{1}{\sqrt{(2\pi)^n |R|}} e^{-\frac{\langle R^{-1}(x-m), (x-m) \rangle}{2}},$$

где  $R = R(t_1, t_2, \dots, t_n)$ ,  $x = (x_1, x_2, \dots, x_n)$ ,  $m = (m(t_1), m(t_2), \dots, m(t_n))$ .

Легко проверяется, что заданные таким образом маргинальные распределения удовлетворяют свойству согласованности. Значит, по теореме Колмогорова существует гауссовская система с заданным математическим ожиданием и ковариацией.

Таким образом, теорема Колмогорова помогает доказать существование случайных процессов  $X_t, t \in [0, 1]$ . В данном случае было доказано существование гауссовского случайного процесса.

## 11 Условные распределения (существование регулярно-го варианта). Интегральные формулы типа полной вероятности. Метод последовательного моделирования случайного вектора. Независимость случайных величин.

### 11.1 Условные распределения (существование регулярно-го варианта). Интегральные формулы типа полной вероятности.

Пусть имеется вероятностное пространство  $(\Omega, \mathcal{F}, \mathbb{P})$ , на нем заданы две случайные величины:  $X$ , принимающая значения в измеримом пространстве  $(\mathcal{X}, \mathcal{A})$ , и  $Y$ , принимающая значения в измеримом пространстве  $(\mathcal{Y}, \mathcal{B})$ . По совместному распределению  $\mathbb{P}_{XY}$  можно определить маргинальные (см. 10 главу), оно полностью характеризует поведение случайного вектора  $(X, Y)$ . Известно, что маргинальные распределения  $\mathbb{P}_X$  и  $\mathbb{P}_Y$  не учитывают зависимость между случайными величинами. Однако пользоваться совместным распределением не всегда удобно.

Естественно было бы задать *условное распределение  $Y$  при условии  $X = x$*  по формуле

$$\mathbb{P}_{Y|X=x}(B) = \mathbb{P}(Y \in B \mid X = x) = \frac{\mathbb{P}(X = x, Y \in B)}{\mathbb{P}(X = x)} = \frac{\mathbb{P}_{XY}(\{x\}, B)}{\mathbb{P}_X(\{x\})}. \quad (49)$$

Однако величина  $\mathbb{P}_X(\{x\})$  может быть не определена (если  $\{x\} \notin \mathcal{A}$ ) или равняться нулю. Поэтому расширим данное понятие на более общий случай.

**Определение 92.** *Переходным ядром*<sup>104</sup> называется функция  $Q(x, B): (\mathcal{X}, \mathcal{B}) \rightarrow [0, 1]$  (по смыслу равная  $\mathbb{P}_{Y|X=x}(B)$ ), удовлетворяющая свойствам:

1. при каждом фиксированном  $x$  функция  $Q(x, \cdot)$  является вероятностной мерой на  $(\mathcal{Y}, \mathcal{B})$ ;
2. при каждом фиксированном  $B$  функция  $Q(\cdot, B)$  является измеримой, т.е.  $\forall c \{x : Q(x, B) < c\} \in \mathcal{A}$  (иными словами,  $Q(\cdot, B)$  является  $(\mathcal{A}, \mathcal{B}_{[0,1]})$ -измеримой, где  $\mathcal{B}_{[0,1]}$  — борелевская  $\sigma$ -алгебра на отрезке  $[0, 1]$ );
3. совместное распределение  $X$  и  $Y$  находится по формуле, которая является *интегральным аналогом формулы полной вероятности*:

$$\mathbb{P}_{XY}(A, B) = \int_A Q(x, B) \mathbb{P}_X(dx) = \int_A \mathbb{P}_{Y|X=x}(B) \mathbb{P}_X(dx).^{105} \quad (50)$$

**Утверждение 12.** *Если существует функция  $Q(x, B)$ , удовлетворяющая свойствам (1)–(2), то функция  $\mathbb{P}_{XY}$ , задаваемая по формуле (50), является вероятностной мерой на  $(\mathcal{X} \otimes \mathcal{Y}, \mathcal{A} \otimes \mathcal{B})$ .*

**Доказательство.**

1. неотрицательность:  $\mathbb{P}_{XY}(C) \geq 0 \forall C \in \mathcal{A} \otimes \mathcal{B}$ .

Очевидно выполнено.

<sup>104</sup>или регулярным условным распределением  $Y$

<sup>105</sup>рассматриваемый интеграл существует в силу требования (2):  $Q(x, B)$  измерима и ограничена, а значит интегрируема по Лебегу

2. аддитивность:  $\mathbb{P}_{XY}(C_1 + C_2) = \mathbb{P}_{XY}(C_1) + \mathbb{P}_{XY}(C_2) \quad \forall C_1, C_2 \in \mathcal{A} \otimes \mathcal{B}$ .

Перепишем формулу (50)<sup>106</sup>:

$$\mathbb{P}_{XY}(A, B) = \int_A \mathbb{P}_X(dx) \int_B Q(x, dy).$$

Напомним, что  $Q(x, \cdot)$  — вероятностная мера на  $(\mathcal{Y}, \mathcal{B})$ , поэтому запись корректна.

По теореме Фубини можно записать

$$\mathbb{P}_{XY}(A, B) = \int_{A \times B} Q(x, dy) \mathbb{P}_X(dx).$$

Аддитивность меры  $\mathbb{P}_{XY}$  следует из аддитивности интеграла Лебега.

3. непрерывность:  $C_n \uparrow C \Rightarrow \mathbb{P}_{XY}(C_n) \rightarrow \mathbb{P}_{XY}(C) \quad \forall C_n, C \in \mathcal{A} \otimes \mathcal{B}$ .

Непрерывность меры  $\mathbb{P}_{XY}$  следует из абсолютной непрерывности интеграла Лебега.

4. нормированность:  $\mathbb{P}_{XY}(\mathcal{X} \otimes \mathcal{Y}) = 1$ .

Нормированность очевидно следует из нормированности вероятностных мер  $\mathbb{P}_X$  и  $Q(x, \cdot)$   $\square$

### Замечание 1.

**Теорема 21** (Фубини). 1. Пусть  $f(x, y)$  интегрируема на  $X \times Y$ , тогда

$$I(x) = \int_Y f(x, y) d\mu_y$$

существует для почти всех  $x$ ,  $I(x)$  интегрируема на  $X$  и

$$\int_{X \times Y} f(x, y) d\mu_x \otimes \mu_y = \int_X \int_Y f(x, y) d\mu_y d\mu_x.$$

2. Пусть  $f(x, y) \geq 0$  измеримая функция и существует интеграл

$$\int_X \int_Y f(x, y) d\mu_y d\mu_x.$$

Тогда  $f(x, y)$  интегрируема на  $X \times Y$  и

$$\int_{X \times Y} f(x, y) d\mu_x \otimes \mu_y = \int_X \int_Y f(x, y) d\mu_y d\mu_x.$$

**Замечание 2.** Свойство абсолютной непрерывности интеграла Лебега: если  $f$  интегрируема, то

$$\forall \varepsilon > 0 \quad \exists \delta > 0: \mu(A) < \delta \Rightarrow \left| \int_A f d\mu \right| < \varepsilon.$$

---

<sup>106</sup>  $d\mathbb{P}(x)$  и  $\mathbb{P}(dx)$  — эквивалентные обозначения

**Теорема 22.** Пусть имеются две случайные величины:  $X$ , принимающая значения в измеримом пространстве  $(\mathcal{X}, \mathcal{A})$ , и  $Y$ , принимающая значения в измеримом пространстве  $(\mathcal{Y}, \mathcal{B})$ . Пусть также имеется их совместное распределение  $\mathbb{P}_{XY}$ , и пространство  $(\mathcal{Y}, \mathcal{B})$  обладает «хорошими» свойствами (например, является польским с борелевской  $\sigma$ -алгеброй). Тогда существует переходное ядро  $Q(x, B)$ , причем оно единственно с точностью до почти наверное по мере  $\mathbb{P}_X$ .

Эту теорему мы примем без доказательства.

**Замечание 1.** Теорема неконструктивна, она говорит только о существовании переходного ядра.

**Замечание 2.** При фиксированном  $B$  существование  $Q(x, B)$  является следствием теоремы Радона–Никодима.

Пусть имеются две меры  $\nu$  и  $\mu$ , определенные на общем измеримом пространстве  $(\Omega, \mathcal{F})$ .

**Определение 93.** Мера  $\nu$  называется **абсолютно непрерывной** относительно меры  $\mu$ , если

$$\mu(E) = 0 \Rightarrow \nu(E) = 0 \quad \forall E \in \mathcal{F}.$$

**Теорема 23** (Радон–Никодим). Если мера  $\nu$  абсолютно непрерывна относительно меры  $\mu$ , то существует интегрируемая функция  $f$  (называемая производной Радона–Никодима от меры  $\nu$  по мере  $\mu$ ), такая что

$$\nu(A) = \int_A f d\mu, \quad \forall A \in \mathcal{F}.$$

Очевидно, что мера  $\mathbb{P}_{XY}(\cdot, B)$  абсолютно непрерывна относительно  $\mathbb{P}_X(\cdot)$  ( $\mathbb{P}_X(A) = 0 \Rightarrow \mathbb{P}_{XY}(A, B) = 0$ ), значит,  $Q(x, B)$  является производной Радона–Никодима, о существовании которой говорит теорема.

Рассмотрим примеры условных вероятностей для некоторых классов случайных величин.

#### 1. дискретные случайные величины

Случайная величина  $X$  называется дискретной, если существует счетное множество  $D = \{x_1, \dots, x_n, \dots\}$ , такое что  $\mathbb{P}_X(D) = 1$  ( $\mathbb{P}_X(x_i) \neq 0 \forall i$ ). Переходное ядро примет вид:

$$\mathbb{P}_{Y|X=x}(B) = \begin{cases} \frac{\mathbb{P}_{XY}(\{x\}, B)}{\mathbb{P}_X(\{x\})}, & x \in D, \\ Q(x, B), & x \notin D. \end{cases}$$

Здесь  $Q(x, B)$  — любая функция, удовлетворяющая первым двум свойствам в определении переходного ядра. Так как  $\mathbb{P}_X(\{x\}) = 0 \forall x \notin D$ , то третье свойство выполнено автоматически.

#### 2. абсолютно непрерывные случайные величины

Случайная величина  $X$  называется абсолютно непрерывной, если существует такая интегрируемая на  $\mathcal{X}$  функция  $p(x)$ , называемая плотностью случайной величины  $X$ , что  $\mathbb{P}_X(A) = \int_A p(x) d\mu_x$ . Здесь  $d\mu_x$  — мера Лебега в пространстве  $\mathcal{X}$ .

$$p_{Y|X=x}(y) = \begin{cases} \frac{p_{XY}(x, y)}{p_X(x)}, & p_X(x) > 0, \\ q(x, y), & p_X(x) = 0. \end{cases}$$

Здесь  $q(x, y)$  — произвольная функция, такая что  $\int_B q(x, y) d\mu_y = Q(x, B)$  является переходным ядром.

**Упражнение.** Показать, что определенные таким образом условные вероятности обладают свойствами переходного ядра.

1. дискретные случайные величины

- (a) при каждом фиксированном  $x$  функция  $\mathbb{P}_{Y|X=x}(B)$  является вероятностной мерой на  $(\mathcal{Y}, \mathcal{B})$ : проверяется непосредственно.
- (b) при каждом фиксированном  $B$  функция  $\mathbb{P}_{Y|X=x}(B)$  является измеримой: это следует из того, что  $\mathbb{P}_{Y|X=x}(B)$  лишь в счетном числе точек отлична от функции  $Q(x, B)$ , которая является измеримой.

(c)

$$\int_A \mathbb{P}_{Y|X=x}(B) d\mathbb{P}_X(x) = \sum_{x_i \in D \cap A} \mathbb{P}_{XY}(x_i, B) = \mathbb{P}_{XY}(A, B).$$

2. абсолютно непрерывные случайные величины

Условной вероятностью в данном случае будет являться функция

$$\mathbb{P}_{Y|X=x}(B) = \int_B p_{Y|X=x}(y) d\mu_y.$$

- (a) при каждом фиксированном  $x$  функция  $\mathbb{P}_{Y|X=x}(B)$  является вероятностной мерой на  $(\mathcal{Y}, \mathcal{B})$ : следует из свойств интеграла Лебега.
- (b) при каждом фиксированном  $B$  функция  $\mathbb{P}_{Y|X=x}(B)$  является измеримой: интеграл Лебега от измеримой функции есть функция абсолютно непрерывная (см. теорему 17), а значит, измеримая.

(c)

$$\begin{aligned} \int_A \mathbb{P}_{Y|X=x}(B) p_X(x) d\mu_x &= \int_A \left( \int_B \frac{p_{XY}(x, y)}{p_X(x)} d\mu_y \right) p_X(x) d\mu_x = \\ &= \int_A \left( \int_B p_{XY}(x, y) d\mu_y \right) p_X(x) d\mu_x = \mathbb{P}_{XY}(A, B). \end{aligned}$$

**Упражнение.** Доказать, что если совместное распределение случайных величин является абсолютно непрерывным с плотностью  $p_{XY}(x, y)$ , то маргинальное распределение будет также абсолютно непрерывным с плотностью

$$p_X(x) = \int_{\mathcal{Y}} p_{XY}(x, y) d\mu_y.$$

$$\mathbb{P}_X(A) = \mathbb{P}_{XY}(A, \mathcal{Y}) = \int_{A \times \mathcal{Y}} p_{XY}(x, y) d\mu_x \otimes \mu_y = \int_A \left( \int_{\mathcal{Y}} p_{XY}(x, y) d\mu_y \right) d\mu_x.$$

Последнее равенство выполнено в силу теоремы Фубини. Значит,

$$p_X(x) = \int_{\mathcal{Y}} p_{XY}(x, y) d\mu_y.$$

## 11.2 Метод последовательного моделирования случайного вектора.

Приведем теперь теоремы, являющуюся в некотором смысле аналогом теоремы Колмогорова, только теперь множество индексов не более, чем счетно (в теореме Колмогорова множество индексов бесконечно и любой мощности), причем вместо маргинальных распределений даны переходные ядра. Заметим также, что данная теорема имеет конструктивное доказательство (в отличие от теоремы Колмогорова).

**Теорема 24** (Ионеску Тулча). Пусть имеется семейство случайных величин  $\{X_i\}$ ,  $i \in I$ ,  $I$  — множество индексов, не более чем счетное. Случайная величина  $X_i$  принимает значения из измеримого пространства  $(\mathcal{X}_i, \mathcal{A}_i)$ . Пусть задана вероятностная мера  $Q_1(\cdot)$  на  $\mathcal{X}_1$ , а также переходные ядра<sup>107</sup>

$$Q_2(x_1, \cdot), Q_3((x_1, x_2), \cdot), \dots, Q_n((x_1, x_2, \dots, x_{n-1}), \cdot), \dots$$

Тогда существует  $Q$  — совместное распределение семейства случайных величин  $\{X_i\}$ ,  $i \in I$ , на  $(\bigotimes_{i \in I} \mathcal{X}_i, \bigotimes_{i \in I} \mathcal{A}_i)$ .

**Доказательство.** Как уже было показано, совместное распределение случайных величин  $X_1, X_2$  можно записать в виде

$$\mathbb{P}_{X_1 X_2}(A_1, A_2) = \int_{A_1 \times A_2} Q_1(dx_1) Q_2(x_1, dx_2).$$

Значит, совместное распределение случайных величин  $X_1, X_2, X_3$  можно записать в виде

$$\begin{aligned} \mathbb{P}_{X_1 X_2 X_3}(A_1, A_2, A_3) &= \int_{A_1 \times A_2} Q_3((x_1, x_2), A_3) d\mathbb{P}_{X_1 X_2}(x_1, x_2) = \\ &= \int_{A_1 \times A_2} Q_3((x_1, x_2), A_3) Q_1(dx_1) Q_2(x_1, dx_2) = \int_{A_1 \times A_2 \times A_3} Q_1(dx_1) Q_2(x_1, dx_2) Q_3((x_1, x_2), dx_3). \end{aligned}$$

Таким образом, получаем формулу для совместного распределения  $n$  случайных величин:

$$\mathbb{P}_{X_1 X_2 \dots X_n}(A_1, A_2, \dots, A_n) = \int_{A_1 \times A_2 \times \dots \times A_n} Q_1(dx_1) Q_2(x_1, dx_2) \dots Q_n((x_1, x_2, \dots, x_{n-1}), dx_n).$$

Получаем, что имеем всевозможные конечномерные маргинальные распределения. Если  $I$  конечно, то все доказано. Если  $I$  бесконечно (счетно), то по теореме Колмогорова  $\exists Q$  — совместное распределение семейства случайных величин  $\{X_i\}$ ,  $i \in I$ , на  $(\bigotimes_{i \in I} \mathcal{X}_i, \bigotimes_{i \in I} \mathcal{A}_i)$ .

**Метод последовательного моделирования случайного вектора.** Пусть необходимо смоделировать случайный вектор с компонентами  $(X_1, X_2, \dots, X_n)$ , причем известно распределение первой компоненты  $Q_1(\cdot)$  и переходные ядра

$$Q_2(x_1, \cdot), Q_3((x_1, x_2), \cdot), \dots, Q_n((x_1, x_2, \dots, x_{n-1}), \cdot).$$

Тогда сначала генерируется значение случайной величины  $X_1$  с распределением  $Q_1(\cdot)$ . Затем полученное значение  $x_1$  подставляется в переходное ядро  $Q_2(x_1, \cdot)$  и генерируется случайная величина  $X_2$  с распределением  $Q_2(x_1, \cdot)$ . И т.д.

<sup>107</sup> $Q_i((x_1, \dots, x_{i-1}), \cdot)$  — условное распределение по вектору  $(x_1, \dots, x_{i-1})$  на  $\mathcal{X}_i$



### 11.3 Независимость случайных величин.

**Определение 94.** Случайные величины  $X$  (принимаяющая значения из  $(\mathcal{X}, \mathcal{A})$ ) и  $Y$  (принимаяющая значения из  $(\mathcal{Y}, \mathcal{B})$ ) называются **независимыми**, если

$$\mathbb{P}_{Y|X=x}(B) = \mathbb{P}_Y(B) \quad \forall B \in \mathcal{B}, \forall x \in \mathcal{X}. \quad (51)$$

Недостаток этого определения в том, что оно несимметрично. Поэтому чаще пользуются таким определением:

**Определение 95.** Случайные величины  $X$  (принимаяющая значения из  $(\mathcal{X}, \mathcal{A})$ ) и  $Y$  (принимаяющая значения из  $(\mathcal{Y}, \mathcal{B})$ ) называются **независимыми**, если

$$\mathbb{P}_{XY}(A, B) = \mathbb{P}_X(A)\mathbb{P}_Y(B) \quad \forall A \in \mathcal{A}, \forall B \in \mathcal{B}. \quad (52)$$

Равенство (52) можно интерпретировать как определение независимости событий  $\{X \in A\}$  и  $\{Y \in B\}$  для  $\forall A \in \mathcal{A}, \forall B \in \mathcal{B}$ .

Для произвольного количества случайных величин вводится следующее определение:

**Определение 96.** Пусть  $\{X_i\}, i \in I$ , — совокупность случайных величин ( $X_i$  принимает значения из  $(\mathcal{X}_i, \mathcal{A}_i)$ ,  $I$  — набор индексов любой мощности). Случайные величины  $X_i$  называются **независимыми в совокупности**, если для любого конечного набора чисел  $i_1, i_2, \dots, i_n, i_j \in I$ :

$$\mathbb{P}_{X_{i_1} X_{i_2} \dots X_{i_n}}(A_{i_1}, A_{i_2}, \dots, A_{i_n}) = \mathbb{P}_{X_{i_1}}(A_{i_1}) \mathbb{P}_{X_{i_2}}(A_{i_2}) \dots \mathbb{P}_{X_{i_n}}(A_{i_n}) \quad \forall A_i \in \mathcal{A}_i, \quad (53)$$

т.е.  $X_{i_1}, X_{i_2}, \dots, X_{i_n}$  независимы в совокупности.

**Упражнение.** Дать определения независимости случайных величин в терминах условных распределений и показать, что определения будут эквивалентными.

**Определение 97.** Пусть  $\{X_i\}, i \in I$ , — совокупность случайных величин ( $X_i$  принимает значения из  $(\mathcal{X}_i, \mathcal{A}_i)$ ,  $I$  — набор индексов любой мощности). Случайные величины  $X_i$  называются **независимыми в совокупности**, если для любого конечного набора чисел  $i_1, i_2, \dots, i_n, i_j \in I$ :

$$Q_{i_1, i_2, \dots, i_n}((x_{i_2}, \dots, x_{i_n}), A_{i_1}) = \mathbb{P}_{X_{i_1}|X_{i_2}=x_{i_2}, \dots, X_{i_n}=x_{i_n}}(A_{i_1}) = \mathbb{P}_{X_{i_1}}(A_{i_1}), \quad \forall A_{i_1} \in \mathcal{A}_{i_1}, \forall x_i \in \mathcal{X}_i. \quad (54)$$

Определение 96  $\Rightarrow$  определение 97

С одной стороны,

$$\mathbb{P}_{X_{i_1} \dots X_{i_n}}(A_{i_1}, \dots, A_{i_n}) = \int_{A_{i_2} \times \dots \times A_{i_n}} \mathbb{P}_{X_{i_1}|X_{i_2}=x_{i_2}, \dots, X_{i_n}=x_{i_n}}(A_{i_1}) \mathbb{P}_{X_{i_2} \dots X_{i_n}}(dx_{i_2}, \dots, dx_{i_n}).$$

С другой стороны,

$$\begin{aligned} \mathbb{P}_{X_{i_1} \dots X_{i_n}}(A_{i_1}, \dots, A_{i_n}) &= \mathbb{P}_{X_{i_1}}(A_{i_1}) \dots \mathbb{P}_{X_{i_n}}(A_{i_n}) = \mathbb{P}_{X_{i_1}}(A_{i_1}) \cdot \mathbb{P}_{X_{i_2} \dots X_{i_n}}(A_{i_2}, \dots, A_{i_n}) = \\ &= \int_{A_{i_2} \times \dots \times A_{i_n}} \mathbb{P}_{X_{i_1}}(A_{i_1}) \mathbb{P}_{X_{i_2} \dots X_{i_n}}(dx_{i_2}, \dots, dx_{i_n}). \end{aligned} \quad (55)$$

**Утверждение 13.** Если  $f(x)$  измерима,  $\int_D f(x) d\mu_x = 0 \quad \forall D$  измеримого, то  $f(x) = 0$  почти всюду.

**Доказательство.** Возьмем в качестве  $D = \{f(x) \geq 0\}$  — это множество измеримо (в силу измеримости  $f(x)$ ). На этом множестве функция  $f(x)$  неотрицательна и

$$\int_D f(x) d\mu_x = 0.$$

Значит,  $f(x) = 0$  почти всюду.  $\square$

Из этого утверждения следует, что

$$\mathbb{P}_{X_{i_1}|X_{i_2}=x_2,\dots,X_{i_n}=x_n}(A_1) = \mathbb{P}_{X_{i_1}}(A_1).$$

Напомним, что переходное ядро определено с точностью до почти всюду по соответствующей мере. Поэтому равенства с участием условного распределения в данных выше определениях нужно понимать соответствующе.

Определение 97  $\Rightarrow$  определение 96

Индукция по  $n$ .

$n = 2$ .

$$\mathbb{P}_{X_{i_1}X_{i_2}}(A_{i_1}, A_{i_2}) = \int_{A_{i_2}} P_{X_{i_1}|X_{i_2}=x_{i_2}}(A_{i_1}) \mathbb{P}_{X_{i_2}}(dx_{i_2}) = \mathbb{P}_{X_{i_1}}(A_{i_1}) \mathbb{P}_{X_{i_2}}(A_{i_2}).$$

$n - 1 \rightarrow n$ .

$$\begin{aligned} \mathbb{P}_{X_{i_1}\dots X_{i_n}}(A_{i_1}, \dots, A_{i_n}) &= \int_{A_{i_2} \times \dots \times A_{i_n}} \mathbb{P}_{X_{i_1}|X_{i_2}=x_{i_2}, \dots, X_{i_n}=x_{i_n}}(A_{i_1}) \mathbb{P}_{X_{i_2}\dots X_{i_n}}(dx_{i_2}, \dots, dx_{i_n}) = \\ &= \mathbb{P}_{X_{i_1}}(A_{i_1}) \mathbb{P}_{X_{i_2}\dots X_{i_n}}(A_{i_2}, \dots, A_{i_n}) = \mathbb{P}_{X_{i_1}}(A_{i_1}) \mathbb{P}_{X_{i_2}}(A_{i_2}) \dots \mathbb{P}_{X_{i_n}}(A_{i_n}). \end{aligned}$$

Обратимся теперь к **случайным процессам**. Пусть имеется измеримое пространство  $(\mathcal{X}, \mathcal{A})$ . Рассмотрим декартову степень множества  $\mathcal{X}$ :

$$\bigotimes_{\alpha \in T} \mathcal{X} = \mathcal{X}^T.$$

Напомним, что множество  $\mathcal{X}^T$  мы понимаем в смысле

$$\mathcal{X}^T = \{f : T \rightarrow \mathcal{X}\}.$$

На этом множестве рассмотрим  $\sigma$ -алгебру, порожденную цилиндрическими множествами:

$$\bigotimes_{\alpha \in T} \mathcal{A} = \mathcal{A}^T.$$

Заметим, что это достаточно бедная  $\sigma$ -алгебра, т.е. она содержит в себе мало информации о модели.

Множество индексов  $T$  будем трактовать как время.

**Определение 98.** Случайным процессом  $X_t$ ,  $t \in T$ , называется случайная величина, определенная на  $(\mathcal{X}^T, \mathcal{A}^T)$ . Другими словами, случайным процессом называется континуальное семейство случайных величин  $X^T = \{X_t, t \in T\}$ .

Используют различные обозначения случайного процесса, когда хотят акцентировать внимание на той или иной его природе: если пишут  $X_t(\omega)$ , то рассматривают случайный процесс как набор случайных величин, если пишут  $X_\omega(t)$ , то рассматривают случайный процесс как случайную функцию. При фиксированном  $\omega$  функцию  $X_\omega(t)$  называют **траекторией процесса**.

**Определение 99.** Процессы  $X_t$  и  $X'_t$  называют **эквивалентными** (обозначение:  $X_t \sim X'_t$ ), если

$$\forall t \in T \quad \mathbb{P}(X_t = X'_t) = 1. \quad (56)$$

Про эквивалентные процессы  $X_t$  и  $X'_t$  говорят, что один является модификацией другого

Условие (56) означает, что все конечномерные распределения<sup>108</sup> эквивалентных случайных процессов совпадают<sup>109</sup>, а значит, по теореме Колмогорова совпадают их распределения на пространстве  $(\mathcal{X}^T, \mathcal{A}^T)$  (пространстве Колмогорова).

Условие (56) является достаточно слабым. Иногда рассматривают более сильное условие:

$$\mathbb{P}(\forall t \in T \quad X_t = X'_t) = 1. \quad (57)$$

Процессы, обладающие этим свойством, являются траекторно неразличимыми.

**Пример** (Процессы, удовлетворяющие свойству (56), но не удовлетворяющие (57)).

$$\begin{aligned} X_t &\equiv 0, \\ X'_t &= X_t + D(t + \tau). \end{aligned}$$

Здесь  $D$  — функция Дирихле,  $\tau$  — случайная величина, независимая с  $X_t$ , имеющая абсолютно непрерывное распределение. Фиксируем произвольное  $t \in T$ .

$$\mathbb{P}(X_t = X'_t) = \mathbb{P}(D(t + \tau) = 0) = \mathbb{P}(\tau \in \{\mathbb{R} \setminus \mathbb{Q} - t\}) = 1,$$

т.к. вероятность того, что абсолютно непрерывная случайная величина примет значение из множества меры нуль равна нулю.

С другой стороны,

$$\mathbb{P}(\forall t \in T \quad X_t = X'_t) = \mathbb{P}(\forall t \in T \quad D(t + \tau) = 0) = \mathbb{P}(\forall t \in T \quad t + \tau \in \{\mathbb{R} \setminus \mathbb{Q}\}) = 0.$$

Различие объясняется тем, что при фиксированном  $\omega$   $\tau$  становится определенным числом и фиксируется траектория обоих процессов, причем  $X_\omega(t) \equiv 0$ , а  $X'_\omega(t) = D(t + \tau)$  — разрывная траектория; с другой стороны  $D(t + \tau) \stackrel{\text{п.в.}}{=} 0$  при фиксированном  $t$ .

Таким образом, в нашей модели получилось, что различные процессы имеют одинаковое распределение. Разрешить эту проблему можно различными способами:

1. расширить  $\sigma$ -алгебру;
2. отождествить эквивалентные процессы;

<sup>108</sup>Распределение случайного вектора  $(X_{t_1}, \dots, X_{t_n})$  при любых  $t_1, \dots, t_n \in T$  называется **конечномерным распределением** случайного процесса  $X_t$ ,  $t \in T$ .

<sup>109</sup>Фиксируем любые  $t_1, \dots, t_n \in T$ . Событие  $B_i = \{X_{t_i} = X'_{t_i}\}$  тривиально, т.к.  $\mathbb{P}(B_i) = 1$ . Тривиальное событие независимо с любым событием (см. 2 главу), значит  $\mathbb{P}(B_1 \cap \dots \cap B_n) = \mathbb{P}(B_1) \cdot \dots \cdot \mathbb{P}(B_n) = 1$ . Получаем:  $\mathbb{P}_{X_{t_1} \dots X_{t_n}}(A_1, \dots, A_n) = \mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n) = \mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n) \cdot \mathbb{P}(B_1 \cap \dots \cap B_n) = \mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n, X_{t_1} = X'_{t_1}, \dots, X_{t_n} = X'_{t_n}) = \mathbb{P}(X'_{t_1} \in A_1, \dots, X'_{t_n} \in A_n, X_{t_1} = X'_{t_1}, \dots, X_{t_n} = X'_{t_n}) = \mathbb{P}(X'_{t_1} \in A_1, \dots, X'_{t_n} \in A_n) = \mathbb{P}_{X'_{t_1} \dots X'_{t_n}}(A_1, \dots, A_n)$ .

3. рассматривать в качестве траекторий случайных процессов не всевозможные функции, а только удовлетворяющие определенным свойствам.

Мы выберем третий способ и ограничимся функциями, которые не имеют разрывов второго рода.

**Определение 100.** **Пространством Скорохода** называется множество функций без разрывов второго рода, т.е. тех, у которых в любой точке существует предел справа и слева:

$$D[a, b] = \{f: [a, b] \rightarrow \mathbb{R}: \forall x \exists \lim_{\substack{t \rightarrow x-0 \\ t \rightarrow x+0}} f(t)\}.$$

Если рассматривать только процессы с траекториями из пространства Скорохода, а не из всего  $\mathcal{X}^T$ , то описанных выше проблем не будет.

Заметим, что если  $T = 0, 1, 2, \dots$ , то эквивалентные процессы также являются траекторно неразличимыми.

## 12 Пуассоновское поле. Условия хаотичности, однородности, стационарности.

Пусть в пространстве  $\mathbb{R}^n$  раскиданы точки. Предположения:

1. Точки раскиданы случайно (хаотично).
2. Однородность: в любой области пространства точки раскиданы похожим образом.
3. Ординарность: в одном месте не может быть слишком много точек (не слишком густо).

Заметим, что все три условия записаны неформально (пока).

Пример: капли дождя на листе бумаги (случайное счетное множество точек на плоскости).

Вместо множества точек в пространстве  $\mathbb{R}^n$  будем рассматривать считающую меру, определенную в  $\mathbb{R}^n$ .

**Определение 101.** *Считающая мера* (в собственном смысле) — мера, сосредоточенная на не более, чем счетном количестве точек, и приписывающая каждой такой точке (в которой она сосредоточена) единичную массу.

Пусть  $S$  — множество (не более чем счетное) точек в пространстве  $\mathbb{R}^n$ . Тогда считающая мера  $\nu$  (в собственном смысле) определяется следующим образом:

$$\nu(A) = \#A \cap S \text{ (количество точек во множестве } A), \quad \forall A \subseteq \mathbb{R}^n.$$

Получаем, что задание  $S$  означает задание меры  $\nu$  и наоборот (что задать  $S$ , что задать  $\nu$  — одно и то же). Нам приятнее работать с мерой  $\nu$ , поэтому в дальнейшем будем рассматривать случайные считающие меры.

Раз заговорили о случайности, то надо подумать о построении модели Колмогорова вероятностного пространства, где  $\Omega$  — множество всевозможных считающих мер в  $\mathbb{R}^n$ . Ясно, что пространство  $\Omega$  непрерывно, т.к. оно не изоморфно дискретному, а значит в качестве класса событий  $\mathcal{F}$  нельзя взять множество всех подмножеств (см. 3 главу). Множество считающих мер можно сделать метрическим (ввести на нем метрику), а значит в

качестве  $\mathcal{F}$  можно взять борелевскую  $\sigma$ -алгебру, причем утверждается, что на ней можно задать вероятностную меру (см. 3 главу). Более детальное построение вероятностного пространства опустим (поверим, что модель непротиворечива).

Если не накладывать условие ординарности, то в одно место может попасть несколько точек (в одну точку несколько капель). В таком случае вводят понятие обобщенной считающей меры, учитывающую кратность точки.

**Определение 102.** *Считающая мера* (в расширенном смысле) — мера, сосредоточенная на не более, чем счетном количестве точек, и приписывающая каждой такой точке (в которой она сосредоточена) положительное целое число.

Пусть задано измеримое пространство  $(E, \mathcal{E})$ .

Введем понятие поля<sup>110</sup>.

**Определение 103.** *Поле* — случайное счетное множество точек в пространстве  $E$ .

Рассмотрим случайную считающую меру  $\nu(A)$  на измеримом пространстве  $(E, \mathcal{E})$ , равную количеству точек во множестве  $A \in \mathcal{E}$  и полностью характеризующую поле. Как уже говорилось ранее, можно рассматривать поле в собственном смысле, когда кратных точек нет и  $\nu$  — считающая мера в собственном смысле, а можно рассматривать поле в расширенном смысле, когда кратность точек учитывается и  $\nu$  — считающая мера в расширенном смысле. Если фазовое пространство, где находятся точки, конечно, то логично рассматривать поле в расширенном смысле. В континуальном фазовом пространстве иногда также рассматривают поле в расширенном смысле, но чаще все-таки берут считающую меру в собственном смысле.

$\nu(A)$  — это по сути случайная величина. Обозначим через  $\mathbb{E}(\nu(A)) = \mu(A)$  — среднее число точек, попавших во множество  $A$ .  $\mu(A)$ ,  $A \in \mathcal{E}$  называется *мерой интенсивности случайного поля* ( $\mu(A)$  является счетно-аддитивной мерой, заданной на измеримом пространстве  $(E, \mathcal{E})$ ; это следует из того, что  $\nu(A)$  — счетно-аддитивная мера, и свойств математического ожидания).

Пусть на измеримом пространстве  $(E, \mathcal{E})$  задана некоторая другая мера  $\kappa$ , причем мера интенсивности абсолютно непрерывна относительно этой меры:  $\mu \ll \kappa$ . Тогда по теореме Радона-Никодима существует производная одной меры по другой (производная Радона-Никодима) (определена с точностью до почти всюду по мере  $\kappa$ ):

$$\lambda(x) = \frac{d\mu}{d\kappa}(x), \quad x \in E$$

такая, что

$$\mu(A) = \int_A \lambda(x) \kappa(dx).$$

Функция  $\lambda(x)$  называется *интенсивностью поля (по отношению к мере  $\kappa$ )*.

Положим  $E = \mathbb{R}^n$ ,  $\mathcal{E} = \mathcal{B}^n$  — борелевская  $\sigma$ -алгебра.

Пусть тогда  $\kappa$  — мера Лебега, заданная на  $\mathbb{R}^n$ . Напомним, что мера Лебега — единственная мера, инвариантная относительно сдвигов (поворотов, отражений, т.е. группы движения); мера Лебега определена с точностью до мультипликативной константы (с точки зрения здравого смысла, это объясняется возможностью выбора единицы измерения).

<sup>110</sup>Понятие “поля” по разному понимают в литературе. Во времена Колмогорова не было  $\sigma$ -алгебры, поэтому  $\mathcal{F}$  как элемент вероятностного пространства называли полем. Иногда также говорят, что поле — это случайная функция многих переменных (если одна переменная, то это случайный процесс)  $\rightarrow$  многомерное / одномерное поле. Мы этими интерпретациями понятия поля пользоваться не будем.

Итак, рассмотрим этот частный случай:  $(\mathbb{R}^n, \mathcal{B}^n)$ . Пока не утверждаем, что  $\mu \ll \kappa$ , где  $\kappa$  — мера Лебега. Также будем предполагать, что мера  $\nu$  — считающая мера в расширенном смысле, т.е. рассматриваем расширенное случайное поле.

Пусть выполнены следующие предположения относительно рассматриваемого расширенного случайного поля (ограничения на распределения случайных величин  $\nu(A)$ ,  $A \in \mathcal{E}$ ) (формализация условий хаотичности, однородности и ординарности):

1. Свойство “хаотичности”: для  $\forall A_1, \dots, A_n$ ,  $A_i \subset \mathbb{R}^n$ ,  $A_i \cap A_j = \emptyset$ ,  $i \neq j \implies$  случайные величины  $\nu(A_1), \dots, \nu(A_n)$  независимы<sup>111</sup>.
2. Свойство однородности по пространству (стационарности)<sup>112</sup>: для  $\forall A \subset \mathbb{R}^n$  и  $x \in \mathbb{R}^n$  распределения случайных величин  $\nu(A)$  и  $\nu(A+x)$  одинаковы (другими словами, распределение числа точек при сдвиге множества не меняется).
3. Свойство ординарности<sup>113</sup> (попадание точек в малое множество маловероятно, но возможно):
  - а)  $\mathbb{P}(\nu(A) > 0) \rightarrow 0$  при  $\kappa(A) \rightarrow 0$  (в малом множестве мало точек);
  - б)  $\mathbb{P}(\nu(A) > 1 | \nu(A) > 0) \rightarrow 0$  при  $\kappa(A) \rightarrow 0$  (если какие-то точки и попали в малое множество, то, скорее всего, только одна).

Заметим, что третье условие отсекает кратность, т.е. на самом деле мы рассматриваем поле в собственном смысле.

Так введенное поле называют **пуассоновским полем (однородным)**. Его название объясняет следующая теорема:

**Теорема 25** (без доказательства). *Если для поля выполнены перечисленные три условия (“хаотичности”, однородности и ординарности), то случайная считающая мера  $\nu(A)$  имеет распределение Пуассона с параметром  $\lambda \cdot \kappa(A)$ :*

$$\nu(A) \sim \text{Pois}(\lambda \cdot \kappa(A)).$$

**Замечание 1.** Наличие коэффициента  $\lambda$  не удивительно, ведь мера Лебега определена с точностью до мультипликативной константы.

**Замечание 2.** Эта теорема указывает еще один способ введения пуассоновского распределения (подробнее о пуассоновском распределении см. главу 15).

Из свойств пуассоновского распределения получаем, что  $\mathbb{E}(\nu(A)) = \mu(A) = \lambda \cdot \kappa(A)$ . Таким образом, мера интенсивности  $\mu$  также является мерой Лебега (что и не удивительно, ведь условие однородности является, фактически, характеристикой меры Лебега), причем параметр  $\lambda$  является интенсивностью пуассоновского поля (по отношению к мере  $\kappa$ ). Заметим, что интенсивность пуассоновского поля на  $\mathbb{R}^n$  постоянна.

**Замечание.** Можно дать более общее определение пуассоновского поля. Пусть на измеримом пространстве  $(E, \mathcal{E})$  задана обобщенная случайная считающая мера  $\nu(A)$ , удовлетворяющая условию “хаотичности” (это свойство не связано со спецификой  $\mathbb{R}^n$ ) и имеющая пуассоновское распределение с параметром  $\mu(A)$  ( $\nu(A) \sim \text{Pois}(\mu(A))$ ), где  $\mu(A)$  — мера

<sup>111</sup>т.е. то, что происходит в одной области, не влияет на то, что происходит в другой, а значит, не может, например, быть такого распределения для некоторого  $A$ :  $\mathbb{P}(\nu(A) = k) = \begin{cases} \frac{1}{n+1}, & k = 0, 1, \dots, n, \\ 0, & k > n. \end{cases}$

<sup>112</sup>это условие связано со спецификой  $\mathbb{R}^n$  и тем, что  $\kappa$  — мера Лебега (мера  $\kappa$  используется дальше, в третьем условии)

<sup>113</sup>в формулировке свойства ординарности используется специфика  $\mathbb{R}^n$ , однако, в отличие от второго свойства, это условие можно переформулировать (этого мы делать не будем) и убрать специфику  $\mathbb{R}^n$

интенсивности:  $\mathbb{E}(\nu(A)) = \mu(A)$  ( $\mu(A)$  — произвольная счетно-аддитивная функция множеств). Тогда говорят, что задано пуассоновское поле.

Скажем несколько слов в защиту непротиворечивости модели (строгое построение вероятностного пространства опустим). С одной стороны

$$\nu(A_1) \sim \text{Pois}(\mu(A_1)), \quad \nu(A_2) \sim \text{Pois}(\mu(A_2)) \quad \forall A_1 \cap A_2 = \emptyset;$$

$$\nu(A_1 + A_2) \sim \text{Pois}(\mu(A_1 + A_2)).$$

С другой стороны

$$\nu(A_1 + A_2) = \nu(A_1) + \nu(A_2), \quad \mu(A_1 + A_2) = \mu(A_1) + \mu(A_2).$$

Значит, в силу независимости  $\nu(A_1)$  и  $\nu(A_2)$ , получим

$$\mathbb{P}(\nu(A_1) + \nu(A_2) = k) = \sum_{i=0}^k \mathbb{P}(\nu(A_1) = k - i) \mathbb{P}(\nu(A_2) = i), \text{ т.е.}$$

$$\nu(A_1) + \nu(A_2) \sim \text{Pois}(\mu(A_1)) * \text{Pois}(\mu(A_2)) \quad (\text{“} * \text{” — операция свертки}).$$

Тогда получаем, что

$$\text{Pois}(\mu(A_1) + \mu(A_2)) = \text{Pois}(\mu(A_1)) * \text{Pois}(\mu(A_2)).$$

Но это верно для пуассоновского распределения. Противоречия не получаем.

Отметим, что случай  $\mathbb{R}^n$  важен тем, что для него “пуассоновость” не задавали специально, а вывели из свойств “хаотичности”, однородности и ординарности.

### Поведение пуассоновского поля на полупрямой.

Рассмотрим полупрямую  $\mathbb{R}_+ = [0, +\infty)$ . На ней случайным образом раскиданы точки. Интервалы между соседними точками обозначим через  $\tau_i$   $i = 0, 1, \dots$ . Случайные величины  $\tau_i$  независимы, одинаково распределены и имеют экспоненциальное распределение с параметром  $\lambda$ :  $\tau_i \sim \text{Exp}(\lambda)$ ,  $i = 0, 1, \dots$ . Покажем, что в результате получится пуассоновское поле ( $\lambda$  — интенсивность поля). Ясно, что свойства “хаотичности” и однородности выполняется в силу характеристики экспоненциального распределения как единственного абсолютно непрерывного распределения, обладающего свойством отсутствия памяти (процесс восстановления является марковским, см. 13 главу). Свойство ординарности выполняется очевидно.

Заметим, что верно и обратное. Если дано пуассоновское поле на полупрямой, то распределение интервалов  $\tau_i$  будет экспоненциальным с параметром  $\lambda$  — интенсивность поля ( $N(t) = \nu([0, t])$ ):

$$\begin{aligned} \mathbb{P}(\tau_n \geq t) &= \{\text{однородность}\} = \mathbb{P}(\tau_0 \geq t) = \mathbb{P}(N(t) - N(0) = 0) = \\ &= \mathbb{P}(N(t) = 0) = \{N(t) \sim \text{Pois}(\lambda t)\} = e^{-\lambda t}. \end{aligned}$$

Процесс  $N(t)$ ,  $t \in [0, \infty)$ , называется **пуассоновским процессом**.

**Моделирование произвольного пуассоновского поля с конечной мерой интенсивности.**

$$\mathbb{E}(\nu(A)) = \mu(A) < \infty, \quad \forall A \implies$$

$$\mathbb{P}(\nu(A) = \infty) = 0, \quad \mathbb{P}(\nu(A) < \infty) = 1.$$

Таким образом, рассматриваем случай, когда с вероятностью 1 число точек в пространстве будет конечно (мы же не можем сгенерировать бесконечное число точек, поэтому приходится накладывать такое ограничение).

*Алгоритм моделирования пуассоновского поля с конечной мерой интенсивности:*

Пусть  $\mu(E) < \infty$ .



1. Моделируем общее число точек, попавшее в  $E$ :  $\nu(E) \sim \text{Pois}(\mu(E)) \rightarrow N = \nu(E)$ . Если  $N > 0$ , то переходим к следующему шагу (иначе finish).
2. Разыгрываем  $N$  независимых одинаково распределенных случайных величин  $X_1, \dots, X_N$  на  $(E, \mathcal{E})$  с распределением  $Q(A) = \mathbb{P}(X_i \in A) = \frac{\mu(A)}{\mu(E)}$ .

Интуитивно понятно, что получится пуассоновское поле. Строгое доказательство опустим.

Аналогично можно смоделировать пуассоновское поле на любом множестве  $D \subset E$ , в случае, если  $\mu(D) < \infty$  (при этом возможно  $\mu(E) = \infty$ ).

В случае с полупрямой получаем два варианта моделирования пуассоновского поля.

Комбинируя данный метод моделирования однородного пуассоновского поля и метод элиминации фон Неймана, можно получить неоднородное пуассоновское поле.

## 13 Марковское свойство. Уравнение Колмогорова-Чепмена. Представление для цепей Маркова (стохастическая динамическая система). Однородность по времени, стационарность, эргодичность. Винеровский процесс. Процесс Орнштейна-Уленбека.

### 13.1 Марковское свойство. Уравнение Колмогорова-Чепмена.

Напомним определение случайного процесса, данного в конце 11 главы.

**Определение 104.** *Случайным процессом*<sup>114</sup>  $X_t, t \in T$ , называется случайная величина, определенная на  $(\mathcal{X}^T, \mathcal{A}^T)$ . Другими словами, случайным процессом называется континуальное семейство случайных величин  $X^T = \{X_t, t \in T\}$ , определенных на общем фазовом пространстве  $(\mathcal{X}, \mathcal{A})$ , где множество индексов  $T$  трактуется как время.

Рассмотрим случайный процесс  $X_t, t \in T, X_t \in \mathcal{X}, (\mathcal{X}, \mathcal{A})$ . Время  $T$  может быть как дискретно, так и непрерывно:

1.  $T = \{0, 1, 2, \dots\}$ ;
2.  $T = [0, +\infty)$ .

Введем понятие марковского процесса неформально.

**Определение 105.** Случайный процесс  $X_t, t \in T$  называется *марковским*<sup>115</sup> *процессом* (обладающим *марковским свойством*), если поведение процесса в будущем зависит от прошлого только через настоящее; другими словами: поведение процесса после момента  $t$  определяется не всей его предысторией, а лишь значением, которое процесс принял в момент  $t$ ; другими словами: “будущее” и “прошлое” условно независимы при фиксированном “настоящем”; другими словами: если *точно* известно состояние процесса в момент времени  $t$ , то условное поведение (распределение) процесса в будущем полностью определено (т.е. “хвост прошлого” можно откинуть без ущерба для информации о будущем).

<sup>114</sup>Чаще всего под случайным процессом понимают семейство случайных величин, определенных на  $(\mathbb{R}, \mathcal{B})$ , а под случайной функцией — более общее понятие: семейство случайных элементов, определенных на общем измеримом пространстве  $(E, \mathcal{E})$ .

<sup>115</sup>Марков Андрей Андреевич (старший)

Можно провести аналогию с задачей Коши в прямом времени: по начальным данным можно однозначно восстановить траекторию системы, не имея никакой информации о ее поведении в прошлом. А вот систем с запаздыванием не обладает марковским свойством.

Вообще говоря, формально любой процесс можно сделать марковским, усложнив фазовое пространство путем добавления в  $X_t$  всей информации о прошлом ( $X_t \rightarrow$  вся траектория системы до момента времени  $t$ ).

**Пример.** (Процесс восстановления) Рассмотрим работающий прибор, который время от времени выходит из строя (отказывает). В момент отказа прибор немедленно заменяется новым. Будем считать, что продолжительность исправной работы  $n$ -го прибора  $\tau_n$ ,  $n = 0, 1, \dots$ , является случайной величиной, причем все величины  $\tau_n$  независимы и одинаково распределены. Моменты времени, когда один из приборов выходит из строя, называют *моментами восстановления*. При этом считается, что 0 не является моментом восстановления. Можно нарисовать ось времени, и на ней откладывать интервалы времени непрерывной исправной работы прибора (см. рис. ()). Положим  $N_t$  равным количеству моментов восстановления до момента времени  $t$ .

Итак, приходим к случайному процессу  $N_t$ ,  $t \in [0, +\infty)$ , который называется **считающим процессом восстановления**. Заметим, что если распределение  $\tau_n$  экспоненциально, то получим пуассоновское поле на прямой (см. 12 главу) ( $N_t$  — пуассоновский процесс).

В общем случае для определения в момент времени  $\theta$  всех дальнейших моментов восстановления, т.е. поведения процесса  $N_t$  при  $t > \theta$ , необходима и достаточна информация о последнем времени восстановления (см. рис. ()) (в силу независимости  $\tau_n$ ). Другими словами, зная, что к моменту времени  $\theta$  произошло  $N_\theta$  сбоев, распределение случайной величины, равной моменту следующего сбоя, можно определить только зная время, прошедшее с момента последнего отказа оборудования. Получаем, что в общем случае процесс восстановления не является марковским.

Однако есть частный случай процесса восстановления, когда он обладает марковским свойством.

**Утверждение 14.** *Процесс восстановления  $N_t$  является марковским тогда и только тогда, когда распределение  $\tau_n$  экспоненциально.*

Фактически это характеристика экспоненциального распределения (единственное непрерывное распределение, обладающее свойством отсутствия памяти).

Доказательство.

Свойство отсутствия памяти (марковское свойство) для процесса восстановления:

$$\begin{aligned} \mathbb{P}(N_{t+s} - N_t = 0 \mid N_\tau = n_\tau, \tau \leq t) &= \mathbb{P}(N_{t+s} - N_t = 0 \mid N_t = n_t), \quad \forall s \geq 0, \quad \Longleftrightarrow \\ \mathbb{P}\left(\sum_{k=0}^{n_t+1} \tau_k > s+t \mid \sum_{k=0}^{n_t} \tau_k = t - \tau, \tau_{n_t+1} > \tau\right) &= \mathbb{P}\left(\sum_{k=0}^{n_t+1} \tau_k > s+t \mid \sum_{k=0}^{n_t} \tau_k \leq t, \sum_{k=0}^{n_t+1} \tau_k > t\right), \quad \forall s \geq 0, \forall \tau \geq 0, \\ \Longleftrightarrow \mathbb{P}(\tau_{n_t+1} > s + \tau \mid \tau_{n_t+1} > \tau) &= \mathbb{P}(\tau_{n_t+1} > s), \quad \forall s \geq 0, \forall \tau \geq 0. \end{aligned}$$

Обозначим  $G(x) = \mathbb{P}(\tau_n > x)$ . Тогда

$$\begin{aligned} \mathbb{P}(\tau_{n_t+1} > s + \tau \mid \tau_{n_t+1} > \tau) &= \mathbb{P}(\tau_{n_t+1} > s), \quad \forall s \geq 0, \forall \tau \geq 0 \quad \Longleftrightarrow \\ \frac{G(s + \tau)}{G(\tau)} &= G(s), \quad \forall s \geq 0, \forall \tau \geq 0 \quad \Longleftrightarrow \quad G(s + \tau) = G(\tau)G(s), \quad \forall s \geq 0, \forall \tau \geq 0. \end{aligned}$$

Заметим, что  $G(x) = 1 - F_{\tau_n}(x)$ ,  $G(x) \in [0, 1]$  — абсолютно непрерывная невозрастающая функция (т.к. распределение  $\tau_n$ , очевидно, абсолютно непрерывное). Пусть  $\tau = 0$ , тогда

$G(s) = G(0)G(s)$ . По физическому смыслу задачи  $G(s) > 0 \forall s \geq 0$ , значит  $G(0) = 1$ . Учитывая, что абсолютно непрерывная функция дифференцируема почти всюду (в качестве производной имеем в виду производную Радона-Никодима), получим:

$$G(s + \tau) = G(\tau)G(s) \iff \ln G(s + \tau) = \ln G(\tau) + \ln G(s) \iff \{\tilde{G}(x) = \ln G(x)\} \iff \tilde{G}(s + \tau) = \tilde{G}(\tau) + \tilde{G}(s).$$

Из этого следует, что

$$\tilde{G}'(x) = \lim_{\delta x \rightarrow 0} \frac{\tilde{G}(x + \delta x) - \tilde{G}(x)}{\delta x} = \lim_{\delta x \rightarrow 0} \frac{\tilde{G}(\delta x)}{\delta x} = \tilde{G}'(0) \implies \tilde{G}'(x) = \text{const} \implies \tilde{G}(x) = cx + b.$$

Учитывая, что  $\tilde{G}(0) = 0$ , легко показать, что

$$\tilde{G}(s + \tau) = \tilde{G}(\tau) + \tilde{G}(s) \iff \tilde{G}(x) = \lambda x.$$

Следовательно,  $G(x) = 1 - F_{\tau_n}(x) = e^{\lambda x}$ , т.е.  $\tau_n$  имеет экспоненциальное распределение (что и т.д.).

Поясним на примере. Представьте себе, что вы пришли на автобусную остановку. Время вашего ожидания автобуса сильно зависит от того, когда приезжал последний автобус (если среднее время между приходами автобусов равно 3 мин, то при прошествии 2 мин с момента ухода последнего автобуса ориентировочное время ожидания составит 1 мин). Таким образом, получили процесс, не обладающий марковским свойством, при этом распределение времени между приходами автобусов, очевидно, не экспоненциально.

Пример марковского процесса восстановления: девушка звонит по телефону (если она уже говорила 2 часа, то время ее дальнейшего разговора абсолютно не зависит от этого факта).

Расширить фазовое пространство так, чтобы любой процесс восстановления стал марковским, можно такой образом:  $N_t \rightarrow (N_t, t_{\text{последнего сбоя}})$ .

**Формализация марковского свойства.** Существует множество неэквивалентных формализаций марковского свойства (т.к., например, не всегда существует регулярный вариант условного распределения (переходные ядра), а значит надо делать дополнительные технические предположения). В книге А.Д.Вентцеля “Курс теории случайных процессов” есть 64 варианта определения марковского свойства.

Приведем одну из формализаций марковского свойства. Рассмотрим для определенности непрерывный случай, т.е.  $X_t, t \in [0, +\infty)$ .

**Определение 106.** Случайный процесс  $X_t, t \in T$  называется **марковским процессом** (обладающим **марковским свойством**), если  $\forall t$ :

$$\forall 0 \leq t_1 < t_2 < \dots < t_n < t \quad \forall h > 0 \quad \mathbb{P}_{X_{t+h}|X_t=x, X_{t_1}=x_1, X_{t_2}=x_2, \dots, X_{t_n}=x_n} \stackrel{\text{п.в.}}{=} \mathbb{P}_{X_{t+h}|X_t=x}. \quad (58)$$

**Замечание 1.** В определении говорится о равенстве почти всюду по мере  $\mathbb{P}_{X_{t_1}X_{t_2}\dots X_{t_n}}$ , т.е. о равенстве с точностью до почти всех  $x_1, \dots, x_n$ . Поясним это. В правой и левой частях написаны, фактически, переходные ядра, а переходное ядро определено почти всюду (см. 11 главу).

**Замечание 2.** Точную информацию о прошлом  $X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_n} = x_n$  можно заменить на неполную информацию  $X_{t_1} \in A_1, X_{t_2} \in A_2, \dots, X_{t_n} \in A_n$ . Однако важно, что о состоянии в момент времени  $t$  (настоящее) мы имеем точную информацию ( $X_t = x$ ). Фактически, равенство (58) означает, что любую информацию о прошлом можно откинуть.

Обозначим

$$\mathbb{P}_{X_{t+h}|X_t=x}(A) = \mathbb{P}^{t,t+h}(x, A)$$

— вероятность перехода из состояния  $x$  в момент времени  $t$  во множество  $A$  в момент времени  $t + h$  (переходное ядро).

Для процессов с дискретным временем  $X_t$ ,  $t = 0, 1, 2, \dots$ , применима теорема Тулча: если известно распределение  $X_0$  и переходные ядра  $\mathbb{P}_{X_1|X_0=x_0}$ ,  $\mathbb{P}_{X_2|X_1=x_1, X_0=x_0}, \dots$ , то существует совместное распределение  $\mathbb{P}_{X_1 \dots X_n \dots}$ , т.е. модель случайного процесса непротиворечива (т.к. можно построить вероятностное пространство).

Аналогичная теорема может быть сформулирована и для случайных процессов с непрерывным временем путем комбинации теоремы Тулча и теоремы Колмогорова (для любых счетных подмножеств бесконечного множества  $T$  должны быть выполнены условия теоремы Тулча). Таким образом, понятие случайного процесса корректно.

В случае наличия марковского свойства для полного задания процесса достаточно знать  $\mathbb{P}_{X_0}$ ,  $\mathbb{P}^{k-1,k}(x, A)$  для всех  $k$  (для дискретного времени) и  $\mathbb{P}_{X_0}$ ,  $\mathbb{P}^{t,t+h}(x, A)$  для всех  $t$  и  $h$  (для непрерывного времени).

В теореме Тулча была доказана формула для совместного распределения  $n$  случайных величин:

$$\mathbb{P}_{X_1 X_2 \dots X_n}(A_1, A_2, \dots, A_n) = \int_{A_1 \times A_2 \times \dots \times A_n} Q_1(dx_1) Q_2(x_1, dx_2) \dots Q_n((x_1, x_2, \dots, x_{n-1}), dx_n).$$

Учитывая марковское свойство процесса, можем сказать, что  $\forall 0 \leq t_1 < t_2 < \dots < t_k$  справедливо

$$\mathbb{P}(X_{t_1} \in A_1, \dots, X_{t_k} \in A_k) = \int_{\mathcal{X}} \left[ \int_{A_1} \mathbb{P}^{0,t_1}(x_0, dx_1) \int_{A_2} \mathbb{P}^{t_1,t_2}(x_1, dx_2) \dots \int_{A_k} \mathbb{P}^{t_{k-1},t_k}(x_{k-1}, dx_k) \right] \mathbb{P}_{X_0}(dx_0),$$

при этом  $\mathbb{P}^{0,0}(x, A) \stackrel{\text{обозн.}}{=} I(x, A) = \mathbf{I}_A(x)$  — единичное ядро (вырожденное распределение, сосредоточенное в точке  $x$ ,  $I(x, \cdot) = \delta_x$  — мера Дирака).

Таким образом, по  $\mathbb{P}_{X_0}$ ,  $\mathbb{P}^{t,t+h}(x, A)$  можно определить всевозможные конечномерные распределения, а значит, по теореме Колмогорова<sup>116</sup>, восстановить весь процесс. Однако это касается только марковских процессов, в случае отсутствия марковского свойства при таком задании конечномерных распределений могут не выполняться условия согласованности.

Для любых процессов выполнен интегральный аналог формулы полной вероятности (из определения условной вероятности):  $\forall 0 \leq s < t < u$

$$\int_{\mathcal{X}} \mathbb{P}_{X_t|X_s=x}(dy) \mathbb{P}_{X_u|X_s=x, X_t=y}(A) = \mathbb{P}_{X_u|X_s=x}(A) \quad (\text{интегрирование по } y),$$

или если применить введенные обозначения:

$$\int_{\mathcal{X}} \mathbb{P}^{s,t}(x, dy) \mathbb{P}^{s,t,u}(x, y, A) = \mathbb{P}^{s,u}(x, A).$$

Для марковских процессов  $\mathbb{P}^{s,t,u}(x, y, A) = \mathbb{P}^{t,u}(y, A)$ , поэтому для них данная выше формула примет немного другой вид:  $\forall 0 \leq s < t < u$

$$\int_{\mathcal{X}} \mathbb{P}^{s,t}(x, dy) \mathbb{P}^{t,u}(y, A) = \mathbb{P}^{s,u}(x, A) \quad (59)$$

<sup>116</sup>заметим, что ограничения на фазовое пространство в теореме Колмогорова по смыслу такие же, какие нужны для существования переходных ядер

— *уравнение Колмогорова-Чепмена.*

При условии, что определены переходные ядра, уравнение Колмогорова-Чепмена является необходимым условием того, что процесс обладает марковским свойством.

**Утверждение 15.** *Из уравнения Колмогорова-Чепмена (59) следует согласованность конечномерных распределений, заданных следующим образом:*

$$Q_{t_1, \dots, t_k}(A_1 \times A_2 \times \dots \times A_k) = \int_{\mathcal{X}} \left[ \int_{A_1} \mathbb{P}^{0, t_1}(x_0, dx_1) \int_{A_2} \mathbb{P}^{t_1, t_2}(x_1, dx_2) \dots \int_{A_k} \mathbb{P}^{t_{k-1}, t_k}(x_{k-1}, dx_k) \right] \mathbb{P}_{X_0}(dx_0), \quad (60)$$

**Доказательство.** Условие согласованности: для любых конечных наборов  $T_1, T_2, T_3$  таких, что  $T_1 \subseteq T_2 \subseteq T_3 \subset T$ , выполнено:

$$\pi_{T_3}^{T_1} \mathbb{P}_{T_3} = \pi_{T_2}^{T_1} \mathbb{P}_{T_2} = \mathbb{P}_{T_1}.$$

Легко видеть, что для доказательства согласованности достаточно показать, что при выполнении уравнения Колмогорова-Чепмена справедливы следующие равенства (остальные равенства условия согласованности доказываются аналогично предложенным трем):

$$\begin{aligned} Q_{t_1, \dots, t_{k-1}, t_k}(A_1 \times A_2 \times \dots \times A_{k-1} \times \mathcal{X}) &= Q_{t_1, \dots, t_{k-1}}(A_1 \times A_2 \times \dots \times A_{k-1}), \\ Q_{t_1, \dots, t_{i-1}, t_i, t_{i+1}, \dots, t_{k-1}, t_k}(A_1 \times A_2 \times \dots \times A_{i-1} \times \mathcal{X} \times A_{i+1} \times \dots \times A_k) &= \\ = Q_{t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_{k-1}, t_k}(A_1 \times A_2 \times \dots \times A_{i-1} \times A_{i+1} \times \dots \times A_k), \\ Q_{t_1, t_2, \dots, t_k}(\mathcal{X} \times A_2 \times \dots \times A_k) &= Q_{t_2, \dots, t_k}(A_2 \times \dots \times A_k). \end{aligned}$$

где  $Q$  определяется по формуле (60).

Докажем первое равенство.

$$\begin{aligned} &Q_{t_1, \dots, t_{k-1}, t_k}(A_1 \times A_2 \times \dots \times A_{k-1} \times \mathcal{X}) = \\ &= \int_{\mathcal{X}} \left[ \int_{A_1} \mathbb{P}^{0, t_1}(x_0, dx_1) \int_{A_2} \mathbb{P}^{t_1, t_2}(x_1, dx_2) \dots \int_{A_{k-1}} \mathbb{P}^{t_{k-2}, t_{k-1}}(x_{k-2}, dx_{k-1}) \int_{\mathcal{X}} \mathbb{P}^{t_{k-1}, t_k}(x_{k-1}, dx_k) \right] \mathbb{P}_{X_0}(dx_0) = \\ &= \int_{\mathcal{X}} \left[ \int_{A_1} \mathbb{P}^{0, t_1}(x_0, dx_1) \int_{A_2} \mathbb{P}^{t_1, t_2}(x_1, dx_2) \dots \int_{A_{k-1}} \mathbb{P}^{t_{k-2}, t_{k-1}}(x_{k-2}, dx_{k-1}) \mathbb{P}^{t_{k-1}, t_k}(x_{k-1}, \mathcal{X}) \right] \mathbb{P}_{X_0}(dx_0) = \\ &= \int_{\mathcal{X}} \left[ \int_{A_1} \mathbb{P}^{0, t_1}(x_0, dx_1) \int_{A_2} \mathbb{P}^{t_1, t_2}(x_1, dx_2) \dots \int_{A_{k-1}} \mathbb{P}^{t_{k-2}, t_{k-1}}(x_{k-2}, dx_{k-1}) \right] \mathbb{P}_{X_0}(dx_0) = \\ &= Q_{t_1, \dots, t_{k-1}}(A_1 \times A_2 \times \dots \times A_{k-1}). \end{aligned}$$

Докажем второе равенство. Уравнение Колмогорова-Чепмена можно переписать в виде:

$$\begin{aligned} &\int_{\mathcal{X}} \mathbb{P}^{s, t}(x, dy) \mathbb{P}^{t, u}(y, dz) = \mathbb{P}^{s, u}(x, dz) \implies \\ &\int_{\mathcal{X}} \mathbb{P}^{s, t}(x, dy) \int_A \mathbb{P}^{t, u}(y, dz) f(z) = \int_A \mathbb{P}^{s, u}(x, dz) f(z) \quad \forall \text{ ограниченной измеримой функции } f(z) \end{aligned}$$

Следовательно,

$$\begin{aligned}
& Q_{t_1, \dots, t_{i-1}, t_i, t_{i+1}, \dots, t_{k-1}, t_k}(A_1 \times A_2 \times \dots \times A_{i-1} \times \mathcal{X} \times A_{i+1} \times \dots \times A_k) = \\
& = \int_{\mathcal{X}} \left[ \int_{A_1} \mathbb{P}^{0, t_1}(x_0, dx_1) \dots \int_{A_{i-1}} \mathbb{P}^{t_{i-2}, t_{i-1}}(x_{i-2}, dx_{i-1}) \int_{\mathcal{X}} \mathbb{P}^{t_{i-1}, t_i}(x_{i-1}, dx_i) \cdot \right. \\
& \quad \left. \int_{A_{i+1}} \mathbb{P}^{t_i, t_{i+1}}(x_i, dx_{i+1}) \dots \int_{A_k} \mathbb{P}^{t_{k-1}, t_k}(x_{k-1}, dx_k) \right] \mathbb{P}_{X_0}(dx_0) = \\
& = \int_{\mathcal{X}} \left[ \int_{A_1} \mathbb{P}^{0, t_1}(x_0, dx_1) \dots \int_{A_{i-1}} \mathbb{P}^{t_{i-2}, t_{i-1}}(x_{i-2}, dx_{i-1}) \int_{A_{i+1}} \mathbb{P}^{t_{i-1}, t_{i+1}}(x_{i-1}, dx_{i+1}) \cdot \right. \\
& \quad \left. \int_{A_{i+2}} \mathbb{P}^{t_{i+1}, t_{i+2}}(x_{i+1}, dx_{i+2}) \dots \int_{A_k} \mathbb{P}^{t_{k-1}, t_k}(x_{k-1}, dx_k) \right] \mathbb{P}_{X_0}(dx_0) = \\
& = Q_{t_1, \dots, t_{i-1}, t_{i+1}, \dots, t_{k-1}, t_k}(A_1 \times A_2 \times \dots \times A_{i-1} \times A_{i+1} \times \dots \times A_k).
\end{aligned}$$

Доказательство третьего равенства полностью аналогично доказательству второго.

Введем понятие композиции (свертки) переходных ядер.

Рассмотрим три измеримых пространства  $(C, \mathcal{C})$ ,  $(D, \mathcal{D})$ ,  $(E, \mathcal{E})$  и случайные элементы  $X: \Omega \rightarrow C$ ,  $Y: \Omega \rightarrow D$ ,  $Z: \Omega \rightarrow E$ . Обозначим через  $Q$  — переходное ядро из  $(C, \mathcal{C})$  в  $(D, \mathcal{D})$  (т.е.  $Q(x, A) = \mathbb{P}_{Y|X=x}(A)$ ), а через  $R$  — переходное ядро из  $(D, \mathcal{D})$  в  $(E, \mathcal{E})$  (т.е.  $R(y, B) = \mathbb{P}_{Z|Y=y}(B)$ ). Тогда **композицией (сверткой) переходных ядер**  $Q$  и  $R$  называется переходное ядро<sup>117</sup>  $Q \cdot R$  (обозначения могут быть разными) из  $(C, \mathcal{C})$  в  $(E, \mathcal{E})$ . Можно записать формально:

$$Q \cdot R(x, dz) = \int_{\mathcal{D}} Q(x, dy) R(y, dz).$$

Используя понятие композиции ядер, можно переписать уравнение Колмогорова-Чепмена в виде:

$$\mathbb{P}^{s, t} \cdot \mathbb{P}^{t, u} = \mathbb{P}^{s, u}.$$

**Замечание 1.** Переходные ядра можно рассматривать как линейные операторы на пространстве измеримых ограниченных функций:

$$Qf(x) = \int Q(x, dy) f(y).$$

Значит, к ним применима теория линейных операторов.

**Замечание 2.** В случае, когда  $\mathcal{X}$  — дискретное фазовое пространство (не более, чем счетно), переходные ядра процесса имеют вид:

$$\mathbb{P}^{t, t+h}(x, y) = \mathbb{P}(X_{t+h} = y \mid X_t = x).$$

<sup>117</sup> вообще говоря, не обязательно относящееся к случайным элементам  $X$  и  $Z$

**Утверждение 16.** Композиция ядер соответствует перемножению матриц переходных вероятностей.

Пронумеруем все элементы фазового пространства. Тогда переходному ядру  $\mathbb{P}^{t,t+h}$  соответствует матрица (возможно бесконечная)  $A^{t,t+h}$ , причем  $a_{i,j}^{t,t+h} = \mathbb{P}^{t,t+h}(x_i, x_j)$ . Операция композиции для дискретного пространства:

$$Q \cdot R(x_i, x_j) = \sum_{k=1}^{\infty(n)} Q(x_i, x_k) R(x_k, x_j).$$

Утв. доказано.

## 13.2 Однородность по времени.

Рассмотрим частный случай марковских процессов — однородные по времени марковские процессы. Заметим, что можно также рассматривать однородность по фазовому пространству, мы этим заниматься не будем.

**Определение 107.** Марковский процесс называется *однородным по времени*, если  $\mathbb{P}^{t,t+h}(x, A)$  не зависит от  $t$ .

Возвращаясь к аналогии с задачей Коши, можно сказать, что однородность (по времени) марковского процесса, соответствующего задаче Коши, означает автономность (стационарность) системы (правая часть не зависит от времени).

Пуассоновский процесс является однородным по времени.

Для однородных по времени марковских процессов переходное ядро  $\mathbb{P}^{t,t+h}$  обозначается через  $\mathbb{P}^h$  в силу его независимости от  $t$ :

$$\mathbb{P}^{t,t+h} = \mathbb{P}^h.$$

Тогда из уравнения Колмогорова-Чепмена получим:

$$\mathbb{P}^{0,h_1} \cdot \mathbb{P}^{h_1,h_1+h_2} = \mathbb{P}^{0,h_1+h_2} \implies \mathbb{P}^{h_1} \cdot \mathbb{P}^{h_2} = \mathbb{P}^{h_1+h_2}.$$

Напомним, что множество  $(\mathcal{X}, *)$  называется *группой*, если бинарная операция  $*$ :  $\mathcal{X} \times \mathcal{X} \rightarrow \mathcal{X}$  ассоциативна, обладает нейтральным элементом (единицей)<sup>118</sup> ( $\exists e \in \mathcal{X} : a * e = e * a = a, \forall a \in \mathcal{X}$ ) и любой элемент  $a \in \mathcal{X}$  обладает симметричным (обратным) элементом ( $\forall a \in \mathcal{X} \exists a' \in \mathcal{X} : a * a' = a' * a = e$ ). Если бинарная операция лишь ассоциативна, то множество  $(\mathcal{X}, *)$  называется *полугруппой*. Заметим, что если в полугруппе с единицей существует обратный элемент для любого  $a \in \mathcal{X}$ , то это уже группа.

**Утверждение 17.** Множество всех переходных ядер из  $(\mathcal{X}, \mathcal{A})$  в  $(\mathcal{X}, \mathcal{A})$  образуют полугруппу с единицей относительно операции композиции.

Доказательство. Покажем ассоциативность.

$$\begin{aligned} (Q \cdot R) \cdot S(x, dw) &= \int_{\mathcal{X}} \left( \int_{\mathcal{X}} Q(x, dy) R(y, dz) \right) S(z, dw) = \{\text{теорема Фубини}\} = \\ &= \int_{\mathcal{X}} Q(x, dy) \left( \int_{\mathcal{X}} R(y, dz) S(z, dw) \right) = Q \cdot (R \cdot S). \end{aligned}$$

<sup>118</sup>единица (если существует) определена единственным образом



В роли единичного элемента выступает тождественное ядро  $I(x, \cdot) = \delta_x$  — мера Дирака:

$$Q \cdot I(x, dz) = \int_{\mathcal{X}} Q(x, dy) I(y, dz) = \int_{\mathcal{X}} Q(x, dy) \delta_y(dz) = Q(x, dz),$$

$$QI = IQ = Q.$$

Заметим, что такая полугруппа переходных ядер не обязательно является коммутативной. Как было показано выше, операция композиции соответствует перемножению матриц переходных вероятностей в случае дискретного фазового пространства, а значит коммутативность ядер соответствует коммутативности матриц.

Рассмотрим подмножество множества всех переходных ядер из  $(\mathcal{X}, \mathcal{A})$  в  $(\mathcal{X}, \mathcal{A})$ , состоящее из переходных ядер однородного по времени марковского процесса с фазовым пространством  $(\mathcal{X}, \mathcal{A})$ :  $\{\mathbb{P}^h, h \geq 0\}$ . Это однопараметрическая коммутативная подгруппа относительно операции композиции (в силу равенства  $\mathbb{P}^{h_1} \cdot \mathbb{P}^{h_2} = \mathbb{P}^{h_1+h_2}$ ,  $\mathbb{P}^0(x, \cdot) = I(x, \cdot)$  — единица).

Соотношение  $\mathbb{P}^{h_1} \cdot \mathbb{P}^{h_2} = \mathbb{P}^{h_1+h_2}$  говорит о том, что коммутативные подгруппы с единицей  $\{\mathbb{P}^h, h \geq 0; \cdot\}$  и  $\{[0, +\infty); +\}$  гомеоморфны<sup>119</sup> ( $\mathbb{P}^{h_1} \cdot \mathbb{P}^{h_2} = \mathbb{P}^{h_1+h_2}$  — гомеоморфизм, сохраняющий групповую операцию).

Напомним, что марковский процесс задается с помощью  $\mathbb{P}_{X_0}, \mathbb{P}^{t, t+h}(x, A)$ . Значит, однородный марковский процесс задается с помощью:

1. переходных вероятностей  $\{\mathbb{P}^t, t \in T\}$ ,
2.  $\alpha$  — начального распределения.

Если при этом  $\mathcal{X}$  дискретно и конечно:  $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ , то достаточно для задания однородного марковского процесса знать матрицы переходных вероятностей  $\mathbb{P}^t \in \mathbb{R}^{n \times n}$  ( $\mathbb{P}_{i,j}^t$  — вероятность перехода из состояния  $x_i$  в состояние  $x_j$  за время  $t$ , т.е. элементы матрицы — меры одноточечных множеств).

### 13.3 Представление для цепей Маркова (стохастическая динамическая система).

**Определение 108.** *Цепь Маркова (марковская цепь)* — марковский процесс  $X_t$ ,  $t \in T$ ,  $X_t \in \mathcal{X}$ ,  $(\mathcal{X}, \mathcal{A})$ , причем время  $T$  дискретно, т.е.  $T = 0, 1, 2, \dots$ . Если при этом еще и фазовое пространство  $\mathcal{X}$  дискретно, т.е. конечное или счетное множество, а  $\mathcal{A} = 2^{\mathcal{X}}$ , то такой марковский процесс называют **дискретной цепью Маркова** (среди дискретных цепей выделяют конечные цепи, т.е. те, для которых фазовое пространство конечно).

Для цепи Маркова есть специальная универсальная форма представления в виде **случайной динамической системы** (в непрерывном случае такоо представления нет).

Итак, рассмотрим марковскую цепь  $X_t$ ,  $t \in T = 0, 1, 2, \dots$ ,  $X_t \in \mathcal{X}$ ,  $(\mathcal{X}, \mathcal{A})$ . Элементы фазового пространства марковской цепи называют состояниями динамической системы ( $x \in \mathcal{X}$  — *состояние*). Соответственно само фазовое пространство  $\mathcal{X}$  называют *пространством состояний*. Если пространство  $(\mathcal{X}, \mathcal{A})$  удовлетворяет некоторым топологическим свойствам (например, польское с борелевской  $\sigma$ -алгеброй), то цепь Маркова  $X_0, X_1, X_2, \dots$  представима в виде случайной динамической системы:

$$X_{t+1} = f_{t+1}(X_t, Y_{t+1}), \quad (61)$$

<sup>119</sup>заметим, что они не изоморфны, т.к. никто не сказал, что ядра  $\mathbb{P}^{h_1}$  и  $\mathbb{P}^{h_2}$  различны при  $h_1 \neq h_2$

где  $Y_1, Y_2, \dots$  — последовательность независимых одинаково распределенных случайных величин. Заметим, что распределение  $Y_k$  можно взять любым (например, равномерным), т.к. одно сводится к другому методом обращения функции распределения.

Если цепь Маркова *однородна по времени*, то она представима в виде:

$$X_{t+1} = f(X_t, Y_{t+1}). \quad (62)$$

**Теорема 26.** *Любая цепь Маркова представима в виде случайной динамической системы (61) (в случае однородности функция  $f$  не зависит от времени) ((61) — это общий вид цепи Маркова). Другими словами, случайный процесс с дискретным временем представим в виде (61)  $\iff$  он марковский, т.е. это марковская цепь.*

**Доказательство.** Пусть процесс представим в виде (61). Тогда очевидно, что  $\forall 0 \leq t_1 < t_2 < \dots < t_n < t \quad \forall h > 0 \mathbb{P}_{X_{t+h}|X_t=x, X_{t_1}=x_1, X_{t_2}=x_2, \dots, X_{t_n}=x_n} = \mathbb{P}_{X_{t+h}|X_t=x}$ . Пришли к определению марковского процесса.

Обратно. Пусть дана марковская цепь. Тогда она полностью характеризуется  $\mathbb{P}_{X_0}, \mathbb{P}^{t,t+h}, k = 1, 2, \dots$ . Рассмотрим процесс, заданный случайной динамической системой вида (61) с начальным распределением  $\mathbb{P}_{X_0}$ . Покажем, что выбором функции  $f_t$  можно переходные ядра этого процесса свести к заданным переходным ядрам исходного процесса, т.к. выбором функции  $f_t$  сделать процессы эквивалентными. Действительно, всегда можно выбрать такие  $f_t$ , что  $\mathbb{P}(X_{t+1} \in A | X_t = x) = \mathbb{P}(f_{t+1}(X_t, Y_{t+1}) \in A | X_t = x) = \mathbb{P}^{t,t+1}(x, A) \quad \forall t, x, A$ . Остальные переходные ядра (для  $h > 1$ ) совпадают в силу уравнения Колмогорова-Чепмена:  $\mathbb{P}^{t,t+1} \cdot \dots \cdot \mathbb{P}^{t+h-1,t+h} = \mathbb{P}^{t,t+h}$ .

Получаем преобразование пространства состояний  $\mathcal{X}$ :

$$x \mapsto f_t(x, y),$$

другими словами, разыгрываем  $y$  в соответствии с выбранным распределением  $\mathbb{P}_{Y_k}$  и применяем к выбранному состоянию  $x$  и сгенерированной величине  $y$  функцию  $f_t$ , получая в результате новое состояние.

Если функция  $f_t$  не зависит от своего второго аргумента (т.е. от  $y$ ), то получаем частный случай цепи Маркова:  $X_{t+1} = f_{t+1}(X_t)$  — обычная (неслучайная) динамическая система.

**Пример** (Случайное блуждание).

$$X_{t+1} = X_t + Y_{t+1}, \quad \mathcal{X} = \mathbb{R}^n, \quad \mathcal{A} = \mathcal{B}^n.$$

Другими словами,  $X_t = X_0 + \sum_{s=1}^t Y_s$  — сумма независимых одинаково распределенных случайных векторов с точностью до начального сдвига.

Заметим, что это однородная цепь Маркова.

## 13.4 Стационарность.

Рассмотрим понятие *стационарного процесса (в узком смысле)*.

Пусть случайный процесс  $X_t, t \in T$  ( $T = [0, +\infty)$  либо  $T = \{1, 2, 3, \dots\}$ , хотя можно рассматривать и  $T = \mathbb{R}$  либо  $T = \mathbb{Z}$ ) принимает значения в абстрактном, вообще говоря, пространстве  $(\mathcal{X}, \mathcal{A})$ .

**Определение 109.** *Стационарный процесс (в узком смысле)* — случайный процесс, траектория которого с вероятностью 1 принадлежит функциональному пространству, инвариантному относительно сдвига по времени.

Запишем определение стационарного процесса более формально. Траекторию процесса обозначим через  $X_\omega(t)$ . Введем оператор сдвига по времени:

$$(\theta_h X_\omega(\cdot))(t) = X_\omega(t+h), \quad t \in T.$$

Примерами функциональных пространств, инвариантных относительно сдвига, могут служить пространство непрерывных функций  $C[0, \infty)$  и пространство Скорохода.

Пусть  $\mathcal{M}$  — множество всех траекторий процесса  $X_t$ .  $\mathcal{M}$  можно рассматривать как пространство элементарных событий (каждому исходу  $\omega$  ставится в соответствие траектория  $X_\omega(t)$ ). Это пространство можно метризовать, а потом ввести на нем борелевскую  $\sigma$ -алгебру  $\mathfrak{M}$ . Итак, получили измеримое пространство  $(\mathcal{M}, \mathfrak{M})$ . На нем можно ввести вероятностную меру, соответствующую совместному распределению процесса. Получили пространство Колмогорова. Инвариантность (замкнутость) этого пространства относительно сдвига означает  $\theta_h \mathcal{M} \subseteq \mathcal{M}$  почти всюду, т.е. при сдвиге пространство  $(\mathcal{M}, \mathfrak{M}, \mathbb{P}_\mathcal{M})$  переходит практически в себя (с сохранением меры!).

**Определение 110.** *Стационарный процесс (в узком смысле)* — случайный процесс, для которого  $\theta_h X = X \ \forall h$ , т.е. сдвинутый процесс имеет такое же распределение, как и исходный, т.е. эквивалентен исходному (распределение процесса инвариантно относительно сдвига).

Заметим, что если брать  $T = \mathbb{R}$ , то условие стационарности становится сильнее.

**Замечание.** Из того, что распределения  $\theta_h X$  и  $X$  совпадают, не следует, что  $X_\omega(t) = X_\omega(t+h)$ , т.к. из совпадения распределений не следует равенство случайных элементов. Поэтому говорят не о совпадении траекторий при сдвиге, а о совпадении характера траекторий.

Итак, дадим формальное определение стационарного процесса (в узком смысле).

**Определение 111.** *Стационарный процесс (в узком смысле)* — случайный процесс, для которого все конечномерные распределения любого сдвинутого процесса совпадают с конечномерными распределениями исходного процесса.

Покажем эквивалентность определений, т.е. то, что из совпадения конечномерных распределений исходного и сдвинутого процессов следует совпадение совместных распределений этих процессов. Действительно, по теореме Колмогорова по согласованным конечномерным распределениям (в данном случае согласованность не вызывает сомнений) можно однозначно восстановить совместное распределение процесса.

Отметим, что из определения стационарности следует совпадение одномерных распределений исходного и сдвинутого процессов.

Рассмотрим теперь понятие *стационарного процесса (в широком смысле)*.

Рассмотрим случайный процесс  $X_t, t \in T$ , причем  $\mathcal{X} = \mathbb{R}, \mathcal{A} = \mathcal{B}$ .

**Определение 112.** Случайный процесс  $X_t, t \in T$ , называется *стационарным процессом (в широком смысле)*, если

1.  $\mathbb{E}X_t^2 < \infty \ \forall t \in T$ ,  $\mathbb{E}X_t$  не зависит от  $t$ ;
2. ковариационная функция процесса  $R_X(s, t) = \text{cov}(X_t, X_s)$  удовлетворяет следующему условию:  $\forall t, s : t \leq s$  справедливо  $R_X(s, t) = \text{cov}(X_t, X_{t+(s-t)}) = \text{cov}(X_0, X_{0+(s-t)}) = \phi(s-t)$ ; другими словами, в силу симметричности  $R_X(s, t) = R_X(t, s)$ ,  $R_X(s, t) = \phi(|s-t|) \ \forall s, t$  (т.е. ковариационная матрица зависит только от модуля разности своих аргументов).

Заметим, что, в силу  $R_X(s, t) = \mathbb{E}X_t X_s - \mathbb{E}X_t \mathbb{E}X_s = \phi(|s - t|)$  и  $\mathbb{E}X_t$  не зависит от  $t$ , получаем  $\mathbb{E}X_t X_s = \psi(|s - t|)$ , а значит  $\mathbb{E}X_t^2 = \mathbb{E}X_t X_t = \psi(|t - t|)$  не зависит от  $t$  для стационарного процесса. Значит условие  $\mathbb{E}X_t^2 < \infty$  достаточно проверить в одной точке.

Очевидно, что из стационарности в узком смысле следует, стационарность в широком смысле (при условии  $\mathbb{E}X_t^2 < \infty$ ). Заметим, что из условия  $\mathbb{E}X_t^2 < \infty$  следует, что  $\mathbb{E}X_t$  определено и конечно.

**Утверждение 18.** *Гауссовский процесс, стационарный в широком смысле, является стационарным процессом в узком смысле.*

**Доказательство.** Для гауссовского процесса любое конечномерное распределение является гауссовским. Следовательно, достаточно показать, что при любом сдвиге  $h$  для любых  $t_1 < t_2 < \dots < t_n$  математическое ожидание  $\mathbb{E}(X_{t_1}, X_{t_2}, \dots, X_{t_n}) = (\mathbb{E}X_{t_1}, \mathbb{E}X_{t_2}, \dots, \mathbb{E}X_{t_n})$  и ковариационная функция  $R_X(t_i, t_j)$  не изменяются:  $\mathbb{E}X_{t_i} = \mathbb{E}X_{t_i+h}$ ,  $R_X(t_i, t_j) = R_X(t_i + h, t_j + h)$ . Первое равенство следует из первого условия стационарности в широком смысле:  $\mathbb{E}X_t$  не зависит от  $t$ . Второе равенство следует из второго условия стационарности в широком смысле:  $R_X(s, t) = \phi(|s - t|) \forall s, t \implies R_X(t_i + h, t_j + h) = \phi(|t_i - t_j|) = R_X(t_i, t_j)$ .

Больше мы к понятию стационарности в широком смысле возвращаться не будем. Далее везде под стационарностью будем подразумевать стационарность в узком смысле.

Это была общая теория стационарных процессов.

Возвратимся к марковским процессам. Очевидно, что стационарный марковский процесс является однородным по времени. Ясно, что обратное неверно (пример: случайное блуждание, распределения  $X_0$  и  $X_1 = X_0 + Y_1$  не обязаны совпадать). Когда же однородный марковский процесс является стационарным?

Рассмотрим однородный марковский процесс  $X_t, t \in T, (\mathcal{X}, \mathcal{A})$ . Необходимым условием стационарности является равенство одномерных распределений, в частности  $\mathbb{P}_{X_0}$  и  $\mathbb{P}_{X_t} \forall t$ . По определению переходного ядра

$$\mathbb{P}_{X_t}(A) = \int_{\mathcal{X}} \mathbb{P}^t(x, A) \mathbb{P}_{X_0}(dx).$$

Введем понятие инвариантной меры.

Как уже говорилось ранее, переходные ядра можно рассматривать как линейные операторы на пространстве измеримых ограниченных функций:

$$Qf(x) = \int Q(x, dy) f(y).$$

А еще переходные ядра можно рассматривать как линейные операторы на пространстве мер:

$$\mu Q(A) = \int \mu(dx) Q(x, A).$$

Таким образом, имеем преобразования:

$$f \mapsto Qf, \quad \mu \mapsto \mu Q.$$

Заметим, что можно рассматривать и  $\sigma$ -конечные меры, т.е.  $\mu Q(A)$  может быть равным и бесконечности для некоторых  $A$ .

**Определение 113.** Функция  $f$  *инвариантна* относительно переходного ядра  $Q$ , если  $Qf = f$ .

**Определение 114.** Мера  $\mu$  *инвариантна* относительно переходного ядра  $Q$ , если  $\mu Q = \mu$ .

Инвариантные мера/функция являются аналогом собственных чисел с единичным собственным значением (собственная мера, собственная функция).

Заметим, что инвариантная мера может быть  $\sigma$ -конечной.

**Упражнение.** Проверить, что мера Лебега инвариантна относительно переходных ядер, соответствующих случайному блужданию в  $\mathbb{R}^n$ .

????????????????????

**Замечание.** Для переходных ядер, соответствующих случайному блужданию, не существует конечной инвариантной меры, тождественно не равной нулю, за исключением случая вырожденного распределения  $Y_k$ . Действительно, если такая мера существует для некоторого распределения  $Y_k$ , значит существует инвариантная вероятностная (стохастическая) мера  $\mathbb{P}$ , отличающаяся от заданной конечной инвариантной меры лишь на константу. Положим  $\mathbb{P}_{X_0} = \mathbb{P}$ . Тогда  $\mathbb{P}_{X_t} = \mathbb{P}_{X_0}$ , т.е. распределения  $X_t$  и  $X_0$  совпадают. Значит,  $X_0 + \sum_{k=1}^t Y_k$  и  $X_0$  одинаково распределены. Такое может быть только тогда, когда распределение  $Y_k$  вырожденное, т.е.  $\mathbb{P}(Y_k = a) = 1$ .

**Упражнение.** Ясно, что если мера  $\mu$  инвариантна относительно переходного ядра  $Q$ , то мера  $C\mu$  ( $C$  — константа) также инвариантна относительно  $Q$ . Привести пример переходного ядра, для которого существует неединственная инвариантная вероятностная (нормированная) мера.

Рассмотрим цепь Маркова с конечным числом состояний. ?????????????????????????????

Таким образом, нормированная инвариантная мера может:

- не существовать;
- быть не единственной.

Пусть для рассматриваемого однородного марковского процесса существует нормированная (вероятностная, стохастическая) мера относительно всех его переходных ядер  $\mathbb{P}^t$ . Обозначим ее через  $\pi$ . Таким образом,  $\pi \mathbb{P}^t = \pi$ ,  $\forall t$ ,  $\pi(\mathcal{X}) = 1$ . Пусть начальное распределение совпадает с этой мерой:  $\mathbb{P}_{X_0} = \pi$  ( $X_0 \sim \pi$ ). Из определения переходного ядра  $\mathbb{P}_{X_t} = \mathbb{P}_{X_0} \mathbb{P}^t$  ( $X_t \sim \mathbb{P}_{X_0} \mathbb{P}^t$ ). Значит, получим:

$$\mathbb{P}_{X_t} = \pi \mathbb{P}^t = \pi = \mathbb{P}_{X_0}.$$

**Теорема 27.** Если начальное распределение однородного марковского процесса есть вероятностная инвариантная мера относительно всех переходных ядер процесса, то этот процесс является стационарным.

**Доказательство.** Покажем, что все конечномерные распределения такого процесса инвариантны относительно сдвига. Итак, рассмотрим распределение  $\mathbb{P}_{(X_{t_0}, \dots, X_{t_n})}$ . Покажем, что оно совпадает с распределением  $\mathbb{P}_{(X_{t_0+h}, \dots, X_{t_n+h})}$ . Действительно, выше было показано, что  $\mathbb{P}_{X_{t_0}} = \mathbb{P}_{X_{t_0+h}}$ . С учетом однородности, получаем совпадение условных распределений, а значит, по теореме Тулча, совпадают и совместные распределения. Что и т.д.

Заметим, что для стационарного однородного марковского процесса могут существовать разные стационарные режимы (стационарный режим — это начальное распределение, другое название — стационарное распределение), т.к. нормированная инвариантная мера может быть неединственной.

Теорему можно переформулировать: если для однородного марковского процесса имеет место стационарное распределение, то этот процесс является стационарным.

Пуассоновский процесс не является стационарным.

Напомним, что однородный марковский процесс, определенный на конечном фазовом пространстве  $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ , задается с помощью:

1. матриц переходных вероятностей  $\{\mathbb{P}^t, t \in T\}$  ( $\mathbb{P}_{i,j}^t$  — вероятность перехода из состояния  $x_i$  в состояние  $x_j$  за время  $t$ ),
2.  $\alpha$  — начального распределения.

Тогда действие переходных ядер  $\mathbb{P}^t$  как линейных операторов на функцию  $f$  соответствует умножению матрицы  $\mathbb{P}^t$  на вектор-столбец  $f = (f(x_1), \dots, f(x_n))^T$ , а на меру  $\mu$  — умножению вектора-строки  $\mu = (\mu(x_1), \dots, \mu(x_n))^T$  на матрицу  $\mathbb{P}^t$ .

Уравнение Колмогорова-Чепмена для этого процесса:  $\mathbb{P}^t \cdot \mathbb{P}^s = \mathbb{P}^{s+t}$ , где “ $\cdot$ ” — это операция перемножения матриц.

Рассматриваемый процесс является стационарным, если  $\alpha = \alpha \mathbb{P}^t \forall t$  (умножение строки-распределения на матрицу).

Если время дискретно, то для проверки стационарности достаточно проверить  $\alpha = \alpha \mathbb{P}^1$ , т.к. из уравнения Колмогорова-Чепмена

$$\alpha = \alpha \mathbb{P}^1 \implies \alpha \mathbb{P}^{s+1} = \alpha \mathbb{P}^1 \mathbb{P}^s = \alpha \mathbb{P}^s \implies \alpha \mathbb{P}^{s+1} = \alpha \mathbb{P}^s = \dots = \alpha.$$

Таким образом  $\alpha$  — стационарная мера  $\iff \alpha$  — собственный вектор матрицы  $\mathbb{P}^1$ . Такой собственный вектор всегда существует (может быть неединственен), значит для конечного числа состояний всегда существует стационарный режим. ?????????????

### 13.5 Эргодичность.

Углубляться в теория эргодичности мы не будем, а дадим общее представление (на метаязыке) о том, что понимается под эргодичностью.

Свойство *эргодичности* (как его трактуют физики): среднее по ансамблю совпадает (близко) со средним по времени.

Пояснение свойства эргодичности: доля времени, которое процесс проводит во множестве  $A$  за время  $t$ , (среднее по времени) при  $t \rightarrow \infty$  приближается (в каком-то смысле) к вероятности нахождения процесса в этом множестве при  $t \rightarrow \infty$  (среднее по ансамблю):

$$\frac{1}{t} \sum_{s=1}^t \mathbf{I}_A(X_s) \text{ приближается при } t \rightarrow \infty \text{ к } \pi(A);$$

$$\frac{1}{t} \int_1^t \mathbf{I}_A(X_s) ds \text{ приближается при } t \rightarrow \infty \text{ к } \pi(A).$$

Свойство эргодичности выполнено, конечно же, не всегда. К тому же мы не указали вид сходимости.

Рассмотрим марковский процесс, для которого выполнено свойство эргодичности (в эргодической теории марковских процессов обычно рассматривают сходимость по норме полной вариации). Тогда  $\mathbb{P}_{X_t} \rightarrow \pi$  при  $t \rightarrow \infty$ . Выше было показано, что для однородных марковских процессов  $\mathbb{P}_{X_t} = \alpha \mathbb{P}^t$ , где  $\alpha$  — начальное распределение. Значит,  $\alpha \mathbb{P}^t \rightarrow \pi$  при  $t \rightarrow \infty$ . Тогда для стационарных процессов  $\pi = \alpha$  — единственная инвариантная нормированная мера. Заметим, что неединственной она быть не может:  $\pi_1 \mathbb{P}^t = \pi_1 \rightarrow \pi$ ,  $\pi_2 \mathbb{P}^t = \pi_2 \rightarrow \pi$ , т.е.  $\pi_1 = \pi_2 = \pi$ .



### 13.6 Винеровский процесс.

Рассмотрим произвольный случайный процесс  $X_t$ ,  $t \in T = [0, +\infty)$ ,  $(\mathcal{X}, \mathcal{A})$  (пространство  $\mathcal{X}$ , вообще говоря, произвольное линейное пространство, не обязательно  $\mathbb{R}^n$ ).

**Определение 115.** Процесс  $X_t$ ,  $t \in T = [0, +\infty)$ , называется **процессом с независимыми приращениями**, если для любых  $0 \leq t_1 < t_2 < \dots < t_n$  случайные элементы (со значениями в  $\mathcal{X}$ )  $X_{t_0}$ ,  $X_{t_1} - X_{t_0}$ ,  $\dots$ ,  $X_{t_n} - X_{t_{n-1}}$  независимы в совокупности.

**Утверждение 19.** Процесс с независимыми приращениями обладает марковским свойством.

Доказательство. Определение марковского процесса:  $\forall t$ :

$$\forall 0 \leq t_1 < t_2 < \dots < t_n < t \quad \forall h > 0 \quad \mathbb{P}_{X_{t+h}|X_t=x, X_{t_1}=x_1, X_{t_2}=x_2, \dots, X_{t_n}=x_n} = \mathbb{P}_{X_{t+h}|X_t=x}.$$

По определению процесса с независимыми приращениями случайные элементы  $X_{t_i}$ ,  $X_t - X_{t_i}$ ,  $X_{t+h} - X_t$  независимы, а значит,  $X_{t+h} - X_t$  и  $X_{t_i}$  независимы. Следовательно,

$$\mathbb{P}_{X_{t+h}|X_t=x, X_{t_i}=x_i} = \mathbb{P}_{X_{t+h}-X_t+x|X_t=x, X_{t_i}=x_i} = \mathbb{P}_{X_{t+h}-X_t+x|X_t=x} = \mathbb{P}_{X_{t+h}|X_t=x},$$

откуда и следует марковское свойство.

Заметим, что процесс с независимыми приращениями однородным, вообще говоря, не является.

Переходные ядра процесса с независимыми приращениями имеют вид:???????

**Определение 116.** Процесс  $X_t$ ,  $t \in T = [0, +\infty)$ , называется **однородным по времени процессом**, если распределения приращений  $X_{t+h} - X_t$  не зависят от  $t$  для  $\forall h > 0$ .

Однородный по времени процесс с независимыми приращениями называется *процессом с независимыми однородными приращениями*.

**Замечание.** Рассмотрим марковский процесс. Тогда приращения  $X_{t+h} - X_t$  не зависят от  $t \iff \mathbb{P}^{t, t+h}(x, A)$  не зависит от  $t$  (т.е. то определение однородности по времени, которое мы давали для марковского процесса, эквивалентно приведенному выше определению однородных по времени процессов в случае наличия марковского свойства). Докажем это. ?????????

**Замечание.** Пусть  $X_t$  — процесс с независимыми однородными приращениями, случайная величина  $X_{t+h} - X_t$  имеет распределение  $Q_h$  ( $\mathbb{P}_{X_{t+h}-X_t} = Q_h$ ). В силу доказанного выше марковского свойства такого процесса выполнено уравнение Колмогорова-Чепмена:

$$\int_{\mathcal{X}} \mathbb{P}^{t, t+h_1}(x, dy) \mathbb{P}^{t+h_1, t+h_1+h_2}(y, A) = \mathbb{P}^{t, t+h_1+h_2}(x, A).$$

Приращения процесса независимы и однородны, значит:

$$\begin{aligned} \mathbb{P}^{t, t+h}(x, A) &= \mathbb{P}(X_{t+h} \in A \mid X_t = x) = \mathbb{P}(X_{t+h} - X_t \in A - x \mid X_t = x) = \\ &= \{\text{независимость приращений}\} = \mathbb{P}(X_{t+h} - X_t \in A - x) = Q_h(A - x). \end{aligned}$$

Уравнение Колмогорова-Чепмена перепишется:

$$\int_{\mathcal{X}} Q_{h_1}(dy - x) Q_{h_2}(A - y) = Q_{h_1+h_2}(A - x).$$



Обозначим  $B = A - x$ ,  $z = y - x$ . Тогда

$$\int_{\mathcal{X}} Q_{h_1}(dz) Q_{h_2}(B - z) = Q_{h_1+h_2}(B).$$

Таким образом, уравнение Колмогорова-Чепмена для процесс с независимыми однородными приращениями имеет вид:

$$Q_{h_1} * Q_{h_2} = Q_{h_1+h_2} \quad (``*'' \text{ — операция свертки приращений}). \quad (63)$$

Следствие:  $Q_h$  — безгранично делимое распределение (т.е.  $\forall n \in \mathbb{N}$   $Q_h$  является  $n$ -кратной сверткой некоторого распределения  $\tilde{Q}_n$ :  $Q_h = \tilde{Q}_n^{*n}$ , в данном случае в качестве распределения  $\tilde{Q}_n$  берется  $Q_{\frac{h}{n}}$ ).

Заметим, что равенство (63), вообще говоря, очевидно и так, ведь распределение суммы независимых случайных величин есть свертка распределений.

Напомним, что в пространстве Скорохода эквивалентные процессы траекторно совпадают.

**Теорема 28.** *Процесс с независимыми однородными приращениями (ПНОП) может быть реализован в пространстве Скорохода  $D[0, \infty) = \{f: [0, \infty) \rightarrow \mathcal{X}: \forall \tau \exists \lim_{\substack{t \rightarrow \tau-0 \\ t \rightarrow \tau+0}} f(t)\}$  функций без разрывов второго рода. Другими словами: если дан ПНОП (т.е. фиксированы все конечномерные распределения), то существует его модификация в пространстве Скорохода. Другими словами: в пространстве Скорохода всегда можно задать любой ПНОП.*

Итак, рассмотрим “хорошую” модификацию ПНОП, т.е. модификацию, принадлежащую пространству Скорохода<sup>120</sup>.

**Определение 117.** *Процессом броуновского движения* называется процесс с независимыми однородными приращениями, модификация которого в пространстве Скорохода имеет непрерывную траекторию (т.е. отсутствие скачков).

**Утверждение 20.** *Для ПНОП условие непрерывности траекторий эквивалентно условию  $X_{t+h} - X_t \sim Q_h = \mathcal{N}(m(h), b(h))$ .*

Это утверждение мы примем без доказательства.

Как было показано ранее, для ПНОП справедливо соотношение  $Q_{h_1} * Q_{h_2} = Q_{h_1+h_2}$ .

---

<sup>120</sup>приведем аналогию с плотностями: при рассмотрении абсолютно непрерывного распределения берем непрерывный вариант плотности

Тогда:

$$\begin{aligned}
m(h_1) &= \mathbb{E}(X_{h_1} - X_0) = \int x Q_{h_1}(dx), \\
m(h_2) &= \mathbb{E}(X_{h_2} - X_0) = \int x Q_{h_2}(dx), \\
m(h_1 + h_2) &= \mathbb{E}(X_{h_1+h_2} - X_0) = \int x Q_{h_1+h_2}(dx) = \{Q_{h_1} * Q_{h_2} = Q_{h_1+h_2}\} = \\
&= \int x \int Q_{h_1}(dx - y) Q_{h_2}(dy) = \{z = x - y\} = \iint (z + y) Q_{h_1}(dz) Q_{h_2}(dy) = \\
&= \int z \int Q_{h_1}(dz) Q_{h_2}(dy) + \int y \int Q_{h_1}(dz) Q_{h_2}(dy) = \\
&= \int z Q_{h_1}(dz) \int Q_{h_2}(dy) + \int y Q_{h_1}(dy) \int Q_{h_2}(dz) = \\
&= \int z Q_{h_1}(dz) + \int y Q_{h_1}(dy) = m(h_1) + m(h_2); \\
b(h_1) &= \mathbb{V}\text{ar}(X_{h_1} - X_0) = \mathbb{E}(X_{h_1} - X_0 - m(h_1))^2 = \int (x - m_{h_1})^2 Q_{h_1}(dx), \\
b(h_2) &= \mathbb{V}\text{ar}(X_{h_2} - X_0) = \mathbb{E}(X_{h_2} - X_0 - m(h_2))^2 = \int (x - m_{h_2})^2 Q_{h_2}(dx), \\
b(h_1 + h_2) &= \mathbb{V}\text{ar}(X_{h_1+h_2} - X_0) = \mathbb{E}(X_{h_1+h_2} - X_0 - m(h_1 + h_2))^2 = \int (x - m_{h_1+h_2})^2 Q_{h_1+h_2}(dx) = \\
&= \int (x - m_{h_1} - m_{h_2})^2 \int Q_{h_1}(dx - y) Q_{h_2}(dy) = \{z = x - y\} = \\
&= \iint (y + z - m_{h_1} - m_{h_2})^2 Q_{h_1}(dz) Q_{h_2}(dy) = \\
&= \iint [(y - m_{h_2})^2 + (z - m_{h_1})^2 + 2(y - m_{h_2})(z - m_{h_1})] Q_{h_1}(dz) Q_{h_2}(dy) = \\
&= \int (y - m_{h_2})^2 Q_{h_2}(dy) + \int (z - m_{h_1})^2 Q_{h_1}(dz) + 2 \int (y - m_{h_2}) Q_{h_2}(dy) \int (z - m_{h_1}) Q_{h_1}(dz) = \\
&= b(h_1) + b(h_2).
\end{aligned}$$

Итак, из  $Q_{h_1} * Q_{h_2} = Q_{h_1+h_2}$  следует, что  $m(h_1 + h_2) = m(h_1) + m(h_2)$ ,  $b(h_1 + h_2) = b(h_1) + b(h_2)$  (это можно вывести проще из факта однородности и независимости приращений). Если наложить дополнительное условие измеримости<sup>121</sup> функции математического ожидания  $m(h)$  (функция дисперсии  $b(h)$  в силу своей неотрицательности монотонна:  $b(h_1 + h_2) = b(h_1) + b(h_2) \geq b(h_1)$ , а значит измерима), то решение данных уравнений единственно (с точностью до мультипликативной константы)<sup>122</sup>:  $m(h) = \mu h$ ,  $b(h) = \sigma^2 h$ .

Итак, для ПНОП с непрерывными траекториями в пространстве Скорохода (процесса броуновского движения)  $X_{t+h} - X_t \sim Q_h = \mathcal{N}(\mu h, \sigma^2 h)$ .

**Определение 118.** *Винеровским процессом (стандартным броуновским движением)* называется процесс с независимыми однородными приращениями, имеющий непрерывные траектории, причем  $\mu = 0$ ,  $\sigma = 1$ ,  $\mathbb{P}(X_0 = 0) = 1$ .

<sup>121</sup>вообще, вопрос о единственности решения зависит от аксиоматики теории множеств, т.е. если принять аксиому детерминированности, то неизмеримых функций просто не существует, так как не существует неизмеримых множеств; если принять аксиому выбора, то существует второе неизмеримое решение

<sup>122</sup>это мы доказывать не будем

Как известно, ПНОП является однородным по времени марковским процессом. Исследуем процесс броуновского движения на наличие стационарных режимов. Пусть  $X_0$  имеет вырожденное распределение:  $\mathbb{P}(X_0 = x) = 1$ . Тогда, в силу  $X_{t+h} - X_t \sim Q_h = \mathcal{N}(\mu h, \sigma^2 h)$ , получим  $X_t = (X_t - X_0) + X_0 \sim \mathcal{N}(\mu t + x, \sigma^2 t)$ . Даже если  $\mu = 0$  и математическое ожидание постоянно, при  $\sigma > 0$  дисперсия растет в любом случае, что противоречит определению стационарного процесса. Если распределение  $X_0$  гауссовское  $X_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$ , то в силу независимости приращений  $X_t = (X_t - X_0) + X_0 \sim \mathcal{N}(\mu t + \mu_0, \sigma^2 t + \sigma_0^2)$ , получаем ту же картину. Таким образом, только если распределение  $Q_h$  сосредоточено в нуле ( $\mathbb{P}(X_{t+h} = X_t) = 1$ ), то получаем стационарный процесс, причем стационарных режимов бесконечно много (для любой точки  $x$   $\mathbb{P}(X_0 = x) = 1$  — стационарный режим).

Таким образом, винеровский процесс стационарным не является.

Можно дать другое определение винеровского процесса.

**Определение 119.** *Винеровским процессом* называется гауссовский процесс со средним 0 и ковариационной функцией  $k(s, t) = \text{cov}(X_s, X_t) = \min(s, t)$ .

Покажем эквивалентность определений винеровского процесса.

1. Рассмотрим ПНОП с  $X_{t+h} - X_t \sim Q_h = \mathcal{N}(0, h)$ ,  $\mathbb{P}(X_0 = 0) = 1$ .

Рассмотрим конечномерные распределения этого процесса.

Для  $0 = t_0 < t_1 < t_2 < \dots < t_n$  рассмотрим случайный вектор  $\delta X = (X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}})$ . Он имеет плотность распределения

$$\begin{aligned} p_{\delta X}(y_1, y_2, \dots, y_n) &= \{\text{независимость приращений}\} = \prod_{i=1}^n p_{X_{t_i} - X_{t_{i-1}}}(y_i) = \\ &= \{X_t - X_s \sim \mathcal{N}(0, t - s)\} = \prod_{i=1}^n \left[ \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} e^{-\frac{y_i^2}{2(t_i - t_{i-1})}} \right] = \\ &= \prod_{i=1}^n \left[ \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \right] e^{-\sum_{i=1}^n \frac{y_i^2}{2(t_i - t_{i-1})}}. \end{aligned}$$

Учтем, что  $\mathbb{P}(X_0 = 0) = 1$ .

$$\begin{aligned} p_{\delta X}(y_1, y_2, \dots, y_n) &= {}^{123}\mathbb{P}(X_{t_1} - X_{t_0} = y_1, X_{t_2} - X_{t_1} = y_2, \dots, X_{t_n} - X_{t_{n-1}} = y_n) = \\ &= \mathbb{P}(X_{t_1} = y_1, X_{t_2} = y_2 + y_1, \dots, X_{t_n} = y_n + y_{n-1}) = \\ &= p_{(X_{t_1}, \dots, X_{t_n})}(y_1, y_2 + y_1, \dots, y_n + y_{n-1}). \end{aligned}$$

Тогда получим

$$\begin{aligned} p_{(X_{t_1}, \dots, X_{t_n})}(x_1, x_2, \dots, x_n) &= p_{\delta X}(x_1, x_2 - x_1, \dots, x_n - x_{n-1}) = \\ &= \prod_{i=1}^n \left[ \frac{1}{\sqrt{2\pi(t_i - t_{i-1})}} \right] e^{-\sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{2(t_i - t_{i-1})}} = \\ &= \frac{1}{\sqrt{(2\pi)^n \det(R)}} e^{-\frac{(x-m)' R^{-1} (x-m)}{2}}, \end{aligned}$$

где  $x = (x_1, x_2, \dots, x_n)'$ ,  $m = 0$ ,  $R = \{\min(t_i, t_j)\}_{i,j=1}^n$ .

Получили гауссовский процесс со средним 0 и ковариационной функцией  $\text{cov}(X_s, X_t) = \min(s, t)$ .

<sup>123</sup>да, запись не совсем корректна, но суть-то верна;

2. Рассмотрим гауссовский процесс  $X_t$  со средним 0 и ковариационной функцией  $k(s, t) = \min(s, t)$ .

Тогда  $X_0 \sim \mathcal{N}(0, 0)$ , т.е.  $\mathbb{P}(X_0 = 0) = 1$ .

Покажем независимость приращений  $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ . В силу нормальности процесса достаточно показать попарную некоррелированность приращений:

$$\begin{aligned} & \text{cov}(X_{t_i} - X_{t_{i-1}}, X_{t_j} - X_{t_{j-1}}) = \\ &= \mathbb{E}[(X_{t_i} - X_{t_{i-1}} - \mathbb{E}(X_{t_i} - X_{t_{i-1}}))(X_{t_j} - X_{t_{j-1}} - \mathbb{E}(X_{t_j} - X_{t_{j-1}}))] = \\ &= \{\mathbb{E}X_t = 0\} = \mathbb{E}[(X_{t_i} - X_{t_{i-1}})(X_{t_j} - X_{t_{j-1}})] = \\ &= \mathbb{E}(X_{t_i}X_{t_j}) - \mathbb{E}(X_{t_{i-1}}X_{t_j}) - \mathbb{E}(X_{t_i}X_{t_{j-1}}) + \mathbb{E}(X_{t_{i-1}}X_{t_{j-1}}) = \\ &= k(t_i, t_j) - k(t_{i-1}, t_j) - k(t_i, t_{j-1}) + k(t_{i-1}, t_{j-1}) = \{\text{без ограничения общности } i > j\} = \\ &= t_j - t_j - t_{j-1} + t_{j-1} = 0. \end{aligned}$$

Ясно, что если  $X_t \sim \mathcal{N}(0, t)$ , то  $X_{t+h} - X_t \sim \mathcal{N}(0, h)$ , т.к.  $\text{Var}(X_{t+h} - X_t) = \mathbb{E}(X_{t+h} - X_t)^2 = \text{Var}X_{t+h} + \text{Var}X_t - 2\mathbb{E}X_{t+h}X_t = t + h + t - 2\text{cov}(X_{t+h}, X_t) = t + h + t - 2t = h$ . Отсюда следует однородность процесса.

Эквивалентность определений доказана.

## 13.7 Процесс Орнштейна-Уленбека.

**Определение 120.** *Процессом Орнштейна-Уленбека* называется марковский гауссовский стационарный процесс.

Можно также дать другие эквивалентные определения процесса Орнштейна-Уленбека. Например, через математическое ожидание и дисперсию гауссовского процесса или через переходные вероятности  $\mathbb{P}^{t, t+h}$  и начальное распределение для марковского процесса.

Рассмотрим понятие симметричности процесса, т.е. вопрос о сохранении свойств процесса, если время пускают в обратную сторону. Другими словами, сравним процессы  $X_t$ ,  $t \in [0, T]$ , и  $Y_t = X_{T-t}$ ,  $t \in [0, T]$ .

Напомним, что марковское свойство может быть сформулировано симметрично относительно будущего и прошлого: если известно настоящее, то условные (относительно настоящего) распределения процесса в прошлом и будущем не зависимы. Поэтому марковское свойство при обращении времени сохраняется: если  $X_t$  — марковский процесс, то и  $Y_t$  — марковский процесс.

А вот свойство однородности процесса при таком преобразовании не сохраняется. Например, винеровский процесс: процесс в прямом времени стартует из нуля, значит, процесс в обратном времени должен прийти в ноль; процесс “загоняют” в ноль переходные вероятности  $\mathbb{P}^{t, t+h}$ , которые поэтому зависят от  $t$ .

Свойства стационарности и гауссовости при обращении времени, очевидно, сохраняются.

Получаем, что если  $X_t$  — марковский гауссовский стационарный процесс (процесс Орнштейна-Уленбека), то и  $Y_t$  — марковский гауссовский стационарный процесс (процесс Орнштейна-Уленбека). Заметим, что параметры процесса (математическое ожидание  $m$ , дисперсия  $\sigma^2$  ( $X_t \sim \mathcal{N}(m, \sigma^2)$ ), параметр релаксации  $\lambda$  (скорость убывания ковариационной функции  $k(0, t) = \text{cov}(X_0, X_t)$  при росте  $t$ ) при этом также сохраняются. Таким образом, процесс Орнштейна-Уленбека является симметричным, т.е. переходные вероятности в обратном времени совпадают с переходными вероятностями в прямом времени.

## 14 Виды сходимости случайных величин: по вероятности, в среднем, почти наверное, по распределению - слабая и по вариации.

Рассматривают два класса сходимостей случайных величин:

- I. Виды сходимостей, характеризующие близость случайных величин (сходимость почти наверное, по вероятности, в среднем порядка  $p$ , почти равномерно).
- II. Виды сходимостей, характеризующие близость распределений<sup>124</sup> (слабая сходимость, сходимость по норме полной вариации).

Из близости случайных величин вытекает близость в некотором смысле их распределений (это будет показано при рассмотрении минимальной метрики), однако при наличии близости распределений некорректно, вообще говоря, говорить о близости случайных величин, т.к. случайные величины с одинаковыми распределениями могут быть определены на разных пространствах, т.е. быть несравнимыми. Следующий пример показывает, что даже если случайные величины определены на одном пространстве, близость распределений отнюдь не влечет близость самих случайных величин.

**Пример.** Рассмотрим вероятностное пространство  $(\Omega, \mathcal{F}, \mathbb{P})$ :  $\Omega = [0, 1]$ ,  $\mathcal{F} = \mathcal{B}_{[0,1]}$ ,  $\mathbb{P}$  — мера Лебега на  $[0, 1]$ .<sup>125</sup> Рассмотрим последовательность случайных величин  $X_n$ ,  $n = 1, 2, \dots$ :

$$X_{2k-1}(\omega) = \begin{cases} 1, & \omega \in [0, \frac{1}{2}], \\ 0, & \omega \in (\frac{1}{2}, 1], \end{cases} \quad X_{2k}(\omega) = \begin{cases} 0, & \omega \in [0, \frac{1}{2}], \\ 1, & \omega \in (\frac{1}{2}, 1]. \end{cases}$$

Все случайные величины имеют одинаковое распределение (Бернулли с параметром  $p = \frac{1}{2}$ ), поэтому сходимость их распределений в любом смысле очевидна, однако сами случайные величины не сходятся ни в каком смысле (см. ниже).

Итак, сходимость распределений не влечет в общем случае сходимость случайных величин, однако существует теорема Скорохода, точная формулировка которой будет приведена ниже, которая говорит, что если распределения близки в слабой топологии, то можно указать вероятностное пространство и случайные величины, которые будут иметь заданные распределения и будут близки.

### I. Сходимости случайных величин.

#### 1. Почти наверное.

Последовательность случайных величин  $X_n$  сходится почти наверное к случайной величине  $X$ , если

$$\mathbb{P}(\omega: X_n(\omega) \xrightarrow{n \rightarrow \infty} X(\omega)) = 1.$$

#### 2. По вероятности

Последовательность случайных величин  $X_n$  сходится по вероятности к случайной величине  $X$ , если

$$\forall a > 0 \quad \mathbb{P}(|X_n - X| > a) \xrightarrow{n \rightarrow \infty} 0.$$

<sup>124</sup>не являются в строгом смысле сходимостями случайных величин

<sup>125</sup>это пространство будем всегда рассматривать по умолчанию

3. В среднем порядка  $p$ .

Последовательность случайных величин  $X_n$  сходится в среднем порядка  $p \geq 1$  к случайной величине  $X$ , если

$$\mathbb{E}|X_n - X|^p \xrightarrow{n \rightarrow \infty} 0.$$

Сходимость случайных величин в среднем порядка  $p \geq 1$  есть сходимость их как функций в пространстве  $L_p$ .

4. Почти равномерно.

Последовательность случайных величин  $X_n$  сходится почти равномерно к случайной величине  $X$ , если

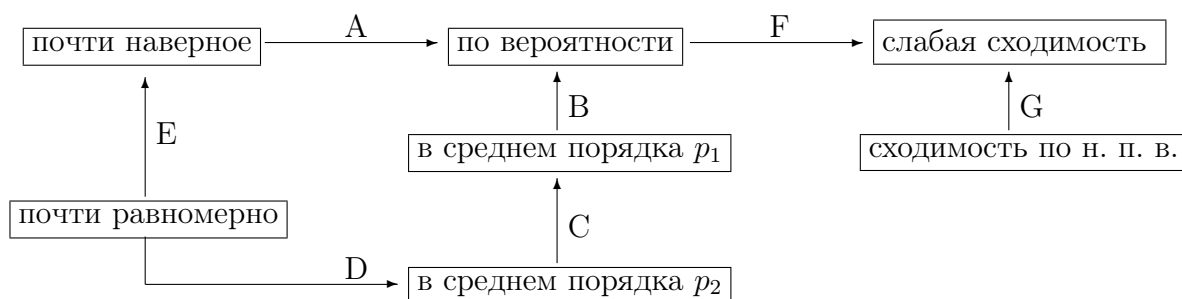
$$\text{vgrainmax} |X_n - X| \xrightarrow{n \rightarrow \infty} 0.$$

**Определение 121.** **Существенным супремумом** случайной величины  $X$  называется величина (конечное число или  $+\infty$ )

$$\text{vgrainmax} X = \inf \{a \in \mathbb{R} : \mathbb{P}(\omega : X > a) = 0\}.$$

Сходимость случайных величин почти равномерно есть сходимость их как функций в пространстве  $L_\infty$ .

Приведенный ниже рисунок иллюстрирует связь между различными видами сходимостей. Из него видно, что сходимость почти равномерно является самой сильной, а слабая сходимость — самой слабой.



Рассмотрим связи между различными видами сходимостей случайных величин.

А: Из сходимости почти наверное следует сходимость по вероятности: доказывалось в курсе функционального анализа. В обратную сторону это не выполнено. Однако из любой последовательности, сходящейся по вероятности, можно выделить сходящуюся почти наверное подпоследовательность (доказывалось в курсе функционального анализа).

**Пример** (Серия бегущих импульсов). В каждой серии импульс движется слева направо, становясь с каждой серией тоньше.

Обозначим  $I_n^k$  — интервал, на котором  $k$ -я случайная величина из  $n$ -й серии равна 1. Очевидно,  $|I_n^k| = \frac{1}{2^n}$ .

—  $X_n \xrightarrow{n \rightarrow \infty} X \equiv 0$  по вероятности.

$$\mathbb{P}(|X_n - X| > a) = \mathbb{P}(X_n = 1) = |I_n^k| = \frac{1}{2^n} \xrightarrow{n \rightarrow \infty} 0.$$



Рис. 10: 1 серия.

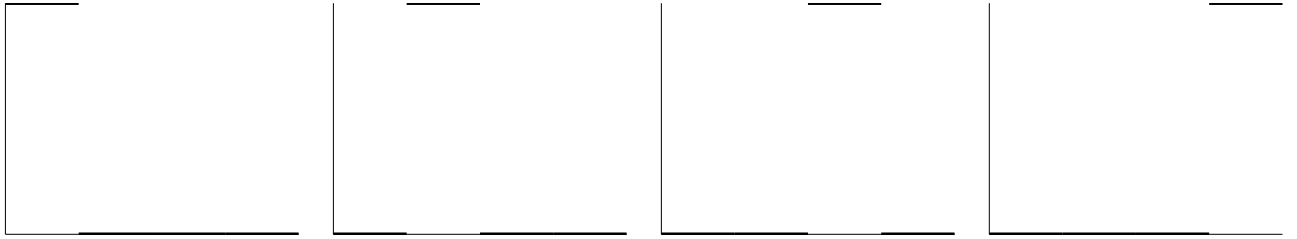


Рис. 11: 2 серия.

- $X_n$  не сходится к  $X \equiv 0$  почти наверное. Действительно, для любого  $\omega$  в каждой серии существует случайная величина, принимающая значение 0 на этом  $\omega$ , и существует случайная величина, принимающая значение 1 на этом  $\omega$ , значит,  $\mathbb{P}(\omega: X_n(\omega) \xrightarrow{n \rightarrow \infty} X(\omega)) = 0$ .

Однако можно выделить подпоследовательность случайных величин, сходящуюся почти наверное к нулю, например, первый импульс в каждой серии.

В: Из сходимости в среднем порядка  $p \geq 1$  следует сходимость по вероятности. Это следует из неравенства Маркова:

$$\mathbb{P}(|X_n - X| > a) \leq \frac{\mathbb{E}|X_n - X|^p}{a^p}.$$

В обратную сторону неверно, т.к. во-первых, соответствующее математическое ожидание может не существовать. Но даже если оно существует, можно привести пример, когда имеет место сходимость по вероятности и не имеет места сходимость в среднем порядка  $p$ :

$$X_n = \begin{cases} n^2, & \omega \in [0, \frac{1}{n}], \\ 0, & \omega \in (\frac{1}{n}, 1]. \end{cases}, \quad X \equiv 0.$$

$\mathbb{P}(|X_n| > a) = \mathbb{P}(X_n = n^2) = \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0$ , значит, имеет место сходимость по вероятности, однако,  $\mathbb{E}|X_n|^p = n^{2p-1} \xrightarrow{n \rightarrow \infty} \infty$ , значит, сходимости в среднем порядка  $p \geq 1$  нет.

С: Из сходимости в среднем порядка  $p_2 > p_1$  следует сходимость в среднем порядка  $p_1 \geq 1$ . Это вытекает из следующего утверждения:

**Утверждение 21.** Функция

$$p \rightarrow \|X\|_p = (\mathbb{E}|X|^p)^{\frac{1}{p}}.$$



является неубывающей ( $p \geq 1$ ).

**Доказательство.** Рассмотрим произвольные  $p_2 > p_1 \geq 1$ . Воспользуемся неравенством Йенсена: для любой выпуклой функции  $g(x)$

$$g(\mathbb{E}Y) \leq \mathbb{E}g(Y).$$

Возьмем  $Y = |X_n - X|^{p_1}$ ,  $g(x) = x^{\frac{p_2}{p_1}}$ . Получим

$$(\mathbb{E}|X_n - X|^{p_1})^{\frac{p_2}{p_1}} \leq \mathbb{E}|X_n - X|^{p_2}.$$

Извлекая корень степени  $p_2$  из этого неравенства, получаем требуемое утверждение.  $\square$ .

В обратную сторону неверно: например, последовательность

$$X_n = \begin{cases} \frac{1}{\sqrt{\omega}}, & \omega \in [0, \frac{1}{n}], \\ 0, & \omega \in (\frac{1}{n}, 1] \end{cases}$$

сходится к нулю в среднем порядка 1, но не сходится в среднем порядка 2.

D: Из сходимости почти равномерно следует сходимость в среднем порядка  $p \geq 1$ . Это следует из того, что

$$\|X\|_p \leq \|X\|_\infty = \text{vgrainmax } X.$$

Действительно, из определения существенного супремума следует, что  $|X| \leq \text{vgrainmax } X$  почти всюду, значит

$$\|X\|_p = \left( \int_{\Omega} |X(\omega)|^p d\mathbb{P}(\omega) \right)^{\frac{1}{p}} \leq \left( \int_{\Omega} \|X(\omega)\|_\infty^p d\mathbb{P}(\omega) \right)^{\frac{1}{p}} = \|X\|_\infty \mathbb{P}(\Omega)^{\frac{1}{p}} = \|X\|_\infty.$$

В обратную сторону неверно: например, последовательность

$$X_n = \begin{cases} n, & \omega \in \left[0, \frac{1}{n^{p+1}}\right], \\ 0, & \omega \in \left(\frac{1}{n^{p+1}}, 1\right] \end{cases}$$

сходится к нулю в среднем порядка  $p$ , но не сходится почти равномерно.

G: Из сходимости почти равномерно следует сходимость почти наверное.

Рассмотрим множество

$$A = \{\omega: X_n(\omega) \not\rightarrow X(\omega)\}.$$

Запишем определение того, что  $X_n(\omega) \not\rightarrow X(\omega)$ :

$$\exists k: \forall N \exists n \geq N \quad |X_n(\omega) - X(\omega)| > \frac{1}{k}.$$

Значит, множество  $A$  можно представить в виде

$$A = \bigcup_k \bigcap_N \bigcup_{n \geq N} \left\{ |X_n(\omega) - X(\omega)| > \frac{1}{k} \right\} \quad k, N, n \in \mathbb{N}.$$

Покажем, что из почти равномерной сходимости следует, что  $\mathbb{P}(A) = 0$ .

Из того, что

$$\inf\{a \in \mathbb{R}: \mathbb{P}(|X_n - X| > a) = 0\} \xrightarrow{n \rightarrow \infty} 0$$

следует, что существует такая последовательность  $a_n \rightarrow 0$ , что  $\mathbb{P}(|X_n - X| > a_n) = 0$ . Фиксируем некоторое произвольное  $k$ . Для него существует номер  $N$ , такой что  $\forall n \geq N \ a_n < \frac{1}{k}$ , из чего следует

$$\mathbb{P}(|X_n - X| > \frac{1}{k}) \leq \mathbb{P}(|X_n - X| > a_n) = 0.$$

Из свойств вероятности:

$$\mathbb{P}\left(\bigcup_{n \geq N} \left\{|X_n(\omega) - X(\omega)| > \frac{1}{k}\right\}\right) \leq \sum_{n \geq N} \mathbb{P}\left(\left\{|X_n(\omega) - X(\omega)| > \frac{1}{k}\right\}\right) = 0. \quad (64)$$

Обозначим

$$B(N) = \bigcup_{n \geq N} \left\{|X_n(\omega) - X(\omega)| > \frac{1}{k}\right\}.$$

Очевидно,  $B(N) \supset B(N+1)$ . Из непрерывности вероятности

$$\mathbb{P}\left(\bigcap_N \bigcup_{n \geq N} \left\{|X_n(\omega) - X(\omega)| > \frac{1}{k}\right\}\right) = \lim_{N \rightarrow \infty} \mathbb{P}\left(\bigcup_{n \geq N} \left\{|X_n(\omega) - X(\omega)| > \frac{1}{k}\right\}\right) = 0.$$

Далее, применяя те же рассуждения для  $k$ , что и в (64), получаем

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_k \bigcap_N \bigcup_{n \geq N} \left\{|X_n(\omega) - X(\omega)| > \frac{1}{k}\right\}\right) = 0.$$

В обратную сторону неверно: например, последовательность  $X_n = \omega^n$  сходится почти наверное к нулю, но не сходится почти равномерно.

I. Сходимости распределений. Пусть распределения заданы на измеримом пространстве  $(E, \mathcal{E})$ .

#### 1. Слабая сходимость.

Пусть  $(E, \mathcal{E})$  — польское. Последовательность распределений  $Q_n$  сходится слабо к распределению  $Q$ , если для любой непрерывной и ограниченной функции  $f(x)$

$$\int f(x) Q_n(dx) \xrightarrow{n \rightarrow \infty} \int f(x) Q(dx). \quad (65)$$

**Замечание.** Можно рассматривать функции  $Q_n(x)$ ,  $Q(x)$  как элементы пространства  $BV(E)$  — пространства функций ограниченной вариации.  $BV(E)$  является сопряженным пространством к  $C(E)$  — пространству функций, непрерывных и ограниченных на  $E$ . Тогда данное выше определение можно рассматривать как определение сходимости функций  $Q_n$  в  $*$ -слабой топологии ( $*$ -слабую сходимость).

**Утверждение 22.**  $Q_n$  сходится к  $Q$  слабо тогда и только тогда, когда

$$\forall B: Q(\partial B) = 0 \quad Q_n(B) \xrightarrow{n \rightarrow \infty} Q(B).$$

**Доказательство.** Пусть  $Q_n \rightarrow Q$  слабо. Рассмотрим  $\forall B: Q(\partial B)$  и функцию  $f = \mathbb{I}_B$  — индикатор множества  $B$ . Она имеет разрыв только на границе множества  $B$ , т.е. на множестве меры нуль, а значит, является непрерывной почти всюду, и кроме того, ограниченной. Условие (65) для нее имеет вид  $Q_n(B) \xrightarrow{n \rightarrow \infty} Q(B)$ . В обратную сторону доказательство можно найти в книге Ширяев «Вероятность» (оно более сложно технически).  $\square$

**Утверждение 23.** Пусть  $E = \mathbb{R}$ ,  $\mathcal{E} = \mathcal{B}$ . Тогда слабая сходимость эквивалентна поточечной сходимости функций распределения  $F_n(x) \rightarrow F(x)$  во всех точках непрерывности  $F(x)$ .

**Доказательство.**

$$F_n(x) = Q_n((-\infty; x)) = Q_n(B).$$

$\partial B = x$ .  $x$  — точка непрерывности функции  $F(x)$  тогда и только тогда, когда  $Q(x) = 0$ . Значит, из предыдущего утверждения следует, что

$$F_n(x) = Q_n((-\infty; x)) = Q_n(B) \xrightarrow{n \rightarrow \infty} Q(B) = F(x).$$

В обратную сторону доказательство можно найти в книге Ширяев «Вероятность» (оно более сложно технически).  $\square$

**Теорема 29** (Скороход). Пусть  $Q_n \xrightarrow{n \rightarrow \infty} Q$  слабо. Тогда можно указать такое вероятностное пространство  $(\Omega, \mathcal{F}, \mathbb{P})$  и случайные величины  $X_n \sim Q_n$ ,  $X \sim Q$ , что  $X_n \xrightarrow{n \rightarrow \infty} X$  по вероятности.<sup>126</sup>

**Упражнение.** Доказать теорему Скорохода конструктивно для случая  $E = \mathbb{R}$ ,  $\mathcal{F} = \mathcal{B}$ .

Если имеются распределения, то можно построить функции распределения  $F_n, F$ :

$$F_n(x) = Q_n(-\infty; x), \quad F(x) = Q(-\infty; x).$$

Пусть  $\xi \sim U[0, 1]$ . Построим случайные величины  $X_n, X$  методом обращения функции распределения:

$$X_n(\omega) = F_n^{-1}(\xi(\omega)), \quad X(\omega) = F^{-1}(\xi(\omega)).$$

Здесь имеются в виду квазиобратные функции. Заметим, что «полочки» функции  $F$  соответствуют скачкам функции  $F^{-1}$ , и наоборот. И «полочек» и скачков у функции  $F$  не более счетного числа. Рассмотрим, кроме того, график функции распределения и соединим разрывы вертикальными отрезками. Получившуюся непрерывную кривую будем в течение данного доказательства для удобства называть «графиком» функции распределения.

Покажем, что  $X_n$  сходится к  $X$  почти равномерно, т.е. что  $\mathbb{P}(\omega: X_n \not\rightarrow X) = 0$ . Для этого покажем, что  $F_n^{-1}$  сходится к  $F^{-1}$  во всех точках непрерывности  $F^{-1}$ , т.е. во всех точках строгой монотонности  $F$ . Т.к. разрывов у  $F^{-1}$  не более счетного числа, и  $\mathbb{P}(\omega: \xi(\omega) = y_i, i = 1, 2, \dots) = 0$  ( $y_i$  — разрывы  $F^{-1}$ ), то этим будет доказано требуемое утверждение.

<sup>126</sup>В книге Булинский, Ширяев «Теория случайных процессов» утверждается, что имеет место сходимость почти наверное

Итак, возьмем  $y$  — точку непрерывности  $F^{-1}$ , пусть  $x = F^{-1}(y)$ , тогда  $F(x - 0) \leq y \leq F(x + 0)$ .

Теперь выберем произвольное  $\varepsilon$ . Мы хотим показать, что начиная с некоторого номера  $n$

$$|F_n^{-1}(y) - F^{-1}(y)| < \varepsilon. \quad (66)$$

Т.к. у  $F$  счетное число точек разрыва, то в  $\varepsilon$ -окрестности точки  $x$  существуют точки непрерывности  $F$   $x_-$ ,  $x_+$ :

$$x_- < x < x_+.$$

Т.к.  $x$  — точка строгой монотонности  $F$ , то

$$F(x_-) < F(x - 0) \leq F(x + 0) < F(x_+).$$

Далее, для того чтобы функции  $F_n$  сходились в этих точках к значениям функции  $F$ , необходимо, чтобы было выполнено

$$F_n(x + \varepsilon) > F(x + 0) \geq y, \quad F_n(x - \varepsilon) < F(x - 0) \leq y, \quad (67)$$

начиная с некоторого  $n$ . Иначе, если не выполнено, например, первое неравенство, то

$$F_n(x_+) \leq F_n(x + \varepsilon) \leq F(x + 0) < F(x_+)$$

для всех  $n$ , и нет сходимости  $F_n$  к  $F$  в точке  $x_+$ .

Легко проверяется, что (67) переписывается в виде

$$F_n^{-1}(y) < F^{-1}(y) + \varepsilon, \quad F_n^{-1}(y) > F^{-1}(y) - \varepsilon,$$

что эквивалентно выполнению (66).  $\square$

## 2. Сходимость по норме полной вариации.

**Определение 122.** Зарядом  $\mu$ , заданным на  $(E, \mathcal{E})$ , называется функция  $\mu: \mathcal{E} \rightarrow \mathbb{R}$ , т.ч.

$$\forall A = \sum_{k=1}^{\infty} A_k, \quad A_k \in \mathcal{E} \quad \mu(A) = \sum_{k=1}^{\infty} \mu(A_k).$$

Таким образом, заряд можно рассматривать как знакопеременную меру.

Рассмотрим линейный функционал  $\tilde{\mu}(f)$  на множестве измеримых равномерно ограниченных функций с нормой

$$\|f\|_{\infty} = \sup_{x \in E} |f(x)|.$$

$$\tilde{\mu}(f) = \int f d\mu.$$

**Норму полной вариации** заряда  $\mu$  определим как норму функционала  $\tilde{\mu}(f)$ :

$$\|\mu\| = \sup_{\|f\|_{\infty} \leq 1} \left| \int f d\mu \right|. \quad (68)$$

Итак, сходимость по норме полной вариации есть сходимость в данной норме.

**Определение 123.** Множество  $A$  называется *положительно (отрицательно) определенным относительно заряда  $\mu$* , если

$$\forall B \subset A, B \in \mathcal{E} \quad \mu(B) \geq 0 \quad (\mu(B) \leq 0).$$

Из курса функционального анализа известно, что пространство  $E$  можно представить в виде

$$E = E^+ + E^-,$$

где  $E^+, E^- \in \mathcal{E}$ ,  $E^+$  — положительно определенное множество,  $E^-$  — отрицательно определенное множество относительно  $\mu$ . Сам заряд  $\mu$  можно при этом представить в виде

$$\mu = \mu^+ + \mu^-,$$

где  $\mu^+, \mu^-$  —  $\sigma$ -аддитивные меры,  $\mu^+$  принимает ненулевые значения только на множествах из  $E^+$ , а  $\mu^-$  — на множествах из  $E^-$ .

Видно при этом, что супремум в (68) достигается на  $f = \mathbb{I}_{E^+} - \mathbb{I}_{E^-}$ .

**Упражнение.** Доказать, что если  $Q, P$  — вероятностные меры, т.е.  $\int \mathbb{I}_E dQ = 1$ ,  $\int \mathbb{I}_E dP = 1$ , то

$$\|Q - P\| = 2 \sup_{A \in \mathcal{E}} |Q(A) - P(A)|.$$

Обозначим  $\mu = Q - P$ . Очевидно, что

$$\|\mu\| = \mu(E^+) - \mu(E^-).$$

Из того, что  $Q, P$  — вероятностные меры, следует, что  $\mu(E^+) = -\mu(E^-)$ :

$$\mu(E^+) = Q(E^+) - P(E^+) = 1 - Q(E^-) - 1 + P(E^-) = -\mu(E^-).$$

Далее

$$\sup_{A \in E} \mu(A) = \sup_{A \in E^+} \mu(A) = \mu(E^+) = -\mu(E^-) = \sup_{A \in E^-} (-\mu(A)) = \sup_{A \in E} (-\mu(A)),$$

откуда следует, что  $\sup_{A \in E} \mu(A) = \sup_{A \in E} (-\mu(A)) = \sup_{A \in E} |\mu(A)|$ .

Значит,

$$\|\mu\| = \mu(E^+) - \mu(E^-) = \sup_{A \in E} \mu(A) + \sup_{A \in E} (-\mu(A)) = 2 \sup_{A \in E} |\mu(A)|,$$

что и требовалось доказать.  $\square$

Норма полной вариации может быть определена без всяких топологических предположений о вероятностном пространстве. Однако если пространство польское с борелевской  $\sigma$ -алгеброй, то из сходимости по норме полной вариации следует слабая сходимость (это доказывает стрелку G). Действительно, если  $\|Q_n - Q\| \rightarrow 0$ , то из приведенного выше утверждения следует, что  $Q_n(B) \rightarrow Q(B) \quad \forall B \in \mathcal{E}$ , а не только для множеств с «хорошей» границей, как в случае слабой сходимости.

Изучим теперь вопрос о метризуемости сходимостей.

**Определение 124.** Сходимость случайных величин называется **метризуемой**, если в пространстве  $E$  можно ввести метрику  $\rho(X, Y)$ , такую что

$$\rho(X_n, X) \xrightarrow{n \rightarrow \infty} 0 \iff X_n \xrightarrow{n \rightarrow \infty} X$$

в смысле рассматриваемой сходимости.

Сходимость распределений называется **метризуемой**, если в пространстве вероятностных распределений на  $(E, \mathcal{E})$  можно ввести метрику  $\rho(Q, P)$ , такую что

$$\rho(Q_n, Q) \xrightarrow{n \rightarrow \infty} 0 \iff Q_n \xrightarrow{n \rightarrow \infty} Q$$

в смысле рассматриваемой сходимости.

**Утверждение 24.** Все рассмотренные виды сходимостей случайных величин и распределений метризуемы, за исключением сходимости почти наверное.

**Доказательство.**

1. Сходимость в среднем порядка  $p$ .

Соответствующая метрика порождается нормой в  $L_p$ :

$$\rho(X, Y) = \|X - Y\|_p = (\mathbb{E}|X - Y|^p)^{\frac{1}{p}}.$$

Необходимо, однако, рассматривать величины с точностью до почти наверное, т.е. классы случайных величин, иначе не будет выполнено свойство

$$\rho(X, Y) = 0 \iff X = Y.$$

2. Сходимость почти равномерно.

Соответствующая метрика порождается нормой в  $L_\infty$ :

$$\rho(X, Y) = \|X - Y\|_\infty = \text{vraimax} |X - Y|.$$

Аналогичное замечание относительно классов случайных величин.

3. Сходимость по вероятности.

Соответствующая метрика есть метрика Ки Фан:

$$\rho(X, Y) = \inf\{a > 0: \mathbb{P}(|X - Y| > a) < a\}$$

Докажем, что это метрика. Достаточно проверить неравенство треугольника (остальные свойства очевидны). Рассмотрим произвольное  $\varepsilon$ . Из определения инфимума следует, что существуют такие  $a, b$ , что

$$\begin{aligned} \mathbb{P}(|X - Z| > a) < a, & \quad a < \rho(X, Z) + \frac{\varepsilon}{2}, \\ \mathbb{P}(|Y - Z| > b) < b, & \quad b < \rho(Y, Z) + \frac{\varepsilon}{2}. \end{aligned}$$

$$\begin{aligned} \mathbb{P}(|X - Y| > a + b) &\leq \mathbb{P}(|X - Z| + |Y - Z| > a + b) \leq \mathbb{P}(\{|X - Z| > a\} \cup \{|Y - Z| > b\}) \leq \\ &\leq \mathbb{P}(|X - Z| > a) + \mathbb{P}(|Y - Z| > b) < a + b < \rho(X, Z) + \rho(Y, Z) + \varepsilon. \end{aligned}$$

Значит,

$$\rho(X, Y) < \rho(X, Z) + \rho(Y, Z) + \varepsilon.$$

Из произвольности  $\varepsilon$  следует, что

$$\rho(X, Y) \leq \rho(X, Z) + \rho(Y, Z),$$

что и требовалось доказать.

Покажем, что  $\rho(X_n, X) \rightarrow 0 \iff X_n \rightarrow X$  по вероятности.

$\implies$ :  $\rho(X_n, X) \rightarrow 0$ . Значит, существует последовательность  $a_n$ , таких что  $a_n \rightarrow 0$  и  $\mathbb{P}(|X_n - X| > a_n) < a_n$ . Из того, что  $a_n \rightarrow 0$  следует, что

$$\forall a > 0, \forall k \exists N: \forall n \geq N \quad a_n < \frac{1}{k}, \quad a_n < a.$$

Значит,

$$\mathbb{P}(|X_n - X| > a) \leq \mathbb{P}(|X_n - X| > a_n) < a_n \leq \frac{1}{k}.$$

$\impliedby$ :  $\forall a > 0 \quad \mathbb{P}(|X_n - X| > a) \rightarrow 0$ . Значит,

$$\forall k \exists N: \forall n \geq N \quad \mathbb{P}(|X_n - X| > \frac{1}{k}) < \frac{1}{k}.$$

Следовательно,  $\rho(X_n, X) = \inf\{a > 0: \mathbb{P}(|X - Y| > a) < a\} < \frac{1}{k}$ .

#### 4. Слабая сходимость

**Определение 125.** Пусть задана метрика  $\rho(X, Y)$ , описывающая близость случайных величин. Тогда **индуцированной или минимальной по отношению к  $\rho$  метрикой** называется

$$r(Q, P) = \inf\{\rho(X, Y): X \sim Q, Y \sim P\}, \quad (69)$$

где инфимум берется по всем совместным распределениям  $P_{XY}$ , таким что его маргинальные распределения  $P_X \sim Q, P_Y \sim P$ .

Докажем, что это метрика. Будем обозначать  $\rho(X, Y) |_{P_{XY}}$  — расстояние между случайными величинами  $X$  и  $Y$ , если их совместное распределение  $P_{XY}$ . Рассмотрим произвольное  $\varepsilon$ . Из определения минимальной метрики следует, что существуют такие распределения  $P_{XZ}, P_{YZ}$ , что соответствующие им маргинальные распределения есть  $P_X, P_Y, P_Z$  и

$$\rho(X, Z) |_{P_{XZ}} < r(P_X, P_Z) + \frac{\varepsilon}{2}, \quad \rho(Y, Z) |_{P_{YZ}} < r(P_Y, P_Z) + \frac{\varepsilon}{2}.$$

Положим  $P_{XY} = P_X \cdot P_Y$  ( $X, Y$  — независимы). Тогда

$$r(P_X, P_Y) \leq \rho(X, Y) |_{P_{XY}} \leq \rho(X, Z) |_{P_{XZ}} + \rho(Y, Z) |_{P_{YZ}} \leq \rho(X, Z) + \rho(Y, Z) + \varepsilon.$$

В силу произвольности  $\varepsilon$  неравенство треугольника выполнено. Минимальная метрика метризует слабую сходимость.



**Пример.** Пусть близость случайных величин описывается метрикой в  $L_2$ . Тогда для определения  $\rho(X, Y)$  достаточно знать совместное распределение  $P_{XY}$  по теореме о замене переменных под знаком интеграла:

$$\rho(X, Y) = (\mathbb{E}(|X - Y|^2))^{\frac{1}{2}} = \left( \int |x - y|^2 P_{XY}(dx, dy) \right)^{\frac{1}{2}}.$$

Совместное распределение  $P_{XY}$ , на котором достигается минимум в (69), строится следующим образом. Разыгрывается случайная величина  $\xi \sim U[0, 1]$ .  $X$  и  $Y$  строятся методом обращения функции распределения, используя один аргумент:

$$X(\omega) = F_X^{-1}(\xi(\omega)), \quad Y(\omega) = F_Y^{-1}(\xi(\omega)).$$

Минимальной метрикой по отношению к метрике Ки Фан является метрика Леви-Прохорова, своеобразный аналог метрики Хаусдорфа. Это утверждение составляет содержание теоремы Штрассена.<sup>127</sup>

Пусть  $B \in \mathcal{E}$ ,  $\varepsilon > 0$ .  $\rho$  — некоторая метрика в пространстве  $E$ . (Когда мы обсуждали метрику Ки Фан, мы говорили о случайных величинах, т.е. считали, что  $E = \mathbb{R}$ ,  $\mathcal{E} = \mathbb{B}$ , и использовали стандартную метрику  $\rho(x, y) = |x - y|$ .) Обозначим

$$B^\varepsilon = \{x \in E : \rho(x, B) < \varepsilon\}, \quad \text{где } \rho(x, B) = \inf\{\rho(x, y), y \in B\}.$$

Тогда метрика Леви-Прохорова в пространстве вероятностных распределений на  $E$  вводится следующим образом:

$$\pi(Q, P) = \inf\{\varepsilon > 0 : P(B) \leq Q(B^\varepsilon) + \varepsilon, Q(B) \leq P(B^\varepsilon) + \varepsilon \text{ для } B \in \mathcal{E}\}.$$

Тем самым доказана стрелка  $F$ : если последовательность случайных величин сходится по вероятности, то она сходится в метрике Ки Фан, значит, соответствующая последовательность распределений сходится в метрике Леви-Прохорова, а значит, имеет место слабая сходимость.

Кроме того, на понятии минимальной метрики можно основать доказательство теоремы Скорохода в общем случае. Если имеет место слабая сходимость распределений  $Q_n \rightarrow Q$ , то  $\pi(Q_n, Q) \rightarrow 0$ , а значит, существуют совместные распределения  $P_{X_n, X}$ , т.ч.  $\rho(X_n, X) \rightarrow 0$ , где  $\rho$  — метрика Ки Фан, следовательно  $X_n \rightarrow X$  по вероятности.

## 5. Сходимость по норме полной вариации.

Соответствующая метрика порождается нормой полной вариации:

$$\rho(Q, P) = \|Q - P\|.$$

□

**Упражнение.** Доказать, что сходимость почти наверное не метризуема. Использовать утверждение о том, что из любой сходящейся по вероятности последовательности можно выделить сходящуюся почти наверное подпоследовательность

В заключение рассмотрим понятие топологического носителя распределения, тесно связанное с понятием существенного супремума.

<sup>127</sup>см. Булинский, Ширяев «Теория случайных процессов»

**Определение 126.** Точкой роста функции распределения называется точка, любая окрестность которой имеет положительную вероятность (меру).

**Определение 127.** Топологическим носителем распределения называется множество точек роста функции распределения.

Свойства топологического носителя:

1. Топологический носитель замкнут.

Действительно, пусть последовательность точек  $x_n$  из носителя распределения сходится к некоторой точке  $x$ . Тогда для любого  $\varepsilon$  существует точка  $x_n$  и число  $\tilde{\varepsilon}$ , что  $\tilde{\varepsilon}$ -окрестность точки  $x_n$  содержится в  $\varepsilon$ -окрестности точки  $x$ . Значит,  $\varepsilon$ -окрестность точки  $x$  имеет положительную меру.

2. Топологический носитель распределения есть наименьшее замкнутое множество полной меры.

Докажем, что дополнение к топологическому носителю есть наибольшее открытое множество меры ноль. Все точки, не являющиеся точками роста функции распределения, содержатся на ее «полочках». Мера каждой полочки — ноль, их число не более чем счетно, значит, мера рассматриваемого множества — ноль. Пусть существует открытое множество мощности большей, чем рассматриваемое. Значит, оно содержит, по крайней мере, одну точку роста, а вместе с ней, в силу открытости, и некоторую ее окрестность, и следовательно, имеет положительную меру.

3. Супремум топологического носителя есть наименьшее значение аргумента функции распределения, при котором функция распределения равна 1.

**Замечание.** Что такое  $\sup \emptyset$  — вообще говоря, вопрос конвенции. Нам удобно взять  $\sup \emptyset = +\infty$ .

Пусть  $x$  — супремум топологического носителя. В силу замкнутости носителя,  $x$  ему принадлежит. Множество  $(x, +\infty)$  имеет меру ноль по предыдущему свойству.

$$0 = \mathbb{P}(x, +\infty) = 1 - F(x). \quad \implies \quad F(x) = 1.$$

Пусть  $y < x$ ,  $F(y) = 1$ , тогда рассмотрим  $(\frac{x-y}{2})$ -окрестность точки  $x$ , получим, что она имеет нулевую меру.

Отсюда следует, что  $\text{vrai max } X$ , который представляет из себя наименьшую почти наверное верхнюю грань множества значений случайной величины  $X$  и есть супремум топологического носителя распределения  $X$ .