

# Слабая сходимость распределений. Основы теории и некоторые приложения

В.В. Некруткин

кафедра статистического моделирования <http://statmod.ru>, матмех СПбГУ  
Материал к специальному курсу, 2018 г.

## Содержание

|          |                                                                                 |           |
|----------|---------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Введение</b>                                                                 | <b>3</b>  |
| 1.1      | О сходимости распределений . . . . .                                            | 3         |
| 1.2      | О расстоянии по вариации . . . . .                                              | 3         |
| 1.3      | О слабой сходимости . . . . .                                                   | 6         |
| <b>2</b> | <b>Топологические и метрические пространства</b>                                | <b>10</b> |
| <b>3</b> | <b>Меры в измеримых метрических пространствах</b>                               | <b>13</b> |
| 3.1      | Измеримые топологические пространства . . . . .                                 | 13        |
| 3.2      | Распределения, определенные на борелевских $\sigma$ -алгебрах . . . . .         | 15        |
| <b>4</b> | <b>Слабая сходимость распределений. Основная теорема</b>                        | <b>18</b> |
| 4.1      | Определение и вспомогательные утверждения . . . . .                             | 18        |
| 4.2      | Основная теорема . . . . .                                                      | 20        |
| <b>5</b> | <b>Слабая сходимость и отображения</b>                                          | <b>22</b> |
| 5.1      | Сохранение слабой сходимости при отображениях . . . . .                         | 22        |
| 5.2      | Сходимость интегралов. Интеграл Римана . . . . .                                | 24        |
| <b>6</b> | <b>Равномерная интегрируемость функций и слабая сходимость</b>                  | <b>26</b> |
| 6.1      | Функции, равномерно интегрируемые с семейством распределений . . . . .          | 26        |
| 6.2      | Сходимость интегралов от неограниченных функций . . . . .                       | 28        |
| <b>7</b> | <b>Классы функций, определяющие слабую сходимость</b>                           | <b>30</b> |
| 7.1      | Общее утверждение . . . . .                                                     | 30        |
| 7.2      | Примеры: сходимость моментов и характеристических функций . . . . .             | 31        |
| 7.3      | О метризации слабой сходимости . . . . .                                        | 36        |
| <b>8</b> | <b>Сходимость по вариации и теорема Шеффе</b>                                   | <b>37</b> |
| 8.1      | Расстояние по вариации и полная вариация заряда . . . . .                       | 37        |
| 8.2      | Интегральные представления $\rho_{\text{var}}$ . Теорема Шеффе . . . . .        | 39        |
| <b>9</b> | <b>Несколько предельных теорем, используемых в статистике</b>                   | <b>41</b> |
| 9.1      | Сходимость по вероятности к константе. Общие утверждения . . . . .              | 41        |
| 9.2      | Замена мешающего параметра на его состоятельную оценку . . . . .                | 42        |
| 9.2.1    | Общее утверждение . . . . .                                                     | 42        |
| 9.2.2    | О доверительных интервалах . . . . .                                            | 44        |
| 9.2.3    | Сходимость к нормальному распределению . . . . .                                | 46        |
| 9.3      | Предельная теорема критерия $\chi^2$ . . . . .                                  | 52        |
| 9.3.1    | Многомерная теорема Леви . . . . .                                              | 52        |
| 9.3.2    | Предельная теорема критерия $\chi^2$ . . . . .                                  | 53        |
| 9.4      | Сохранение асимптотической нормальности при гладком отображении . . . . .       | 56        |
| 9.4.1    | Теорема о сохранении асимптотической нормальности . . . . .                     | 56        |
| 9.4.2    | Доверительный интервал для условной вероятности . . . . .                       | 58        |
| 9.4.3    | Выборочные характеристики, выражающиеся через первые и вторые моменты . . . . . | 60        |
|          | <b>Список литературы</b>                                                        | <b>64</b> |

# 1 Введение

## 1.1 О сходимости распределений

Пусть  $\mathcal{P}_1$  и  $\mathcal{P}_2$  — два распределения в  $\mathbb{R}^d$  с борелевской  $\sigma$ -алгеброй  $\mathcal{B}_d$ . Интуитивно ясно, что какие-то  $\mathcal{P}_1$  и  $\mathcal{P}_2$  «похожи» друг на друга, а какие-то — совсем не похожи. Например, при маленьких  $\varepsilon$  равномерное распределение  $U(-\varepsilon, \varepsilon)$  на отрезке  $[-\varepsilon, \varepsilon]$  «похоже» на распределение  $\delta_0$ , сосредоточенное в нуле, распределение  $U(0, 1 + \varepsilon)$  близко к  $U(0, 1)$ , а распределение  $U(0, 1)$  сильно отличается от  $\text{Exp}(\lambda)$  при любом  $\lambda > 0$ .

Хочется формализовать эти понятия — «похоже» или «не похоже». Стандартный математический путь для этого — задание на множестве интересующих нас объектов метрики, измеряющей расстояние между объектами. Здесь, однако, не все так просто.

Даже в  $\mathbb{R}$ , где вроде бы есть единственная «естественная» метрика  $\rho(x, y) = |x - y|$ , иногда удобно использовать альтернативную метрику<sup>1</sup>

$$\rho_1(x, y) = \frac{|x - y|}{1 + |x - y|},$$

которая эквивалентна  $\rho$  в том смысле, что  $x_n \xrightarrow{\rho_1} x$  тогда и только тогда, когда  $x_n \xrightarrow{\rho} x$ , но обладает тем свойством, что  $\rho_1(x, y) < 1$  при любых  $x, y \in \mathbb{R}$ . А можно рассматривать, например, дискретную метрику

$$\rho_2(x, y) = \begin{cases} 1 & x \neq y, \\ 0 & x = y, \end{cases}$$

которая, конечно, уже не эквивалентна  $\rho$ .

При переходе к  $\mathbb{R}^d$  ситуация еще усложняется. Скажем, в  $\mathbb{R}^2$  можно задать уже как минимум 3 «естественные» (и эквивалентные) метрики — евклидову  $\rho^{(2)}$ , равномерную  $\rho^{(\infty)}$  и метрику  $\rho^{(1)}$ :

$$\begin{aligned} \rho^{(2)}(\mathbf{x}, \mathbf{y}) &= \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}, & \rho^{(\infty)}(\mathbf{x}, \mathbf{y}) &= \max(|x_1 - x_2|, |y_1 - y_2|), \\ \rho^{(1)}(\mathbf{x}, \mathbf{y}) &= |x_1 - x_2| + |y_1 - y_2|, & \text{где } \mathbf{x} &= (x_1, x_2), \mathbf{y} = (y_1, y_2). \end{aligned}$$

В зависимости от обстоятельств оказывается удобно использовать одну или другую из них.

Итак, задание «естественной» метрики на некотором множестве объектов — задача не простая и не однозначная. В нашем случае она осложняется тем, что интересующие нас объекты сами по себе сложны — это распределения (то есть некоторые функции), имеющие областью определения борелевские подмножества  $\mathbb{R}^d$ .

Пожалуй, самое простое в этой ситуации — определить равномерную метрику на множестве  $\Pi_d$  распределений в  $\mathbb{R}^d$  аналогично тому как определяется расстояние между двумя непрерывными функциями в пространстве  $C(0, 1)$ . Введем соответствующую терминологию и дадим соответствующее определение.

## 1.2 О расстоянии по вариации

**Определение 1.1.** (*Расстояние по вариации*). Пусть  $\Pi_d$  — множество распределений, определенных на борелевской  $\sigma$ -алгебре  $\mathcal{B}_d$  подмножеств  $\mathbb{R}^d$  и  $\mathcal{P}, \mathcal{Q} \in \Pi_d$ . Расстоянием по вариации между  $\mathcal{P}$  и  $\mathcal{Q}$  называется число

$$\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) = \sup_{B \in \mathcal{B}_d} |\mathcal{P}(B) - \mathcal{Q}(B)|. \quad (1.1)$$

Говорят, что последовательность  $\mathcal{P}_n \in \Pi_d$  сходится по вариации<sup>2</sup> к  $\mathcal{P} \in \Pi_d$ , если  $\rho_{\text{var}}(\mathcal{P}_n, \mathcal{P}) \rightarrow 0$  при  $n \rightarrow \infty$ . Этот факт кратко записывается как  $\mathcal{P}_n \xrightarrow{\text{var}} \mathcal{P}$ .

<sup>1</sup>Докажите, что это метрика.

<sup>2</sup>В другой терминологии — *сильно сходится*.

**Замечание 1.1.** 1. Формула (1.1) задает метрику в множестве  $\Pi_d$ .<sup>3</sup> 2. Ясно, что  $\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) \leq 1$ .

Какими свойствами обладает расстояние по вариации? Один из ответов на этот вопрос дает теорема Шеффе.<sup>4</sup>

**Теорема 1.1.** (Теорема Шеффе).

1a. Если распределения  $\mathcal{P}$  и  $\mathcal{Q}$  в  $\mathbb{R}^d$  абсолютно непрерывны с плотностями  $p$  и  $q$ , то

$$\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) = \frac{1}{2} \int_{\mathbb{R}^d} |p(x) - q(x)| dx. \quad (1.2)$$

1b. Если распределения  $\mathcal{P}$  и  $\mathcal{Q}$  дискретны и сосредоточены на одном и том же не более чем счетном множестве и имеют таблицы распределений

$$\mathcal{P}: \begin{pmatrix} x_1 & \dots & x_\ell & \dots \\ p_1 & \dots & p_\ell & \dots \end{pmatrix} \quad \text{и} \quad \mathcal{Q}: \begin{pmatrix} x_1 & \dots & x_\ell & \dots \\ q_1 & \dots & q_\ell & \dots \end{pmatrix},$$

то

$$\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) = \frac{1}{2} \sum_{\ell} |p_\ell - q_\ell|. \quad (1.3)$$

2a. Пусть  $\mathcal{P}_n, \mathcal{P}$  абсолютно непрерывные распределения в  $\mathbb{R}^d$  с плотностями  $p_n$  и  $p$ . Если  $p_n \rightarrow p$  почти всюду по мере Лебега, то  $\mathcal{P}_n \xrightarrow{\text{var}} \mathcal{P}$ .

2b. Если  $\mathcal{P}_n, \mathcal{P}$  — дискретные распределения<sup>5</sup> с вероятностями  $p_\ell^{(n)}$  и  $p_\ell$  соответственно. Если  $p_\ell^{(n)} \rightarrow p_\ell$  при всех  $\ell$ , то  $\mathcal{P}_n \xrightarrow{\text{var}} \mathcal{P}$ .

Первое утверждение теоремы Шеффе показывает, что для абсолютно непрерывных распределений расстояние по вариации имеет очень простой вид: с точностью до множителя  $1/2$  это всего лишь расстояние в  $\mathbb{L}_1$  между плотностями соответствующих распределений.

С другой стороны, формула (1.2) дает возможность вычислять расстояние по вариации. Например, нетрудно посчитать, что расстояние по вариации между распределениями  $U(0, 1 + \varepsilon)$  и  $U(0, 1)$  при положительных<sup>6</sup>  $\varepsilon$  равно  $\varepsilon/(1 + \varepsilon)$ , в то время как  $\rho_{\text{var}}(U(0, 1), \text{Exp}(\lambda)) = e^{-\lambda}$  при  $\lambda \leq 1$ .<sup>7</sup>

Второе утверждение теоремы Шеффе позволяет доказывать (при благоприятных обстоятельствах) сам факт сходимости по вариации. Например, из того, что плотность

$$s_n(x) = \frac{\Gamma((n+1)/2)}{\sqrt{\pi n} \Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

распределения Стюдента  $S_n$  с параметром  $n$  (так называемым «числом степеней свободы») поточно стремится<sup>8</sup> при  $n \rightarrow \infty$  к плотности распределения стандартного нормального закона сразу же следует, что  $S_n \xrightarrow{\text{var}} N(0, 1)$ .<sup>9</sup>

Все это, несомненно, весьма привлекательные свойства. Рассмотрим, однако, расстояние по вариации (1.1) между распределениями  $\mathcal{P} = \mathcal{P}_\varepsilon = U(-\varepsilon, \varepsilon)$  и  $\mathcal{Q} = \delta_0$ . Оказывается, оно равно 1 для любого положительного  $\varepsilon$ . Действительно, взяв  $B = \{0\}$ , получим, что  $\mathcal{P}_\varepsilon(B) = 0$ , а  $\mathcal{Q}(B) = 1$ . Это

<sup>3</sup>Докажите это.

<sup>4</sup>Здесь она формулируется отдельно для абсолютно непрерывных и дискретных распределений в  $\mathbb{R}^d$ . Более общий случай с доказательством см. в разделе 8.2, а также в [1, с. 306] или в [3, гл. III §9 лемма 2]

<sup>5</sup>Сосредоточенные на одном множестве и том же не более чем счетном множестве.

<sup>6</sup>А при отрицательных?

<sup>7</sup>А при  $\lambda > 1$ ?

<sup>8</sup>Проверьте!

<sup>9</sup>В практической статистике принято считать, что при  $n > 30$  различиями между распределениями  $S_n$  и  $N(0, 1)$  можно пренебречь. Интересно, как такой выбор  $n$  описывается в терминах расстояния по вариации между этими распределениями?

выглядит странно: при  $\varepsilon \downarrow 0$  распределения  $U(-\varepsilon, \varepsilon)$  все больше и больше концентрируются около нуля, а расстояние между этими распределениями и  $\delta_0$  остается равным 1!

Более того, расстояние по вариации между любым абсолютно непрерывным и любым дискретным распределением тоже равно единице (достаточно в качестве  $B$  взять носитель дискретного распределения).

И еще более того: если взять  $\mathcal{P} = \mathcal{P}_\varepsilon = \delta_\varepsilon$  и  $\mathcal{Q} = \delta_0$ , то снова окажется, что  $\rho_{\text{var}}(\mathcal{P}_\varepsilon, \mathcal{Q}) = 1$  для любого сколь угодно малого  $\varepsilon > 0$ . Таким образом получается, что случайные величины  $\xi_\varepsilon \equiv \varepsilon$  сходятся<sup>10</sup> при  $\varepsilon \rightarrow 0$  к случайной величине  $\xi \equiv 0$ , а их распределения не сходятся.<sup>11</sup>

Это представляется неестественным и, следовательно, метрика Определения 1.1 является мало пригодной для описания близости распределений на всем множестве  $\Pi_d$ . Попробуем понять, как ее можно модифицировать, чтобы ликвидировать описанные неприятности.

Оказывается, имеет место следующее утверждение (доказательство см. в разделе 8, а также в [3, гл. III §9 лемма 1]).

**Предложение 1.1.**

$$\sup_{B \in \mathcal{B}_d} |\mathcal{P}(B) - \mathcal{Q}(B)| = \frac{1}{2} \sup_{|f| \leq 1} \left| \int_{\mathbb{R}^d} f d\mathcal{P} - \int_{\mathbb{R}^d} f d\mathcal{Q} \right|, \quad (1.4)$$

где супремум берется по всем  $\mathcal{B}_d$ -измеримым функциям  $f$ , по модулю меньшим 1.<sup>12</sup>

Равенство (1.4) наводит на следующие соображения. Поскольку метрика  $\rho_{\text{var}}(\mathcal{P}, \mathcal{Q})$  является слишком «жесткой» (она приписывает слишком большое отличие похожим распределениям), ее хочется «смягчить». Сделать это можно, сузив класс функций, по которым берется супремум в интегральном представлении (1.4) расстояния  $\rho_{\text{var}}(\mathcal{P}, \mathcal{Q})$ .

Как это естественно сделать? В множество рассматриваемых в (1.4) функций входят функции, сильно меняющиеся при любом сколь угодно малом изменении аргумента. Например, туда входят индикаторы измеримых множеств. Но именно индикаторы (которые фактически стоят в определении метрики  $\rho_{\text{var}}(\mathcal{P}, \mathcal{Q})$ ) и доставили нам все неприятности! Давайте же наложим на множество функций ограничения, не позволяющие им сильно меняться при небольших изменениях аргументов!

Приведем формализацию этих соображений.

**Определение 1.2.** (BL-метрика). Пусть  $\Pi_d$  — множество распределений, определенных на борелевской  $\sigma$ -алгебре  $\mathcal{B}_d$  подмножеств  $\mathbb{R}^d$  и  $\mathcal{P}, \mathcal{Q} \in \Pi_d$ . Тогда BL-расстоянием между  $\mathcal{P}$  и  $\mathcal{Q}$  называется число

$$\rho_{\text{BL}}(\mathcal{P}, \mathcal{Q}) = \sup_{f \in \text{BL}(1,1)} \left| \int_{\mathbb{R}^d} f d\mathcal{P} - \int_{\mathbb{R}^d} f d\mathcal{Q} \right|, \quad (1.5)$$

где  $\text{BL}(M, L)$  — множество  $\mathcal{B}_d$ -измеримых функций  $f$ , удовлетворяющих условию  $|f| \leq M$  и условию Липшица с постоянной  $L$ .

**Замечание 1.2.** 1. Легко видеть, что формула (1.5) действительно определяет метрику.<sup>13</sup>

2. Поскольку

$$2\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) \geq \rho_{\text{BL}}(\mathcal{P}, \mathcal{Q}),$$

то из сходимости по вариации следует сходимость распределений в BL-метрике. Далее мы увидим, что обратное, вообще говоря, неверно.

<sup>10</sup>В любом разумном смысле.

<sup>11</sup>В этом примере расстояние по вариации аналогично дискретной метрике в  $\mathbb{R}$ .

<sup>12</sup>Исходя из равенства (1.4), за определение расстояния по вариации иногда берут  $\sup_{|f| \leq 1} \left| \int_{\mathbb{R}^d} f d\mathcal{P} - \int_{\mathbb{R}^d} f d\mathcal{Q} \right|$ , то есть  $2 \sup_{B \in \mathcal{B}_d} |\mathcal{P}(B) - \mathcal{Q}(B)|$ . Доказательство равенства (1.4) приведено в разделе 8.2.

<sup>13</sup>Покажите это.

К сожалению, для BL-метрики не существует конструктивного представления в стиле теоремы Шеффе. Однако, если рассматривать не саму метрику  $\rho_{BL}(\mathcal{P}, \mathcal{Q})$ , а сходимость по этой метрике, то ситуация становится более оптимистичной. А именно, имеет место следующая теорема.

Обозначим  $C_b(\mathbb{R}^d)$  множество непрерывных ограниченных функций, определенных на  $\mathbb{R}^d$ .

**Теорема 1.2.** Пусть распределения  $\mathcal{P}$  и  $\mathcal{P}_n$  ( $n \geq 1$ ) определены на измеримом пространстве  $(\mathbb{R}^d, \mathcal{B}_d)$ . Тогда сходимость  $\rho_{BL}(\mathcal{P}_n, \mathcal{P}) \rightarrow 0$  эквивалентна тому, что

$$\int_{\mathbb{R}^d} f d\mathcal{P}_n \rightarrow \int_{\mathbb{R}^d} f d\mathcal{P} \quad (1.6)$$

для любой  $f \in C_b(\mathbb{R}^d)$ .

Доказательство Теоремы 1.2 достаточно сложно<sup>14</sup> (в отличие от Теоремы 1.1 и Предложения 1.1), и поэтому саму сходимость (1.6) обычно берут в качестве определения, а метрику  $\rho_{BL}$  рассматривают только при доказательстве теоретических результатов. Начиная с этого места мы будем поступать именно таким образом.

### 1.3 О слабой сходимости

**Определение 1.3.** (Слабая сходимость распределений). Говорят, что последовательность распределений  $\mathcal{P}_n$ , заданных на измеримом пространстве  $(\mathbb{R}^d, \mathcal{B}_d)$ , слабо сходится к распределению  $\mathcal{P}$ , если (1.6) выполняется для любой функции  $f \in C_b(\mathbb{R}^d)$ .

Факт слабой сходимости  $\mathcal{P}_n$  к  $\mathcal{P}$  записывается в виде  $\mathcal{P}_n \Rightarrow \mathcal{P}$ .<sup>15</sup>

**Замечание 1.3.** Из Теоремы 1.2, конечно же, следует, что сходимости  $\mathcal{P}_n \Rightarrow \mathcal{P}$  и  $\mathcal{P}_n \Rightarrow \mathcal{Q}$  влекут за собой равенство  $\mathcal{P} = \mathcal{Q}$ , то есть единственность слабого предела. Этот факт, однако, можно доказать и непосредственно, пользуясь только Определением 1.3.

Суммируем некоторые важные свойства слабой сходимости в следующем предложении. Рассмотрим последовательность распределений  $\mathcal{P}_n$  и распределение  $\mathcal{P}$ , определенные на борелевской  $\sigma$ -алгебре  $\mathcal{B}_d$  подмножеств  $\mathbb{R}^d$ . Обозначим соответствующие характеристические функции  $\varphi_n, \varphi$  и функции распределения —  $F_n, F$ .

**Предложение 1.2.** Эквивалентны следующие утверждения.

1.  $\mathcal{P}_n \Rightarrow \mathcal{P}$ .
2.  $F_n(x) \rightarrow F(x)$  для любой точки  $x \in \mathbb{R}^d$ , являющейся точкой непрерывности функции  $F$ .
3.  $\varphi_n(t) \rightarrow \varphi(t)$  для любого  $t \in \mathbb{R}^d$ .
4.  $\mathcal{P}_n(B) \rightarrow \mathcal{P}(B)$  при любом  $B \in \mathcal{B}_d$  таком, что  $\mathcal{P}(\partial B) = 0$ , где  $\partial B$  — граница множества  $B$ .

**Замечание 1.4.** 1. Второй и третий пункт Предложения 1.2 дают возможность проверять слабую сходимость конкретной последовательности распределений.

Что касается четвертого пункта, он носит скорее теоретический характер. Например, сравнение этого пункта с определением сходимости по вариации еще раз показывает, что из сильной сходимости распределений следует слабая сходимость.<sup>16</sup>

<sup>14</sup>С ним можно ознакомиться в [3, гл. III §7].

<sup>15</sup>Иногда пишут  $\mathcal{P}_n \xrightarrow{w} \mathcal{P}$ . Используются и другие обозначения.

<sup>16</sup>Доказательство эквивалентности первого и второго утверждения можно найти (по крайней мере в одномерном варианте) в большинстве более или менее продвинутых учебниках по теории вероятностей. В данном курсе доказательство этого утверждения опущено. Эквивалентность первого и третьего пунктов получена в разделе 7.2 как пример применения теоремы о классе функций, определяющих слабую сходимость. Эквивалентность первого и четвертого утверждения доказана в разделе 4.2.

Действительно, согласно определению сходимости по вариации сходимость  $\mathcal{P}_n(B) \rightarrow \mathcal{P}(B)$  должна выполняться<sup>17</sup> для любых измеримых  $B$ , в то время как при слабой сходимости на множества  $B$  накладываются ограничения.

2. В части достаточности утверждение об эквивалентности слабой сходимости и поточечной сходимости характеристических функций может быть усилено. А именно, можно доказать, что если  $\varphi_n(t) \rightarrow \psi(t)$  для любого  $t \in \mathbb{R}^d$  и  $\psi$  — непрерывная в нуле функция, то  $\psi$  является характеристической функцией некоторого распределения  $\mathcal{Q}$  и, следовательно,  $\mathcal{P}_n \Rightarrow \mathcal{Q}$ .

**Пример 1.1.** Обсудим результат Предложения 1.2 на простейшем примере. Пусть  $d = 1$ ,  $a_n, a \in \mathbb{R}$ ,  $a_n \neq a$  и  $a_n \rightarrow a$ . Положим  $\mathcal{P}_n = \delta_{a_n}$  и  $\mathcal{P} = \delta_a$ .

1. Для начала заметим, что, поскольку  $a_n \neq a$ , то  $\rho_{\text{var}}(\mathcal{P}_n, \mathcal{P}) = 1$  и  $\mathcal{P}_n$  не сходится к  $\mathcal{P}$  по вариации. В то же время для любой непрерывной функции  $f$

$$\int_{\mathbb{R}^d} f d\mathcal{P}_n = f(a_n) \rightarrow f(a) = \int_{\mathbb{R}^d} f d\mathcal{P},$$

то есть  $\mathcal{P}_n \Rightarrow \mathcal{P}$  по определению.

2. В терминах характеристических функций доказательство слабой сходимости тоже просто:

$$\varphi_n(t) = e^{ita_n} \xrightarrow{n \rightarrow \infty} e^{ita} = \varphi(t).$$

3. С функциями распределения все похитрее, так как там присутствует условие на множество точек  $x$ , для которых должна быть сходимость. Рассмотрим 2 случая:  $a_n \uparrow a$  и  $a_n \downarrow a$ . Так как

$$F_n(x) = \begin{cases} 0 & \text{при } x \leq a_n, \\ 1 & \text{при } x > a_n \end{cases} \quad \text{и} \quad F_n(x) = \begin{cases} 0 & \text{при } x \leq a, \\ 1 & \text{при } x > a, \end{cases}$$

то при  $a_n \downarrow a$  имеет место сходимость  $F_n(x) \rightarrow F(x)$  для любого  $x \in \mathbb{R}$ , а при  $a_n \uparrow a$

$$F_n(x) \rightarrow F_*(x) = \begin{cases} 0 & \text{при } x < a, \\ 1 & \text{при } x \geq a, \end{cases}$$

что совпадает с  $F(x)$  во всех точках, кроме точки  $x = a$ , которая как раз и является единственной точкой разрыва предельной функции распределения  $F$ .

Если же взять такую последовательность  $a_n$ , что  $a_n \downarrow a$  при четных  $n$  и  $a_n \uparrow a$  при нечетных  $n$ , то, очевидно,  $F_n(x) \rightarrow F(x)$  при  $x \neq a$ , в то время как последовательность  $F_n(a)$  предела не имеет.

Этот пример показывает, что (при наличии слабой сходимости распределений) в точках разрыва предельной функции распределения со сходимостью функций распределения может быть все, что угодно. Если же предельная функция распределения непрерывна, то допредельные функции распределения сходятся к ней поточно.<sup>18</sup>

4. Чем же отличается точка  $a$  от других точек на прямой? Перекидывая мостик к четвертому пункту Предложения 1.2, мы видим, что  $a$  — это единственная точка  $x \in \mathbb{R}$ , для которой  $\mathcal{P}(\{x\}) > 0$ .

Более того, если рассмотреть всевозможные интервалы  $B = \langle \alpha, \beta \rangle$ ,<sup>19</sup> то мы увидим, что  $\mathcal{P}_n(B) \rightarrow \mathcal{P}(B)$  при условии, что  $\alpha, \beta \neq a$ . Но множество  $\{\alpha, \beta\}$  — это и есть граница  $\partial B$

<sup>17</sup>Да еще и равномерно по  $B$ .

<sup>18</sup>Более того, в этом случае сходимость равномерна, то есть  $\sup_x |F_n(x) - F(x)| \rightarrow 0$ . Докажите это.

<sup>19</sup>Значки “ $\langle$ ” и “ $\rangle$ ” означают, что нам все равно, принадлежат концы интервала этому множеству или нет.

множества  $B$ . Таким образом мы приходим к иллюстрации (конечно, на простейшем примере и только для  $B = \{\alpha, \beta\}$ ) четвертого пункта Предложения 1.2. То есть становится более-менее понятно, откуда берется условие  $\mathcal{P}(\partial B) = 0$ .

Посмотрим теперь на понятие слабой сходимости с точки зрения случайных величин (а не просто их распределений), ограничиваясь для простоты одномерным случаем  $d = 1$ .

Рассмотрим вероятностное пространство  $(\Omega, \mathcal{F}, P)$  и пусть  $\xi_n, \xi$  — случайные величины, определенные на этом вероятностном пространстве. Обозначим  $\mathcal{P}_n = \mathcal{L}(\xi_n)$ ,  $\mathcal{P} = \mathcal{L}(\xi)$ .

Естественно ожидать, что, если сами случайные величины близки в каком-то смысле как функции  $\omega$ , то и распределения этих случайных величин должны быть близки. Посмотрим, как это можно формализовать в терминах слабой сходимости распределений.

Ясно, что в определении слабой сходимости (1.6) может быть переписано как

$$Ef(\xi_n) \rightarrow Ef(\xi), \quad f \in C_b(\mathbb{R}). \quad (1.7)$$

Такая форма записи определения слабой сходимости наводит на мысль, что из сходимости  $\xi_n \rightarrow \xi$  (понимаемой в некотором разумном смысле) должна следовать слабая сходимость распределений  $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$ .<sup>20</sup> Конечно, так и есть. Проиллюстрируем это с помощью следующей диаграммы.<sup>21</sup>

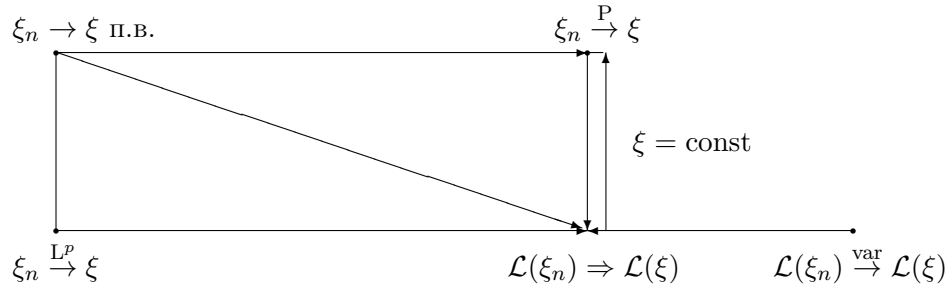


Рис. 1: Соотношения между разными видами сходимости

Рассматриваются 3 вида сходимости случайных величин: сходимость почти всюду<sup>22</sup>, по вероятности<sup>23</sup> и в  $L^p$ .<sup>24</sup> Диаграмма показывает, что каждая из этих сходимостей влечет за собой слабую сходимость соответствующих распределений.<sup>25</sup> Кроме того, для полноты на диаграмме изображен уже обсуждавшийся факт: их сильной сходимости распределений следует их слабая сходимость.

Наконец, вертикальная стрелочка от слабой сходимости к сходимости по вероятности с указанием  $\xi = \text{const}$  показывает, что в этом случае оба вида сходимости (то есть  $\xi_n \xrightarrow{P} a$  и  $\mathcal{L}(\xi_n) \Rightarrow \delta_a$ ) эквивалентны.

При решении задач, связанных со слабой сходимостью распределений, эти соображения могут быть полезны. Действительно, пусть нам нужно доказать, что  $\mathcal{P}_n \Rightarrow \mathcal{P}$ . Если мы построим на некотором вероятностном пространстве такие случайные величины  $\xi_n, \xi$ , что  $\xi_n \rightarrow \xi$  в одном из трех

<sup>20</sup>Иногда вместо  $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$  пишут  $\xi_n \xrightarrow{d} \xi$  и говорят, что  $\xi_n$  сходится к  $\xi$  по распределению. При использовании такой терминологии нужно четко понимать, что речь идет не о сходимости случайных величин  $\xi_n$  к случайной величине  $\xi$ , а о сходимости их распределений. Например, в записи (1.7) ничего не мешает случайным величинам  $\xi_n$  и  $\xi$  быть заданным на разных вероятностных пространствах.

<sup>21</sup>Доказательство этих утверждений можно найти, например, в [3].

<sup>22</sup>Вероятность того, что  $\xi_n \rightarrow \xi$ , равна 1. Часто для такой сходимости применяю термин «почти на верное».

<sup>23</sup>То есть  $P(|\xi_n - \xi| > \varepsilon) \rightarrow 0$  для любого  $\varepsilon > 0$ .

<sup>24</sup>При  $d = 1$  это означает, что  $E|\xi_n - \xi|^p \rightarrow 0$ . Здесь, конечно, предполагается, что  $\xi_n, \xi \in L^p$ . При  $d > 1$  все аналогично.

<sup>25</sup>При желании это можно считать еще одним основанием называть такую сходимость «слабой».

обсуждаемых смыслах, то мы автоматически получим, что  $\mathcal{P}_n$  слабо сходится к  $\mathcal{P}$ . Такой прием носит название *метода одного вероятностного пространства*. Приведем один простой пример.

**Пример 1.2.** Путь распределения  $\mathcal{P}_n$  определяются своими таблицами

$$\mathcal{P}_n : \begin{pmatrix} 0 & \dots & k/n & \dots & (n-1)/n \\ 1/n & \dots & 1/n & \dots & 1/n \end{pmatrix}.$$

Докажем, что  $\mathcal{P}_n \Rightarrow U(0,1)$  четырьмя способами, три<sup>26</sup> из которых связаны с Предложением 1.2, а четвертое реализует идею метода одного вероятностного пространства.

1. Прежде всего, по определению слабой сходимости распределений нужно доказать, что для  $f \in C_b(0,1)$

$$\int_{\mathbb{R}} f d\mathcal{P}_n = \frac{1}{n} \sum_{k=0}^{n-1} f(k/n) \rightarrow \int_0^1 f(x) dx = \int_{\mathbb{R}} f d\mathcal{P}.$$

Это, конечно, ни что иное, как свойство интеграла Римана.<sup>27</sup>

2. Если мы пользуемся вторым пунктом Предложения 1.2, то, взяв  $x \in [0,1]$ ,<sup>28</sup> получаем, что

$$F_n(x) = \frac{[nx]}{n} \rightarrow x = F(x).$$

3. Если использовать для доказательства слабой сходимости характеристические функции, то при любом фиксированном  $t$  и достаточно большом  $n$ <sup>29</sup>

$$\begin{aligned} \varphi_n(t) &= \int_{\mathbb{R}} e^{itx} \mathcal{P}_n(dx) = \frac{1}{n} \sum_{k=0}^{n-1} e^{itk/n} = \begin{cases} 1 & \text{при } t = 0, \\ \frac{e^{it} - 1}{n(e^{it/n} - 1)} & \text{при } t \neq 0 \end{cases} \xrightarrow{n \rightarrow \infty} \\ &\xrightarrow{n \rightarrow \infty} \begin{cases} 1 & \text{при } t = 0, \\ \frac{e^{it} - 1}{it} & \text{при } t \neq 0 \end{cases} = \int_0^1 e^{itx} dx = \int_{\mathbb{R}} e^{itx} \mathcal{P}(dx), \end{aligned}$$

где  $\mathcal{P} = U(0,1)$ .

4. Воспользуемся теперь тем, что из сходимости случайных величин почти наверное следует слабая сходимость их распределений. А именно, рассмотрим  $\xi \in U(0,1)$  и положим  $\xi_n = [n\xi]/n$ .<sup>30</sup> Тогда при  $k = 0, \dots, n-1$

$$P(\xi_n = k/n) = P([n\xi] = k) = 1/n$$

и, следовательно,  $\mathcal{L}(\xi_n) = \mathcal{P}_n$ . С другой стороны,

$$\frac{[n\xi]}{n} = \frac{n\xi - \{n\xi\}}{n} = \xi - \frac{\{n\xi\}}{n} \rightarrow \xi$$

при  $n \rightarrow \infty$ . Поэтому  $\xi_n \rightarrow \xi$  п.в. и, следовательно,  $\mathcal{P}_n \Rightarrow \mathcal{P}$ .

:

<sup>26</sup>Как уже говорилось, четвертый пункт Предложения 1.2 мало пригоден для решения подобных задач.

<sup>27</sup>Здесь мы столкнулись с тем редким случаем, когда факт слабой сходимости может быть проверен прямо по определению.

<sup>28</sup>При других  $x$  сходимость функций распределения очевидна.

<sup>29</sup>Чтобы  $t/n$  не было целым числом при фиксированном  $t \neq 0$ .

<sup>30</sup>Тем самым  $\xi_n$  и  $\xi$  определены на одном вероятностном пространстве.

## 2 Топологические и метрические пространства

В этом разделе приводится сводка определений и результатов, относящихся к стандартным понятиям метрических и топологических пространств. Поэтому доказательства этих результатов не приводятся.

**1. Топология, порожденная метрикой.** Пусть  $D$  — некоторое множество. Семейство  $\mathcal{C}$  его подмножеств называется *топологией*, если оно замкнуто относительно любых объединений и конечных пересечений, а также если  $D, \emptyset \in \mathcal{C}$ . Каждый элемент семейства  $\mathcal{C}$  называется *открытым множеством*, а его дополнение — *замкнутым множеством*.<sup>31</sup> Пара  $(D, \mathcal{C})$  называется *топологическим пространством*.

Если  $G \subset D$  и  $\mathcal{C}_G = \{C \cap G, C \in \mathcal{C}\}$ , то пара  $(G, \mathcal{C}_G)$  является топологическим пространством с топологией  $\mathcal{C}_G$ , *индуцированной* топологией  $\mathcal{C}$ .

Подмножество  $\mathcal{C}_0 \subset \mathcal{C}$  называется *базой* топологии  $\mathcal{C}$ , если каждое  $A \in \mathcal{C}$  представляется в виде объединения множеств из  $\mathcal{C}_0$ . Конечно, каждая топология полностью определяется любой своей базой.

Пусть  $(D, \rho)$  — некоторое метрическое пространство. Рассмотрим множество открытых (в метрике  $\rho$ ) шаров  $B_r(x)$  со всевозможными центрами  $x \in D$  и всевозможными радиусами  $r \geq 0$ . Если определить  $\mathcal{C}$  как семейство всевозможных объединений шаров  $B_r(x)$  с  $r \geq 0$  и  $x \in D$ , то, как нетрудно видеть, мы получим некоторую топологию, которая *порождается* метрикой  $\rho$ .<sup>32</sup> В этом случае тройка  $(D, \rho, \mathcal{C})$  будет называться *метрическим топологическим пространством*. Конечно, разные метрики в  $D$  могут порождать одно и то же топологическое пространство  $(D, \mathcal{C})$ . Тогда эти метрики являются *топологически эквивалентными*.

**2. Сепарабельные пространства.** Если в топологическом пространстве  $(D, \mathcal{C})$  у топологии  $\mathcal{C}$  существует счетная<sup>33</sup> база  $\mathcal{C}_0$ , то будем говорить, что пространство  $(D, \mathcal{C})$  является *топологически сепарабельным*.<sup>34</sup> В этом случае любое открытое множество по определению является объединением счетного числа множеств из  $\mathcal{C}_0$ .

Если в метрическом пространстве  $(D, \rho)$  существует счетное всюду плотное множество  $D_0$ , то такое пространство называется *метрически сепарабельным*. Хорошо известно следующее утверждение.

**Предложение 2.1.** Пусть  $(D, \rho, \mathcal{C})$  — метрическое топологическое пространство. Тогда пространство  $(D, \mathcal{C})$  является топологически сепарабельным тогда и только тогда, когда пространство  $(D, \rho)$  является метрически сепарабельным. При этом множество открытых шаров  $B_r(x_i)$ , где  $x_i$  — точки счетного всюду плотного множества, а  $r$  пробегает множество неотрицательных рациональных чисел, является счетной базой топологии  $\mathcal{C}$ .

**3. Непрерывные отображения.** Пусть  $(D_1, \mathcal{C}_1)$  и  $(D_2, \mathcal{C}_2)$  — два топологических пространства и  $f : D_1 \mapsto D_2$ . Если отображение  $f$  обладает тем свойством, что  $f^{-1}(A) \in \mathcal{C}_1$  для любого  $A \in \mathcal{C}_2$ , то это отображение называется *топологически непрерывным*. Этот факт записывается как  $f : (D_1, \mathcal{C}_1) \mapsto (D_2, \mathcal{C}_2)$ .

<sup>31</sup>Свойства замкнутых множеств очевидны из определения: пересечение любого числа замкнутых множеств замкнуто, как и объединение конечного числа замкнутых множеств.

<sup>32</sup>Иначе говоря, в этом случае множество открытых шаров объявляется базой топологии  $\mathcal{C}$ .

<sup>33</sup>Там, где это не вызывает недоразумений, мы будем употреблять термин «счетный» в значении «не более чем счетный».

<sup>34</sup>Более общепринято говорить, что в этом случае  $(D, \mathcal{C})$  удовлетворяет второй аксиоме счетности. Здесь, однако, нам удобна несколько другая терминология.

Если  $(D_1, \rho_1)$  и  $(D_2, \rho_2)$  — два метрических пространства, то отображение  $f : D_1 \mapsto D_2$  является метрически непрерывным, если для любой последовательности  $x_n \in D_1$ , сходящейся в метрике  $\rho_1$  к некоторой точке  $x \in D_1$ , имеет место сходимость  $f(x_n) \xrightarrow{\rho_2} f(x)$ .<sup>35</sup>

**Предложение 2.2.** Пусть  $(D_1, \rho_1, \mathcal{C}_1)$  и  $(D_2, \rho_2, \mathcal{C}_2)$  — два метрических топологических пространства. Тогда отображение  $f : D_1 \mapsto D_2$  является топологически непрерывным тогда и только тогда, когда оно метрически непрерывно.

**4. Компактные пространства и множества.** Топологическое пространство  $(D, \mathcal{C})$  называется (топологически) *компактным*,<sup>36</sup> если из каждого открытого покрытия  $D$  можно выделить конечное подпокрытие. Подмножество  $G \subset D$  называется *топологически компактным*, если компактным является пространство  $(G, \mathcal{C}_G)$ , где  $\mathcal{C}_G$  — топология подмножеств  $G$ , индуцированная  $\mathcal{C}$ .

Если  $(D, \rho)$  — метрическое пространство, то подмножество  $G \subset D$  называется *метрически компактным*, если из любой последовательности  $\{x_n, x_n \in G\}$  можно выделить подпоследовательность, сходящуюся в метрике  $\rho$  к некоторой точке  $x \in G$ .

**Предложение 2.3.** Пусть  $(D, \rho, \mathcal{C})$  — метрическое топологическое пространство и  $G \subset D$ . Для того, чтобы  $G$  было топологически компактным необходимо и достаточно, чтобы оно было метрически компактным.

Пусть  $(D, \rho)$  — метрическое пространство. Подмножество  $G \subset D$  называется *полным*, если любая фундаментальная<sup>37</sup> последовательность точек  $x_n \in G$  имеет предел, принадлежащий  $G$ . Если само множество  $D$  является полным, то говорят о *полном метрическом пространстве*. Ясно, что любое замкнутое подмножество полного метрического пространства полно.

Множество  $S \subset G \subset D$  называется  $\varepsilon$ -сетью для  $G$ , если для любого  $x \in G$  найдется  $y \in S$  такое, что  $\rho(x, y) < \varepsilon$ . Если множество  $G$  обладает конечной  $\varepsilon$ -сетью для любого  $\varepsilon$ , то оно называется *вполне ограниченным*.<sup>38</sup>

**Предложение 2.4.** Пусть  $(D, \rho, \mathcal{C})$  — метрическое топологическое пространство и  $G \subset D$ . Для того, чтобы  $G$  было компактом, необходимо и достаточно, чтобы оно было вполне ограниченным и полным.

**5. Произведение пространств.** Пусть  $(D_1, \mathcal{C}_1), \dots, (D_n, \mathcal{C}_n)$  — некоторые топологические пространства. Обозначим  $D = D_1 \times \dots \times D_n$  и введем в  $D$  топологию  $\mathcal{C}$  следующим образом. А именно, рассмотрим всевозможные произведения  $C_1 \times \dots \times C_n$  с  $C_i \in \mathcal{C}_i$  и объявим семейство всех таких произведений базой топологии  $\mathcal{C}$ . Сама топология  $\mathcal{C}$  при этом носит название *топологии произведения*.

**Предложение 2.5.** Пусть  $\mathcal{C}_1^{(0)}, \dots, \mathcal{C}_n^{(0)}$  — базы топологий  $\mathcal{C}_1, \dots, \mathcal{C}_n$  соответственно. Тогда семейство множеств

$$\mathcal{C}_0 = \{C_1^{(0)} \times \dots \times C_n^{(0)}, C_1^{(0)} \in \mathcal{C}_1^{(0)}, \dots, C_n^{(0)} \in \mathcal{C}_n^{(0)}\}$$

является базой топологии произведения. Если  $\mathcal{C}_1^{(0)}, \dots, \mathcal{C}_n^{(0)}$  — счетные базы топологий  $\mathcal{C}_1, \dots, \mathcal{C}_n$ , то топологическое пространство  $(D, \mathcal{C})$  является сепарабельным, а  $\mathcal{C}_0$  — его счетной базой.

Пусть теперь  $(D_1, \rho_1, \mathcal{C}_1), \dots, (D_n, \rho_n, \mathcal{C}_n)$  — метрические топологические пространства и  $x_i, y_i \in D_i$ . Обозначим  $\mathbf{x} = (x_1, \dots, x_n)$  и  $\mathbf{y} = (y_1, \dots, y_n)$ . Ясно, что  $\mathbf{x}, \mathbf{y} \in D = D_1 \times \dots \times D_n$ .

<sup>35</sup>Этот факт можно записывать как  $f : (D_1, \rho_1) \mapsto (D_2, \rho_2)$ .

<sup>36</sup>Более коротко — топологическим компактом.

<sup>37</sup>то есть такая, что  $\rho(x_n, x_m) \rightarrow 0$  при  $n, m \rightarrow \infty$ .

<sup>38</sup>Докажите, что замыкание вполне ограниченного множества является вполне ограниченным.

**Предложение 2.6.** Метрика  $r_\infty(\mathbf{x}, \mathbf{y}) = \max_{1 \leq i \leq n} \rho_i(x_i, y_i)$  порождает топологию произведения в  $D$ .<sup>39</sup>

*Доказательство.* Обозначим  $\mathcal{C}_\infty$  топологию, порожденную в  $D$  метрикой  $r_\infty$ . Рассмотрим всевозможные открытые шары  $B_s(\mathbf{x})$  в пространстве  $(D, r_\infty)$ . Множество таких шаров образует базу топологии  $\mathcal{C}_\infty$ . В то же время

$$B_s(\mathbf{x}) = \{\mathbf{y} : r_\infty(\mathbf{x}, \mathbf{y}) < s\} = \{(y_1, \dots, y_n) : \rho_i(x_i, y_i) < s\} = \bigotimes_{i=1}^n B_s^{(i)}(x_i),$$

где  $B_s^{(i)}(x_i) = \{z \in D_i : \rho_i(x_i, z) < s\}$ . Из этого следует, что каждый шар  $B_s(\mathbf{x})$  является открытым множеством в топологии произведения, и, следовательно,  $\mathcal{C}_\infty \subset \mathcal{C}$ .<sup>40</sup>

Докажем обратное включение. Рассмотрим множества  $C_i \in \mathcal{C}_i$  и пусть  $x_i \in C_i, i = 1, \dots, n$ . Тогда можно найти такое  $s$ , что  $B_s^{(i)}(x_i) \subset C_i$  при всех  $i$ . Иначе говоря, для любого  $\mathbf{x} \in C_1 \times \dots \times C_n$  существует такое  $s = s(\mathbf{x})$ , что  $B_s(\mathbf{x}) \subset C_1 \times \dots \times C_n$ . Значит, любой элемент базы топологии произведения является объединением открытых (в метрике  $r_\infty$ ) шаров. Поэтому  $\mathcal{C}_\infty \supset \mathcal{C}$ .  $\square$

**6. Одно свойство метрики.** Пусть  $(D, \rho)$  — метрическое пространство. Нам понадобится следующее свойство метрики  $\rho$ .

**Лемма 2.1.** Пусть  $A \subset D$  и  $x \in D$ . Обозначим  $\rho(x, A) = \inf_{z \in A} \rho(x, z)$ . Тогда  $|\rho(x, A) - \rho(y, A)| \leq \rho(x, y)$ .

*Доказательство.* Действительно, для любых  $x, y \in D$

$$\rho(x, A) = \inf_{z \in A} \rho(x, z) \leq \inf_{z \in A} (\rho(x, y) + \rho(z, y)) = \rho(x, y) + \inf_{z \in A} \rho(z, y) = \rho(x, y) + \rho(y, A). \quad (2.1)$$

Меняя в (2.1)  $x$  и  $y$  местами, приходим к неравенству  $|\rho(x, A) - \rho(y, A)| \leq \rho(x, y)$ .  $\square$

**Замечание 2.1.** Пусть  $a \in D$ . Поскольку  $\rho(x, \{a\}) = \rho(x, a)$ , то из этого утверждения следует, что функция  $\rho(\cdot, a)$  является непрерывной.

<sup>39</sup>Конечно,  $r_\infty$  — не единственная естественная метрика, порождающая топологию произведения. В качестве примеров других подобных метрик можно привести  $r_2(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_i \rho_i^2(x_i, y_i)}$  или  $r_1(\mathbf{x}, \mathbf{y}) = \sum_i \rho_i(x_i, y_i)$ . Докажите это.

<sup>40</sup>Так как если  $\mathcal{C}_0^{(1)}$  — база топологии  $\mathcal{C}_1$ ,  $\mathcal{C}_0^{(2)}$  — база топологии  $\mathcal{C}_2$  и  $\mathcal{C}_0^{(1)} \subset \mathcal{C}_0^{(2)}$ , то  $\mathcal{C}_1 \subset \mathcal{C}_2$ .

## 3 Меры в измеримых метрических пространствах

### 3.1 Измеримые топологические пространства

**1. Определение. Измеримость непрерывных отображений.** Пусть  $D$  — некоторое множество. Если оснастить это множество какой-то структурой, то мы получим то, что в математике называется *пространством*. Примеров множество — введем в  $D$  линейную структуру — получим линейное пространство, метрику — метрическое пространство, норму — нормированное пространство, скалярное произведение — гильбертово пространство.<sup>41</sup>

Нас будет интересовать оснащение множества  $D$  семействами его подмножеств. А именно, если рассмотреть пару  $(D, \mathcal{C})$ , где  $\mathcal{C}$  — топология подмножеств  $D$ , то мы придем к топологическому пространству, а если пару  $(D, \mathcal{A})$ , где  $\mathcal{A}$  —  $\sigma$ -алгебра подмножеств  $D$  — то появится измеримое пространство.

Эти два семейства подмножеств отвечают за совсем разные вещи. Когда мы оснастили множество  $D$  некоторой  $\sigma$ -алгеброй  $\mathcal{A}$ , мы получили возможность говорить об измеримости функций и отображений, об интегрируемости или суммируемости функций, пользоваться результатами теории меры и т.д.

Что касается топологии, то после ее введения возникают всевозможные понятия, связанные с непрерывностью. Но формально топология  $\mathcal{C}$  и  $\sigma$ -алгебра  $\mathcal{A}$  — две никак не связанные системы подмножеств множества  $D$ . И, например, в общем случае непрерывная функция в общем случае совсем не обязана быть измеримой. Это, конечно, выглядит странно и неудобно. А поэтому возникает мысль, что топология и  $\sigma$ -алгебра как семейства подмножеств одного множества должны быть согласованы.<sup>42</sup>

Будем проводить это согласование следующим образом. Поставим во главу угла топологию  $\mathcal{C}$ , и дадим следующее определение.

**Определение 3.1.** Пусть  $(D, \mathcal{C})$  — топологическое пространство. Назовем *борелевской  $\sigma$ -алгеброй* минимальную  $\sigma$ -алгебру  $\mathcal{B} = \mathcal{B}_D$ , содержащую все открытые множества.<sup>43</sup> Тогда тройка  $(D, \mathcal{C}, \mathcal{B})$  будет называться *измеримым топологическим пространством*. Если же топология  $\mathcal{C}$  порождена метрикой  $\rho$ , то тройку  $(D, \rho, \mathcal{B})$  естественно назвать *измеримым метрическим пространством*.

**Предложение 3.1.** Пусть  $(D_1, \mathcal{C}_1, \mathcal{B}_1)$  и  $(D_2, \mathcal{C}_2, \mathcal{B}_2)$  — два измеримых топологических пространства. Тогда любое непрерывное отображение  $f : (D_1, \mathcal{C}_1) \mapsto (D_2, \mathcal{C}_2)$  является измеримым относительно борелевских  $\sigma$ -алгебр  $\mathcal{B}_1$  и  $\mathcal{B}_2$ .<sup>44</sup>

*Доказательство.* Обозначим  $\mathcal{B}_2^* = \{B \in \mathcal{B}_2 : f^{-1}(B) \in \mathcal{B}_1\}$  и докажем, что  $\mathcal{B}_2^* = \mathcal{B}_2$ . Действительно, по определению  $\mathcal{B}_2^* \subset \mathcal{B}_2$ . Кроме того,  $\mathcal{B}_2^*$  содержит все открытые множества, так как  $f$  — непрерывное отображение, а  $\mathcal{C}_1 \subset \mathcal{B}_1$ . Наконец,  $\mathcal{B}_2^*$  является  $\sigma$ -алгеброй (и, следовательно совпадает с минимальной  $\sigma$ -алгеброй  $\mathcal{B}_2$ , содержащей все открытые множества). Действительно,  $D_2 \in \mathcal{B}_2^*$ , так как  $f^{-1}(D_2) = D_1 \in \mathcal{B}_1$ . С другой стороны, по свойствам полного прообраза,  $f^{-1}(A^C) = (f^{-1}(A))^C$  для любого  $A \in D_2$  и  $f^{-1}(\cup_\alpha A_\alpha) = \cup_\alpha f^{-1}(A_\alpha)$  для любого семейства  $A_\alpha$  подмножеств  $D_2$ .<sup>45</sup> Поэтому  $\mathcal{B}_2^*$  замкнуто относительно дополнений и счетных объединений.

Предложение доказано. □

**2. Произведения измеримых топологических пространств.** Пусть  $(D_1, \mathcal{A}_1), \dots, (D_n, \mathcal{A}_n)$  — некоторые измеримые пространства. *Произведением  $\sigma$ -алгебр*  $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n$  называется

<sup>41</sup>Конечно, для этого нужно, чтобы пространство уже было линейным.

<sup>42</sup>В евклидовых пространствах такой проблемы не возникает просто потому, что там топология, порожденная евклидовой метрикой и  $\sigma$ -алгебра, порожденная ячейками, оказываются согласованными автоматически.

<sup>43</sup>То есть  $\mathcal{B} = \sigma(\mathcal{C})$ .

<sup>44</sup>Последний факт мы будем записывать как  $f : (D_1, \mathcal{B}_1) \mapsto (D_2, \mathcal{B}_2)$ .

<sup>45</sup>Убедитесь в этом.

$\sigma$ -алгебра подмножеств множества  $D = D_1 \times \dots \times D_n$ , определенная как

$$\mathcal{A} = \sigma(A_1 \times \dots \times A_n, A_i \in \mathcal{A}_i).$$

Тем самым в случае, когда мы имеем дело с измеримыми топологическими пространствами  $(D_1, \mathcal{C}_1, \mathcal{B}_1), \dots, (D_n, \mathcal{C}_n, \mathcal{B}_n)$ , у нас есть два естественных пути построения  $\sigma$ -алгебр в произведении множеств  $D_1 \times \dots \times D_n$ . Первый из них состоит в рассмотрении топологии произведения  $\mathcal{C}$ , а затем в построении борелевской  $\sigma$ -алгебры  $\mathcal{B}^{(1)}$  по этой топологии. Второй — это непосредственное построение произведения  $\mathcal{B}^{(2)}$  борелевских  $\sigma$ -алгебр  $\mathcal{B}_i$ .

Возникает вопрос: всегда ли  $\mathcal{B}^{(1)} = \mathcal{B}^{(2)}$ ? Ответ содержится в Предложении 3.2, но сначала мы докажем следующую простую лемму.

**Лемма 3.1.** *Если  $\mathcal{G}_1$  и  $\mathcal{G}_2$  — две системы подмножеств одного множества  $G$  и такие, что  $\mathcal{G}_1 \subset \mathcal{G}_2$ , то  $\sigma(\mathcal{G}_1) \subset \sigma(\mathcal{G}_2)$ .*

*Доказательство.* Действительно, по определению  $\sigma(\mathcal{G}_2) \supset \mathcal{G}_2 \supset \mathcal{G}_1$ , а  $\sigma(\mathcal{G}_1)$  — минимальная  $\sigma$ -алгебра, содержащая  $\mathcal{G}_1$ . Следовательно,  $\sigma(\mathcal{G}_1) \subset \sigma(\mathcal{G}_2)$ .  $\square$

**Предложение 3.2.** *1. Имеет место включение  $\mathcal{B}^{(1)} \supset \mathcal{B}^{(2)}$ . 2. Если все топологические пространства  $(D_i, \mathcal{C}_i)$  сепарабельны, то  $\mathcal{B}^{(1)} = \mathcal{B}^{(2)}$ .<sup>46</sup>*

*Доказательство.* 1. По определению,  $\mathcal{B}^{(2)} = \sigma(B_1 \times \dots \times B_n, B_i \in \mathcal{B}_i)$ , а

$$\mathcal{B}^{(1)} = \sigma(\mathcal{C}) \supset \sigma(C_1 \times \dots \times C_n, C_i \in \mathcal{C}_i),$$

где  $\mathcal{C}$  — топология произведения.<sup>47</sup> При  $2 \leq i \leq n$  зафиксируем открытые множества  $C_i^* \in \mathcal{C}_i$  и рассмотрим еще одну  $\sigma$ -алгебру  $\mathcal{B}(C_2^*, \dots, C_n^*) = \sigma(C_1 \times C_2^* \times \dots \times C_n^*, C_1 \in \mathcal{C}_1)$ . Ясно, что  $\mathcal{B}(C_2^*, \dots, C_n^*) \subset \mathcal{B}^{(1)}$ .

С другой стороны, так как множества  $C_2^*, \dots, C_n^*$  фиксированы, то

$$\mathcal{B}(C_2^*, \dots, C_n^*) = \mathcal{B}_1 \times C_2^* \times \dots \times C_n^*.$$

Это значит, что для любых  $B_1 \in \mathcal{B}_1$  и  $C_i \in \mathcal{C}_i$  ( $i = 2, \dots, n$ ) множество  $B_1 \times C_2 \times \dots \times C_n$  является элементом  $\mathcal{B}^{(1)}$ .

Далее, зафиксируем множества  $B_1^* \in \mathcal{B}_1$  и  $C_i^* \in \mathcal{C}_i$  ( $i \neq 2$ ) и рассмотрим  $\sigma$ -алгебру

$$\mathcal{B}(B_1^*, C_3^*, \dots, C_n^*) = \sigma(B_1^* \times C_2 \times C_3^* \times \dots \times C_n^*, C_2 \in \mathcal{C}_2).$$

Как и раньше,  $\mathcal{B}^{(1)} \supset \mathcal{B}(B_1^*, C_3^*, \dots, C_n^*) = B_1^* \times \mathcal{B}_2 \times C_3^* \times \dots \times C_n^*$ . Следовательно, для любых  $B_1 \in \mathcal{B}_1$ ,  $B_2 \in \mathcal{B}_2$  и  $C_i \in \mathcal{C}_i$  ( $i = 3, \dots, n$ ) множество  $B_1 \times B_2 \times C_2 \times \dots \times C_n$  является элементом  $\mathcal{B}^{(1)}$ . Продолжая подобные рассуждения дальше, получаем, что любое множество вида  $B_1 \times \dots \times B_n$  с  $B_i \in \mathcal{B}_i$  является элементом  $\mathcal{B}^{(1)}$ . Отсюда (см. Лемму 3.1) сразу следует первое утверждение.

2. Нужно доказать, что в условиях сепарабельности  $\mathcal{B}^{(1)} \subset \mathcal{B}^{(2)}$ . Для этого достаточно (опять см. Лемму 3.1), чтобы любое открытое в топологии произведения множество  $A$  принадлежало  $\mathcal{B}^{(2)}$ .

Так как пространства сепарабельны, то (счетная) база топологии  $\mathcal{C}$  состоит из множеств  $C_1^{(0)} \times \dots \times C_n^{(0)}$ , где  $C_i^{(0)}$  — элемент счетной базы топологии  $\mathcal{C}_i$ . Следовательно, открытое множество  $A$  может быть представлено в виде счетного объединения множеств вида  $C_1^{(0)} \times \dots \times C_n^{(0)}$ . Поскольку каждое такое множество принадлежит  $\mathcal{B}^{(2)}$ , то и  $A \in \mathcal{B}^{(2)}$ .  $\square$

<sup>46</sup>Есть примеры (см. [1, стр. 308]), показывающие, что без предположения о сепарабельности включение  $\mathcal{B}^{(1)} \subset \mathcal{B}^{(2)}$  может не иметь места.

<sup>47</sup>это включение имеет место согласно Лемме 3.1.

### 3.2 Распределения, определенные на борелевских $\sigma$ -алгебрах

Пусть  $(D, \rho, \mathcal{C}, \mathcal{B})$  — измеримое метрическое топологическое пространство и  $\mathcal{P}$  — вероятностная мера, определенная на борелевской  $\sigma$ -алгебре  $\mathcal{B}$ . Имеет место следующее утверждение.

**Предложение 3.3.** *Распределение  $\mathcal{P}$  полностью определяется своими значениями на замкнутых множествах.*

*Доказательство.* Мы докажем, что для любого множества  $B \in \mathcal{B}$  и любого  $\delta > 0$  существует такое открытое множество  $G = G_\delta$  и такое замкнутое множество  $F = F_\delta$ , что

$$F \subset B \subset G \quad \text{и} \quad \mathcal{P}(G \setminus F) < \delta. \quad (3.1)$$

Тем самым утверждение будет доказано. Обозначим  $\mathcal{B}^*$  множество тех  $B \in \mathcal{B}$ , для которых (3.1) выполняется для любого  $\delta$  и докажем, что  $\mathcal{B}^* = \mathcal{B}$ .

Прежде всего,  $D \in \mathcal{B}^*$ , так как множество  $D$  является одновременно открытым и замкнутым, и, следовательно, можно для любого  $\delta$  положить  $F_\delta = G_\delta = D$ . Далее, семейство  $\mathcal{B}^*$  замкнуто относительно дополнений. Это следует из того, что (3.1) эквивалентно соотношениям

$$G^C \subset B^C \subset F^C \quad \text{и} \quad \mathcal{P}(F^C \setminus G^C) = \mathcal{P}(G \setminus F) < \delta.$$

Далее,  $\mathcal{B}^*$  содержит все замкнутые (и, следовательно, все открытые) множества. Действительно, если  $B$  замкнуто, то в качестве  $F$  можно взять само  $B$ , а в качестве  $G$  — достаточно малую открытую  $\varepsilon$ -окрестность  $B_\varepsilon = \{x : \rho(x, B) < \varepsilon\}$  множества  $B$ .<sup>48</sup>

Следовательно, осталось доказать, что  $\mathcal{B}^*$  замкнуто относительно счетных объединений. Пусть  $B_1, \dots, B_n, \dots$  — множества, принадлежащие  $\mathcal{B}^*$  и  $\delta > 0$ . Выберем замкнутые множества  $F_n$  и открытые множества  $G_n$  так, что  $F_n \subset B_n \subset G_n$  и  $\mathcal{P}(G_n \setminus F_n) < \delta/2^{n+1}$ . Положим  $G = \bigcup_n G_n$  и  $F^* = \bigcup_n F_n$ . Тогда, очевидно,

$$\mathcal{P}(G \setminus F^*) = \mathcal{P}\left(\bigcup_n (G_n \setminus F_n)\right) \leq \sum_n \mathcal{P}(G_n \setminus F_n) \leq \sum_n \delta/2^{n+1} = \delta/2.$$

Множество  $G$ , очевидно, является открытым. Что касается множества  $F^*$ , то это, вообще говоря, не так, поскольку это множество является счетным (а не конечным) объединением замкнутых множеств. Тем не менее, можно найти такое  $N$ , что  $\mathcal{P}(F^* \setminus \bigcup_{n=1}^N F_n) \leq \delta/2$ . Положив  $F = \bigcup_{n=1}^N F_n$  и заметив, что  $F$  является замкнутым множеством, получим, что

$$\mathcal{P}(G \setminus F) \leq \mathcal{P}(G \setminus F^*) + \delta/2 \leq \delta.$$

Утверждение доказано. □

В случае, когда метрическое пространство  $(D, \rho)$  обладает рядом дополнительных свойств, утверждение Предложения 3.3 можно усилить. Но сначала дадим одно определение.

**Определение 3.2.** Пусть  $(D, \rho, \mathcal{B})$  — метрическое измеримое пространство,  $\Theta$  — некоторое параметрическое множество, и для каждого  $\theta \in \Theta$  на борелевской  $\sigma$ -алгебре  $\mathcal{B}$  задано распределение  $\mathcal{P}_\theta$ . Семейство распределений  $\{\mathcal{P}_\theta, \theta \in \Theta\}$  называется *плотным*,<sup>49</sup> если для любого  $\varepsilon > 0$  существует такой компакт  $K = K_\varepsilon$ , что  $\inf_\theta \mathcal{P}_\theta(K) > 1 - \varepsilon$ .

<sup>48</sup>Так как  $B$  замкнуто, а  $B_\varepsilon \downarrow B$  при  $\varepsilon \downarrow 0$ , то  $\mathcal{P}(B_\varepsilon \setminus B) \rightarrow 0$ .

<sup>49</sup>Русское слово «плотный» имеет 2 связанных между собой, но все-таки различных значения — «без зазоров» (плотно набитая сумка, уплотнительная застройка, всюду плотное множество, плотно поесть) и «собранный, компактный» (плотный человек). В английском математическом языке этому соответствуют два термина: «dense» (the set  $A$  is dense in  $B$ ) и «tight». В этом определении семейство  $\{\mathcal{P}_\theta, \theta \in \Theta\}$  is tight.

Плотность семейства распределений  $\{\mathcal{P}_\theta, \theta \in \Theta\}$  означает, что все они со сколь угодно большой точностью являются сосредоточенными на одном компакте. Например, если  $D = \mathbb{R}$ , а  $\mathcal{B}$  — борелевская  $\sigma$ -алгебра подмножеств прямой, то семейство распределений  $\{\mathcal{P}_n^{(1)}, n \geq 1\}$  с  $\mathcal{P}_n^{(1)}(\{0\}) = 1 - 1/n$  и  $\mathcal{P}_n^{(1)}(\{n\}) = 1/n$ , очевидно, будет плотным, в то как семейство  $\{\mathcal{P}_n^{(2)}, n \geq 1\}$  с  $\mathcal{P}_n^{(2)}(\{0\}) = 1/n$  и  $\mathcal{P}_n^{(2)}(\{n\}) = 1 - 1/n$  — не будет.<sup>50</sup>

**Предложение 3.4.** В полном сепарабельном метрическом измеримом пространстве<sup>51</sup>  $(D, \rho, \mathcal{B})$  любое конечное семейство распределений  $\mathcal{P}_\theta$  является плотным.

*Доказательство.* Утверждение достаточно доказать для случая, когда семейство распределений состоит из одного единственного распределения  $\mathcal{P}$ .<sup>52</sup>

Рассмотрим счетное всюду плотное в  $D$  множество  $\{x_i, i \geq 1\}$  и окружим каждую точку  $x_i$  открытым шаром  $A_i^{(n)} = B_{1/n}(x_i)$  с центром в точке  $x_i$  и радиусом  $1/n$ . Ясно, что  $\cup_i A_i^{(n)} = D$ . Поэтому для любых  $n$  и  $\varepsilon > 0$  можно найти такое  $N_n$ , зависящее от  $\varepsilon$ , что

$$\mathcal{P}(C_n) > 1 - \varepsilon/2^n, \quad \text{где } C_n = \bigcup_{i=1}^{N_n} A_i^{(n)}.$$

Обозначим  $K_* = \cap_n C_n$  и пусть  $K$  — замыкание  $K_*$ . Тогда

$$\mathcal{P}(K^C) \leq \mathcal{P}(K_*^C) = \mathcal{P}\left(\bigcup_{n \geq 1} C_n^C\right) \leq \sum_{n \geq 1} \mathcal{P}(C_n^C) < \sum_{n \geq 1} \varepsilon 2^{-n} = \varepsilon.$$

С другой стороны, так как  $K$  — замкнутое множество, то оно полно. Кроме того, по построению множество  $K_*$  (и, следовательно, множество  $K$ ) вполне ограничено. Следовательно, (см. Предложение 2.4)  $K$  является компактом.  $\square$

**Следствие 3.1.** В полном метрическом сепарабельном измеримом пространстве любое распределение  $\mathcal{P}$  определяется своими значениями на компактах.

*Доказательство.* Согласно Предложению 3.3 любое распределение определяется своими значениями на замкнутых множествах. Рассмотрим замкнутое множество  $F$  и некоторое  $\varepsilon > 0$ . Тогда (см. Предложение 3.4) существует такой компакт  $K_\varepsilon$ , что  $\mathcal{P}(K_\varepsilon) > 1 - \varepsilon$ . Положим  $F^\varepsilon = F \cap K_\varepsilon$ . Множество  $F^\varepsilon$  является компактом и  $\mathcal{P}(F \setminus F^\varepsilon) \leq \varepsilon$ , откуда и следует требуемое.  $\square$

Снова рассмотрим измеримое метрическое пространство  $(D, \rho, \mathcal{B})$  и обозначим  $C_b(D)$ ,  $C_b^{(u)}(D)$  и  $\text{BL}(D)$ <sup>53</sup> соответственно множества непрерывных ограниченных, равномерно непрерывных ограниченных и ограниченных липшицевских функций с областью определения  $D$ .<sup>54</sup> Конечно,  $C_b(D) \supset C_b^{(u)}(D) \supset \text{BL}(D)$ . Отметим, что согласно Лемме 2.1 функция  $\rho(x, A)$  принадлежит  $\text{BL}(D)$  для любого  $A \subset D$ .

Теперь мы можем доказать следующее утверждение.

**Предложение 3.5.** Пусть  $(D, \rho, \mathcal{B})$  — измеримое метрическое пространство и  $\mathcal{P}$  — некоторое вероятностное распределение, заданное на  $\mathcal{B}$ . Тогда мера  $\mathcal{P}$  полностью определяется значениями всевозможных интегралов  $\int_D g d\mathcal{P}$  при  $g \in \text{BL}(D)$ .

<sup>50</sup>Убедитесь в этом. При каком регулярном поведении последовательностей  $a_n \in \mathbb{R}$  и  $\sigma_n^2 > 0$  последовательность нормальных распределений  $\{N(a_n, \sigma_n^2)\}$  будет плотной?

<sup>51</sup>Такие пространства называют также *польскими* в честь польской топологической школы середины 20-го века.

<sup>52</sup>Почему?

<sup>53</sup>то есть Bounded Lipschitz.

<sup>54</sup>Понятия непрерывных и равномерно (uniform, отсюда обозначение  $C_b^{(u)}(D)$ , где нижний индекс  $b$  означает «ограниченность» — «boundness») непрерывных функций в метрическом пространстве не требуют комментариев (действительно?).

Функция  $f$  принадлежит  $\text{BL}(D)$ , если она ограничена и удовлетворяет условию Липшица:  $|f(x) - f(y)| < L\rho(x, y)$  при некотором  $0 < L < \infty$  для любых  $x, y \in D$ . Постоянная  $L$ , естественно, называется константой Липшица.

*Доказательство.* Рассмотрим функцию

$$\varphi(t) = \begin{cases} 1 & \text{при } t \leq 0, \\ 1 - t & \text{при } t \in (0, 1] \\ 0 & \text{иначе} \end{cases}$$

и при  $u > 0$  положим  $\varphi_u(t) = \varphi(ut)$ . Ясно, что  $\varphi_u(t) = 0$  при  $t \geq 1/u$  (см. рис. 2).

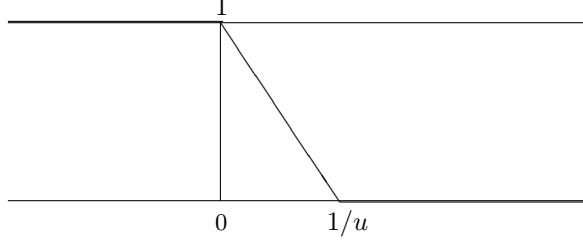


Рис. 2: График функции  $\varphi_u(t)$ .

Нетрудно также видеть, что<sup>55</sup>  $|\varphi_u(t_1) - \varphi_u(t_2)| \leq u|t_1 - t_2|$  и что при  $u \uparrow \infty$

$$\varphi_u(t) \downarrow \begin{cases} 1 & \text{при } t \leq 0, \\ 0 & \text{при } t > 0. \end{cases}$$

Пусть теперь  $F$  — замкнутое множество. Рассмотрим функцию  $f_u(x) = \varphi_u(\rho(x, F))$ . Эта функция обладает следующими свойствами. Во-первых,  $0 \leq f_u(x) \leq 1$  и  $f_u(x) = 1$  при  $x \in F$ . Во-вторых,  $f_u(x) = 0$  при  $\rho(x, F) \geq 1/u$ , то есть функция  $f_u$  обращается в ноль вне  $1/u$ -окрестности множества  $F$ . Затем,  $f_u \in \text{BL}(D)$ , так как

$$|f_u(x) - f_u(y)| = |\varphi_u(\rho(x, F)) - \varphi_u(\rho(y, F))| \leq u|\rho(x, F) - \rho(y, F)| \leq u\rho(x, y)$$

ввиду Леммы 2.1. Наконец, так как  $F$  — замкнутое множество, то  $f_u(x) \downarrow \mathbf{1}_F(x)$  при  $u \rightarrow \infty$ .<sup>56</sup> Поэтому

$$\int_D f_u(x) \mathcal{P}(dx) \downarrow \int_D \mathbf{1}_F(x) \mathcal{P}(dx) = \mathcal{P}(F).$$

Следовательно, для любого замкнутого  $F$  вероятность  $\mathcal{P}(F)$  может быть сколь угодно точно аппроксимирована интегралом от некоторой ограниченной Липшицевой функции. Ввиду Предложения 3.3 утверждение доказано.  $\square$

**Замечание 3.1.** Конечно, утверждение Предложения 3.5 останется верным, если семейство функций  $\text{BL}(D)$  заменить на  $C_b^{(u)}(D)$  или  $C_b(D)$ .

<sup>55</sup>Проверьте это.

<sup>56</sup>А как здесь используется замкнутость  $F$ ? Что будет, если  $F$  не замкнуто?

## 4 Слабая сходимость распределений. Основная теорема

### 4.1 Определение и вспомогательные утверждения

**Определение 4.1.** Пусть  $(D, \rho, \mathcal{B})$  — измеримое метрическое пространство и  $\{\mathcal{P}_n, n \geq 1\}$ ,  $\mathcal{P}$  — некоторые вероятностные меры, определенные на борелевской  $\sigma$ -алгебре  $\mathcal{B}$ . Говорят, что последовательность  $\mathcal{P}_n$  *слабо сходится* к распределению  $\mathcal{P}$  при  $n \rightarrow \infty$  (это записывается как  $\mathcal{P}_n \Rightarrow \mathcal{P}$ ), если для любой функции  $f \in C_b(D)$

$$\int_D f d\mathcal{P}_n \rightarrow \int_D f d\mathcal{P}. \quad (4.1)$$

**Замечание 4.1.** 1. В принципе, требование, чтобы пространство  $(D, \rho, \mathcal{B})$  было метрическим, излишне: достаточно, чтобы оно было измеримым топологическим. Поскольку, однако, в дальнейших рассуждениях о слабой сходимости мы всегда будем иметь дело с метрическими пространствами, мы не будем здесь гнаться за максимальной общностью.

2. Легко видеть, что, если слабый предел последовательности  $\{\mathcal{P}_n, n \geq 1\}$  существует, то он единственен. Действительно, пусть  $\mathcal{P}_n \Rightarrow \mathcal{P}$  и  $\mathcal{P}_n \Rightarrow \mathcal{Q}$ . Тогда из (4.1) следует, что

$$\int_D f d\mathcal{P} = \int_D f d\mathcal{Q}$$

для любой  $f \in C_b(D)$ . Ввиду Предложения 3.5 и Замечания 3.1 из этого следует, что  $\mathcal{P} = \mathcal{Q}$ .

3. Полезно представлять себе, что означает слабая сходимость распределений на языке случайных величин. Пусть  $(\Omega, \mathcal{F}, \mathbb{P})$  — некоторое вероятностное пространство и  $\xi_n : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{B})$  — последовательность случайных величин со значениями в  $(D, \mathcal{B})$ , где  $\mathcal{B}$  — борелевская  $\sigma$ -алгебра, порожденная в  $D$  метрикой  $\rho$ . Тогда распределения  $\mathcal{L}(\xi_n)$  случайных величин  $\xi_n$  слабо сходятся к распределению  $\mathcal{L}(\xi)$  случайной величины  $\xi$ , если  $Ef(\xi_n) \rightarrow Ef(\xi)$  для любой функции  $f \in C_b(D)$ . Конечно, это просто переформулировка Определения 4.1.

Наша цель — найти общие (и содержательные) условия, эквивалентные слабой сходимости распределений. Начнем с одного из них, носящего в основном технический характер.

**Предложение 4.1.** Пусть  $(D, \rho, \mathcal{B})$  — измеримое метрическое пространство и  $\{\mathcal{P}_n, n \geq 1\}$ ,  $\mathcal{P}$  — вероятностные меры, определенные на борелевской  $\sigma$ -алгебре  $\mathcal{B}$ . Тогда эквивалентны следующие утверждения.

1.  $\mathcal{P}_n \Rightarrow \mathcal{P}$ .
2. Для любой функции  $f \in C_b(D)$  имеет место неравенство

$$\limsup_n \int_D f d\mathcal{P}_n \leq \int_D f d\mathcal{P}. \quad (4.2)$$

3. Для любой неотрицательной функции  $f \in C_b(D)$  имеет место неравенство (4.2).

4. Для любой функции  $f \in C_b(D)$  такой, что  $0 \leq f(x) < 1$  при любом  $x \in D$  имеет место неравенство (4.2).

**Доказательство.** Сразу же заметим, что из первого условия следует второе, из второго — третье, а из третьего — четвертое. Кроме того, третье и четвертое условия эквивалентны.<sup>57</sup>

Докажем, что условия 3 и 2 эквивалентны. Действительно, из выполнения неравенства (4.2) для всех неотрицательных непрерывных ограниченных функций следует, что (4.2) выполнено для любой  $f \in C_b(D)$ . Это следует из того, что любая функция  $f \in C_b(D)$  может быть представлена в виде  $f = g + b$ , где  $b$  — постоянная, а  $g$  — непрерывная неотрицательная функция. Тогда

$$\limsup_n \int_D f d\mathcal{P}_n = \limsup_n \int_D g d\mathcal{P}_n + b \leq \int_D g d\mathcal{P} + b = \int_D f d\mathcal{P}.$$

<sup>57</sup> Действительно?

Наконец, если (4.2) выполнено для любой функции  $f \in C_b(D)$ , то это неравенство выполняется и для  $h = -f$ , откуда сразу же получаем, что

$$\limsup_n \int_D (-f) d\mathcal{P}_n = -\liminf_n \int_D f d\mathcal{P}_n \leq -\int_D f d\mathcal{P},$$

то есть  $\liminf_n \int_D f d\mathcal{P}_n \geq \int_D f d\mathcal{P}$ . Вместе с (4.2) это приводит к нужному результату:  $\mathcal{P}_n \Rightarrow \mathcal{P}$ .  $\square$

Для доказательства более интересных условий, эквивалентных слабой сходимости, нам понадобятся несколько вспомогательных утверждений.

**Лемма 4.1.** Пусть  $(D, \rho)$  — метрическое пространство и  $A \subset D$ . Обозначим  $A_\delta = \{x \in D : \rho(x, A) < \delta\}$ . Тогда<sup>58</sup>  $\partial A_\delta \subset \{x \in D : \rho(x, A) = \delta\}$ .

*Доказательство.* Заметим сначала, что множества  $A_\delta$  и  $A^\delta = \{x : \rho(x, A) > \delta\}$  являются открытыми.<sup>59</sup> Возьмем  $z \in \partial A_\delta$ . По определению это означает, что в сколь угодно малой окрестности точки  $z$  есть точки как принадлежащие  $A_\delta$ , так и не принадлежащие  $A_\delta$ . Пусть  $\rho(z, A) < \delta$ , тогда  $z \in A_\delta$ . Следовательно,  $z$  принадлежит внутренности  $A_\delta$ , а не ее границе. Если же  $\rho(z, A) > \delta$ , то  $z \in A^\delta$  и все точки достаточно малой окрестности  $z$  тоже находятся в  $A^\delta$ . Значит, эта окрестность не пересекается с  $A_\delta$ , и противоречие получено.  $\square$

**Лемма 4.2.** Пусть  $(D, \mathcal{A}, \mathcal{Q})$  — некоторое измеримое пространство с вероятностной мерой. Рассмотрим семейство попарно непересекающихся множеств  $A_\theta \in \mathcal{A}$ ,  $\theta \in \Theta$ . Тогда существует не более чем счетное число значений  $\theta$  таких, что  $\mathcal{Q}(A_\theta) > 0$ .<sup>60</sup>

*Доказательство.* При  $i \geq 0$  обозначим через  $n_i$  число таких множеств  $A_\theta$ , что  $2^{-i} \leq \mathcal{Q}(A_\theta) < 2^{-(i-1)}$ . Поскольку множества  $A_\theta$  попарно не пересекаются, то  $n_i \leq 2^i$ . Отсюда немедленно следует требуемое.  $\square$

**Лемма 4.3.** Пусть  $(D, \mathcal{A}, \mathcal{Q})$  — измеримое пространство с распределением  $\mathcal{Q}$ . Рассмотрим функцию  $f : (D, \mathcal{A}) \mapsto (\mathbb{R}, \mathcal{B}_\mathbb{R})$  такую, что  $0 \leq f(x) < 1$  при любом  $x \in D$ . Для натурального  $k \geq 1$  и  $i = 0, \dots, k$  определим множества  $F_i = F_i^{(k)} = \{x : f(x) \geq i/k\}$ . Тогда

$$\frac{1}{k} \sum_{i=0}^{k-1} \mathcal{Q}(F_i) - \frac{1}{k} \leq \int_D f d\mathcal{Q} \leq \frac{1}{k} \sum_{i=0}^{k-1} \mathcal{Q}(F_i). \quad (4.3)$$

*Доказательство.* Прежде всего заметим, что  $F_0 = D$  и  $F_k = \emptyset$ , так что  $\mathcal{Q}(F_0) = 1$  и  $\mathcal{Q}(F_k) = 0$ . Докажем сначала правое из неравенств (4.3). Обозначив  $G_i = \{x : i/k \leq f(x) < (i+1)/k\}$  и используя тот факт, что  $\mathcal{Q}(G_i) = \mathcal{Q}(F_i) - \mathcal{Q}(F_{i+1})$ , получим, что

$$\begin{aligned} \int_D f d\mathcal{Q} &= \sum_{i=0}^{k-1} \int_{G_i} f d\mathcal{Q} \leq \sum_{i=0}^{k-1} \frac{i+1}{k} \mathcal{Q}(G_i) = \sum_{i=0}^{k-1} \frac{i+1}{k} \mathcal{Q}(F_i) - \sum_{i=0}^{k-1} \frac{i+1}{k} \mathcal{Q}(F_{i+1}) = \\ &= \sum_{i=0}^{k-1} \frac{1}{k} \mathcal{Q}(F_i) + \sum_{i=0}^{k-1} \frac{i}{k} \mathcal{Q}(F_i) - \sum_{i=1}^k \frac{i}{k} \mathcal{Q}(F_i) = \frac{1}{k} \sum_{i=0}^{k-1} \mathcal{Q}(F_i), \end{aligned}$$

так как  $\mathcal{Q}(F_0) = 1$ . Аналогичным образом

$$\int_D f d\mathcal{Q} \geq \sum_{i=0}^{k-1} \frac{i}{k} \mathcal{Q}(G_i) = \sum_{i=0}^{k-1} \frac{i}{k} \mathcal{Q}(F_i) - \sum_{i=0}^{k-1} \frac{i}{k} \mathcal{Q}(F_{i+1}) =$$

<sup>58</sup>Здесь  $\partial B$  обозначает границу множества  $B$ .

<sup>59</sup>А почему?

<sup>60</sup>Конечно, это утверждение интересно только тогда, когда  $\Theta$  является более чем счетным.

$$\begin{aligned}
 &= \sum_{i=0}^{k-1} \frac{i}{k} \mathcal{Q}(F_i) - \sum_{i=0}^{k-1} \frac{i+1}{k} \mathcal{Q}(F_{i+1}) + \sum_{i=0}^{k-1} \frac{1}{k} \mathcal{Q}(F_{i+1}) = \\
 &= \frac{1}{k} \sum_{i=0}^{k-1} \mathcal{Q}(F_{i+1}) = \frac{1}{k} \sum_{i=0}^k \mathcal{Q}(F_i) - \frac{\mathcal{Q}(F_0)}{k} = \frac{1}{k} \sum_{i=0}^{k-1} \mathcal{Q}(F_i) - \frac{1}{k}.
 \end{aligned}$$

Доказательство закончено.  $\square$

## 4.2 Основная теорема

**Теорема 4.1.** (Основная теорема.) Пусть  $(D, \rho, \mathcal{B})$  — измеримое метрическое пространство и  $\{\mathcal{P}_n, n \geq 1\}$ ,  $\mathcal{P}$  — вероятностные меры, определенные на борелевской  $\sigma$ -алгебре  $\mathcal{B}$ . Тогда эквивалентны следующие утверждения.

1.  $\mathcal{P}_n \Rightarrow \mathcal{P}$ .
2. Сходимость (4.1) имеет место для любой функции  $f \in \text{BL}(D)$ .
3. Для любого замкнутого множества  $F$  выполнено неравенство  $\limsup_n \mathcal{P}_n(F) \leq \mathcal{P}(F)$ .
4. Для любого открытого множества  $G$  выполнено неравенство  $\liminf_n \mathcal{P}_n(G) \geq \mathcal{P}(G)$ .
5. Для любого  $A \in \mathcal{B}$  такого, что  $\mathcal{P}(\partial A) = 0$ <sup>61</sup> имеет место сходимость  $\mathcal{P}_n(A) \rightarrow \mathcal{P}(A)$ .

*Доказательство.* Начнем с наиболее простых импликаций.

1.  $\rightarrow$  2. Конечно, это следует из того, что  $\text{C}_b(D) \supset \text{BL}(D)$ .

3.  $\leftrightarrow$  4. Оба утверждения получаются простым переходом к дополнению. Действительно, пусть  $G$  — открытое множество. Обозначая  $F = G^C$  получим, что

$$\liminf_n \mathcal{P}_n(G) = \liminf_n (1 - \mathcal{P}_n(F)) = 1 - \limsup_n \mathcal{P}_n(F),$$

и эквивалентность обоих неравенств становится очевидной.

3. и 4.  $\rightarrow$  5. Пусть  $A \in \mathcal{B}$  и  $\mathcal{P}(\partial A) = 0$ . Обозначим  $\bar{A}$  замыкание множества  $A$  и  $A^0$  внутренность  $A$ . Ясно, что  $\mathcal{P}(\bar{A}) = \mathcal{P}(A^0) = \mathcal{P}(A)$ .<sup>62</sup> Тогда, последовательно применяя неравенства пунктов 3 и 4, получим, что

$$\limsup_n \mathcal{P}_n(A) \leq \limsup_n \mathcal{P}_n(\bar{A}) \leq \mathcal{P}(\bar{A}) = \mathcal{P}(A) = \mathcal{P}(A^0) \leq \liminf_n \mathcal{P}_n(A^0) \leq \liminf_n \mathcal{P}_n(A). \quad (4.4)$$

Следовательно, в (4.4) все неравенства можно заменить на равенства и  $\mathcal{P}_n(A) \rightarrow \mathcal{P}(A)$ .

5.  $\rightarrow$  3. Пусть  $F$  — замкнутое множество. При  $\delta > 0$  рассмотрим множества

$$F_\delta = \{x \in D : \rho(x, F) < \delta\}.$$

Эти множества, конечно же, открыты.<sup>63</sup>

Из Леммы 4.1 следует, что границы  $\partial F_\delta$  множеств  $F_\delta$  попарно не пересекаются. Обозначим  $\Delta \subset (0, \infty)$  множество тех  $\delta$ , для которых  $\mathcal{P}(\partial F_\delta) = 0$ . Согласно Лемме 4.2 множество  $(0, \infty) \setminus \Delta$  не более чем счетно и поэтому можно найти сколь угодно малое  $\delta \in \Delta$ . Кроме того,  $F_\delta \downarrow F$  при  $\delta \downarrow 0$ .

Зафиксировав  $\varepsilon > 0$ , выберем  $\delta \in \Delta$  такое, что  $\mathcal{P}(F_\delta) \leq \mathcal{P}(F) + \varepsilon$ . Тогда, так как  $\lim_n \mathcal{P}_n(F_\delta) = \mathcal{P}(F_\delta)$ , то

$$\limsup_n \mathcal{P}_n(F) \leq \limsup_n \mathcal{P}_n(F_\delta) = \mathcal{P}(F_\delta) \leq \mathcal{P}(F) + \varepsilon.$$

Поскольку  $\varepsilon$  — произвольное положительное число, то утверждение доказано.

<sup>61</sup>Часто такие множества  $A$  называют множествами  $\mathcal{P}$ -непрерывности. Поскольку граница  $\partial A$  — замкнутое множество, то  $\partial A \in \mathcal{B}$ .

<sup>62</sup>Здесь мы пользуемся тем, что  $\mathcal{P}(A^0) \leq \mathcal{P}(A) \leq \mathcal{P}(\bar{A})$  и что  $\mathcal{P}(\bar{A} \setminus A^0) = \mathcal{P}(\partial A) = 0$ .

<sup>63</sup>Действительно?

2.  $\rightarrow$  3. Пусть  $F$  — замкнутое множество. Рассмотрим функции  $f_u$ , определенные в Предложении 3.5. Как уже доказано в Предложении 3.5,  $f_u \in \text{BL}(D)$ , причем  $f_u \downarrow \mathbf{1}_F$  при  $u \rightarrow \infty$ . Выбрав  $u$  столь большим, что  $\int_D f_u d\mathcal{P} \leq \mathcal{P}(F) + \varepsilon$ , приходим к неравенству

$$\limsup_n \mathcal{P}_n(F) \leq \limsup_n \int_D f_u d\mathcal{P}_n = \int_D f_u d\mathcal{P} \leq \mathcal{P}(F) + \varepsilon,$$

что и заканчивает доказательство.

3.  $\rightarrow$  1. Эта последняя импликация, которую нужно проверить. Согласно Предложению 4.1 достаточно доказать, что для любой  $f \in C_b(D)$  такой, что  $f(x) \in [0, 1]$  при любом  $x \in D$  имеет место неравенство (4.2). Выберем  $k > 1$  и введем, как и в Лемме 4.3, множества  $F_i = \{x : f(x) \geq i/k\}$ , которые в нашем случае являются замкнутыми. Воспользовавшись последовательно правой (с  $\mathcal{Q} = \mathcal{P}_n$ ) и левой (с  $\mathcal{Q} = \mathcal{P}$ ) частью неравенства (4.3), получим, что

$$\limsup_n \int_D f d\mathcal{P}_n \leq \limsup_n \frac{1}{k} \sum_{i=0}^k \mathcal{P}_n(F_i) \leq \frac{1}{k} \sum_{i=0}^k \limsup_n \mathcal{P}_n(F_i) \leq \frac{1}{k} \sum_{i=0}^k \mathcal{P}(F_i) \leq \int_D f d\mathcal{P} + 1/k.$$

Ввиду произвольности  $k$  утверждение доказано.  $\square$

**Замечание 4.2.** Для конкретных измеримых метрических пространств  $(D, \rho, \mathcal{B})$  можно найти и другие условия, эквивалентные слабой сходимости  $\mathcal{P}_n \Rightarrow \mathcal{P}$ . Так, для евклидовых пространств  $\mathbb{R}^d$ , оснащенных стандартными борелевскими  $\sigma$ -алгебрами  $\mathcal{B}_d$ , такими условиями являются сходимости соответствующих функций распределений  $F_n(\mathbf{x}) \rightarrow F(\mathbf{x})$  в любой точке непрерывности  $\mathbf{x}$  предельной функции распределения  $F$  или поточечная сходимость характеристических функций  $\varphi_n$  распределений  $\mathcal{P}_n$  к характеристической функции  $\varphi$  распределения  $\mathcal{P}$ .

## 5 Слабая сходимость и отображения

Введем сначала следующие обозначения. Пусть  $(D_1, \mathcal{A}_1)$  и  $(D_2, \mathcal{A}_2)$  — измеримые пространства и  $f : (D_1, \mathcal{A}_1) \mapsto (D_2, \mathcal{A}_2)$ . Тогда любому распределению  $\mathcal{P}$ , определенному в  $(D_1, \mathcal{A}_1)$ , будет соответствовать распределение  $\mathcal{P}f^{-1}$  в  $(D_2, \mathcal{A}_2)$ , заданное равенством

$$\mathcal{P}f^{-1}(A) = \mathcal{P}(f^{-1}(A)) = \mathcal{P}(x : f(x) \in A), \quad A \in \mathcal{A}_2. \quad (5.1)$$

**Замечание 5.1.** Обозначение  $\mathcal{P}f^{-1}$  связано с первым равенством в (5.1).

Определение (5.1) имеет прозрачный вероятностный смысл. Проинтерпретируем  $(D_1, \mathcal{A}_1, \mathcal{P})$  как некоторое вероятностное пространство. Иначе говоря, рассмотрим вероятностное пространство  $(\Omega, \mathcal{F}, \mathbb{P})$  с  $\Omega = D_1$ ,  $\mathcal{F} = \mathcal{A}_1$  и  $\mathbb{P} = \mathcal{P}$ . Тогда  $\xi = f$  можно рассматривать как случайную величину со значениями в  $(D_2, \mathcal{A}_2)$ . Поэтому  $\mathcal{P}f^{-1}$  (то есть  $\mathcal{P}\xi^{-1}$ ) — это просто распределение случайной величины  $\xi$ , в другом контексте было бы удобно обозначить это распределение как  $\mathcal{P}_\xi$  или как  $\mathcal{L}(\xi)$ .

Разница в обозначениях объясняется тем, что в «обычном» вероятностном пространстве  $(\Omega, \mathcal{F}, \mathbb{P})$  вероятностная мера  $\mathbb{P}$  считается раз и навсегда зафиксированной, в то время как в теории слабой сходимости распределения  $\mathcal{P}$ , определенные на топологическом измеримом пространстве  $(D, \mathcal{A})$ , могут меняться.

Пусть теперь  $(D_1, \mathcal{C}_1, \mathcal{B}_1)$  и  $(D_2, \mathcal{C}_2, \mathcal{B}_2)$  — измеримые топологические пространства (а не просто измеримые),  $f : (D_1, \mathcal{B}_1) \mapsto (D_2, \mathcal{B}_2)$ , и в  $(D_1, \mathcal{B}_1)$  задана последовательность распределений  $\mathcal{P}_n$ , слабо сходящаяся к распределению  $\mathcal{P}$ . Возникает естественный вопрос: при каких условиях на отображение  $f$  можно гарантировать, что  $\mathcal{P}_n f^{-1} \Rightarrow \mathcal{P} f^{-1}$ ? Или: когда слабая сходимость сохраняется при отображениях?

Снова переведем этот вопрос на более привычный язык случайных величин. Пусть  $(\Omega, \mathcal{F}, \mathbb{P})$  — некоторое вероятностное пространство и  $\xi_n : (\Omega, \mathcal{F}) \mapsto (D_1, \mathcal{B}_1)$  — случайные величины со значениями в  $(D_1, \mathcal{B}_1)$ . Обозначим  $\mathcal{P}_n$  распределение  $\xi_n$  и предположим, что  $\mathcal{P}_n \Rightarrow \mathcal{P}$ , где  $\mathcal{P}$  — распределение некоторой случайной величины  $\xi : (\Omega, \mathcal{F}) \mapsto (D_1, \mathcal{B}_1)$ .

Таким образом,  $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$ . Для каких  $f$  имеет место сходимость  $\mathcal{L}(f(\xi_n)) \Rightarrow \mathcal{L}(f(\xi))$ ?

### 5.1 Сохранение слабой сходимости при отображениях

Поскольку само понятие слабой сходимости тесно связано с непрерывностью, естественно ожидать, что и условия, накладываемые на  $f$ , тоже должны быть связаны с этим понятием. Самое простое условие на  $f$  — просто условие непрерывности.

**Предложение 5.1.** Пусть  $f$  — непрерывное отображение  $D_1 \mapsto D_2$ . Тогда из сходимости  $\mathcal{P}_n \Rightarrow \mathcal{P}$  следует сходимость  $\mathcal{P}_n f^{-1} \Rightarrow \mathcal{P} f^{-1}$ .

*Доказательство.* По определению нам нужно доказать, что для любой функции  $g \in C_b(D_2)$  имеет место сходимость

$$\int_{D_2} g d\mathcal{P}_n f^{-1} \rightarrow \int_{D_2} g d\mathcal{P} f^{-1}. \quad (5.2)$$

Но если  $\mathcal{Q}$  — некоторое распределение в  $(D_1, \mathcal{B}_1)$ , то

$$\int_{D_2} g d\mathcal{Q} f^{-1} = \int_{D_1} g(f) d\mathcal{Q} \quad (5.3)$$

(это просто результат замены переменной  $y = f(x)$  в интеграле по мере  $\mathcal{Q}$ ). Поэтому (5.2) переписывается в виде

$$\int_{D_1} g \circ f d\mathcal{P}_n \rightarrow \int_{D_1} g \circ f d\mathcal{P}.$$

Поскольку функция  $g \circ f : D_1 \mapsto \mathbb{R}$  непрерывна (как суперпозиция непрерывных отображений) и ограничена, то утверждение доказано.  $\square$

Перепишем равенство (5.3) в более привычных вероятностных терминах — тогда оно станет просто очевидным. Пусть  $(\Omega, \mathcal{F}, \mathbb{P}) = (D_1, \mathcal{B}_1, \mathcal{Q})$  и  $\xi = f$ . Тогда (5.3) превратится в равенство

$$\int_{D_2} g d\mathcal{P}_\xi = \int_{\Omega} g(\xi) d\mathbb{P} = \mathbb{E}g(\xi),$$

то есть в одно из стандартных свойств математических ожиданий (с заменой  $\mathbb{R}$  или  $\mathbb{R}^d$  на  $D_2$ ).

Сейчас мы докажем гораздо более интересное утверждение о сохранении слабой сходимости при отображениях. Для этого нам понадобится следующая лемма (см. [1, стр. 309]).

**Лемма 5.1.** Пусть  $(D_1, \rho_1)$  и  $(D_2, \rho_2)$  — метрические пространства и  $f : D_1 \mapsto D_2$ . Обозначим  $G_f$  множество точек разрыва отображения  $f$ . Тогда  $G_f$  является измеримым относительно борелевской  $\sigma$ -алгебры  $\mathcal{B}_1$  в  $D_1$ .

*Доказательство.* Для любых  $\varepsilon, \delta > 0$  обозначим

$$A_{\varepsilon, \delta} = \{x \in D_1 : \text{существуют } y, z \in D_1 \text{ такие что } \max(\rho_1(x, y), \rho_1(x, z)) < \delta \text{ и } \rho_2(f(y), f(z)) \geq \varepsilon\}.$$

Нетрудно видеть<sup>64</sup>, что множество  $A_{\varepsilon, \delta}$  является открытым. Тогда<sup>65</sup>

$$G_f = \bigcup_{\varepsilon} \bigcap_{\delta} A_{\varepsilon, \delta},$$

где  $\varepsilon, \delta$  принимают только положительные рациональные значения. Отсюда сразу же следует, что  $G_f \in \mathcal{B}_1$ .  $\square$

**Предложение 5.2.** Пусть  $(D_1, \rho_1, \mathcal{B}_1)$  и  $(D_2, \rho_2, \mathcal{B}_2)$  — метрические измеримые пространства и распределения  $\{\mathcal{P}_n\}_{n \geq 1}, \mathcal{P}$  заданы на борелевской  $\sigma$ -алгебре  $\mathcal{B}_1$ . Рассмотрим отображение  $f : (D_1, \mathcal{B}_1) \mapsto (D_2, \mathcal{B}_2)$ . Если  $\mathcal{P}_n \Rightarrow \mathcal{P}$  и  $\mathcal{P}(G_f) = 0$ ,<sup>66</sup> то  $\mathcal{P}_n f^{-1} \Rightarrow \mathcal{P} f^{-1}$ .

*Доказательство.* Пусть  $F$  — замкнутое множество в  $(D_2, \rho_2)$ . Согласно Основной теореме 4.1 достаточно доказать, что

$$\limsup_n \mathcal{P}_n(f^{-1}(F)) = \mathcal{P}_n f^{-1}(F) \leq \mathcal{P} f^{-1}(F) = \mathcal{P}(f^{-1}(F)).$$

Отметим, что в случае непрерывного отображения  $f$  множество  $f^{-1}(F)$  является замкнутым и поэтому это неравенство автоматически следует из сходимости  $\mathcal{P}_n \Rightarrow \mathcal{P}$ . Тем самым мы получили еще одно доказательство Предложения 5.1.

Ясно, что по той же Теореме 4.1

$$\limsup_n \mathcal{P}_n(f^{-1}(F)) \leq \limsup_n \mathcal{P}_n(\overline{f^{-1}(F)}) \leq \mathcal{P}(\overline{f^{-1}(F)}).$$

С другой стороны,  $\overline{f^{-1}(F)} \setminus f^{-1}(F) \subset G_f$ . Действительно, пусть  $x \in \overline{f^{-1}(F)}$ ,  $x \notin f^{-1}(F)$  (это означает, что  $f(x) \notin F$ ) и  $x$  является точкой непрерывности отображения  $f$ .

Поскольку точка  $x$  является точкой границы множества  $f^{-1}(F)$ , то существует последовательность  $x_m \in f^{-1}(F)$  (что эквивалентно  $f(x_m) \in F$ ), сходящаяся (в метрике  $\rho_1$ ) к  $x$ . Если точка  $x$  — точка непрерывности  $f$ , то из этого следует, что  $f(x_m) \rightarrow f(x)$  в метрике  $\rho_2$ .

Следовательно (так как множество  $F$  замкнуто и  $f(x_m) \in F$ ),  $f(x) \in F$ , что противоречит предположению.

Итак,  $\overline{f^{-1}(F)} \setminus f^{-1}(F) \subset G_f$ . Поскольку  $\mathcal{P}(G_f) = 0$ , то из этого следует, что  $\mathcal{P}(\overline{f^{-1}(F)}) = \mathcal{P}(f^{-1}(F))$ , и предложение доказано.  $\square$

<sup>64</sup> Действительно нетрудно?

<sup>65</sup> Докажите!

<sup>66</sup> В этом случае отображение  $f$  называется  $\mathcal{P}$ -почти всюду непрерывным.

## 5.2 Сходимость интегралов. Интеграл Римана

Из Предложения 5.2 вытекает очень полезное следствие. Пусть  $(D, \rho, \mathcal{B})$  — измеримое метрическое пространство,  $\{\mathcal{P}_n\}_{n \geq 1}, \mathcal{P}$  — распределения, определенные на  $\sigma$ -алгебре борелевских множеств  $\mathcal{B}$ , и  $\mathcal{P}_n \Rightarrow \mathcal{P}$ . По определению это означает, что

$$\int_D f d\mathcal{P}_n \rightarrow \int_D f d\mathcal{P} \quad (5.4)$$

для любой непрерывной ограниченной функции  $f$ . Возникает естественный вопрос — для каких еще функций  $f$  из слабой сходимости  $\mathcal{P}_n$  к  $\mathcal{P}$  будет следовать сходимость интегралов (5.4)? Оказывается, условие непрерывности можно существенно ослабить.

**Следствие 5.1.** Пусть  $(D, \rho, \mathcal{B})$  — измеримое метрическое пространство,  $\{\mathcal{P}_n\}_{n \geq 1}, \mathcal{P}$  — распределения, определенные на  $\sigma$ -алгебре борелевских множеств  $\mathcal{B}$ , и  $\mathcal{P}_n \Rightarrow \mathcal{P}$ . Если функция  $f : (D, \mathcal{B}) \mapsto (\mathbb{R}, \mathcal{B}_R)$  является ограниченной и  $\mathcal{P}$ -почти всюду непрерывной, то имеет место сходимость (5.4).

*Доказательство.* Согласно Предложению 5.2, для любой непрерывной ограниченной функции  $g : \mathbb{R} \mapsto \mathbb{R}$  имеет место сходимость

$$\int_D g \circ f d\mathcal{P}_n \rightarrow \int_D g \circ f d\mathcal{P}.$$

Пусть  $|f| \leq M$ . Возьмем

$$g(x) = \begin{cases} -M & \text{при } x < -M, \\ x & \text{при } |x| \leq M, \\ M & \text{при } x > M. \end{cases}$$

Тогда, очевидно,  $g \circ f = f$  и утверждение доказано.  $\square$

Приведем несколько простых примеров применения Следствия 5.1.

### Пример 5.1.

1. Пусть  $(D, \mathcal{B}) = (\mathbb{R}, \mathcal{B}_R)$  и  $\mathcal{P}_n \Rightarrow \mathcal{P}$ . Обозначим  $F_n, F$  функции распределения, соответствующие  $\mathcal{P}_n, \mathcal{P}$ . Для некоторого  $z \in \mathbb{R}$  введем функцию  $f(x) = \mathbf{1}_{(-\infty, z)}(x)$ , которая имеет единственный разрыв в точке  $z$ . Таким образом,  $G_f = \{z\}$ .

Неравенство  $\mathcal{P}(\{z\}) > 0$  имеет место тогда и только тогда, когда  $z$  является точкой разрыва функции  $F$ . Поэтому из Следствия 5.1 вытекает, что

$$F_n(z) = \int f d\mathcal{P}_n \rightarrow \int f d\mathcal{P} = F(z)$$

в любой точке непрерывности функции распределения  $F$ . Конечно, этот результат и так хорошо известен.

2. Пусть в условиях Следствия 5.1  $\mathcal{P} = \delta_a$ , где  $a$  — некоторая точка, принадлежащая  $D$ . Тогда

$$\int f d\mathcal{P}_n \rightarrow f(a)$$

для любой измеримой ограниченной функции, непрерывной в точке  $a$ .

Следующий пример применения Следствия 5.1 гораздо более интересен.

**Пример 5.2.** Пусть  $D = [0, 1]$  и  $\mathcal{B}$  —  $\sigma$ -алгебра борелевских подмножеств  $D$ . Возьмем  $n > 1$  и рассмотрим последовательность разбиений  $\mathcal{X}_n$  отрезка  $[0, 1]$  точками  $0 = x_0^{(n)} < x_1^{(n)} < \dots < x_n^{(n)} = 1$ . Если  $y_i^{(n)}$  —

некоторые точки, принадлежащие отрезкам  $[x_i^{(n)}, x_{i+1}^{(n)}]$  (здесь  $i = 0 \dots n-1$ ), то любой ограниченной на  $D$  функции  $f$  можно сопоставить *сумму Римана*

$$I_n(f) = \sum_{i=0}^{n-1} f(y_i^{(n)}) \Delta_{in}, \quad \text{где } \Delta_{in} = x_{i+1}^{(n)} - x_i^{(n)}. \quad (5.5)$$

Функция  $f$  называется *интегрируемой по Риману*, если для любой последовательности разбиений  $X_n$ , удовлетворяющей условию  $\max_{0 \leq i < n} \Delta_{in} \rightarrow 0$  при  $n \rightarrow \infty$ , и для любого выбора точек  $y_i^{(n)} \in [x_i^{(n)}, x_{i+1}^{(n)}]$  имеет место сходимость  $I_n(f) \rightarrow I(f)$ , причем предел  $I(f)$  не зависит ни от разбиений  $X_n$ , ни от выбора  $y_i^{(n)}$ . Этот предел называется *интегралом Римана функции  $f$* .

Хорошо известны следующие свойства интеграла Римана:

1. любая непрерывная на отрезке  $[0, 1]$  функция является интегрируемой по Риману;
2. если интеграл Римана функции  $f$  существует, то он совпадает с ее интегралом Лебега.

Опираясь на эти свойства, докажем следующее утверждение.

**Предложение 5.3.** *Любая ограниченная непрерывная почти всюду по мере Лебега функция является интегрируемой по Риману.*

*Доказательство.* Пусть  $f \in C_b(D)$ . Обозначим  $\text{mes}_1$  меру Лебега на отрезке  $[0, 1]$ , то есть равномерное распределение  $U(0, 1)$ ,

Наряду с суммой Римана (5.5) рассмотрим дискретное распределение  $\mathcal{P}_n$ , задаваемое таблицей

$$\mathcal{P}_n : \quad \begin{pmatrix} y_0^{(n)} & \dots & y_\ell^{(n)} & \dots & y_{n-1}^{(n)} \\ \Delta_{0,n} & \dots & \Delta_{\ell,n} & \dots & \Delta_{n-1,n} \end{pmatrix}.$$

Тогда  $I_n(f) = \int_D f d\mathcal{P}_n$  и, поскольку любая функция  $f \in C_b(D)$  интегрируема по Риману, а  $I(f) = \int_0^1 f d\text{mes}_1$ , то сходимость  $I_n(f) \rightarrow I(f)$ , означает, что  $\mathcal{P}_n \Rightarrow \text{mes}_1$ . Применяя Следствие 5.1, получаем нужный результат.  $\square$

**Замечание 5.2.** На самом деле нетрудно<sup>67</sup> (от противного) доказать и обратный результат: если функция  $f$  интегрируема по Риману, то она ограничена и множество ее точек разрыва имеет лебегову меру ноль.

<sup>67</sup> Попробуйте!

## 6 Равномерная интегрируемость функций и слабая сходимость

Результат Следствия 5.1 показывает, как можно ослабить условие непрерывности функции  $f \in C_b(D)$ , чтобы из слабой сходимости  $\mathcal{P}_n \Rightarrow \mathcal{P}$  по-прежнему вытекала сходимость (5.4). Здесь мы займемся ослаблением условий на ограниченность функции  $f$ . Для этого нам потребуется ввести новое понятие.

### 6.1 Функции, равномерно интегрируемые с семейством распределений

Дадим следующее определение.<sup>68</sup>

**Определение 6.1.** Пусть  $(D, \mathcal{A})$  — измеримое пространство и  $\Theta$  — некоторое параметрическое множество. Рассмотрим семейство распределений  $\{\mathcal{P}_\theta, \theta \in \Theta\}$ , определенных в  $(D, \mathcal{A})$  и измеримую функцию  $f : (D, \mathcal{A}) \mapsto (\mathbb{R}, \mathcal{B}_R)$ . Функция  $f$  называется *равномерно интегрируемой с семейством*  $\{\mathcal{P}_\theta, \theta \in \Theta\}$ , если

$$\sup_{\theta \in \Theta} \int_{\{x: |f(x)| > t\}} |f(x)| \mathcal{P}_\theta(dx) \rightarrow 0 \quad \text{при } t \rightarrow +\infty. \quad (6.1)$$

**Замечание 6.1.** Из условия (6.1), конечно, следует, что  $f \in \mathbb{L}^1(d\mathcal{P}_\theta)$  при любом  $\theta$ .

Для дальнейшего будет удобно ввести новые обозначения: для любого  $A \in \mathcal{A}$  и любой измеримой функции  $g$  мы будем писать  $E_\theta(g, A)$  вместо более громоздкого  $\int_A g d\mathcal{P}_\theta$ . Соответственно вместо  $\int g d\mathcal{P}_\theta$  мы будем писать  $E_\theta g$ . В этих обозначениях (6.1) переписывается как

$$\sup_{\theta \in \Theta} E_\theta(|f|, |f| > t) \rightarrow 0. \quad (6.2)$$

Отметим сразу же простейшие свойства равномерно интегрируемых функций.

- Лемма 6.1.** 1. Если функция  $f$  ограничена сверху по модулю некоторой постоянной  $M$ , то она равномерно интегрируема с любым семейством распределений.  
 2. Если множество  $\Theta$  конечно и  $f \in \mathbb{L}^1(d\mathcal{P}_\theta)$  при любом  $\theta$ , то  $f$  равномерно интегрируема с семейством распределений  $\{\mathcal{P}_\theta, \theta \in \Theta\}$ .  
 3. Если функция  $f$  равномерно интегрируема с семейством  $\{\mathcal{P}_\theta, \theta \in \Theta\}$ , то  $\sup_\theta E_\theta |f| < \infty$ .

*Доказательство.* 1. Для доказательства достаточно заметить, что в этом случае левая часть (6.1) будет равна нулю при  $t > M$ .

2. Этот факт (с учетом конечности  $\Theta$ ) следует из теоремы Лебега о мажорируемой сходимости, так как функции  $h_t = |f| \mathbf{1}_{|f| > t}$  стремятся поточечно к нулю при  $t \rightarrow +\infty$  и мажорируются суммируемой функцией  $|f|$ .

3. Действительно,

$$\sup_{\theta \in \Theta} E_\theta |f| \leq \sup_{\theta \in \Theta} E_\theta(|f|, |f| > t) + \sup_{\theta \in \Theta} E_\theta(|f|, |f| \leq t) \leq \sup_{\theta \in \Theta} E_\theta(|f|, |f| > t) + t,$$

что и завершает доказательство.  $\square$

**Замечание 6.2.** Из первых двух утверждений Леммы 6.1 следует, что Определение 6.1 является содержательным в случае, когда семейство  $\{\mathcal{P}_\theta, \theta \in \Theta\}$  бесконечно, а функция  $f$  не является ограниченной.

<sup>68</sup>Материал этого раздела является переложением [2, гл. II §2], где рассматриваются равномерно интегрируемые семейства функций относительно одного распределения.

Приведем один результат о характеристике равномерной интегрируемости функции с семейством распределений. Пусть  $\{\mathcal{P}_\theta, \theta \in \Theta\}$  — семейство распределений, определенных на измеримом пространстве  $(D, \mathcal{A})$  и  $f : (D, \mathcal{A}) \mapsto \mathbb{R}$ .

**Теорема 6.1.** 1. Пусть существует такая измеримая функция  $G : [0, +\infty) \mapsto [0, +\infty)$ , что  $G(t)/t \rightarrow \infty$  при  $t \rightarrow \infty$  и

$$\sup_{\theta} \mathbb{E}_{\theta}(G \circ |f|) < \infty. \quad (6.3)$$

Тогда функция  $f$  равномерно интегрируема с семейством  $\{\mathcal{P}_\theta, \theta \in \Theta\}$ .

2. Пусть функция  $f$  равномерно интегрируема с семейством  $\{\mathcal{P}_\theta, \theta \in \Theta\}$ . Тогда существует такая выпуклая вниз измеримая функция  $G : [0, +\infty) \mapsto [0, +\infty)$ , что  $G(t)/t \rightarrow \infty$  при  $t \rightarrow \infty$  и выполняется неравенство (6.3).

*Доказательство.* 1. Обозначим  $M$  левую часть неравенства (6.3) и заметим, что для любого  $\varepsilon > 0$  при  $t > t_0 = t_0(\varepsilon)$  выполняется неравенство  $t < \varepsilon G(t)$ . Поэтому при  $t > t_0$

$$\mathbb{E}_{\theta}(|f|, |f| > t) \leq \varepsilon \mathbb{E}_{\theta}(G(|f|), |f| > t) \leq \varepsilon \mathbb{E}_{\theta}(G(|f|)) \leq \varepsilon M,$$

что доказывает первое утверждение.

2. Будем искать функцию  $G$  в виде

$$G(t) = \int_0^t g(s) ds,$$

где при  $s \geq 0$

$$g(s) = \sum_{m \geq 0} g_m \mathbb{I}_{(m, m+1]}(s),$$

причем числа  $g_k$  предполагаются целыми, положительными и стремящимися к бесконечности, а  $g_0 = 0$ . Ясно, что эти условия обеспечивают выпуклость вниз положительной функции  $G$ , а также то, что  $G(t)/t \rightarrow \infty$  при  $t \rightarrow \infty$ .<sup>69</sup>

Тогда для любого  $t \geq 0$  и  $k = \lfloor t \rfloor$

$$G(t) = \sum_{m=0}^k g_m + (t - k)g_{k+1} \leq \sum_{m=0}^{k+1} g_m$$

и

$$\begin{aligned} \int G(|f|) d\mathcal{P}_{\theta} &\leq \sum_{k=0}^{\infty} \int_{k < |f| \leq k+1} \left( \sum_{m=0}^{k+1} g_m \right) d\mathcal{P}_{\theta} = \sum_{k=0}^{\infty} \sum_{m=0}^{k+1} g_m \mathcal{P}_{\theta}(k < |f| \leq k+1) = \\ &= \sum_{k=0}^{\infty} \sum_{m=0}^{k+1} g_m \left( \mathcal{P}_{\theta}(|f| > k) - \mathcal{P}_{\theta}(|f| > k+1) \right) = \sum_{m=1}^{\infty} g_m \sum_{k \geq m} \left( \mathcal{P}_{\theta}(|f| > k) - \mathcal{P}_{\theta}(|f| > k+1) \right) = \\ &= \sum_{m=1}^{\infty} g_m \mathcal{P}_{\theta}(|f| > m). \end{aligned}$$

Покажем, что можно подобрать числа  $g_m$  так, что

$$\sup_{\theta} \sum_{m=1}^{\infty} g_m \mathcal{P}_{\theta}(|f| > m) < \infty.$$

<sup>69</sup>Проверьте!

Действительно, так как  $f$  равномерно интегрируема с семейством  $\mathcal{P}_\theta$ , то существует такая целая положительная последовательность  $c_n \rightarrow \infty$  с  $c_1 = 1$ , что равномерно по  $\theta$

$$\sum_n \int_{|f| > c_n} |f| d\mathcal{P}_\theta < \infty.$$

В свою очередь,

$$\begin{aligned} \int_{|f| > c_n} |f| d\mathcal{P}_\theta &= \sum_{k=c_n}^{\infty} \int_{k < |f| \leq k+1} |f| d\mathcal{P}_\theta \geq \sum_{k=c_n}^{\infty} k \left( \mathcal{P}_\theta(|f| > k) - \mathcal{P}_\theta(|f| > k+1) \right) = \\ &= \lim_{N \rightarrow \infty} \sum_{k=c_n}^N k \left( \mathcal{P}_\theta(|f| > k) - \mathcal{P}_\theta(|f| > k+1) \right) = \lim_{N \rightarrow \infty} A_N. \end{aligned}$$

При фиксированном  $N > c_n$

$$\begin{aligned} A_N &= \sum_{k=c_n}^N k \mathcal{P}_\theta(|f| > k) - \sum_{k=c_n}^N k \mathcal{P}_\theta(|f| > k+1) = \\ &= c_n \mathcal{P}_\theta(|f| > c_n) - N \mathcal{P}_\theta(|f| > N+1) + \sum_{k=c_n+1}^N \mathcal{P}_\theta(|f| > k) \geq \\ &\geq \sum_{k=c_n}^N \mathcal{P}_\theta(|f| > k) - N \mathcal{P}_\theta(|f| > N+1). \end{aligned} \tag{6.4}$$

Так как при  $t \rightarrow \infty$

$$0 \leftarrow \sup_{\theta} E_{\theta}(|f|, |f| > t) \geq t \sup_{\theta} \mathcal{P}_{\theta}(|f| > t),$$

то второй член в правой части (6.4) равномерно по  $\theta$  стремится к нулю при  $N \rightarrow \infty$ . Поэтому

$$\begin{aligned} \infty &> \sup_{\theta} \sum_n \int_{|f| > c_n} |f| d\mathcal{P}_\theta \geq \sup_{\theta} \sum_n \sum_{k=c_n}^{\infty} \mathcal{P}_{\theta}(|f| > k) = \\ &= \sup_{\theta} \sum_k \sum_{n: c_n \leq k} \mathcal{P}_{\theta}(|f| > k) = \sup_{\theta} \sum_k g_k \mathcal{P}_{\theta}(|f| > k), \end{aligned}$$

где  $g_k$  — это число тех  $n$ , для которых  $c_n \leq k$ . Таким образом, утверждение доказано.<sup>70</sup>  $\square$

**Замечание 6.3.** Пусть семейство распределений состоит из одного-единственного распределения  $\mathcal{P}$ , а (неограниченная) функция  $f$  суммируема относительно этого распределения. Тогда, конечно, она является равномерно интегрируемой и, согласно второму утверждению Теоремы 6.1, существует суммируемая функция  $f_1 = G \circ |f|$  такая, что  $f_1/|f| \rightarrow \infty$  при  $|f| \rightarrow \infty$ .

Это означает, что свойство функций быть суммируемыми обладает определенным запасом: если неограниченная функция суммируема, то существует «еще более неограниченная» функция, которая тоже суммируема.

## 6.2 Сходимость интегралов от неограниченных функций

Вернемся к слабой сходимости распределений.

<sup>70</sup>Убедитесь в том, что числа  $g_k$  обладают нужными свойствами.

**Теорема 6.2.** Пусть  $(D, \rho, \mathcal{B})$  — измеримое метрическое пространство,  $\{\mathcal{P}_n, n \geq 1\}$  и  $\mathcal{P}$  — распределения, определенные на  $\sigma$ -алгебре  $\mathcal{B}$  борелевских подмножеств множества  $D$ , и  $\mathcal{P}_n \Rightarrow \mathcal{P}$ . Пусть функция  $f : (D, \mathcal{B}) \mapsto (\mathbb{R}, \mathcal{B}_R)$  является  $\mathcal{P}$ -почти всюду непрерывной.

Если  $f$  суммируема относительно  $\mathcal{P}$  и равномерно интегрируема с семейством  $\{\mathcal{P}_n, n \geq 1\}$ , то имеет место сходимость (5.4).

*Доказательство.* Обозначим  $h_N = f \mathbf{1}_{|f|>N}$  и  $f_N \stackrel{\text{def}}{=} f - h_N = f \mathbf{1}_{|f|\leq N}$ . Множество  $G_{f_N}$  точек разрыва функции  $f_N$  удовлетворяет соотношению  $G_{f_N} \subset G_f \cup \Gamma_N$ , где  $\Gamma_N = \{x : |f(x)| = N\}$ , а  $G_f$  — множество точек разрыва функции  $f$ .<sup>71</sup> Поэтому множества  $\Gamma_N$  попарно не пересекаются при разных  $N$ .

Воспользовавшись Леммой 4.2, выберем последовательность  $N \rightarrow \infty$  такую, что  $\mathcal{P}(\Gamma_N) = 0$ . Тогда для любого такого  $N$  функция  $f_N$  является ограниченной и непрерывной  $\mathcal{P}$ -п.в. Поэтому, согласно Следствию 5.1, при  $n \rightarrow \infty$

$$\int_D f_N d\mathcal{P}_n \rightarrow \int_D f_N d\mathcal{P}. \quad (6.5)$$

Далее, при фиксированном  $\varepsilon > 0$  для достаточно больших  $N \geq N_0$  выполняются неравенства

$$\int_D |h_N| d\mathcal{P} = \int_{|f|>N} |f| d\mathcal{P} < \varepsilon \quad \text{и} \quad \sup_n \int_D |h_N| d\mathcal{P}_n = \sup_n \int_{|f|>N} |f| d\mathcal{P}_n < \varepsilon.$$

Первое из этих неравенств следует из того, что функция  $f$  суммируема относительно  $\mathcal{P}$ , а второе — из ее равномерной интегрируемости с семейством  $\{\mathcal{P}_n, n \geq 1\}$ .

Следовательно, ввиду (6.5)

$$\begin{aligned} \limsup_n \int_D f d\mathcal{P}_n &\leq \limsup_n \int_D f_N d\mathcal{P}_n + \limsup_n \int_D h_N d\mathcal{P}_n \leq \\ &\leq \limsup_n \int_D f_N d\mathcal{P}_n + \limsup_n \int_D |h_N| d\mathcal{P}_n \leq \\ &\leq \int_D f_N d\mathcal{P} + \varepsilon \leq \int_D f d\mathcal{P} + \int_D |h_N| d\mathcal{P} + \varepsilon \leq \int_D f d\mathcal{P} + 2\varepsilon \end{aligned}$$

при достаточно больших  $N = N(\varepsilon)$ . Ввиду произвольности  $\varepsilon$  получаем, что

$$\limsup_n \int_D f d\mathcal{P}_n \leq \int_D f d\mathcal{P}. \quad (6.6)$$

Ссылка на доказательство первого пункта Предложения 4.1 заканчивает доказательство.  $\square$

**Замечание 6.4.** Рассмотрим последовательность случайных величин  $\xi_n$  таких, что  $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$ , причем все эти случайные величины имеют конечные вторые моменты. В каких случаях мы можем гарантировать, что  $E\xi_n \rightarrow E\xi$ ? Согласно Теореме 6.2, для этого достаточно,<sup>72</sup> чтобы  $\sup_n E\xi_n^2 < \infty$ .

<sup>71</sup>См. Лемму 4.1 и переход 5.  $\rightarrow$  3. в доказательстве Теоремы 4.1.

<sup>72</sup>Почему?

## 7 Классы функций, определяющие слабую сходимость

В предыдущих разделах мы обсуждали, для каких классов функций  $f$  (кроме, естественно, непрерывных ограниченных) из слабой сходимости  $\mathcal{P}_n \Rightarrow \mathcal{P}$  следует сходимость интегралов

$$\int_D f d\mathcal{P}_n \rightarrow \int_D f d\mathcal{P}. \quad (7.1)$$

Здесь мы рассмотрим обратную задачу. Точнее, мы поставим вопрос следующим образом. Пусть сходимость (7.1) имеет место для некоторого подмножества  $\Phi$  непрерывных ограниченных функций. При каких условиях это гарантирует слабую сходимость  $\mathcal{P}_n \Rightarrow \mathcal{P}$ ?

### 7.1 Общее утверждение

Дадим следующее определение.

**Определение 7.1.** Пусть  $\mathcal{P}_n$  с  $n \geq 1$  и  $\mathcal{P}$  — распределения, определенные на борелевской  $\sigma$ -алгебре  $\mathcal{B}$  измеримого метрического пространства  $(D, \rho, \mathcal{B})$ .

Подмножество  $\Phi$  множества  $C_b(D)$  непрерывных ограниченных функций с областью определения  $D$  называется *классом функций, определяющих слабую сходимость  $\mathcal{P}_n \Rightarrow \mathcal{P}$* , если эта сходимость следует из того, что  $\int_D \psi d\mathcal{P}_n \rightarrow \int_D \psi d\mathcal{P}$  для любой  $\psi \in \Phi$ .

Мы уже встречались с одним классом функций, определяющим слабую сходимость: из основной Теоремы 4.1 следует, что класс  $BL(D)$  ограниченных липшицевых функций определяет слабую сходимость для любых  $\mathcal{P}_n, \mathcal{P}$ . Тем самым класс  $BL(D)$  является универсальным, но, как правило, оказывается слишком широким при решении конкретных задач.

**Замечание 7.1.** Тем не менее и отсюда тоже можно извлекать содержательные выводы. Рассмотрим, например случайные величины  $\xi_n, \xi \in D$  с распределениями  $\mathcal{P}_n, \mathcal{P}$ . Тогда из сходимости  $Ef(\xi_n) \rightarrow Ef(\xi)$ , имеющей место для любой ограниченной липшицевой функции  $f$ , следует слабая сходимость  $\mathcal{P}_n \Rightarrow \mathcal{P}$ . Но

$$|Ef(\xi_n) - Ef(\xi)| \leq E|f(\xi_n) - f(\xi)| \leq L E\rho(\xi_n, \xi)$$

и поэтому, если  $E\rho(\xi_n, \xi) \rightarrow 0$ , то  $\mathcal{P}_n \Rightarrow \mathcal{P}$ .<sup>73</sup>

Подойдем к этой проблеме с другой стороны.

**Теорема 7.1.** 1. Пусть  $\Phi \subset C_b(D)$ . Если для любой  $f \in C_b(D)$  и любого  $\varepsilon > 0$  найдется такая  $\psi_\varepsilon \in \Phi$ , что  $\sup |f - \psi_\varepsilon| < \varepsilon$ , то  $\Phi$  является классом функций, определяющим слабую сходимость  $\mathcal{P}_n \Rightarrow \mathcal{P}$  для любых  $\mathcal{P}_n, \mathcal{P}$ .

2. Пусть  $\Phi \subset C_b(D)$ . Если для любой  $f \in C_b(D)$ , любого  $\varepsilon > 0$  и любого компакта  $K \subset D$  найдется такая функция  $\psi_\varepsilon \in \Phi$ , что

$$\sup_K |f - \psi_\varepsilon| < \varepsilon \quad \text{и} \quad \sup_{K^c} |\psi_\varepsilon| < C,$$

где постоянная  $C = C(f)$  не зависит от  $\varepsilon$ , то  $\Phi$  является классом функций, определяющим слабую сходимость  $\mathcal{P}_n \Rightarrow \mathcal{P}$  для любых  $\mathcal{P}_n, \mathcal{P}$  таких, что семейство распределений  $\{\mathcal{P}, \mathcal{P}_n, n \geq 1\}$  является плотным.<sup>74</sup>

<sup>73</sup>Выведите отсюда, что при  $D = \mathbb{R}$  из сходимости п.н. (или по вероятности)  $\xi_n$  к  $\xi$  следует, что  $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$ . Придумайте несколько аналогов этих фактов для случайных векторов.

<sup>74</sup>См. Определение 3.2 в разделе 3.2.

*Доказательство.* 1. Пусть  $\int_D \psi d\mathcal{P}_n \rightarrow \int_D \psi d\mathcal{P}$  для любой  $\psi \in \Phi$  и  $f \in C_b(D)$ . Зафиксируем  $\varepsilon > 0$  и возьмем функцию  $\psi_\varepsilon \in \Phi$ , аппроксимирующую  $f$  с точностью до  $\varepsilon$ . Поскольку

$$\limsup_n \int_D f d\mathcal{P}_n \leq \limsup_n \int_D (\psi_\varepsilon + \varepsilon) d\mathcal{P}_n = \limsup_n \int_D \psi_\varepsilon d\mathcal{P}_n + \varepsilon = \int_D \psi_\varepsilon d\mathcal{P} + \varepsilon \leq \int_D f d\mathcal{P} + 2\varepsilon.$$

Устремляя  $\varepsilon$  к нулю, получаем нужный результат.

2. Выберем компакт  $K = K_\varepsilon$  так, чтобы  $\mathcal{P}(K^c) < \varepsilon$  и  $\mathcal{P}_n(K^c) < \varepsilon$  для любого  $n$ . Как и раньше, пусть  $f \in C_b(D)$  с  $|f| \leq M = \text{const} < \infty$ . Тогда, взяв функцию  $\psi_\varepsilon \in \Phi$ , аппроксимирующую  $f$  с точностью до  $\varepsilon$  на компакте  $K$ , получим, что

$$\begin{aligned} \limsup_n \int_D f d\mathcal{P}_n &\leq \limsup_n \int_K f d\mathcal{P}_n + \limsup_n \int_{K^c} |f| d\mathcal{P}_n \leq \\ &\leq \limsup_n \int_K f d\mathcal{P}_n + M\varepsilon \leq \limsup_n \int_K \psi_\varepsilon d\mathcal{P}_n + \varepsilon + M\varepsilon \leq \\ &\leq \limsup_n \int_D \psi_\varepsilon d\mathcal{P}_n + \limsup_n \int_{K^c} |\psi_\varepsilon| d\mathcal{P}_n + (M+1)\varepsilon \leq \int_D \psi_\varepsilon d\mathcal{P} + C\varepsilon + (M+1)\varepsilon \leq \\ &\leq \int_K \psi_\varepsilon d\mathcal{P} + \int_{K^c} |\psi_\varepsilon| d\mathcal{P} + (C+M+1)\varepsilon \leq \int_K f d\mathcal{P} + (2C+M+2)\varepsilon \leq \\ &\leq \int_D f d\mathcal{P} + \int_{K^c} |f| d\mathcal{P} + (2C+M+2)\varepsilon \leq \int_D f d\mathcal{P} + 2(C+M+1)\varepsilon. \end{aligned}$$

Тем самым (ввиду произвольности  $\varepsilon$  и первого утверждения Леммы 4.2) и второе утверждение доказано.  $\square$

## 7.2 Примеры: сходимость моментов и характеристических функций

Прокомментируем условия Теоремы 7.1 и приведем соответствующие примеры. Начнем с первого утверждения теоремы. Следующее следствие хорошо иллюстрирует его суть.

**Следствие 7.1.** Рассмотрим единичный куб  $D = [0, 1]^d$ , оснащенный борелевской  $\sigma$ -алгеброй  $\mathcal{B}$ , и распределения  $\mathcal{P}, \mathcal{P}_n, n \geq 1$ , определенные на  $\mathcal{B}$ . Положим для любых целых неотрицательных  $k_1, \dots, k_d$

$$m_{k_1 \dots k_d} = \int_D x_1^{k_1} \dots x_d^{k_d} \mathcal{P}(dx_1 \dots dx_d) \quad \text{и} \quad m_{k_1 \dots k_d}^{(n)} = \int_D x_1^{k_1} \dots x_d^{k_d} \mathcal{P}_n(dx_1 \dots dx_d).$$

Для того, чтобы  $\mathcal{P}_n \Rightarrow \mathcal{P}$ , необходимо и достаточно, чтобы  $m_{k_1 \dots k_d}^{(n)} \rightarrow m_{k_1 \dots k_d}$  при  $n \rightarrow \infty$  для любых  $k_1, \dots, k_d$ .

*Доказательство.* Поскольку функции  $f(x_1, \dots, x_d) = x_1^{k_1} \dots x_d^{k_d}$  являются непрерывными и ограниченными на  $[0, 1]^d$ , то в части необходимости утверждение очевидно. Докажем достаточность сходимости любых моментов для слабой сходимости.

Обозначим  $\Phi$  множество полиномов  $d$  переменных. Из условия следует, что для любого  $\psi \in \Phi$

$$\int_D \psi d\mathcal{P}_n \rightarrow \int_D \psi d\mathcal{P}.$$

С другой стороны, по многомерной аппроксимационной теореме Вейерштрасса любую непрерывную функцию  $d$  переменных можно сколь угодно точно приблизить равномерно на  $D$  некоторым полиномом  $\psi \in \Phi$ . Следовательно, искомая слабая сходимость имеет место.  $\square$

**Замечание 7.2.** На самом деле существует значительно более слабые условия, при которых из сходимости всех моментов последовательности распределений  $\mathcal{P}_n$  к моментам распределения  $\mathcal{P}$  следует слабая сходимость  $\mathcal{P}_n \Rightarrow \mathcal{P}$  (без ограничения на то, что распределения сосредоточены на компакте).

А именно, для этого достаточно, чтобы распределение  $\mathcal{P}$  однозначно определялось своими моментами. В одномерном случае этот факт доказан, например, в [6, гл. IV §11.4].

**Пример 7.1.** (Предельные распределения бета-распределения)<sup>75</sup>

Рассмотрим бета-распределение  $\beta_{\nu_1, \nu_2}$  с параметрами  $\nu_1, \nu_2 > 0$ , заданное своей плотностью<sup>76</sup>

$$\beta_{\nu_1, \nu_2}(x) = \frac{\Gamma(\nu_1 + \nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} x^{\nu_1-1}(1-x)^{\nu_2-1}, \quad 0 < x < 1.$$

Обозначим  $p = \nu_1/(\nu_1 + \nu_2)$  и  $S = \nu_1 + \nu_2$ . Моменты  $\beta$ -распределения легко вычисляются:

$$\begin{aligned} m_k^{(\nu_1, \nu_2)} &= \int_0^1 x^k \beta_{\nu_1, \nu_2}(x) dx = \frac{\Gamma(\nu_1 + \nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} \int_0^1 x^{\nu_1+k-1}(1-x)^{\nu_2-1} dx = \\ &= \frac{\Gamma(\nu_1 + \nu_2)}{\Gamma(\nu_1)\Gamma(\nu_2)} \frac{\Gamma(\nu_1 + k)\Gamma(\nu_2)}{\Gamma(\nu_1 + \nu_2 + k)} = \prod_{i=0}^{k-1} \frac{\nu_1 + i}{\nu_1 + \nu_2 + i} = p \prod_{i=1}^{k-1} \frac{pS + i}{S + i}. \end{aligned}$$

В частности,  $m_1^{(\nu_1, \nu_2)} = p$ .

Пусть теперь  $\nu_1$  и  $\nu_2$  меняются так, что среднее  $p$  остается постоянным,<sup>77</sup> а для  $S$  может быть 2 варианта: а)  $S \rightarrow \infty$  и б)  $S \rightarrow 0$ .

В первом случае  $m_k^{(\nu_1, \nu_2)} \rightarrow p^k$ , и, согласно Следствию 7.1,  $\beta_{\nu_1, \nu_2} \Rightarrow \delta_p$ . Во втором случае  $m_k^{(\nu_1, \nu_2)} \rightarrow p$  и  $\beta_{\nu_1, \nu_2} \Rightarrow \text{Ber}(p)$ .

Прежде чем приводить примеры применения второго утверждения Теоремы 7.1, обсудим не совсем очевидное (хотя и явно используемое при доказательстве) требование плотности семейства распределений  $\{\mathcal{P}, \mathcal{P}_n, n \geq 1\}$ .

Прежде всего, если  $(D, \rho)$  — полное метрическое сепарабельное пространство, то, согласно Предложению 3.4, распределение  $\mathcal{P}$  является плотным, и поэтому достаточно требовать плотности семейства  $\{\mathcal{P}_n, n \geq 1\}$ .<sup>78</sup>

Дадим теперь следующее определение.

**Определение 7.2.** Пусть  $(D, \mathcal{C}, \mathcal{B})$  — измеримое топологическое пространство. Семейство распределений  $\Pi$ , определенных на борелевской  $\sigma$ -алгебре  $\mathcal{B}$ , называется *слабо относительно компактным*,<sup>79</sup> если из каждой последовательности  $\mathcal{P}_n \in \Pi$  можно выделить слабо сходящуюся подпоследовательность.

Конечно, если  $\mathcal{P}_n \Rightarrow \mathcal{P}$ , то семейство  $\{\mathcal{P}_n, n \geq 1\}$  будет относительно компактно.

**Лемма 7.1.** Пусть  $\{\mathcal{P}_n, n \geq 1\}$  — слабо относительно компактная последовательность распределений. Если любая слабо сходящаяся подпоследовательность  $\mathcal{P}_{n'}$  последовательности  $\mathcal{P}_n$  имеет один и тот же слабый предел  $\mathcal{P}$ , то  $\mathcal{P}_n \Rightarrow \mathcal{P}$  при  $n \rightarrow \infty$ .

*Доказательство.* Если  $\mathcal{P}_n \not\Rightarrow \mathcal{P}$ , то существует такая  $f \in C_b(D)$ , что

$$a_n \stackrel{\text{def}}{=} \int_D f d\mathcal{P}_n \not\rightarrow a \stackrel{\text{def}}{=} \int_D f d\mathcal{P}.$$

Поскольку последовательность  $a_n$  ограничена, из этого следует, что  $a_{n'} \rightarrow b \neq a$  для некоторой подпоследовательности  $n'$  натуральных чисел. Выбрав (ввиду относительной компактности) подпоследовательность  $n''$  последовательности  $n'$  так, чтобы  $\mathcal{P}_{n''} \Rightarrow \mathcal{P}$ , сразу же приходим к противоречию.  $\square$

<sup>75</sup>Этот пример используется в некоторых моделях покупательского поведения.

<sup>76</sup>Это действительно плотность?

<sup>77</sup>На самом деле достаточно, чтобы оно стремилось к постоянной.

<sup>78</sup>Более того, достаточно плотности семейства  $\{\mathcal{P}_n, n \geq n_0\}$  при некотором  $n_0 \geq 1$ .

<sup>79</sup>В дальнейшем — просто «относительно компактным».

Имеет место следующее утверждение.

**Теорема 7.2.** (Теорема Прохорова)

Пусть  $(D, \rho)$  — полное сепарабельное метрическое пространство. Тогда семейство распределений  $\mathcal{P}$  является относительно компактным тогда и только тогда, когда оно плотно.

Мы оставляем эту знаменитую теорему без доказательства, хотя в части необходимости оно не является слишком сложным.<sup>80</sup>

Нам сейчас существенно, что в полном метрическом сепарабельном пространстве из слабой сходимости  $\mathcal{P}_n \Rightarrow \mathcal{P}$  благодаря теореме Прохорова будет следовать плотность семейства  $\{\mathcal{P}, \mathcal{P}_n, n \geq 1\}$ . Поскольку Теорема 7.1 направлена на доказательство факта слабой сходимости, условие плотности этого семейства так или иначе там должно проявляться. Во второй пункт Теоремы 7.1 условие плотности введено явно.<sup>81</sup>

Перейдем теперь к примеру применения второго пункта Теоремы 7.1. Начнем со следующих лемм.

**Лемма 7.2.** Рассмотрим случайные вектора  $\bar{\xi}_\theta$  с  $\theta \in \Theta$ , принимающие значения в  $\mathbb{R}^d$ , и обозначим  $\xi_{\theta k}$  координаты вектора  $\bar{\xi}_\theta$ , так что  $\bar{\xi}_\theta = (\xi_{\theta 1}, \dots, \xi_{\theta d})^T$ . Для того, чтобы множество распределений  $\mathcal{P}_\theta = \mathcal{L}(\bar{\xi}_\theta)$  было плотно, необходимо и достаточно, чтобы для любого  $1 \leq k \leq d$  множество распределений  $\mathcal{P}_\theta^{(k)} = \mathcal{L}(\xi_{\theta k})$  было плотно.

*Доказательство.* Прежде всего, так как мы имеем дело с евклидовыми пространствами, нам достаточно<sup>82</sup> в качестве компактов рассматривать замкнутые кубы вида  $K = [-T, T]^d$ . С учетом этого замечания необходимость следует из цепочки

$$\mathcal{P}_\theta(K) = \mathcal{P}_\theta([-T, T]^d) = \mathbb{P}(\bar{\xi}_\theta \in [-T, T]^d) \leq \mathbb{P}(\xi_{\theta k} \in [-T, T]) = \mathcal{P}_\theta^{(k)}([-T, T])$$

(тем самым, если семейство  $\{\mathcal{P}_\theta, \theta \in \Theta\}$  плотно, то и плотно семейство  $\{\mathcal{P}_\theta^{(k)}, \theta \in \Theta\}$  для любого  $k$ ), а достаточность — из

$$\begin{aligned} 1 - \mathcal{P}_\theta([-T, T]^d) &= \mathbb{P}(\bar{\xi}_\theta \notin [-T, T]^d) = \mathbb{P}\left(\bigcup_{k=1}^d \{\xi_{\theta k} \notin [-T, T]\}\right) \leq \\ &\leq \sum_{m=1}^d \mathbb{P}(\xi_{\theta m} \notin [-T, T]) = \sum_{k=1}^d \left(1 - \mathcal{P}_\theta^{(k)}([-T, T])\right). \end{aligned}$$

Утверждение доказано. □

**Лемма 7.3.** Пусть  $\mathcal{P}_n$  — распределения, определенные на борелевских подмножествах  $\mathbb{R}^d$  с характеристическими функциями  $\varphi_n$ . Если функции  $\varphi_n$  поточечно сходятся к непрерывной в нуле функции  $\phi$ , то семейство  $\{\mathcal{P}_n, n \geq 1\}$  является плотным.

*Доказательство.* Ввиду Леммы 7.2 достаточно рассматривать лишь случай  $d = 1$ .<sup>83</sup> Пусть  $u > 0$ . Тогда

$$1 - \frac{1}{2u} \int_{[-u, u]} \varphi_n(t) dt = 1 - \int_{\mathbb{R}} \left( \frac{1}{2u} \int_{[-u, u]} e^{ixt} dt \right) \mathcal{P}_n(dx) = 1 - \int_{\mathbb{R}} \left( \frac{\sin(ux)}{ux} \right) \mathcal{P}_n(dx) =$$

<sup>80</sup>С доказательством теоремы можно ознакомиться в [1, гл. 1 §6]. Для случая  $D = \mathbb{R}$  оно приведено в [3, гл. III §2]. Отметим, что в последнем случае (а также при  $D = \mathbb{R}^d$ ) доказательство достаточности фактически сводится к доказательству теоремы Хелли (см., например, [3, стр. 410] для  $\mathbb{R}$  или [1, стр. 310] в многомерном случае), входящую во все достаточно продвинутые курсы теории вероятностей. Общий случай также опирается на эту теорему.

<sup>81</sup>А в Следствии 7.1 оно присутствует неявно, так как все распределения там сосредоточены на одном компакте  $[0, 1]^d$ .

<sup>82</sup>Почему?

<sup>83</sup>Почему?

$$= \int_{\mathbb{R}} \left(1 - \frac{\sin(ux)}{ux}\right) \mathcal{P}_n(dx) \geq \int_{|x| \geq 1/u} \left(1 - \frac{\sin(ux)}{ux}\right) \mathcal{P}_n(dx) \geq B \mathcal{P}_n([-1/u, 1/u]^c),$$

где  $B = \min_{|z| > 1} (1 - \sin(z)/z)$ .

Кроме того, по теореме Лебега о мажорируемой сходимости

$$\lim_n \int_{-u}^u \varphi_n(t) dt = \int_{-u}^u \phi(t) dt.$$

Поэтому, обозначая  $K = [-1/u, 1/u]$ , получим, что

$$\begin{aligned} B \limsup_n \mathcal{P}_n(K^c) &= B \limsup_n \mathcal{P}_n([-1/u, 1/u]^c) \leq 1 - \liminf_n \frac{1}{2u} \int_{[-u, u]} \varphi_n(t) dt = \\ &= 1 - \frac{1}{2u} \int_{[-u, u]} \lim_n \varphi_n(t) dt = 1 - \frac{1}{2u} \int_{[-u, u]} \phi(t) dt. \end{aligned}$$

Поскольку функция  $\phi$  непрерывна в нуле и  $\phi(0) = 1$ , то выбором  $u$  последнее выражение может быть сделано сколь угодно малым. Тем самым плотность семейства  $\{\mathcal{P}_n, n \geq 1\}$  доказана.<sup>84</sup>  $\square$

Наконец, докажем следующее следствие из Теоремы 7.1.

**Следствие 7.2.** (Многомерная теорема непрерывности для характеристических функций)

Пусть  $\mathcal{P}_n, \mathcal{P}$  — распределения, определенные на борелевских подмножествах  $\mathbb{R}^d$  с характеристическими функциями  $\varphi_n$  и  $\varphi$ .

1. Для того, чтобы  $\mathcal{P}_n \Rightarrow \mathcal{P}$  при  $n \rightarrow \infty$ , необходимо и достаточно, чтобы  $\varphi_n(\bar{t}) \rightarrow \varphi(\bar{t})$  при любом  $\bar{t} \in \mathbb{R}^d$ .

2. Если характеристические функции  $\varphi_n$  распределений  $\mathcal{P}_n$  поточечно сходятся к функции  $\phi$ , непрерывной в нуле, то  $\phi$  является характеристической функцией некоторого распределения  $\mathcal{P}$  и  $\mathcal{P}_n \Rightarrow \mathcal{P}$ .

*Доказательство.* 1. Приступим к доказательству первого пункта теоремы. Необходимость, конечно, очевидна. Для доказательства достаточности будем пользоваться вторым утверждением Теоремы 7.1. Плотность семейства  $\{\mathcal{P}, \mathcal{P}_n, n \geq 1\}$  уже доказана в Лемме 7.3.

В качестве множества  $\Phi$  возьмем множество всевозможных (вещественных) тригонометрических полиномов вида

$$\psi(\bar{x}) = \sum_{j=1}^N c_j e^{i(\bar{t}_j, \bar{x})},$$

где  $\bar{x}, \bar{t}_j \in \mathbb{R}^d$ . Из условия сходимости характеристических функций сразу же следует, что  $\int \psi d\mathcal{P}_n \rightarrow \int \psi d\mathcal{P}$  при  $n \rightarrow \infty$  для любой  $\psi \in \Phi$ .

Возьмем теперь функцию  $f \in C_b(\mathbb{R}^d)$ , компакт  $K = [-T, T]^d$  и обозначим  $f_K$  сужение  $f$  на  $K$ . Далее, при  $T_1 > T$  рассмотрим такое непрерывное продолжение  $f_{K_1}^*$  функции  $f_K$  с  $[-T, T]^d$  на  $K_1 = [-T_1, T_1]^d$ , что  $f_{K_1}^*$  равна нулю на границе компакта  $K_1$  и  $\max |f_{K_1}^*| \leq \max |f|$ .<sup>85</sup> Наконец, продолжим по периодичности (по каждому аргументу) функцию  $f_{K_1}^*$  на все  $\mathbb{R}^d$  и обозначим результат этого продолжения  $f^*$ .

По многомерной теореме Вейерштрасса для тригонометрических полиномов функцию  $f_{K_1}^*$  (и, следовательно, функцию  $f_K$ ) можно равномерно приблизить с любой точностью с помощью тригонометрического полинома  $\phi \in \Phi$ , так что все условия второго утверждения Теоремы 7.1 выполнены.

<sup>84</sup>Почему?

<sup>85</sup>А почему такое продолжение существует?

2. Перейдем к доказательству второго утверждения теоремы. Рассмотрим некоторую подпоследовательность  $\mathcal{P}_{n'}$ , имеющую слабый предел (поскольку последовательность  $\{\mathcal{P}_n, n \geq 1\}$  плотна, то такая подпоследовательность существует благодаря теореме Прохорова). Обозначим этот предел  $\mathcal{P}$  и соответствующую характеристическую функцию —  $\varphi$ . По условию  $\varphi = \phi$  и, следовательно, предельное распределение  $\mathcal{P}$  не зависит от выбора подпоследовательности  $\mathcal{P}_{n'}$ . Ссылка на Лемму 7.1 завершает доказательство.  $\square$

**Замечание 7.3.** 1. Как уже отмечалось, во втором пункте теоремы можно сослаться не на общую теорему Прохорова, а на многомерную теорему Хелли.

2. Доказательство первого пункта теоремы опирается у нас на Теорему 7.1. Если снова воспользоваться теоремой Прохорова, то оно существенно сократится.<sup>86</sup> Сама Теорема 7.1, однако, гораздо проще и прозрачнее теоремы Прохорова (и теоремы Хелли), которая, кстати, приведена здесь без доказательства.

3. Из поточечной сходимости  $\varphi_n$  к  $\varphi$  автоматически следует равномерная сходимость  $\varphi_n \rightarrow \varphi$  на любом компакте.

Из Следствия 7.2 вытекает следующий факт. Пусть  $\mathcal{P}_n, \mathcal{P}$  — распределения в  $\mathbb{R}^d$  и  $k \geq 1$  — целое.

**Следствие 7.3.** Пусть  $k \geq 1$ . Для того, что бы последовательность  $\mathcal{P}_n$  слабо сходилась к  $\mathcal{P}$ , н. и д., чтобы при  $n \rightarrow \infty$

$$\int_{\mathbb{R}^d} f d\mathcal{P}_n \rightarrow \int_{\mathbb{R}^d} f d\mathcal{P}$$

для любой ограниченной  $k$  раз непрерывно дифференцируемой функции.<sup>87</sup>

*Доказательство.* Необходимость очевидна (по определению слабой сходимости). Достаточность следует из того, что функции  $\cos(t, x)$  и  $\sin(t, x)$  при любом  $t \in \mathbb{R}^d$  являются  $k$  раз непрерывно дифференцируемыми для любого  $k$ . Отсюда получается поточечная сходимость характеристических функций распределений  $\mathcal{P}_n$  к характеристической функции распределения  $\mathcal{P}$ , а дальше — ссылка на теорему непрерывности.  $\square$

**Замечание 7.4.** Теорема Прохорова создавалась, конечно, не для того, чтобы с ее помощью доказывать теорему непрерывности или другие относительно простые вероятностные факты. В ней заложена основа для решения гораздо более «продвинутых» задач.

Возьмем, например, в качестве  $D$  пространство  $D = \mathbf{C}([0, 1])$  непрерывных функций на замкнутом отрезке  $[0, 1]$  с равномерной метрикой  $\rho$ .<sup>88</sup>

Стандартным образом мы приходим к конструкции  $(D, \rho, \mathcal{C}, \mathcal{B})$  — метрическому топологическому измеримому пространству, которое, как известно,<sup>89</sup> является польским. То есть мы последовательно имеем дело с открытыми шарами  $B_r(f)$  в метрике  $\rho$  (это множество непрерывных функций, «отстоящих» от заданной  $f$  (то есть «центра шара») не более чем  $r$ ), их всевозможными объединениями — топологией  $\mathcal{C}$ , порожденной равномерной метрикой и, наконец, борелевской  $\sigma$ -алгеброй  $\mathcal{B}$ , порожденной этой топологией.

И, значит, мы можем задавать на  $(D, \mathcal{B})$  различные распределения и изучать вопросы слабой сходимости и использовать все общие доказанные (и постулированные) факты об этой сходимости.

Так вот, не вдаваясь в детали, можно сказать, что теорема Прохорова является одним из решающих результатов, позволяющих получать конструктивные результаты о слабой сходимости распределений в пространствах типа  $\mathbf{C}([0, 1])$ . Некоторые из таких фактов можно найти в книге [1].

<sup>86</sup>Проверьте!

<sup>87</sup>Здесь все происходит в  $\mathbb{R}^d$ .

<sup>88</sup>Если  $f, g \in \mathbf{C}([0, 1])$ , то  $\rho(f, g) = \max_{x \in [0, 1]} |f(x) - g(x)|$ .

<sup>89</sup>Действительно?

### 7.3 О метризации слабой сходимости

Поскольку слабая сходимость распределений определена через сходимость интегралов (4.1), то возникает естественное желание «метризовать» эту сходимость, то есть построить на множестве  $\Pi$  распределений, определенных на одной борелевской  $\sigma$ -алгебре  $\mathcal{B}$ , такую метрику, сходимость по которой эквивалентна слабой сходимости.

Такие метрики действительно существуют, но к сожалению, они не являются достаточно наглядными и, как правило, удобны лишь в теоретических рассуждениях. Не останавливаясь на изучении этих общих метрик (про них можно прочесть в [3, гл. III §7]), мы покажем, как, основываясь на теореме непрерывности для характеристических функций (то есть на Следствии 7.2), построить класс метрик, порождающих слабую сходимость для случая  $D = \mathbb{R}^d$ .

Пусть  $\Pi$  — множество распределений, определенных на борелевских подмножествах пространства  $\mathbb{R}^d$ , а  $c : \mathbb{R} \mapsto \mathbb{R}$  — положительная непрерывная ограниченная функция, стремящаяся к нулю на бесконечности. Кроме того, для  $\bar{t} \in \mathbb{R}^d$  обозначим  $\tau = \|\bar{t}\|$ .

**Теорема 7.3.** 1. Если  $\mathcal{P}, \mathcal{Q} \in \Pi$  с характеристическими функциями  $\varphi_{\mathcal{P}}, \varphi_{\mathcal{Q}}$ , то функция

$$\rho_c(\mathcal{P}, \mathcal{Q}) = \sup_{\bar{t} \in \mathbb{R}^d} |\varphi_{\mathcal{P}}(\bar{t}) - \varphi_{\mathcal{Q}}(\bar{t})| c(\tau)$$

является метрикой в  $\Pi$ .

2. Пусть  $\mathcal{P}_n, \mathcal{P} \in \Pi$  — распределения с характеристическими функциями  $\varphi_n, \varphi$ . Для того, чтобы  $\mathcal{P}_n \Rightarrow \mathcal{P}$ , необходимо и достаточно, чтобы  $\rho_c(\mathcal{P}_n, \mathcal{P}) \rightarrow 0$ .

*Доказательство.* 1. Ясно, что  $\rho_c(\mathcal{P}, \mathcal{Q}) = \rho_c(\mathcal{Q}, \mathcal{P})$ . Кроме того,  $\rho_c(\mathcal{P}, \mathcal{Q}) \geq 0$ , и, так как функция  $c$  положительна, а равенство характеристических функций эквивалентно равенству распределений, то  $\rho_c(\mathcal{P}, \mathcal{Q}) = 0$  тогда и только тогда, когда  $\mathcal{P} = \mathcal{Q}$ . Следовательно, осталось доказать неравенство треугольника.

Для этого сразу же заметим, что функция  $h(\bar{t}) = |\varphi_{\mathcal{P}}(\bar{t}) - \varphi_{\mathcal{Q}}(\bar{t})| c(\tau)$  является непрерывной, ограниченной и стремящейся к нулю при  $\bar{t} \rightarrow \infty$ , поэтому ее супремум достигается в некоторой точке  $\bar{t}_{pq} \in \mathbb{R}^d$ .

Пусть теперь  $\mathcal{P}, \mathcal{Q}, \mathcal{R}$  — распределения с характеристическими функциями  $\varphi_{\mathcal{P}}, \varphi_{\mathcal{Q}}, \varphi_{\mathcal{R}}$ . Тогда в очевидных обозначениях

$$\begin{aligned} \rho_c(\mathcal{P}, \mathcal{Q}) + \rho_c(\mathcal{Q}, \mathcal{R}) &= |\varphi_{\mathcal{P}}(\bar{t}_{pq}) - \varphi_{\mathcal{Q}}(\bar{t}_{pq})| c(\tau_{pq}) + |\varphi_{\mathcal{Q}}(\bar{t}_{qr}) - \varphi_{\mathcal{R}}(\bar{t}_{qr})| c(\tau_{qr}) \geq \\ &\geq |\varphi_{\mathcal{P}}(\bar{t}_{pr}) - \varphi_{\mathcal{Q}}(\bar{t}_{pr})| c(\tau_{pr}) + |\varphi_{\mathcal{Q}}(\bar{t}_{pr}) - \varphi_{\mathcal{R}}(\bar{t}_{pr})| c(\tau_{pr}) = \\ &= \left( |\varphi_{\mathcal{P}}(\bar{t}_{pr}) - \varphi_{\mathcal{Q}}(\bar{t}_{pr})| + |\varphi_{\mathcal{Q}}(\bar{t}_{pr}) - \varphi_{\mathcal{R}}(\bar{t}_{pr})| \right) c(\tau_{pr}) \geq |\varphi_{\mathcal{P}}(\bar{t}_{pr}) - \varphi_{\mathcal{R}}(\bar{t}_{pr})| c(\tau_{pr}) = \rho_c(\mathcal{P}, \mathcal{R}). \end{aligned}$$

2. Если  $\rho_c(\mathcal{P}_n, \mathcal{P}) \rightarrow 0$ , то, поскольку функция  $c$  не обращается в ноль,  $\varphi_n \rightarrow \varphi$  поточечно, и, согласно Следствию 7.2,  $\mathcal{P}_n \Rightarrow \mathcal{P}$ .

Если же  $\mathcal{P}_n \Rightarrow \mathcal{P}$ , то  $\varphi_n \rightarrow \varphi$  не просто поточечно, а еще и равномерно на любом компакте.<sup>90</sup> Поэтому, выбрав  $T > 0$  таким образом, что  $c(\tau) < \varepsilon$  при  $\tau > T$ , мы получим, что

$$\rho_c(\mathcal{P}_n, \mathcal{P}) \leq C \sup_{\|\bar{t}\| \leq T} |\varphi_n(\bar{t}) - \varphi(\bar{t})| + 2\varepsilon,$$

где число  $C$  мажорирует ограниченную функцию  $c$ . Дальнейшее очевидно.  $\square$

<sup>90</sup>Проверьте!

## 8 Сходимость по вариации и теорема Шеффе

О понятии сходимости по вариации и соотношении между этой сходимостью и слабой сходимостью распределений уже рассказано во Введении для случая  $D = \mathbb{R}^d$ . Здесь мы остановимся на этих вопросах подробнее и приведем соответствующие доказательства в общем случае.

### 8.1 Расстояние по вариации и полная вариация заряда

Рассмотрим измеримое пространство  $(D, \mathcal{A})$  и обозначим  $\Pi_D$  множество всевозможных вероятностных мер, определенных на  $\mathcal{A}$ .

**Определение 8.1.** Пусть  $\mathcal{P}, \mathcal{Q} \in \Pi_D$ . Число

$$\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) = \sup_{A \in \mathcal{A}} |\mathcal{P}(A) - \mathcal{Q}(A)| \quad (8.1)$$

называется *расстоянием по вариации* между распределениями  $\mathcal{P}, \mathcal{Q}$ .

Если  $\mathcal{P}_n, \mathcal{P} \in \Pi_D$  и  $\rho_{\text{var}}(\mathcal{P}_n, \mathcal{P}) \rightarrow 0$ , то говорят, что  $\mathcal{P}_n$  *сходится к  $\mathcal{P} \in \Pi_D$  по вариации*.

**Замечание 8.1.** 1. В Определении 8.1 не требуется, чтобы  $\sigma$ -алгебра  $\mathcal{A}$  была борелевской. Она, вообще говоря, произвольна.

2. Нетрудно показать,<sup>91</sup> что  $\rho_{\text{var}}$  действительно определяет метрику в множестве  $\Pi_D$ .

3. Метрическое пространство  $(\Pi_D, \rho_{\text{var}})$ , вообще говоря, не является сепарабельным. Действительно, возьмем  $(D, \mathcal{A}) = (\mathbb{R}, \mathcal{B}_R)$ . Так как  $\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) = 1$  при любых  $\mathcal{P} = \delta_a$  и  $\mathcal{Q} = \delta_b$  с  $a \neq b$ , то в метрическом пространстве  $(\Pi_D, \rho_{\text{var}})$  не существует счетного всюду плотного множества.<sup>92</sup>

Теперь рассмотрим множество  $M_D$  конечных зарядов, определенных на  $\mathcal{A}$ . Если  $\nu \in M_D$ , то  $\nu$  может быть представлен в виде разности положительных конечных мер  $\nu^+$  и  $\nu^-$ , имеющих дизъюнктные носители.

Иначе говоря,<sup>93</sup> существует такое множество  $E \in \mathcal{A}$  и такие меры  $\nu^+$  и  $\nu^-$ , что

$$\nu^+(E^c) = \nu^-(E) = 0 \quad \text{и} \quad \nu(A) = \nu^+(A) - \nu^-(A)$$

для любого  $A \in \mathcal{A}$ . Положим  $|\nu| = \nu^+ + \nu^-$ . Очевидно,  $|\nu|$  является (конечной) положительной мерой.

**Определение 8.2.** Число  $\|\nu\| = |\nu|(D)$  называется *полной вариацией заряда  $\nu$* .

**Замечание 8.2.** Пара  $(M_D, \|\cdot\|)$  представляет собой нормированное пространство.<sup>94</sup>

Следующее утверждение проясняет понятие полной вариации заряда.

**Лемма 8.1.** Пусть  $\nu$  — конечный заряд. Если для некоторой меры  $\mu$  и  $\mathcal{A}$ -измеримой функции  $h$

$$\nu(A) = \int_A h d\mu$$

для любого  $A \in \mathcal{A}$ , то  $\|\nu\| = \int_D |h| d\mu$ .

<sup>91</sup>Покажите!

<sup>92</sup>Действительно?

<sup>93</sup>См, например, [7, стр. 121]. Разложение  $\nu = \nu^+ - \nu^-$  называется *разложением Хана*.

<sup>94</sup>Докажите.

*Доказательство.* Пусть  $E$  — носитель меры  $\mu^+$ . Обозначим

$$h^+(x) = \begin{cases} h(x) & \text{при } x \in E, \\ 0 & \text{при } x \notin E \end{cases} \quad \text{и} \quad h^-(x) = - \begin{cases} 0 & \text{при } x \in E, \\ h(x) & \text{при } x \notin E. \end{cases}$$

Так как  $\nu^+(E^c) = \nu^-(E) = 0$ , то  $h^+$  и  $h^-$  неотрицательны  $\mu$ -п.в., причем  $h = h^+ - h^-$ ,  $|h| = h^+ + h^-$   $\mu$ -п.в. и

$$\nu^+(A) = \int_A h^+ d\mu, \quad \nu^-(A) = \int_A h^- d\mu.$$

Поэтому

$$\|\nu\| = \nu^+(D) + \nu^-(D) = \int_D h^+ d\mu + \int_D h^- d\mu = \int_D (h^+ + h^-) d\mu = \int_D |h| d\mu.$$

Доказательство закончено. □

Займемся связью между расстоянием по вариации между распределениями  $\mathcal{P}, \mathcal{Q}$  и понятием полной вариации заряда. Начнем со следующего вспомогательного утверждения.

**Лемма 8.2.** Пусть  $\nu$  — конечный заряд, удовлетворяющий условию  $\nu(D) = 0$ . Если  $E$  является носителем заряда  $\nu^+$ , то

$$\nu^+(D) = \nu^+(E) = \nu^-(E^c) = \nu^-(D) = \|\nu\|/2.$$

*Доказательство.* Равенства  $\nu^+(D) = \nu^+(E)$  и  $\nu^-(E^c) = \nu^-(D)$  следуют из определения множества  $E$ . Поскольку  $0 = \nu(D) = \nu^+(D) - \nu^-(D)$ , то  $\nu^+(D) = \nu^-(D)$ . Осталось заметить, что по определению  $\|\nu\| = \nu^+(D) + \nu^-(D)$ . □

**Предложение 8.1.** Пусть  $\mathcal{P}, \mathcal{Q} \in \Pi_D$ . Тогда

$$\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) = \frac{1}{2} \|\mathcal{P} - \mathcal{Q}\|.$$

*Доказательство.* Обозначим  $\nu = \mathcal{P} - \mathcal{Q}$  и пусть  $E$  является носителем заряда  $\nu^+$ . Поскольку  $\nu(D) = 0$ , то  $\nu^+(D) = \nu^+(E) = \nu^-(E^c) = \nu^-(D)$ . Далее,

$$\mathcal{P}(A) - \mathcal{Q}(A) = \nu(A) = \nu^+(A) - \nu^-(A) = \nu^+(A \cap E) - \nu^-(A \cap E^c). \quad (8.2)$$

и, следовательно,

$$-\nu^-(A \cap E^c) \leq \mathcal{P}(A) - \mathcal{Q}(A) \leq \nu^+(A \cap E).$$

Поэтому из Леммы 8.2 следует, что

$$|\mathcal{P}(A) - \mathcal{Q}(A)| \leq \max(\nu^+(A \cap E), \nu^-(A \cap E^c)) \leq \max(\nu^+(E), \nu^-(E^c)) = \frac{1}{2} \|\mathcal{P} - \mathcal{Q}\|.$$

Осталось заметить, что при  $A = E$  равенство (8.2) превращается в  $\mathcal{P}(E) - \mathcal{Q}(E) = \nu^+(E)$ , что и заканчивает доказательство. □

## 8.2 Интегральные представления $\rho_{\text{var}}$ . Теорема Шеффе

Во Введении обсуждалась роль интегрального представления (1.4) для понимания разницы между сильной и слабой сходимостью распределений. Здесь мы докажем это представление.

**Предложение 8.2.** *Имеют место следующие равенства:*

1.

$$\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) = \sup_{0 \leq f \leq 1} \left| \int_D f d\mathcal{P} - \int_D f d\mathcal{Q} \right|, \quad (8.3)$$

где супремум в правой части берется по всем неотрицательным  $\mathcal{A}$ -измеримым функциям, не превосходящим единицу.

2.

$$\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) = \frac{1}{2} \sup_{|f| \leq 1} \left| \int_D f d\mathcal{P} - \int_D f d\mathcal{Q} \right|, \quad (8.4)$$

где супремум в правой части берется по всем  $\mathcal{A}$ -измеримым функциям, по модулю не превосходящим единицу.

*Доказательство.* 1. Взяв  $f = \mathbf{1}_A$ , сразу же убеждаемся, что левая часть (8.3) не превосходит правую. Докажем противоположное неравенство. Обозначим  $\nu = \mathcal{P} - \mathcal{Q}$  и  $E$  — носитель меры  $\nu^+$ . Тогда

$$\int_D f d\mathcal{P} - \int_D f d\mathcal{Q} = \int_D f d\nu = \int_D f d\nu^+ - \int_D f d\nu^- = \int_E f d\nu^+ - \int_{E^c} f d\nu^-.$$

Поэтому

$$- \int_{E^c} f d\nu^- \leq \int_D f d\mathcal{P} - \int_D f d\mathcal{Q} \leq \int_E f d\nu^+$$

и, согласно Лемме 8.2 и Предложению 8.1,

$$\left| \int_D f d\mathcal{P} - \int_D f d\mathcal{Q} \right| \leq \max \left( \int_E f d\nu^+, \int_{E^c} f d\nu^- \right) \leq \max (\nu^+(E), \nu^-(E^c)) = \rho_{\text{var}}(\mathcal{P}, \mathcal{Q}).$$

2. Разложим  $f$  на положительную и отрицательную части:  $f = f^+ - f^-$ . Поскольку носители  $f^+$  и  $f^-$  не пересекаются, то обе эти (неотрицательные) функции не превосходят единицы. Поэтому

$$\begin{aligned} \sup_{|f| \leq 1} \left| \int_D f d\mathcal{P} - \int_D f d\mathcal{Q} \right| &= \sup_{|f| \leq 1} \left| \int_D (f^+ - f^-) d\nu \right| \leq \sup_{|f| \leq 1} \left| \int_D f^+ d\nu \right| + \sup_{|f| \leq 1} \left| \int_D f^- d\nu \right| = \\ &= \sup_{0 \leq f^+ \leq 1} \left| \int_D f^+ d\nu \right| + \sup_{0 \leq f^- \leq 1} \left| \int_D f^- d\nu \right| = 2\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) \end{aligned}$$

согласно равенству (8.3).

С другой стороны, если положить  $f = \mathbf{1}_E - \mathbf{1}_{E^c}$ , то получим (снова благодаря Лемме 8.2 и Предложению 8.1), что

$$\int_D f d\nu = \nu^+(E) + \nu^-(E^c) = 2\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}),$$

что и заканчивает доказательство.  $\square$

**Следствие 8.1.** Пусть  $g$  — некоторая  $\mathcal{A}$ -измеримая функция, ограниченная по модулю числом  $M$ . Тогда

$$\left| \int_D g d\mathcal{P} - \int_D g d\mathcal{Q} \right| \leq 2M \rho_{\text{var}}(\mathcal{P}, \mathcal{Q}).$$

*Доказательство.* Обозначим  $g_M = g/M$ . Тогда  $|g_M| \leq 1$ ,

$$\left| \int_D g d\mathcal{P} - \int_D g d\mathcal{Q} \right| \leq M \left| \int_D g_M d\mathcal{P} - \int_D g_M d\mathcal{Q} \right| \leq M \sup_{0 \leq f \leq 1} \left| \int_D f d\mathcal{P} - \int_D f d\mathcal{Q} \right|,$$

и утверждение непосредственно следует из Предложения 8.2.  $\square$

Наконец, докажем в общем виде теорему Шеффе, сформулированную во Введении в виде двух частных случаев.

**Теорема 8.1.** 1. Пусть у распределений  $\mathcal{P}$  и  $\mathcal{Q}$  существуют производные Радона-Никодима  $p$  и  $q$  относительно некоторой меры  $\mu$ . Тогда

$$\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) = \frac{1}{2} \int_D |p(x) - q(x)| \mu(dx). \quad (8.5)$$

2. Пусть у распределений  $\mathcal{P}_n$  и  $\mathcal{P}$  существуют производные Радона-Никодима  $p_n$  и  $p$  относительно некоторой меры  $\mu$ . Если  $p_n \rightarrow p$   $\mu$ -почти всюду, то  $\rho_{\text{var}}(\mathcal{P}_n, \mathcal{P}) \rightarrow 0$ .

*Доказательство.* Первое утверждение теоремы немедленно следует из Леммы 8.1 и Предложения 8.1. Докажем второе утверждение. Обозначим  $\delta_n = p - p_n$ . Поскольку  $\int_D \delta_n d\mu = 0$ , то

$$\int_D \delta_n^+ d\mu = \int_D \delta_n^- d\mu = \frac{1}{2} \int_D |\delta_n| d\mu = \rho_{\text{var}}(\mathcal{P}_n, \mathcal{P})$$

ввиду равенства (8.5). Так как  $\delta_n$  сходится к нулю  $\mu$ -почти всюду и  $0 \leq \delta_n \leq p$  (тоже  $\mu$ -почти всюду), то по теореме Лебега о мажорируемой сходимости получаем нужный результат.  $\square$

**Замечание 8.3.** 1. Если  $(D, \mathcal{A}) = (\mathbb{R}^d, \mathcal{B}_d)$  и  $\mu$  —  $d$ -мерная мера Лебега, то мы приходим к первому утверждению Теоремы 1.1 Введения, если  $\mu$  — считающая мера, то ко второму.

2. В борелевском случае, когда сигма-алгебра  $\mathcal{A}$  порождается некоторой метрикой, можно сравнивать сильную (то есть по вариации) и слабую сходимость распределений. Как и следовало ожидать, из сильной сходимости будет следовать слабая.

Для доказательства достаточно сравнить определение сходимости по вариации с пятым пунктом Основной Теоремы 4.1 раздела 4.2, или определение слабой сходимости с результатом Следствия 8.1.

## 9 Несколько предельных теорем, используемых в статистике

В этом разделе мы остановимся на некоторых вопросах применения теории слабой сходимости распределений к стандартным задачам математической (и прикладной) статистики.

### 9.1 Сходимость по вероятности к константе. Общие утверждения

Хорошо известно определение сходимости по вероятности к константе последовательности случайных величин:  $\xi_n \xrightarrow{P} a$  означает, что  $P(|\xi_n - a| \geq \varepsilon) \rightarrow 0$  при  $n \rightarrow \infty$  для любого  $\varepsilon > 0$ . Это определение<sup>95</sup> естественным образом обобщается на случай случайных величин, принимающих значения в произвольном топологическом измеримом пространстве.

**Определение 9.1.** Пусть  $(D, \mathcal{C}, \mathcal{B})$  — топологическое измеримое пространство (как всегда,  $\mathcal{C}$  обозначает некоторую топологию подмножеств  $D$ , а  $\mathcal{B}$  — соответствующую борелевскую  $\sigma$ -алгебру). Говорят, что последовательность случайных величин  $\xi_n : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{B})$  сходится по вероятности к  $a \in D$  (по-прежнему записывается  $\xi_n \xrightarrow{P} a$ ), если  $P(\xi_n \in U_a) \rightarrow 1$  для любого открытого множества  $U_a$ , содержащего точку  $a$ .

В случае, когда топология  $\mathcal{C}$  порождается метрикой  $\rho$ , ситуация становится более наглядной.

**Предложение 9.1.** Пусть  $(D, \rho, \mathcal{C}, \mathcal{B})$  — метрическое измеримое пространство и  $a \in D$ . Тогда сходимость  $\xi_n \xrightarrow{P} a$  эквивалентна тому, что  $\rho(\xi_n, a) \xrightarrow{P} 0$ .<sup>96</sup>

*Доказательство.* Сходимость  $\rho(\xi_n, a) \xrightarrow{P} 0$  можно переписать<sup>97</sup> как  $P(\xi_n \in B_\varepsilon(a)) \rightarrow 1$  при  $n \rightarrow \infty$ , где  $B_\varepsilon(a)$  — открытый шар радиуса  $\varepsilon$  с центром в точке  $a$ . Тем самым необходимость доказана. Достаточность следует из цепочки

$$P(\xi_n \in U_a) \geq P(\xi_n \in B_\varepsilon(a)) \rightarrow 1,$$

где  $B_\varepsilon(a)$  — некоторый открытый шар радиуса  $\varepsilon$  с центром в точке  $a$ , содержащийся в  $U_a$ .  $\square$

Достоинство Предложения 9.1 состоит в том, что оно позволяет для доказательства сходимости по вероятности использовать наиболее удобную метрику, порождающую топологию  $\mathcal{C}$ . Это можно проиллюстрировать на примере слабой сходимости векторов в  $\mathbb{R}^d$ .<sup>98</sup>

**Предложение 9.2.** Пусть  $\bar{\xi}_n = (\xi_1^{(n)}, \dots, \xi_d^{(n)})^T$  — последовательность случайных векторов в  $\mathbb{R}^d$  и  $\bar{a} = (a_1, \dots, a_d)^T \in \mathbb{R}^d$ . Тогда сходимость  $\bar{\xi}_n \xrightarrow{P} \bar{a}$  эквивалентна покомпонентной сходимости  $\xi_j^{(n)} \xrightarrow{P} a_j$  при  $j = 1, \dots, d$ .

*Доказательство.* Выберем в  $\mathbb{R}^d$  метрику, определенную при  $\bar{x} = (x_1, \dots, x_d)^T$  и  $\bar{y} = (y_1, \dots, y_d)^T$  равенством  $\rho(\bar{x}, \bar{y}) = \max_{1 \leq i \leq d} |x_i - y_i|$ . Как уже говорилось (см. Предложение 3.2 раздела 3.1), эта метрика порождает Борелевскую  $\sigma$ -алгебру в  $\mathbb{R}^d$ . Кроме того, обозначим  $B_\varepsilon(\bar{a})$  шар радиуса  $\varepsilon$  с центром в точке  $\bar{a}$  в метрике  $\rho$ .

Если  $A_i^{(n)} = \{\omega : |\xi_i^{(n)}(\omega) - a_i| < \varepsilon\}$ , то, благодаря такому выбору метрики  $\rho$ ,

$$A_n \stackrel{\text{def}}{=} \bigcap_{i=1}^d A_i^{(n)} = \{\omega : \bar{\xi}_n(\omega) \in B_\varepsilon(\bar{a})\}.$$

<sup>95</sup>Более стандартное определение сходимости по вероятности связано со сходимостью  $P(|\xi_n - a| > \varepsilon) \rightarrow 0$ . Конечно, эти определения эквивалентны. А почему?

<sup>96</sup>Последняя запись корректна, так как функция  $\rho(x, a)$  является непрерывной по  $x$  и, следовательно,  $\mathcal{B}$ -измеримой — см. Замечание 2.1 раздела 2.

<sup>97</sup>А почему?

<sup>98</sup>Сформулируйте и докажите аналог Предложения 9.2 для произведения сепарабельных метрических пространств.

Ясно, что сходимости  $P(A_n) \rightarrow 1$  и  $P(A_i^{(n)}) \rightarrow 1$  для всех  $i$  эквивалентны. Ввиду произвольности  $\varepsilon$  отсюда сразу же следует требуемое.  $\square$

Следующее хорошо известное утверждение оказывается верным не только для евклидовых, но и для произвольных метрических пространств.

**Предложение 9.3.** Пусть  $(D, \mathcal{C}, \rho, \mathcal{B})$  — метрическое измеримое пространство и  $a \in D$ . Тогда сходимость  $\xi_n \xrightarrow{P} a$  эквивалентна тому, что  $\mathcal{L}(\xi_n) \Rightarrow \delta_a$ .

*Доказательство.* 1. Пусть  $\mathcal{P}_n \stackrel{\text{def}}{=} \mathcal{L}(\xi_n) \Rightarrow \mathcal{P} = \delta_a$  и пусть  $U_a$  — открытое множество, содержащее  $a$ . Тогда по четвертому пункту Основной Теоремы 4.1 (см. раздел 4.2)

$$\liminf_n P(\xi_n \in U_a) = \liminf_n \mathcal{P}_n(U_a) \geq \mathcal{P}(U_a) = \delta_a(U_a) = 1.$$

Следовательно,  $P(\xi_n \in U_a) \rightarrow 1$ .

2. Пусть  $\xi_n \xrightarrow{P} a$  и  $f \in \text{BL}(D)$ , причем  $|f| \leq M$  и соответствующая постоянная Липшица равна  $L$ . Тогда

$$\begin{aligned} |Ef(\xi_n) - f(a)| &\leq \int_{\rho(x,a) < \varepsilon} |f(x) - f(a)| \mathcal{P}_n(dx) + \int_{\rho(x,a) \geq \varepsilon} |f(x) - f(a)| \mathcal{P}_n(dx) \leq \\ &\leq L \int_{\rho(x,a) < \varepsilon} \rho(x,a) \mathcal{P}_n(dx) + 2M P(\rho(\xi_n, a) \geq \varepsilon) \leq L\varepsilon + 2M P(\rho(\xi_n, a) \geq \varepsilon). \end{aligned}$$

Первое слагаемое можно сделать малым выбором  $\varepsilon$ , а второе — выбором достаточно большого  $n$  (при уже выбранном  $\varepsilon$ ). Ссылка на первый пункт Основной Теоремы 4.1 заканчивает доказательство.  $\square$

## 9.2 Замена мешающего параметра на его состоятельную оценку

### 9.2.1 Общее утверждение

Теорема 9.1, которую мы обсудим в этом разделе, интересна не только для приложений, она проясняет роль случайных величин, сходящихся по вероятности к нулю, при доказательстве различных предельных теорем.

В математическом анализе бесконечно малые последовательности и функции играют особую роль — их вычленение в сложных выражениях часто помогает находить пределы этих выражений. Случайные величины, сходящиеся по вероятности к нулю, играют аналогичную роль в теории слабой сходимости распределений. Здесь мы приведем несколько важных примеров подобных утверждений, основанных на следующем факте.

Рассмотрим сепарабельные метрические измеримые пространства  $(D_i, \rho_i, \mathcal{C}_i, \mathcal{B}_i)$ ,  $i = 1, 2$ . Построим из них пространство  $(D, \rho, \mathcal{C}, \mathcal{B})$ , где  $D = D_1 \times D_2$ ,  $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2$  — топология произведения,  $\mathcal{B} = \mathcal{B}_1 \times \mathcal{B}_2$  — произведение  $\sigma$ -алгебр  $\mathcal{B}_1$  и  $\mathcal{B}_2$ , а метрика  $\rho$  определяется (например) следующим образом: если  $y_1, y_2 \in D_1$ ,  $z_1, z_2 \in D_2$  и  $\mathbf{x}_1 = (y_1, z_1)$ ,  $\mathbf{x}_2 = (y_2, z_2)$ , то  $\rho(\mathbf{x}_1, \mathbf{x}_2) = \max(\rho_1(y_1, y_2), \rho_2(z_1, z_2))$ .

**Теорема 9.1.** Пусть  $\xi_n, \xi : (\Omega, \mathcal{F}) \mapsto (D_1, \mathcal{B}_1)$ ,  $\eta_n : (\Omega, \mathcal{F}) \mapsto (D_2, \mathcal{B}_2)$  и  $\theta \in D_2$ . Если  $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$  и  $\eta_n \xrightarrow{P} \theta \in D_2$ , то  $\mathcal{L}((\xi_n, \eta_n)) \Rightarrow \mathcal{L}((\xi, \theta))$ .

*Доказательство.* Согласно Основной Теореме достаточно доказать, что  $Ef(\xi_n, \eta_n) \rightarrow Ef(\xi, \theta)$  для любой функции  $f \in \text{BL}(D)$ . Пусть  $|f| \leq M$  и при  $\mathbf{x} = (x_1, x_2)$ ,  $\mathbf{y} = (y_1, y_2)$

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq L\rho(\mathbf{x}, \mathbf{y}) = L \max(\rho_1(x_1, y_1), \rho_2(x_2, y_2)). \quad (9.1)$$

Тогда

$$|Ef(\xi_n, \eta_n) - Ef(\xi, \theta)| \leq |Ef(\xi_n, \eta_n) - Ef(\xi_n, \theta)| + |Ef(\xi_n, \theta) - Ef(\xi, \theta)| = J_1(n) + J_2(n).$$

Так как функция  $f(\cdot, \theta)$  непрерывна и ограничена, то  $J_2(n) \rightarrow 0$  просто по определению слабой сходимости (в пространстве  $(D_1, \rho_1, \mathcal{C}_1, \mathcal{B}_1)$ ).

Что касается последовательности  $J_1(n)$ , то, используя неравенство (9.1) и равенство

$$\rho((x, y), (x, z)) = \rho_2(y, z),$$

получим, что

$$\begin{aligned} |Ef(\xi_n, \eta_n) - Ef(\xi_n, \theta)| &= E(|f(\xi_n, \eta_n) - Ef(\xi_n, \theta)|, \rho_2(\eta_n, \theta) < \varepsilon) + \\ &+ E(|f(\xi_n, \eta_n) - Ef(\xi_n, \theta)|, \rho_2(\eta_n, \theta) \geq \varepsilon) \leq LE(\rho_2(\eta_n, \theta), \rho_2(\eta_n, \theta) < \varepsilon) + 2M P(\rho_2(\eta_n, \theta) \geq \varepsilon). \end{aligned}$$

Дальнейшие рассуждения повторяют окончание доказательства Предложения 9.3.  $\square$

**Замечание 9.1.** Требование сепарабельности метрических пространств  $(D_1, \rho_1)$  и  $(D_2, \rho_2)$  необходимо для того, чтобы произведение  $\sigma$ -алгебр  $\mathcal{B}_1 \times \mathcal{B}_2$  являлось борелевской  $\sigma$ -алгеброй  $\sigma(\mathcal{C}_1 \times \mathcal{C}_2)$  (см. Предложение 3.2 раздела 3.1).

**Замечание 9.2.** В Теореме 9.1 требование сходимости по вероятности к постоянной  $\eta_n \xrightarrow{P} \theta$  (а не просто  $\mathcal{L}(\eta_n) \Rightarrow \mathcal{L}(\eta)$ ) существенно. Легко привести пример, когда  $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$ ,  $\mathcal{L}(\eta_n) \Rightarrow \mathcal{L}(\eta)$ , но  $\mathcal{L}((\xi_n, \eta_n) \not\Rightarrow \mathcal{L}((\xi, \eta))$ . Для этого, например, достаточно рассмотреть  $\alpha \in U(0, 1)$  и положить  $\xi_n \equiv \alpha$  и положить  $\eta_n = \xi_n$  при нечетных  $n$  и  $1 - \eta_n = \xi_n$  при  $n$  четных.

Из Теоремы 9.1 сразу же вытекает следующий полезный факт.

**Следствие 9.1.** Пусть  $(G, r, \mathcal{G})$  — некоторое метрическое пространство с метрикой  $r$  и борелевской  $\sigma$ -алгеброй  $\mathcal{G}$ . Если в условиях Теоремы 9.1 отображение  $g : (D, \rho) \mapsto (G, r)$  непрерывно, то

$$\mathcal{L}(g(\xi_n, \eta_n)) \Rightarrow \mathcal{L}(g(\xi, \theta)).$$

*Доказательство.* Конечно, это всего лишь следствие из теоремы (точнее, Предложения 5.1) о сохранении слабой сходимости при непрерывных отображениях.  $\square$

**Замечание 9.3.** Из утверждения Следствия 9.1, в частности, вытекает,<sup>99</sup> что слабый предел последовательности  $\mathcal{L}(g(\xi_n, \theta))$  совпадает со слабым пределом последовательности  $\mathcal{L}(g(\xi_n, \eta_n))$ .

**Пример 9.1.** Вот простые (и используемые наиболее часто) примеры применения Следствия 9.1.

1. *Аддитивность.*

Если в условиях Теоремы 9.1  $D_1 = D_2 = \mathbb{R}^d$ , а  $\theta$  — нулевой вектор, то  $\mathcal{L}(\xi_n + \eta_n) \Rightarrow \mathcal{L}(\xi)$ . Таким образом, случайные величины (и вектора), сходящиеся по вероятности к нулю, играют роль (аддитивных) бесконечно малых.<sup>100</sup>

Скажем, пусть  $\beta_1, \dots, \beta_n$  — независимые одинаково распределенные случайные величины со средним  $\theta$  и единичной дисперсией. Обозначим (как принято в статистике)  $\bar{x}_n$  среднее арифметическое  $\beta_i$  и положим  $\xi_n = \sqrt{n}(\bar{x}_n - \theta)$ . Тогда, конечно,  $\mathcal{L}(\xi_n) \Rightarrow N(0, 1)$ . Если  $\eta_n = \eta/n$ , где  $\eta$  имеет распределение Коши, то слабый предел последовательности распределений  $\mathcal{L}(\xi_n + \eta_n)$  не изменится,<sup>101</sup> несмотря на то, что у случайной величины  $\xi_n$  среднее и дисперсия совпадают с параметрами предельного распределения  $N(0, 1)$ , а математического ожидания у случайной величины  $\xi_n + \eta_n$  не существует.

<sup>99</sup> А как?

<sup>100</sup> Обратите внимание, что здесь ничего не говорится о зависимости или независимости случайных векторов  $\xi_n$  и  $\eta_n$ ! Если бы они были независимы, то техника характеристических функция позволила бы доказать этот факт очень просто. А как?

<sup>101</sup> Да?

## 2. Мультипликативность.

Если в условиях Теоремы 9.1  $D_1 = \mathbb{R}^d$ , а  $D_2 = \mathbb{R}$ , то  $\mathcal{L}(\xi_n \eta_n) \Rightarrow \mathcal{L}(\theta \xi)$ . В частности, если  $\theta = 0$ , то  $\xi_n \eta_n \xrightarrow{P} 0$ .

Продолжая пример со случайными величинами  $\beta_i$  и ЦПТ, мы мгновенно получим, что

$$\mathcal{L}(\sqrt{n}(\bar{x}_n - \theta)\bar{x}_n) \Rightarrow N(0, \theta^2).$$

В частности, если  $\theta = 0$ , то  $\bar{x}_n^2 \xrightarrow{P} 0$ .

Рассматривая «мультипликативный» вариант Следствия 9.1 с  $\theta = 0$ , можно догадаться, что здесь требование слабой сходимости  $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$  избыточно. Это как в «обычном» анализе: если  $b_n \rightarrow 0$ , то для того, чтобы  $a_n b_n \rightarrow 0$ , нет нужды требовать, чтобы последовательность  $a_n$  имела конечный предел, достаточно, чтобы она была ограниченной. Один из аналогов этого простого факта, относящийся к сходимости по вероятности, может быть сформулирован следующим образом.

**Предложение 9.4.** Пусть  $\xi_n, \eta_n \in \mathbb{R}$ . Если семейство распределений  $\{\mathcal{L}(\xi_n), n \geq 1\}$  плотно, а  $\eta_n \xrightarrow{P} 0$ , то  $\xi_n \eta_n \xrightarrow{P} 0$ .<sup>102</sup>

*Доказательство.* Ясно, что при фиксированном  $\delta > 0$

$$P(|\xi_n \eta_n| \geq \delta) = P(|\xi_n \eta_n| \geq \delta, |\eta_n| < \varepsilon) + P(|\xi_n \eta_n| \geq \delta, |\eta_n| \geq \varepsilon) = I_1(\delta, \varepsilon, n) + I_2(\delta, \varepsilon, n),$$

причем

$$I_1(\delta, \varepsilon, n) \leq P(|\xi_n| \geq \delta/\varepsilon, |\eta_n| < \varepsilon) \leq P(|\xi_n| \geq \delta/\varepsilon),$$

что может быть сделано сколь угодно малым равномерно по  $n$  выбором  $\varepsilon$ .<sup>103</sup> Что касается  $I_2$ , то эта величина стремится к нулю при  $n \rightarrow \infty$  при любом  $\varepsilon$ , так как  $I_2(\delta, \varepsilon, n) \leq P(|\eta_n| \geq \varepsilon)$ , а  $\eta_n \xrightarrow{P} 0$ .  $\square$

**Следствие 9.2.** Если  $|a_n| \rightarrow \infty$ , а семейство  $\{\mathcal{L}(a_n \beta_n), n \geq 1\}$  плотно, то  $\beta_n \xrightarrow{P} 0$ .

*Доказательство.* Обозначим  $\xi_n = a_n \beta_n$  и  $\eta_n = 1/a_n$ , тогда  $\beta_n = \xi_n \eta_n$  и все сводится к применению Предложения 9.4.  $\square$

**Замечание 9.4.** Из теоремы Прохорова<sup>104</sup> следует, что в полном метрическом сепарабельном пространстве любая слабо сходящаяся последовательность распределений плотна. Поэтому утверждение Следствия 9.2 верно, если последовательность распределений  $\{\mathcal{L}(a_n \beta_n)\}$  имеет слабый предел. Этот факт, однако, легко выводится непосредственно из Следствия 9.1.<sup>105</sup>

## 9.2.2 О доверительных интервалах

Одно из общих утверждений, показывающих, как Теорема 9.1 и Следствие 9.1 используются в статистике, может быть описано следующим образом.

Пусть  $x_1, \dots, x_n$  — повторная выборка<sup>106</sup> из распределения  $\mathcal{Q}$ , причем  $\mathcal{Q} = \mathcal{Q}_{\theta, \vartheta}$ , где  $\theta \in \mathbb{R}$ , а  $\vartheta \in \mathbb{R}^d$ .<sup>107</sup> Предположим, что для некоторой оценки  $\hat{\theta}_n$  параметра  $\theta$

$$L(a_n(\hat{\theta}_n - \theta)) \Rightarrow \mathcal{P}_{\theta, \vartheta}, \quad (9.2)$$

<sup>102</sup>1. Обобщите это утверждение на случай  $\xi_n \in \mathbb{R}^d$ . 2. Приведите пример, показывающий, что без условия плотности это утверждение Предложения 9.4, вообще говоря, неверно.

<sup>103</sup>т.к. семейство  $\{\mathcal{L}(\xi_n), n \geq 1\}$  плотно.

<sup>104</sup>Формулировка приведена в разделе 7.2. Доказательство (как уже говорилось) см., например, в [1, гл.1 теор. 6.2].

<sup>105</sup>Убедитесь в этом.

<sup>106</sup>В статистических примерах мы будем использовать стандартные для статистики обозначения, например, обозначать случайные величины не греческими, а латинскими буквами.

<sup>107</sup>Здесь не предполагается, что знание параметров  $\theta, \vartheta$  полностью определяет распределение  $\mathcal{Q}$ .

где  $a_n > 0$ ,  $a_n \rightarrow +\infty$ , а распределение  $\mathcal{P} = \mathcal{P}_{\theta, \vartheta}$  известно с точностью до параметров  $\theta, \vartheta$ .

Если обозначить  $G(x) = G_{\theta, \vartheta}(x)$  функцию распределения, соответствующую распределению  $\mathcal{P}$  и потребовать, чтобы эта функция распределения была непрерывна, то отсюда получим, что для любого  $x \in \mathbb{R}$

$$\mathbb{P}(a_n(\hat{\theta}_n - \theta) < x) \rightarrow G(x) \quad \text{при } n \rightarrow \infty. \quad (9.3)$$

Из (9.3) сразу же следует, что

$$\mathbb{P}(a_n|\hat{\theta}_n - \theta| < x) \rightarrow F(x) \quad \text{при } n \rightarrow \infty, \quad (9.4)$$

где  $F(x) = G(x) - G(-x)$  при  $x \geq 0$ .

Наша цель — построить (асимптотический) доверительный интервал для параметра  $\theta$ .

Легко доказать следующее утверждение. Обозначим для краткости  $\xi_n = a_n|\hat{\theta}_n - \theta|$ .

**Предложение 9.5.** 1. Оценка  $\hat{\theta}_n$  является состоятельной для  $\theta$ .

2. Пусть при любых  $\theta, \vartheta$  функция  $F : [0, +\infty) \mapsto [0, 1)$  является обратимой. Обозначим  $H = H_{\theta, \vartheta} = F_{\theta, \vartheta}^{-1}$ . Тогда

$$\mathcal{L}(F_{\theta, \vartheta}(\xi_n)) \Rightarrow \mathcal{U}(0, 1) \quad (9.5)$$

и для любого  $\gamma \in (0, 1)$

$$\mathbb{P}(|\hat{\theta}_n - \theta| < a_n^{-1}H_{\theta, \vartheta}(\gamma)) \rightarrow \gamma. \quad (9.6)$$

*Доказательство.* 1. Из того, что функция  $F$  является непрерывной, следует, что сходимость в (9.4) равномерная. Поэтому для любого  $\varepsilon > 0$

$$|\mathbb{P}(|\hat{\theta}_n - \theta| < \varepsilon) - F(a_n\varepsilon)| = |\mathbb{P}(a_n|\hat{\theta}_n - \theta| < a_n\varepsilon) - F(a_n\varepsilon)| \rightarrow 0$$

при  $n \rightarrow \infty$ . Поскольку при этом  $F(a_n\varepsilon) \rightarrow 1$ , то все доказано.

2. Рассмотрим случайную величину  $\xi$  с функцией распределения  $F$ . Поскольку  $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$ , а функция  $F$  непрерывна, то  $\mathcal{L}(F(\xi_n)) \Rightarrow \mathcal{L}(F(\xi))$ . Ну, а случайная величина  $F(\xi)$  равномерно распределена на  $[0, 1]$ ,<sup>108</sup> откуда и следует (9.5).

Что касается (9.6), то эта сходимость сразу же следует из (9.5):

$$\gamma \leftarrow \mathbb{P}(F_{\theta, \vartheta}(\xi_n) < \gamma) = \mathbb{P}(\xi_n < H_{\theta, \vartheta}(\gamma)) = \mathbb{P}(|\hat{\theta}_n - \theta| < H_{\theta, \vartheta}(\gamma)/a_n).$$

Утверждение доказано. □

Если функция  $F$  не зависит от  $\theta, \vartheta$  (то есть полностью нам известна), то мы легко можем построить асимптотический доверительный интервал уровня  $\gamma$  для параметра  $\theta$ :

$$\mathbb{P}(|\hat{\theta}_n - \theta| < a_n^{-1}H(\gamma)) \approx \gamma \quad \text{при больших } n.$$

На самом деле это довольно редкий случай. Приведем один пример (он связан с ЦПТ), когда это так.

Пусть неизвестный параметр  $\theta$  — это среднее  $\mathbb{E}x_1$ , а распределение случайной величины  $x_1 - \theta$  известно.<sup>109</sup> Следовательно, дисперсия  $\sigma^2 \stackrel{\text{def}}{=} D(x_1) = D(x_1 - \theta)$  — тоже известная величина, и взяв  $\hat{\theta}_n = \bar{x}_n$ , получим,<sup>110</sup> что при  $a_n = \sqrt{n}$  и  $x \geq 0$

$$\mathcal{L}(a_n|\hat{\theta}_n - \theta| < x) \rightarrow F(x) = F_\sigma(x), \quad (9.7)$$

<sup>108</sup> На самом деле последний факт верен без требования обратимости функции распределения  $F$ , достаточно лишь ее непрерывности. Убедитесь в этом.

<sup>109</sup> Тогда параметр  $\theta$  называют *параметром сдвига*.

<sup>110</sup> Проверьте!

где  $F(x) = \Phi_\sigma(x) - \Phi_\sigma(-x)$ , а  $\Phi_\sigma$  — функция распределения нормального закона с нулевым средним и (известной) дисперсией  $\sigma^2$ .

Что же делать, если функция распределения  $F$  зависит от  $\theta$  и еще какого-то (тоже неизвестного) параметра  $\vartheta$  (ну, или от одного из них)?

Пример зависимости  $F$  от  $\vartheta$  у нас уже на самом деле приведен: если в формуле (9.7) дисперсия  $\sigma^2$  неизвестна, то  $\vartheta = \sigma$ .

Пример, когда  $F$  зависит от  $\theta$  тоже легко построить: для этого достаточно предположить, что  $x_i \in \Pi(\theta)$ . Тогда, так как  $D\xi_i = \theta$ , то в правой части той же формулы (9.7) будет стоять  $F_\theta$ .

**Предложение 9.6.** Пусть функция  $F(x) = F_{\theta, \vartheta}(x)$  является непрерывной по «трехмерному» аргументу  $(x; \theta, \vartheta)$ . Если  $\hat{\vartheta}_n$  — состоятельная оценка  $\vartheta$ , то при  $n \rightarrow \infty$

$$P(|\hat{\theta}_n - \theta| < a_n^{-1} H_{\hat{\theta}_n, \hat{\vartheta}_n}(\gamma)) \rightarrow \gamma \quad (9.8)$$

для любого  $\gamma \in (0, 1)$ .

*Доказательство.* Применяя Следствие 9.1 (точнее, Замечание 9.3 к этому следствию) к слабой сходимости (9.5), сразу же получаем, что

$$\mathcal{L}(F_{\hat{\theta}_n, \hat{\vartheta}_n}(\xi_n) \Rightarrow U(0, 1).$$

Отсюда (9.8) получается точно так же, как (9.6) из (9.5) □

Таким образом, в этом случае асимптотический доверительный интервал уровня  $\gamma$  для параметра  $\theta$  имеет вид

$$P(|\hat{\theta}_n - \theta| < a_n^{-1} H_{\hat{\theta}_n, \hat{\vartheta}_n}(\gamma)) \approx \gamma \quad \text{при больших } n. \quad (9.9)$$

**Замечание 9.5.** 1. Мы здесь рассматривали задачу построения (реализуемого) асимптотического доверительного интервала для неизвестного параметра  $\theta$ . Конечно, этот же интервал (и, более того, аналогичные построения) можно использовать и для задач проверки гипотез.

2. Условие, что параметры  $\theta$  и  $\vartheta$  могут принимать любые значения из  $\mathbb{R}$  и  $\mathbb{R}^d$  соответственно, конечно, являются ограничительными. Легко видеть, что это условие можно ослабить, фактически не меняя доказательства: можно потребовать, чтобы эти параметры принадлежали некоторым открытым подмножествам  $\Theta_1$  и  $\Theta_2$  прямой и  $\mathbb{R}^d$ . Тогда, однако, накладываются дополнительные условия на оценки этих параметров, которые тоже должны принимать значения в этих множествах.

Иначе говоря, утверждение Предложения 9.6 с ограничениями  $\theta \in \Theta_1$  и  $\nu \in \Theta_2$  сохранится, если дополнительно потребовать, чтобы при всех достаточно больших  $n$  выполнялись равенства  $P(\hat{\theta}_n \in \Theta_1) = P(\hat{\vartheta}_n \in \Theta_2) = 1$ . В следующем разделе мы обсудим эту проблему.

**Замечание 9.6.** Другая проблема, связанная с построением асимптотических доверительных интервалов, относится к необходимому объему выборки  $n$  для того, чтобы левая часть (9.9) достаточно хорошо аппроксимировала число  $\gamma$ . При этом величина  $n$  определяется не только скоростью сходимости в (9.6), но и выбором состоятельной оценки  $\hat{\theta}_n$  мешающего параметра  $\theta$ , которая вносит “дополнительную случайность” в решаемую задачу.

### 9.2.3 Сходимость к нормальному распределению

Перейдем теперь к непосредственному использованию Теоремы 9.1 (точнее, Следствия 9.1 или, еще точнее, Предложения 9.6) для целей математической статистики.<sup>111</sup>

<sup>111</sup>См. примеры в разделе 9.4.

А именно, будем считать, что предельное распределение  $\mathcal{P}$  в (9.2) имеет вид  $N(0, \sigma^2)$  с  $\sigma > 0$ . Тем самым мы полагаем  $\vartheta = \sigma$  с  $\Theta_2 = (0, \infty)$ , считаем, что  $\mathcal{P}$  не зависит от  $\theta$  и, для простоты, не накладываем никаких условий на  $\theta$ .

Если обозначит  $\Phi_\sigma$  функцию распределения  $N(0, \sigma^2)$ , то, как нетрудно видеть,  $\Phi_\sigma(x) = \Phi_1(x/\sigma)$  и поэтому (9.4) превращается в<sup>112</sup>

$$P(a_n|\hat{\theta}_n - \theta| < x) \rightarrow F_\sigma(x) = \Phi_1(x/\sigma) - \Phi_1(-x/\sigma) \quad \text{при } n \rightarrow \infty.$$

Ясно, что  $H_\sigma(y) = F_\sigma^{-1}(y) = \sigma H_1(y)$  при  $y \in [0, 1)$ , и поэтому (9.6) приобретает вид

$$P\left(|\hat{\theta}_n - \theta| < a_n^{-1}\sigma H_1(\gamma)\right) \rightarrow \gamma.$$

Поскольку функция  $F_\sigma(x)$  является непрерывной при  $\sigma > 0$  и  $x > 0$ , то из Предложения сразу же вытекает, что, если  $\hat{\sigma}_n$  является состоятельной и положительной<sup>113</sup> оценкой  $\sigma$ , то

$$P\left(|\hat{\theta}_n - \theta| < a_n^{-1}\hat{\sigma}_n H_1(\gamma)\right) \rightarrow \gamma, \quad (9.10)$$

Тем самым асимптотический доверительный интервал надежности  $\gamma$  для параметра  $\theta$  имеет вид

$$(\hat{\theta}_n - \hat{\sigma}_n H_1(\gamma)/a_n, \hat{\theta}_n + \hat{\sigma}_n H_1(\gamma)/a_n), \quad (9.11)$$

а ширина его пропорциональна  $1/a_n$ .

Тот факт, что функция  $H_\sigma$  очень просто (линейно) зависит от  $\sigma$ , позволяет получить результат Теоремы 9.3 более прямолинейно, без использования Предложения 9.6, а непосредственно опираясь на более общую Теорему 9.1.

А именно, этот результат сразу же<sup>114</sup> будет вытекать из следующего утверждения.

**Теорема 9.2.** Пусть  $\mathcal{L}(a_n(\hat{\theta}_n - \theta)) \Rightarrow N(0, \sigma^2)$ . Если  $\hat{\sigma}_n \xrightarrow{P} \sigma > 0$  и  $P(\hat{\sigma}_n > 0) = 1$  при всех достаточно больших  $n$ , то имеет место сходимость

$$\mathcal{L}(a_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n) \Rightarrow N(0, 1). \quad (9.12)$$

*Доказательство.* Будем для простоты считать, что  $\hat{\sigma}_n > 0$  при всех  $\omega$  и  $n$ .<sup>115</sup>

Положим в обозначениях Следствия 9.1  $\xi_n = a_n(\hat{\theta}_n - \theta)$ ,  $\theta = \sigma$ ,  $\eta_n = \hat{\sigma}_n$  и  $g(x, y) = x/y$ , где  $x \in \mathbb{R}$  и  $y > 0$ . Тогда  $g(\xi_n, \theta) = a_n(\hat{\theta}_n - \theta)/\sigma$  и  $g(\xi_n, \eta_n) = a_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n$ . Кроме того, рассмотрим случайную величину  $\xi \in N(0, \sigma^2)$ .

Так как  $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi) = N(0, \sigma^2)$ ,  $\mathcal{L}(g(\xi_n, \theta)) \Rightarrow \mathcal{L}(g(\xi, \theta)) = N(0, 1)$ ,  $0 < \eta_n \xrightarrow{P} \theta$  и функция  $g$  является непрерывной, то  $\mathcal{L}((\xi_n, \eta_n)) \Rightarrow \mathcal{L}(g(\xi, \theta))$  по Следствию 9.1.  $\square$

Можно привести еще одно, совсем «примитивное» доказательство Теоремы (9.2), опирающееся на мультипликативный случай Примера 9.1. Действительно, так как

$$\frac{a_n(\hat{\theta}_n - \theta)}{\hat{\sigma}_n} = \frac{a_n(\hat{\theta}_n - \theta)}{\sigma} \frac{\sigma}{\hat{\sigma}_n},$$

то для получения результата нам достаточно доказать, что  $\eta_n \stackrel{\text{def}}{=} \sigma/\hat{\sigma}_n \xrightarrow{P} 1$ . Докажем этот факт «в лоб».

<sup>112</sup>Конечно,  $\Phi_1$  — это нечто иное, как функция распределения  $N(0, 1)$ .

<sup>113</sup>для достаточно больших  $n$  с вероятностью 1

<sup>114</sup>Да?

<sup>115</sup>То, что  $n$  предполагаются достаточно большим, не имеет значения, так как нас интересует случай  $n \rightarrow \infty$ . Что касается положительности  $\hat{\sigma}_n$  почти всюду, то можно изменить  $\hat{\sigma}_n$  на множестве меры ноль так, чтобы они все были строго положительны, доказать нужный нам предельный результат, а потом «вернуться» к настоящим  $\hat{\sigma}_n$  воспользовавшись тем, что все используемые распределения от таких замен не меняются.

Очевидно,<sup>116</sup>  $1/\eta_n = \hat{\sigma}_n/\sigma \xrightarrow{P} 1$ . Далее, при малых  $\varepsilon > 0$

$$\begin{aligned} P(|\eta_n - 1| < \varepsilon) &= P(1 - \varepsilon < \sigma/\hat{\sigma}_n < 1 + \varepsilon) = P(1/(1 + \varepsilon) < \hat{\sigma}_n/\sigma < 1/(1 - \varepsilon)) = \\ &= P(1/(1 + \varepsilon) - 1 < \hat{\sigma}_n/\sigma - 1 < 1/(1 - \varepsilon) - 1) = P(-\varepsilon/(1 + \varepsilon) < \hat{\sigma}_n/\sigma - 1 < \varepsilon/(1 - \varepsilon)) \geq \\ &\geq P(|\hat{\sigma}_n/\sigma - 1| < \varepsilon/(1 + \varepsilon)), \end{aligned}$$

что стремится к единице ввиду сходимости  $\hat{\sigma}_n/\sigma$  к 1 по вероятности.

Простейший пример применения этой конструкции связан с теоремой П. Леви.

Пусть  $x_1, \dots, x_n, \dots$  — независимые одинаково распределенные случайные величины, имеющие среднее  $a$  и положительную конечную дисперсию  $\sigma^2$ . Если обозначить  $\bar{x}_n = (x_1 + \dots + x_n)/n$ , то в этих терминах классическая теорема Поля Леви утверждает, что

$$\mathcal{L}(\sqrt{n}(\bar{x}_n - a)/\sigma) \Rightarrow N(0, 1). \quad (9.13)$$

Из (9.13) следует, что для любого  $\gamma \in (0, 1)$

$$P(|\bar{x}_n - a| < z_\gamma \sigma / \sqrt{n}) \rightarrow \gamma, \quad (9.14)$$

где  $z_\gamma$  — решение уравнения

$$\Phi(z) - \Phi(-z) = \gamma, \quad (9.15)$$

а  $\Phi$  — функция распределения стандартного нормального закона.<sup>117</sup>

При известной дисперсии равенство (9.14) позволяет строить (асимптотический) доверительный интервал уровня  $\gamma$  для неизвестного среднего  $a$ , этот интервал имеет вид

$$(\bar{x}_n - z_\gamma \sigma / \sqrt{n}, \bar{x}_n + z_\gamma \sigma / \sqrt{n}).$$

Однако на практике обычно величина  $\sigma^2$  неизвестна, и общая рекомендация в этом случае состоит в замене в доверительном интервале  $\sigma$  на  $\hat{\sigma}_n$ , где  $\hat{\sigma}_n$  — какая-то состоятельная (то есть стремящаяся по вероятности к  $\sigma$ ) и положительная оценка  $\sigma$ . Тем самым новый доверительный интервал приобретает вид

$$(\bar{x}_n - z_\gamma \hat{\sigma}_n / \sqrt{n}, \bar{x}_n + z_\gamma \hat{\sigma}_n / \sqrt{n}), \quad (9.16)$$

что в точности соответствует (9.11).

Обычный выбор  $\hat{\sigma}_n$  — это выборочный стандарт

$$\bar{s}_n = \sqrt{\sum_{i=1}^n (x_i - \bar{x}_n)^2 / n}$$

или «исправленный» выборочный стандарт

$$\hat{s}_n = \sqrt{\frac{n}{n-1}} \bar{s}_n.$$

Такие рассуждения нуждаются, конечно, в строгом обосновании, которое во многом уже было сделано. А именно, уже доказан следующий результат.

<sup>116</sup> Действительно очевидно?

<sup>117</sup> В предыдущих обозначениях  $z_\gamma = H_1(\gamma)$ .

**Теорема 9.3.** Если  $\hat{\sigma}_n \xrightarrow{P} \sigma > 0$  и  $P(\hat{\sigma}_n > 0) = 1$  при всех достаточно больших  $n$ , то

$$\mathcal{L}(\sqrt{n}(\bar{x}_n - a)/\hat{\sigma}_n) \Rightarrow N(0, 1);$$

2. Для любого  $\gamma \in (0, 1)$

$$P(\bar{x}_n - z_\gamma \hat{\sigma}_n / \sqrt{n} < a < \bar{x}_n + z_\gamma \hat{\sigma}_n / \sqrt{n}) \rightarrow \gamma. \quad (9.17)$$

**Замечание 9.7.** Первый пункт Теоремы 9.3 можно назвать *модифицированной теоремой П. Леви*.

Как уже говорилось, стандартным вариантом выбора  $\hat{\sigma}_n^2$  является выборочная дисперсия  $\bar{s}_n^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2 / n$  или «исправленная» (несмещенная) выборочная дисперсия  $\hat{s}_n^2 = n\bar{s}_n^2 / (n - 1)$ . Покажем, что эти статистики можно использовать в качестве  $\hat{\sigma}_n^2$  в Теореме 9.2.

**Предложение 9.7.** Если распределение  $\mathcal{L}(x_1)$  непрерывно,<sup>118</sup> то для  $\hat{\sigma}_n = \bar{s}_n = \sqrt{\bar{s}_n^2}$  и  $\hat{\sigma}_n = \hat{s}_n = \sqrt{\hat{s}_n^2}$  выполняются условия Теоремы 9.2.

*Доказательство.* Поскольку выборочная дисперсия является дисперсией выборочного распределения, то  $\bar{s}_n^2 = (x_1^2 + \dots + x_n^2) / n - \bar{x}_n^2$ . Ясно, что  $\xi_n \stackrel{\text{def}}{=} (x_1^2 + \dots + x_n^2) / n \xrightarrow{P} Ex_1^2$  и  $\bar{x}_n \xrightarrow{P} Ex_1 = a$ . Поэтому (функция  $z - y^2$  непрерывна и снова используется Следствие 9.1)  $\bar{s}_n^2 = \xi_n - \bar{x}_n^2 \xrightarrow{P} Ex_1^2 - a^2 = \sigma^2$ .

Далее, равенство  $\bar{s}_n^2 = 0$  эквивалентно равенству  $x_1 = \dots = x_n$ , а последнее событие имеет нулевую вероятность, так как распределение  $\mathcal{L}(x_1)$  не имеет атомов. Тем самым утверждение для выборочной дисперсии доказано. Поскольку исправленная выборочная дисперсия отличается от  $\bar{s}_n^2$  лишь множителем, стремящемся к 1 при  $n \rightarrow \infty$ , то (снова применяя Следствие 9.1), получаем требуемое и для  $\hat{s}_n$ .  $\square$

**Замечание 9.8.** Доказательство Предложения 9.7 сознательно выбрано таким образом, чтобы оно (в части сходимости по вероятности) опиралось только на факты слабой сходимости распределений (в форме Следствия 9.1). Если выйти за эти рамки и использовать другую вероятностную технику, то состоятельность выборочной дисперсии можно доказать гораздо короче.

Так, из усиленного закона больших чисел следует, что почти наверное  $\bar{x}_n \rightarrow Ex_1$  и  $\xi_n \rightarrow Ex_1^2$ . Поэтому  $\bar{s}_n^2 = \xi_n - (\bar{x}_n)^2$  сходится к  $Ex_1^2 - (Ex_1)^2 = \sigma^2$  (тоже почти наверное), а из сходимости почти наверное следует сходимость по вероятности.

Модифицированная теорема П. Леви иллюстрирует общую схему применения Теоремы 9.1 в статистических исследованиях. В задаче построения (асимптотического) доверительного интервала для среднего при неизвестной дисперсии мы имеем дело с двумя параметрами (точнее — двумя характеристиками распределения  $\mathcal{L}(x_1)$ ). Один из них — само неизвестное среднее — является объектом нашего изучения, а второй — дисперсия — является «мешающим». Действительно, если мы знаем значение дисперсии, нет проблемы с построением доверительного интервала, не знаем — проблемы возникают. Теорема 9.1 и Следствие 9.1 показывают, как решается эта проблема: если заменить мешающий параметр его состоятельной оценкой, то предельное распределение, на основе которого строится доверительный интервал (или критерий проверки некоторой гипотезы) не меняется. Конечно, для правильного применения этого принципа нужно сначала проверить выполнение условий Теоремы 9.1 и Следствия 9.1.

Вернемся к модифицированной теореме Леви и обсудим, что делать, если условие  $P(\hat{\sigma}_n > 0) = 1$  не выполняется. Для конкретности, пусть  $\hat{\sigma}_n = \bar{s}_n$ . Представим себе, что  $x_i$  — это «бернуллиевские» случайные величины с параметром  $p$ . Тогда для любого  $p \in (0, 1)$  существует положительная вероятность того, что все  $x_i$  равны нулю (или все  $x_i$  равны 1). В этом случае  $\bar{s}_n = 0$ , выражение

<sup>118</sup>То есть если у этого распределения нет атомов.

в левой части (9.12) просто не имеет смысла, а доверительный интервал (9.16) вырождается в одноточечное множество.<sup>119</sup>

Что делать в этом случае? Ответ может быть дан на языке практики, языке чистой (формальной) теории и, так сказать, внутри некоторой логической концепции. Формально-теоретический ответ самый простой. Поскольку  $\bar{s}_n \xrightarrow{P} \sigma > 0$ , то для любой (неслучайной) строго положительной бесконечно малой  $\alpha_n$  случайная величина  $\hat{\sigma}_n^{(1)} = \bar{s}_n + \alpha_n$ <sup>120</sup> тоже сходится по вероятности к  $\sigma$  и при этом является положительной. Так давайте заменим  $\bar{s}_n$  на  $\bar{s}_n + \alpha_n$  и на этом успокоимся!

Конечно, это решение не является удовлетворительным с практической точки зрения хотя бы потому, что слишком велик произвол в выборе  $\alpha_n$ . Действительно, можно взять  $\alpha_n = 1/n$ , а можно положить  $\alpha_n = 10^6 / \ln(n)$ , предельное распределение будет одинаковым, но доверительные интервалы при фиксированном  $n$  могут быть совершенно не похожи.

На практике равенство  $\bar{s}_n$  нулю, скорее всего, означает, что имеющийся объем выборки  $n$  недостаточен, и нужно его увеличивать (если, конечно, это возможно). Существует и другие варианты. Представим себе, например, что выбранная для описания исследуемого явления модель не является адекватной (для нашего примера — что бернуллиевские случайные величины  $x_i$  являются зависимыми и/или не являются одинаково распределенными). Тогда мы вообще не имеем права пользоваться теоремой П. Леви и нужно с самого начала разбираться в постановке задачи.<sup>121</sup>

Мы же ставим вопрос следующим образом: как изменить формулировку предельной Теоремы 9.2, чтобы она включала в себя и случай  $P(\hat{\sigma}_n = 0) > 0$ ?

Сначала введем нужное обозначение и докажем простую лемму.

**Определение 9.2.** Пусть  $(\Omega, \mathcal{F}, P)$  — некоторое вероятностное пространство,  $A \in \mathcal{F}$  и  $P(A) > 0$ . Рассмотрим измеримое пространство  $(D, \mathcal{B})$  и случайную величину  $\xi : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{B})$ . Тогда условное распределение  $\xi$  при условии  $A$  (пишется —  $\mathcal{L}_\cdot$ ,  $A$  или  $\mathcal{L}(\xi | A)$ ) определяется равенством

$$\mathcal{L}_\xi(B|A) = P(\xi \in B | A) = \frac{P(\xi \in B, A)}{P(A)}, \quad B \in \mathcal{B}.$$

**Лемма 9.1.** Пусть  $\sigma$ -алгебра  $\mathcal{B}$  является борелевской, то есть порождается некоторой топологией подмножеств множества  $D$ . Рассмотрим  $A_n \in \mathcal{F}$  такие, что  $P(A_n) \rightarrow 1$ . Пусть, кроме того, распределения  $\mathcal{P}_n$  случайных величин  $\xi_n : (\Omega, \mathcal{F}) \mapsto (D, \mathcal{B})$  слабо сходятся к распределению  $\mathcal{P}$ . Тогда  $\mathcal{L}(\xi_n | A_n) \Rightarrow \mathcal{P}$ .

*Доказательство.* Нужно доказать, что для любой функции  $f \in C_b(D)$

$$\frac{E(f(\xi_n), A_n)}{P(A_n)} \rightarrow \int_D f d\mathcal{P}.$$

Пусть  $|f| \leq M$ . Так как  $P(A_n) \rightarrow 1$ , то результат следует из цепочки неравенств

$$\begin{aligned} \left| E(f(\xi_n), A_n) - \int_D f d\mathcal{P} \right| &\leq \left| E f(\xi_n) - \int_D f d\mathcal{P} \right| + |E(f(\xi_n), A_n^C)| \leq \\ &\leq \left| E f(\xi_n) - \int_D f d\mathcal{P} \right| + M(1 - P(A_n)). \end{aligned}$$

Лемма доказана. □

<sup>119</sup>В общем случае такая возможность может реализоваться тогда и только тогда, когда случайные величины  $x_i$  принимают какое-то значение с положительной вероятностью.

<sup>120</sup>Или  $\hat{\sigma}_n^{(2)} = \min(\bar{s}_n, \alpha_n)$

<sup>121</sup>Все дело, конечно, в том, что используемая оценка дисперсии может принимать значения, непозволительные для самой дисперсии: у нас предполагается, что  $\sigma^2 > 0$ , а  $\bar{s}_n^2$  может оказаться нулем.

Подобные ситуации не редкость на практике, и испытания Бернулли — всего лишь пример. Представьте себе, что Вы оцениваете по выборке некоторую характеристику распределения, которая по своему содержательному смыслу должна быть больше 1. А «естественная» (и состоятельная) оценка этой характеристики оказывается меньше 1. Что тогда делать? Ведь содержательно интерпретировать результат невозможно.

Перейдем теперь к соответствующему варианту Теоремы 9.2.

**Теорема 9.4.** Если в условиях Теоремы 9.2  $\hat{\sigma}_n \xrightarrow{P} \sigma > 0$  и  $P(\hat{\sigma}_n > 0) \rightarrow 1$ , то

$$\mathcal{L}\left(a_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n \mid \hat{\sigma}_n > 0\right) \Rightarrow N(0, 1). \quad (9.18)$$

*Доказательство.* Обозначим  $\beta_n = a_n(\hat{\theta}_n - \theta)/\sigma$ , так что  $\mathcal{L}(\beta_n) \Rightarrow N(0, 1)$ , и положим  $\eta_n = \beta_n \nu_n$ , где

$$\nu_n = \begin{cases} \sigma/\hat{\sigma}_n, & \hat{\sigma}_n > 0, \\ 1, & \hat{\sigma}_n \leq 0. \end{cases}$$

Так как  $\mathcal{L}(a_n(\hat{\theta}_n - \theta)/\hat{\sigma}_n \mid \hat{\sigma}_n > 0) = \mathcal{L}(\eta_n \mid \hat{\sigma}_n > 0)$ , то по Лемме 9.1 достаточно показать, что  $\mathcal{L}(\eta_n) \Rightarrow N(0, 1)$ . В свою очередь, это будет следовать из того, что  $\nu_n \xrightarrow{P} 1$ . Докажем последнюю сходимость. Ясно, что

$$P(|\nu_n - 1| < \varepsilon) = P(|\sigma/\hat{\sigma}_n - 1| < \varepsilon, \hat{\sigma}_n > 0) + P(\hat{\sigma}_n \leq 0) = I_1(n) + I_2(n).$$

По условию  $I_2(n) \rightarrow 0$ . Далее, если число  $\varepsilon > 0$  достаточно мало, то

$$\begin{aligned} I_1(n) &= P(1 - \varepsilon < \sigma/\hat{\sigma}_n < 1 + \varepsilon, \hat{\sigma}_n > 0) = P(1/(1 + \varepsilon) < \hat{\sigma}_n/\sigma < 1/(1 - \varepsilon), \hat{\sigma}_n > 0) \leq \\ &\leq P(-\sigma\varepsilon/(1 + \varepsilon) < \hat{\sigma}_n - \sigma < \sigma\varepsilon/(1 - \varepsilon)). \end{aligned}$$

Последнее выражение, очевидно, стремится к 1, так как  $\hat{\sigma}_n \xrightarrow{P} \sigma$ . □

Тем самым мы получаем условный асимптотический интервал надежности  $\gamma$ :

$$P(\hat{\theta}_n - z_\gamma \hat{\sigma}_n/a_n < \theta < \hat{\theta}_n + z_\gamma \hat{\sigma}_n/a_n \mid \hat{\sigma}_n > 0) \rightarrow \gamma,$$

где  $\gamma \in (0, 1)$ , а величина  $z_\gamma$  определена в (9.15).

### 9.3 Предельная теорема критерия $\chi^2$

Критерий  $\chi^2$  является одним из самых популярных при начальном изучении статистики. Тем не менее, с его обоснованием происходят странные вещи. В «слабых» курсах это (например, для экономистов) это обоснование, естественно, опускают. А в «продвинутых» им пренебрегают, считая очевидным.

На самом деле в доказательстве предельной теоремы, на которой основывается этот критерий используются несколько фактов о слабой сходимости распределений: многомерная ЦПТ, утверждение о сохранении слабого гауссовского предела при линейном преобразовании и (общее) утверждение о сохранении слабой сходимости при непрерывном отображении.<sup>122</sup>

Начнем с многомерной ЦПТ, которую можно назвать *многомерной теоремой П. Леви*.

#### 9.3.1 Многомерная теорема Леви

**Теорема 9.5.** Пусть  $\bar{\xi}_1, \dots, \bar{\xi}_n, \dots$  — последовательность  $d$ -мерных независимых одинаково распределенных векторов со средним  $\bar{a}$  и ковариационной матрицей  $\Sigma$ . Тогда

$$\mathcal{L}((\bar{\xi}_1 + \dots + \bar{\xi}_n - n\bar{a})/\sqrt{n}) \Rightarrow N_d(\mathbf{0}, \Sigma). \quad (9.19)$$

*Доказательство.* Не умаляя общности, можно считать, что  $\bar{a} = \mathbf{0}$ . Обозначим  $\varphi(\bar{t}) = \varphi(t_1, \dots, t_d)$  характеристическую функцию случайного вектора  $\bar{\xi}_1 = (\xi_1, \dots, \xi_d)^T$ . Тогда, как нетрудно показать,

$$\left. \frac{\partial \varphi(\bar{t})}{\partial t_j} \right|_{\bar{t}=\mathbf{0}} = iE\xi_j = 0 \quad \text{и} \quad \left. \frac{\partial \varphi(\bar{t})}{\partial t_j \partial t_k} \right|_{\bar{t}=\mathbf{0}} = -E\xi_j \xi_k = -\sigma_{jk},$$

где  $\sigma_{jk}$  — элемент матрицы  $\Sigma$ . Отсюда сразу же следует, что при  $\bar{t} \rightarrow \mathbf{0}$

$$\varphi(\bar{t}) = 1 - (\Sigma \bar{t}, \bar{t})/2 + o(|\bar{t}|^2). \quad (9.20)$$

Пусть  $\varphi_n$  — характеристическая функция вектора  $(\bar{\xi}_1 + \dots + \bar{\xi}_n)/\sqrt{n}$ . Тогда, очевидно,  $\varphi_n(\bar{t}) = \varphi^n(\bar{t}/\sqrt{n})$ , и из (9.20) следует, что для фиксированного  $\bar{t}$

$$\varphi_n(\bar{t}) = \left(1 - n^{-1}(\Sigma \bar{t}, \bar{t})/2 + o(n^{-1})\right)^n \rightarrow e^{-(\Sigma \bar{t}, \bar{t})/2}$$

при  $n \rightarrow \infty$ , что и доказывает утверждение. □

Приведем два содержательных примера использования Теоремы 9.5.

**Пример 9.2.** *Предельная теорема для независимых испытаний с  $m$  исходами.*

Пусть  $x_1, \dots, x_n$  — повторная независимая выборка из распределения

$$\mathcal{P} : \begin{pmatrix} 1 & 2 & \dots & m \\ p_1 & p_2 & \dots & p_m \end{pmatrix}, \quad (9.21)$$

причем  $p_i > 0$  при всех  $i$ . Обозначим  $k_i$  число  $x_j$ , равных  $i$ . Очевидно,  $k_1 + \dots + k_m = n$ . Если  $m = 2$ , то мы фактически имеем дело с испытаниями Бернулли, и для этого случая хорошо известна предельная теорема Муавра-Лапласа,<sup>123</sup> которая утверждает, что  $\mathcal{L}((k - np)/\sqrt{n}) \Rightarrow N(0, p(1-p))$ , где  $p = p_1$  и  $k = k_1$ . Выведем из многомерной теоремы П. Леви аналогичный результат для  $m > 2$ .

Построим  $m$ -мерные вектора  $\bar{\xi}_1, \dots, \bar{\xi}_n$  следующим образом. Пусть  $e_1, \dots, e_m$  —  $m$ -мерные орты, записанные в естественном порядке.<sup>124</sup> Положим

$$\bar{\xi}_i = (\xi_i^{(1)}, \dots, \xi_i^{(m)})^T = e_k \quad \text{при} \quad x_i = k, \quad 1 \leq k \leq m.$$

<sup>122</sup>Не так уж мало!

<sup>123</sup>Фактически — частный случай одномерной теоремы П. Леви.

<sup>124</sup>То есть  $e_1 = (1, 0, \dots, 0)^T$ ,  $e_2 = (0, 1, \dots, 0)^T$ , ...,  $e_m = (0, 0, \dots, 1)^T$ .

Конечно, вектора  $\bar{\xi}_i$  независимы и одинаково распределены, причем  $(k_1, \dots, k_m)^T = \sum_{i=1}^n \bar{\xi}_i$ .

Поскольку при фиксированном  $j$  случайные величины  $\xi_i^{(j)}$  представляют собой испытания Бернулли с вероятностью успеха  $p_j$ , а  $\xi_i^{(j)} \xi_i^{(k)} = 0$  при  $j \neq k$ , то  $E\bar{\xi}_i = (p_1, \dots, p_m)^T$  и

$$\sigma_{jk} \stackrel{\text{def}}{=} \text{Cov}(\xi_i^{(j)}, \xi_i^{(k)}) = E\xi_i^{(j)} \xi_i^{(k)} - p_j p_k = \begin{cases} p_j(1-p_j) & \text{при } k=j, \\ -p_j p_k & \text{при } k \neq j. \end{cases} \quad (9.22)$$

Поэтому многомерная теорема П. Леви выглядит в данном случае следующим образом:

$$\mathcal{L}\left(\left((k_1, \dots, k_m) - n(p_1, \dots, p_m)\right)^T / \sqrt{n}\right) \Rightarrow N(\mathbf{0}, \Sigma_m), \quad (9.23)$$

где  $\Sigma_m$  — матрица размера  $m \times m$  с элементами  $\sigma_{jk}$ .

**Замечание 9.9.** Поскольку  $k_1 + \dots + k_m = n$ , то ковариационная матрица  $\Sigma_m$  является вырожденной, но, если убрать из нее последний столбец и последнюю строку, то вырожденность пропадет.<sup>125</sup> Этот факт можно наглядно проиллюстрировать, рассматривая случай  $m = 2$ , то есть испытания Бернулли. Тогда  $k_1$  будет иметь смысл числа успехов, а  $k_2$  — числа неудач в  $n$  испытаниях Бернулли с вероятностью успеха  $p = p_1$ . Здесь матрица  $\Sigma_2$  приобретает вид

$$\Sigma_2 = \begin{pmatrix} p_1(1-p_1) & -p_1(1-p_1) \\ -p_1(1-p_1) & p_1(1-p_1) \end{pmatrix},$$

то есть имеет ранг, равный 1.

В то же время, если рассматривать только успехи, то мы приходим к стандартной теореме Муавра-Лапласа, где предельное нормальное распределение, конечно, не является вырожденным.

**Пример 9.3.** Асимптотическое распределение начальных моментов.

Рассмотрим независимые одинаково распределенные случайные величины  $x_1, \dots, x_n$  в предположении, что  $E x_i^{2k} < \infty$  для некоторого  $k > 1$ . Составим вектора  $\bar{\xi}_i = (x_i, x_i^2, \dots, x_i^k)^T$ . Очевидно, что  $E\bar{\xi}_i \stackrel{\text{def}}{=} \bar{a} = (m_1, m_2, \dots, m_k)^T$ , где  $m_j = E x_i^j$ . Столь же очевидно, что элемент  $\sigma_{st}$  ковариационной матрицы  $\Sigma$  вектора  $\bar{\xi}_i$  равен  $m_{s+t} - m_s m_t$ , где  $1 \leq s, t \leq k$ .

Обозначим  $\hat{m}_j^{(n)}$  начальный выборочный<sup>126</sup> момент  $j$ -го порядка, построенный по выборке  $x_i$ :

$$\hat{m}_j^{(n)} = \frac{1}{n} \sum_{i=1}^n x_i^j$$

и обозначим  $\bar{\eta}_n = (\hat{m}_1^{(n)}, \dots, \hat{m}_k^{(n)})^T$ . Теорема 9.5 в этом примере утверждает, что

$$\mathcal{L}(\sqrt{n}(\bar{\eta}_n - \bar{a})^T) \Rightarrow N(\mathbf{0}, \Sigma)$$

и является теоремой об асимптотическом распределении начальных выборочных моментов.

### 9.3.2 Предельная теорема критерия $\chi^2$

На основе сходимости (9.23) доказывается знаменитый критерий  $\chi^2$  для проверки простой гипотезы о дискретном распределении. Докажем сначала одну простое утверждение.

**Предложение 9.8.** Пусть  $\bar{\xi}_n \in \mathbb{R}^d$  — случайные вектора, причем  $\mathcal{L}(\bar{\xi}_n) \Rightarrow N_d(\bar{a}, \Sigma)$ . Рассмотрим (детерминированные) вектор  $\bar{b} \in \mathbb{R}^k$  и матрицу  $A : \mathbb{R}^d \mapsto \mathbb{R}^k$  и положим  $\bar{\eta}_n = A\bar{\xi}_n + \bar{b}$ . Тогда  $\mathcal{L}(\bar{\eta}_n) \Rightarrow N_k(\bar{b} + A\bar{a}, A\Sigma A^T)$ .

<sup>125</sup>Убедитесь как в первом, так и во втором. А как будет выглядеть предельная теорема, соответствующая этой операции?

<sup>126</sup>Слово «выборочный» здесь используется в следующем смысле. Пусть  $x_1, \dots, x_n$  повторная выборка из распределения  $\mathcal{P}$  и  $\theta$  — функционал, определенный на некотором подмножестве распределений, включающем  $\mathcal{P}$  и все конечные дискретные распределения. Если  $\tilde{\mathcal{P}}_n$  — эмпирическое распределение, построенное по выборке  $x_1, \dots, x_n$ , то  $\tilde{\theta}_n \stackrel{\text{def}}{=} \theta(\tilde{\mathcal{P}}_n)$  является выборочным аналогом характеристики  $\theta(\mathcal{P})$ . Таким образом, например, выборочная дисперсия — это дисперсия выборочного распределения.

*Доказательство.* Обозначим  $\varphi_n$  и  $\psi_n$  характеристические функции случайных векторов  $\bar{\xi}_n$  и  $\bar{\eta}_n$  соответственно. Тогда  $\varphi_n(\bar{t}) \rightarrow e^{i(\bar{t}, \bar{a})} e^{-(\Sigma \bar{t}, \bar{t})/2}$  для  $\bar{t} \in \mathbb{R}^d$  и  $\psi_n(\bar{s}) = e^{i(\bar{s}, \bar{b})} \varphi_n(A^T \bar{s})$  для  $\bar{s} \in \mathbb{R}^k$ .

Поскольку  $A^T \bar{s} \in \mathbb{R}^d$ , то отсюда ясно, что

$$\psi_n(\bar{s}) \rightarrow e^{i(\bar{s}, \bar{b})} e^{i(A^T \bar{s}, \bar{a})} e^{-(\Sigma A^T \bar{s}, A^T \bar{s})/2} = e^{i(\bar{s}, \bar{b} + A\bar{a})} e^{-(A\Sigma A^T \bar{s}, \bar{s})/2}.$$

Утверждение доказано.  $\square$

Вернемся теперь к критерию  $\chi^2$  и проведем необходимые построения,<sup>127</sup>

**Теорема 9.6.** *В рамках Примера 9.2*

$$\mathcal{L}(\eta_1^2 + \dots + \eta_m^2) \Rightarrow \chi^2(m-1), \quad (9.24)$$

где  $\eta_i = (k_i - np_i)/\sqrt{np_i}$ , а под  $\chi^2(K)$  понимается распределение суммы квадратов  $K$  независимых  $N(0, 1)$ -распределенных случайных величин.

*Доказательство.* Обозначим  $\xi_i = (k_i - np_i)/\sqrt{n}$  и положим  $\bar{\xi}_n = (\xi_1, \dots, \xi_m)^T$  и  $\bar{\eta}_n = (\eta_1, \dots, \eta_m)^T$ . Согласно (9.23),  $\mathcal{L}(\bar{\xi}_n) \Rightarrow N(\mathbf{0}, \Sigma_\xi)$ , где элементы матрицы  $\Sigma_\xi : \mathbb{R}^m \mapsto \mathbb{R}^m$  определены в (9.22).

Поскольку  $\eta_i = \xi_i/\sqrt{p_i}$ , то  $\bar{\eta} = A\bar{\xi}_n$ , где квадратная матрица  $A$  является диагональной с числами  $1/\sqrt{p_i}$  на диагонали.

Поэтому  $\mathcal{L}(\eta_n) \Rightarrow N(\mathbf{0}, \Sigma_\eta)$ , где  $\Sigma_\eta = A\Sigma_m A^T$ . Нетрудно видеть, что элементы  $\sigma'_{ij}$  матрицы  $\Sigma_\eta$  имеют вид

$$\sigma'_{jk} = \frac{\sigma_{jk}}{\sqrt{p_j p_k}} = \begin{cases} 1 - p_j & \text{при } j = k; \\ -\sqrt{p_j p_k} & \text{при } j \neq k. \end{cases}$$

Иначе говоря,  $\Sigma_\eta = \mathbf{I} - XX^T$ , где  $\mathbf{I}$  — единичная матрица, а  $X = (\sqrt{p_1}, \dots, \sqrt{p_m})^T$ .

Снова сделаем линейное преобразование — теперь с вектором  $\bar{\eta}_n$ . Возьмем ортонормированную матрицу  $C : \mathbb{R}^m \mapsto \mathbb{R}^m$ , первая строка которой есть  $(\sqrt{p_1}, \dots, \sqrt{p_m})$ , и положим  $\bar{\beta}_n = (\beta_1, \dots, \beta_m)^T = C\bar{\eta}_n$ . Заметим, что

$$\beta_1 = \sum_{j=1}^m \sqrt{p_j}(k_j - np_j)/\sqrt{np_j} = \frac{1}{\sqrt{n}} \sum_{j=1}^m (k_j - np_j) = 0.$$

Асимптотически вектор  $\bar{\beta}_n$  будет иметь гауссовское распределение с нулевым средним и ковариационной матрицей

$$\Sigma_\beta = C(\mathbf{I} - XX^T)C^T = C\mathbf{I}C^T - CXX^T C^T = \mathbf{I} - CXX^T C^T.$$

Так как все строки матрицы  $C$ , начиная со второй, ортогональны вектору  $X$ , а скалярное произведение первой строки и вектора  $X$  дает единицу, то  $CX = (1, 0, \dots, 0)^T$  и

$$\Sigma_\beta = \mathbf{I} - CXX^T C^T = \mathbf{I} - \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} (1 \ 0 \ \dots \ 0) = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix},$$

<sup>127</sup> см. [8, гл. III §1].

то есть в пределе оказывается, что случайные величины  $\beta_2, \dots, \beta_m$  являются не просто гауссовскими, но к тому же еще и независимыми с нулевыми средними и единичными дисперсиями. Учитывая, что  $\beta_1 = 0$ , и применяя тот факт, что слабая сходимость сохраняется при непрерывных отображениях, получаем, что

$$\mathcal{L}(\beta_1^2 + \dots + \beta_m^2) \Rightarrow \chi^2(m-1).$$

Осталось заметить, что<sup>128</sup>  $\eta_1^2 + \dots + \eta_m^2 = \beta_1^2 + \dots + \beta_m^2$ , так что сходимость (9.24) действительно имеет место.  $\square$

**Замечание 9.10.** При  $m = 2$

$$\eta_1^2 + \eta_2^2 = \frac{(k - np)^2}{np} + \frac{(n - k - n(1 - p))^2}{n(1 - p)} = \frac{(k - np)^2}{np} + \frac{(k - np)^2}{n(1 - p)} = \frac{(k - np)^2}{np(1 - p)},$$

так что в этом случае сходимость (9.24) элементарно<sup>129</sup> следует из предельной теоремы Муавра-Лапласа.

На утверждении Теоремы 9.6 и основан статистический критерий  $\chi^2$  проверки простой гипотезы о том, что исходная выборка получена из распределения (9.21).<sup>130</sup>

**Замечание 9.11.** Существует еще один, более сложный вариант критерия  $\chi^2$ , когда нулевая гипотеза не является простой и вероятности  $p_i$  в распределении (9.21) зависят от некоторого числа параметров. Конечно, тогда эти параметры должны каким-то образом оцениваться. Мы здесь не рассматриваем этот случай, связанный с методами их оценки.

<sup>128</sup>Так как матрица  $C$  ортонормированная.

<sup>129</sup>Да?

<sup>130</sup>Стандартный вид этого критерия состоит в том, что гипотеза отвергается, если  $\eta_1^2 + \dots + \eta_m^2$  принимает слишком большие значения. А почему критическая область устроена именно так? Является ли этот критерий состоятельным и почему?

## 9.4 Сохранение асимптотической нормальности при гладком отображении

Теорема о сохранении асимптотической нормальности довольно широко распространена, наглядна и часто используется как в теоретических построениях, так и на практике.

### 9.4.1 Теорема о сохранении асимптотической нормальности

Предположим, что мы измеряем  $n$  раз некоторый угол  $a$ . Результаты измерения — независимая повторная выборка  $x_1, \dots, x_n$  со средним  $a$  и положительной дисперсией  $\sigma^2$ . Если нас интересует сам угол  $a$ , то, в принципе, нету проблем с построением (асимптотического) доверительного интервала для этой величины. Что, однако, делать, если нас интересует не  $a$ , а  $\cos(a)$  или, в общем случае,  $f(a)$ ?

Прежде, чем обсуждать подобные проблемы во всей их полноте, прикинем, какой должен быть ответ в этой конкретной задаче. Предположим, что  $f$  непрерывна в точке  $a$ . Ясно, что естественной (выборочной) оценкой величины  $f(a)$  является  $f(\bar{x}_n)$ , причем  $f(\bar{x}_n) \xrightarrow{P} f(a)$ .

Пусть теперь функция  $f$  является достаточной гладкой, тогда при больших  $n$  следует ожидать, что  $f(\bar{x}_n) - f(a) \approx f'(a)(\bar{x}_n - a)$ . Так как  $\mathcal{L}(\sqrt{n}(\bar{x}_n - a)) \Rightarrow N(0, \sigma^2)$ , то мы приходим к естественному предположению, что в этом случае  $\mathcal{L}(\sqrt{n}(f(\bar{x}_n) - f(a))) \Rightarrow N(0, \sigma_1^2)$ , где  $\sigma_1 = |f'(a)|\sigma$ . Тогда, если производная  $f'$  непрерывна в точке  $a$  и  $f'(a) \neq 0$ , то в качестве доверительно интервала уровня  $\gamma$  для  $f(a)$  можно взять интервал<sup>131</sup>

$$(f(\bar{x}_n) - |f'(\bar{x}_n)| \bar{s}_n z_\gamma / \sqrt{n}, f(\bar{x}_n) + |f'(\bar{x}_n)| \bar{s}_n z_\gamma / \sqrt{n}),$$

где  $z_\gamma$  — решение уравнения (9.15).

Конечно, эти нестрогие соображения можно сделать точными, в том числе и для многомерного случая. Заметим, что для линейного отображения  $f$  у нас уже есть результат, доказанный в Предложении 9.8.

Действительно, если  $\mathcal{L}(\bar{\xi}_n) \Rightarrow N(\bar{a}, \Sigma)$  и  $\bar{\eta}_n = A\bar{\xi}_n + \bar{b}$ , то, согласно Предложению 9.8,

$$\mathcal{L}(\bar{\eta}_n) \Rightarrow N(\bar{b} + A\bar{a}, A\Sigma A^T).$$

Перейдем теперь общему утверждению о сохранении асимптотической нормальности при гладком отображении (см., например, [10, теор. 1.12 и след. 1.1] или [11, раздел. 3.3]).<sup>132</sup>

Пусть  $\bar{\eta}_n$  — последовательность  $d$ -мерных случайных векторов,  $U$  — открытое подмножество  $\mathbb{R}^d$  и  $\bar{a} \in U$ . Рассмотрим отображение  $f = (f_1, \dots, f_k)^T : U \mapsto \mathbb{R}^k$  и предположим, что это отображение является дифференцируемым в точке  $a$ . Обозначим  $\Delta f_i \in \mathbb{R}^d$  градиент функции  $f_i$  и положим  $\Delta_f = (\Delta f_1 : \dots : \Delta f_k)$ . Очевидно, матрица  $\Delta_f$  имеет  $k$  столбцов и  $d$  строк.

**Теорема 9.7.** Пусть для некоторой последовательности  $c_n \rightarrow +\infty$  имеет место сходимость

$$\mathcal{L}(c_n(\bar{\eta}_n - \bar{a})) \Rightarrow N(\mathbf{0}, \Sigma). \quad (9.25)$$

Тогда  $P(\bar{\eta}_n \in U) \rightarrow 1$  и

$$\mathcal{L}(c_n(f(\bar{\eta}_n) - f(\bar{a})) \mid \bar{\eta}_n \in U) \Rightarrow N(\mathbf{0}, \Sigma_f), \quad (9.26)$$

где  $\Sigma_f = \Delta_f^T(\bar{a}) \Sigma \Delta_f(\bar{a})$ .

*Доказательство.* Ввиду сходимостей  $c_n \rightarrow \infty$  и (9.25) из Следствия 9.2 и Замечания 9.4 сразу же выводится, что  $\bar{\eta}_n \xrightarrow{P} \bar{a}$ . Отсюда следует, что  $P(\bar{\eta}_n \in U) \rightarrow 1$ .

<sup>131</sup>Здесь, конечно, предполагается, что  $\bar{s}_n > 0$  и  $f'(\bar{x}_n) \neq 0$ .

<sup>132</sup>В англоязычной литературе это утверждение называется «дельта-метод».

Остаток доказательства мы приведем в предположении  $P(\bar{\eta}_n \in U) = 1$ . Переход к общему случаю не представляет труда<sup>133</sup> и проводится тем же приемом, что и в Теореме 9.4.

Из разложения функции  $f$  в ряд Тейлора в точке  $\bar{a}$  следует, что

$$c_n(f(\bar{\eta}_n) - f(\bar{a})) = \Delta_f^T(\bar{a}) (c_n(\bar{\eta}_n - \bar{a})) + \bar{\mathcal{E}}(\|\bar{\eta}_n - \bar{a}\|) c_n \|\bar{\eta}_n - \bar{a}\| = J_1(n) + J_2(n),$$

где  $\bar{\mathcal{E}}(\|\bar{x}\|) \rightarrow \mathbf{0}$  при  $\|\bar{x}\| \rightarrow 0$ . Согласно утверждению Предложения 9.8  $\mathcal{L}(J_1(n)) \Rightarrow N(\mathbf{0}, \Sigma_f)$ . Что касается  $J_2$ , то, ввиду (9.25) и непрерывности функции  $\|\bar{x}\|$ , распределения случайных величин  $c_n \|\bar{\eta}_n - \bar{a}\|$  имеют слабый предел. Поэтому  $\|\bar{\eta}_n - \bar{a}\| \xrightarrow{P} 0$  и  $\bar{\mathcal{E}}(\|\bar{\eta}_n - \bar{a}\|) \xrightarrow{P} \mathbf{0}$ . Следовательно,  $J_2 \xrightarrow{P} \mathbf{0}$  и утверждение доказано.  $\square$

**Замечание 9.12.** Наиболее употребительное применение Теоремы 9.7 связано с многомерной теоремой П. Леви, доказанной в разделе 9.3.

Обозначив в условиях Теоремы 9.5  $\bar{\eta}_n = (\bar{\xi}_1 + \dots + \bar{\xi}_n)/n$  и взяв  $c_n = \sqrt{n}$ , мы видим, что (9.19) приобретает вид (9.25), и поэтому можно применять Теорему 9.7.

Приведем теперь несколько простых примеров.

**Пример 9.4.** *Одномерный случай.*

В случае, когда  $d = k = 1$ , (9.26) приобретает вид

$$\mathcal{L}(c_n(f(\eta_n) - f(a)) \mid \eta_n \in U) \Rightarrow N(0, (f'(a))^2 \sigma^2).$$

Следовательно, если  $f'(a) = 0$ , то имеет место сходимость  $\mathcal{L}(c_n(f(\eta_n) - f(a)) \mid \eta_n \in U) \Rightarrow \delta_0$ .

Например, пусть  $x_1, \dots, x_n$  — независимая повторная выборка со средним  $a$  и положительной дисперсией  $\sigma^2$ . Обозначим  $\bar{x}_n = (x_1 + \dots + x_n)/n$ . По теореме П. Леви  $\mathcal{L}(\sqrt{n}(\bar{x}_n - a)) \Rightarrow N(0, \sigma^2)$ . Поэтому  $\mathcal{L}(\sqrt{n}(\cos(\bar{x}_n) - \cos(a))) \Rightarrow N(0, \sin^2(a) \sigma^2)$ . Если  $a = 0$ , то в результате получается, что  $\sqrt{n}(\cos(\bar{x}_n) - 1) \xrightarrow{P} 0$ , в то время как  $\mathcal{L}(\sqrt{n} \sin(\bar{x}_n)) \Rightarrow N(0, \sigma^2)$ .

Этот пример показывает, что матрица  $\Sigma_f$  не обязана быть невырожденной,<sup>134</sup> более того, она может оказаться даже нулевой.

**Пример 9.5.** *Стабилизация предельной дисперсии.*

Как уже говорилось, проблемы с построением доверительных интервалов возникают тогда, когда предельное распределение, возникающее, например, в (9.4), зависит от неизвестных параметров. В общем случае эта проблема решается с помощью состоятельных оценок этих параметров. Иногда, однако, ее можно решить просто подходящим нелинейным преобразованием.

Рассмотрим повторную выборку  $x_1, \dots, x_n$  из распределения  $\text{Ber}(p)$ , где  $p$  близко к нулю или 1. Несмотря на то, что построение асимптотического доверительного интервала для  $p$  не представляет труда — все основывается на теореме Муавра-Лапласа:

$$\mathcal{L}(\sqrt{n}(\bar{x}_n - p)) \Rightarrow N(0, p(1 - p)), \quad (9.27)$$

и например, один из вариантов границ такого интервала имеет вид

$$\bar{x}_n \pm z_\gamma \sqrt{\bar{x}_n(1 - \bar{x}_n)/n}, \quad (9.28)$$

при маленьких  $p$  задача оказывается не такой уж простой.

С интуитивной точки зрения это понятно: при очень маленьких вероятностях успеха нужно иметь очень большой объем выборки, чтобы достаточно точно оценить вероятность успеха. С формальной точки зрения этот эффект объясняется тем, что на практике доверительный интервал для маленького положительного параметра должен иметь длину, существенно меньшую значения этого

<sup>133</sup>Убедитесь в этом.

<sup>134</sup>Приведите такой пример при  $d, k > 1$  и невырожденной матрицы  $\Sigma$ .

параметра.<sup>135</sup> Поскольку в доверительном интервале (9.28)  $\sqrt{|\bar{x}_n|} < |\bar{x}_n|$ , то может потребоваться большой объем выборки  $n$  чтобы, например, нижняя граница интервала была больше нуля.

В этом смысле может помочь один прием, называемый приемом *стабилизации дисперсии*.<sup>136</sup>

А именно, снова положив  $d = k = 1$  в (9.26) и взяв  $f(x) = \arcsin \sqrt{x}$ , мы получим, что в нашем случае (9.27) перейдет в

$$\mathcal{L}(\sqrt{n}(\arcsin \sqrt{\bar{x}_n} - \arcsin \sqrt{p})) \Rightarrow N(0, 1/4),$$

так как при  $x \in (0, 1)$

$$(\arcsin \sqrt{x})' = \frac{1}{\sqrt{1-x}} \frac{1}{2\sqrt{x}} = \frac{1}{2\sqrt{x(1-x)}}.$$

Тем самым исчезает проблема оценивания предельной дисперсии и доверительный интервал для нового параметра  $\arcsin \sqrt{p}$  приобретает вид

$$\arcsin \sqrt{\bar{x}_n} - 0.5z_\gamma/\sqrt{n} < \arcsin \sqrt{p} < \arcsin \sqrt{\bar{x}_n} + 0.5z_\gamma/\sqrt{n}.$$

Теперь, решая эти неравенства относительно  $p$ , получим, что при больших  $n$

$$\left(\sin(\arcsin \sqrt{\bar{x}_n} - 0.5z_\gamma/\sqrt{n})\right)^2 < p < \left(\sin(\arcsin \sqrt{\bar{x}_n} + 0.5z_\gamma/\sqrt{n})\right)^2,$$

с вероятностью, близкой к  $\gamma$ .

Как видно отсюда, «стабилизация» дисперсии заключается в выборе такого преобразования, после которого предельная дисперсия становится независимой от неизвестных параметров.<sup>137</sup>

#### 9.4.2 Доверительный интервал для условной вероятности

**Пример 9.6.** *Доверительный интервал для условной вероятности.*

Предположим, что мы проводим следующий опрос. Сначала каждому из  $n$  человек задаем вопрос, на который он может ответить «Да» или «Нет». После этого тем из опрашиваемых, которые ответили «Да» на первый вопрос, задается второй вопрос, на который тоже можно ответить либо «Да», либо «Нет». Нас интересуют статистические свойства доли тех опрашиваемых, которые ответили положительно 2 раза среди тех, кто ответил «Да» на первый вопрос.

Формализуем описанную задачу. Все пространство элементарных событий разбивается на 3 подмножества:  $A_0$  (человек ответил «Нет» на первый вопрос),  $A_{11}$  — оба положительных ответа и  $A_{10}$  — первый ответ положительный, второй — отрицательный. Обозначив  $A_1 = A_{11} \cup A_{10}$ , мы видим, что нас на самом деле интересует условная вероятность  $P(A_{11} | A_1)$ .

Если обозначить  $p_{11} = P(A_{11})$  и  $p_{10} = P(A_{10})$ , то окажется, что нам нужно оценить величину

$$p \stackrel{\text{def}}{=} P(A_{11} | A_1) = \frac{P(A_{11})}{P(A_1)} = \frac{p_{11}}{p_{11} + p_{10}}.$$

Описанная схема опроса полностью отвечает Примеру 9.2, если в таблице распределения (9.21) положить  $m = 3$ ,  $p_1 = p_{11}$ ,  $p_2 = p_{10}$ ,  $p_3 = 1 - p_{11} - p_{10}$  и закодировать числом 1 ответ «Да-Да», числом 2 ответ «Да-Нет» и числом 3 ответ «Нет».

<sup>135</sup>Вряд ли имеет смысл оценивать число 0.01 с точностью 0.1.

<sup>136</sup>Достаточно подробный обзор методов построения доверительных интервалов для неизвестной вероятности успеха в испытаниях Бернулли можно найти в [9].

<sup>137</sup>Пусть  $x_1, \dots, x_n$  — повторная выборка из показательного распределения  $\text{EXP}(\lambda)$ , где  $\lambda$  — неизвестный параметр. Для этого случая ЦПТ имеет вид  $\mathcal{L}(\sqrt{n}(\bar{x}_n - 1/\lambda)) \Rightarrow N(0, 1/\lambda^2)$ . Как в этом случае построить «естественный» доверительный интервал для  $\lambda$  и как его «стабилизировать»? Тот же вопрос — для выборки из распределения Пуассона.

Поэтому, если обозначить  $k_{11}$  и  $k_{10}$  число людей, ответивших «Да-Да» и «Да-Нет» соответственно, то (9.23) переписывается как

$$\mathcal{L}\left(\sqrt{n}((k_{11}/n, k_{10}/n) - (p_{11}, p_{10}))^T\right) \Rightarrow N(\mathbf{0}, \Sigma), \quad (9.29)$$

где матрица  $\Sigma$  имеет вид

$$\Sigma = \begin{pmatrix} p_{11}(1 - p_{11}) & -p_{11}p_{10} \\ -p_{11}p_{10} & p_{10}(1 - p_{10}) \end{pmatrix}.$$

Поскольку  $\hat{p}_{11}^{(n)} \stackrel{\text{def}}{=} k_{11}/n$  и  $\hat{p}_{10}^{(n)} \stackrel{\text{def}}{=} k_{10}/n$  являются хорошими<sup>138</sup> оценками вероятностей  $p_{11}$  и  $p_{10}$  соответственно, то «естественной» (на самом деле — выборочной) оценкой условной вероятности  $p$  является

$$\hat{p}_n = \frac{k_{11}/n}{k_{11}/n + k_{10}/n} = \frac{k_{11}}{k_{11} + k_{10}}. \quad (9.30)$$

Учитывая (9.29), (9.30) и Замечание 9.12, мы можем применить Теорему 9.7 с  $d = 2$ ,  $k = 1$ ,  $\bar{\eta}_n = (k_{11}/n, k_{10}/n)^T$ ,  $\bar{a} = (p_{11}, p_{10})^T$ ,  $c_n = \sqrt{n}$ ,  $U = \{(x, y) : x, y > 0\}$  и  $f(x, y) = x/(x + y)$ .

Поскольку

$$\frac{\partial f(x, y)}{\partial x} = \frac{y}{(x + y)^2} \quad \text{и} \quad \frac{\partial f(x, y)}{\partial y} = -\frac{x}{(x + y)^2},$$

то

$$\Delta_f(\bar{a}) = \frac{1}{(p_{11} + p_{10})^2} \begin{pmatrix} p_{10} \\ -p_{11} \end{pmatrix}.$$

Нетрудно посчитать,<sup>139</sup> что

$$\Delta_f^T(\bar{a}) \Sigma \Delta_f(\bar{a}) = \frac{p_{11}p_{10}}{(p_{11} + p_{10})^3}.$$

Поэтому

$$\mathcal{L}\left(\sqrt{n}\left(\frac{k_{11}}{k_{11} + k_{10}} - \frac{p_{11}}{p_{11} + p_{10}}\right) \middle| k_{11}, k_{10} > 0\right) \Rightarrow N\left(0, \frac{p_{11}p_{10}}{(p_{11} + p_{10})^3}\right).$$

Заменяя (как это описано в Разделе 9.2) асимптотическую дисперсию  $p_{11}p_{10}/(p_{11} + p_{10})^3$  на ее состоятельную оценку

$$\hat{p}_{11}^{(n)}\hat{p}_{10}^{(n)}/(\hat{p}_{11}^{(n)} + \hat{p}_{10}^{(n)})^3 = n k_{11}k_{10}/(k_{11} + k_{10})^3,$$

получаем предельную теорему, пригодную для построения доверительного интервала.<sup>140</sup>

$$\mathcal{L}\left((\hat{p}_n - p) \sqrt{\frac{(k_{11} + k_{10})^3}{k_{11}k_{10}}} \middle| k_{11}, k_{10} > 0\right) \Rightarrow N(0, 1).$$

<sup>138</sup>выборочными, состоятельными, несмещенными, эффективными,...

<sup>139</sup>проверьте!

<sup>140</sup>И как выглядит этот интервал?

### 9.4.3 Выборочные характеристики, выражающиеся через первые и вторые моменты

Некоторые полезные характеристики распределений выражаются только через первые и вторые (начальные) моменты этих распределений. Сюда относятся, например, дисперсия, стандартное отклонение, ковариация, коэффициент корреляции, коэффициенты линейной регрессии и т.д. Легко строятся выборочные аналоги этих характеристик, распределения которых (в том числе и допредельные) хорошо изучены в случае, когда выборка делается из нормального распределения.<sup>141</sup> Эти распределения используются в различных статистических пакетах как стандартные. Случаи, когда выборка отлична от нормальной, гораздо менее популярны, хотя и в этой ситуации можно получать предельные распределения, пригодные для практики.<sup>142</sup>

В этом разделе мы приведем дальнейшие примеры, иллюстрирующие Теорему 9.7. Но сначала докажем одно общее утверждение, иллюстрирующую роль коэффициентов асимметрии и эксцесса в предельных теоремах.

**Лемма 9.2.** Пусть  $x_1, \dots, x_n$  — повторная независимая выборка, причем  $E x_1^4 < \infty$ . Обозначим, как и в Примере 9.3,  $m_1 = E x_1$ ,  $m_2 = E x_1^2$ ,  $\hat{m}_1^{(n)} = \bar{x}_n$  и  $\hat{m}_2^{(n)} = \sum_{j=1}^n x_j^2/n$ . Рассмотрим функцию  $f$  двух аргументов, гладкую в окрестности точки  $(m_1, m_2)$ .

Если  $E(x_1 - m_1)^3 = 0$  и  $E(x_1 - m_1)^4 = 3(m_2 - m_1^2)^2$ , то дисперсия предельного распределения случайной величины

$$\xi_n \stackrel{\text{def}}{=} \sqrt{n} \left( f(\hat{m}_1^{(n)}, \hat{m}_2^{(n)}) - f(m_1, m_2) \right)$$

будет такой же, как если бы выборка  $x_1, \dots, x_n$  была из нормального распределения со средним  $m_1$  и дисперсией  $\sigma^2 = m_2 - m_1^2$ .

*Доказательство.* Рассмотрим Пример 9.3 с  $k = 2$ . Результат этого примера утверждает, что

$$\mathcal{L} \left( \sqrt{n} \left( (\hat{m}_1^{(n)}, \hat{m}_2^{(n)})^T - (m_1, m_2)^T \right) \right) \Rightarrow N(\mathbf{0}, \Sigma), \quad (9.31)$$

где  $m_i = E x_i$ ,  $\hat{m}_1^{(n)} = \bar{x}_n$ ,  $\hat{m}_2^{(n)} = \sum_{j=1}^n x_j^2/n$  и

$$\Sigma = \begin{pmatrix} m_2 - m_1^2 & m_3 - m_1 m_2 \\ m_3 - m_1 m_2 & m_4 - m_2^2 \end{pmatrix}. \quad (9.32)$$

Согласно Теореме 9.7,  $\mathcal{L}(\xi_n) \Rightarrow N(0, \sigma_f^2)$ , где  $\sigma_f^2$  выражается (кроме первых частных производных функции  $f$ , сосчитанных в точке  $(m_1, m_2)$ ), через первые 4 начальных момента распределения  $\mathcal{L}(x_1)$ . Легко видеть,<sup>143</sup> что равенства  $E(x_1 - m_1)^3 = 0$  и  $E(x_1 - m_1)^4 = 3(m_2 - m_1^2)^2$ , выполняемые для любого нормального распределения, однозначно определяют моменты  $m_3$  и  $m_4$ , откуда и следует требуемое.  $\square$

Перейдем теперь непосредственно к примерам, используя, если это необходимо, обозначения Примера 9.3 с  $k = 2$ .

**Пример 9.7.** Асимптотическое распределение выборочной дисперсии.

Пусть  $x_1, \dots, x_n$  — повторная независимая выборка из распределения с конечным 4-м моментом.

Рассмотрим выборочную дисперсию  $\bar{s}_n^2 = \hat{m}_2^{(n)} - (\hat{m}_1^{(n)})^2$  как оценку генеральной дисперсии  $\sigma^2 = m_2 - m_1^2$ . Как уже говорилось, имеют место (9.31) и (9.32).

<sup>141</sup>См., например, [5, гл. 3].

<sup>142</sup>Скажем, проверка некоррелированности в статистических пакетах, как правило, основана на предположении о нормальности двумерного распределения. А что делать, если распределения явно не гауссовские (например, дискретные или сильно несимметричные?)

<sup>143</sup>Убедитесь!

Если задать функцию  $f(z_1, z_2) = z_2 - z_1^2$  с областью определения  $\mathbb{R} \times (0, +\infty)$ , то окажется, что  $f(m_1, m_2) = \sigma^2$  и  $f(\hat{m}_1^{(n)}, \hat{m}_2^{(n)}) = \bar{s}_n^2$ .

Далее используем равенство (9.26) с  $\Sigma_f = \Delta_f^T(\bar{a}) \Sigma \Delta_f(\bar{a})$ , где  $\Sigma$  определена в (9.32). Поскольку  $\Delta_f(z_1, z_2) = (-2z_1, 1)^T$ , то, как нетрудно подсчитать,<sup>144</sup>

$$\Delta_f^T(\bar{a}) \Sigma \Delta_f(\bar{a}) = m_4 - m_2^2 + 8m_2m_1^2 - 4m_1m_3 - 4m_1^4 = D(x_1 - Ex_1)^2,$$

где  $\bar{a} = (m_1, m_2)^T$ . Таким образом,

$$\mathcal{L}(\sqrt{n}(\bar{s}_n^2 - \sigma^2)) \Rightarrow N(0, D(x_1 - Ex_1)^2). \quad (9.33)$$

В случае, когда  $E(x - m_1)^4 = 3\sigma^4$ , предельная дисперсия  $D(x_1 - Ex_1)^2$  имеет вид  $2\sigma^4$ .<sup>145</sup> Основываясь на (9.33), нетрудно построить асимптотический доверительный интервал для  $\sigma^2$ .<sup>146</sup>

**Пример 9.8.** *Выборочный стандарт.*

В условиях Примера 9.7

$$\mathcal{L}(\sqrt{n}(\bar{s}_n - \sigma)) \Rightarrow N(0, D(x_1 - Ex_1)^2/4\sigma^2). \quad (9.34)$$

Здесь в качестве исходной предельной теоремы естественно взять (9.33) и положить  $f(z) = \sqrt{z}$ . Так как  $f'(z) = 1/\sqrt{4z}$ , то отсюда сразу же приходим к (9.34).

В гауссовском случае асимптотическая дисперсия равна  $\sigma^2/2$ . Постройте асимптотический доверительный интервал для  $\sigma$ .

**Пример 9.9.** *Совместное распределение выборочных среднего и дисперсии.*

В условиях Примера 9.7

$$\mathcal{L}(\sqrt{n}(\bar{x}_n - Ex_1, \bar{s}_n^2 - D(x_1 - Ex_1)^2)^T) \Rightarrow N(\mathbf{0}, \Sigma_f), \quad (9.35)$$

где  $\Sigma_f$  — ковариационная матрица случайного вектора  $(x_1 - Ex_1, (x_1 - Ex_1)^2)^T$ .

Действительно, рассмотрим отображение  $f: \mathbb{R}^2 \mapsto \mathbb{R}^2$ , задаваемое формулой

$$f(z_1, z_2) = (f_1(z_1, z_2), f_2(z_1, z_2))^T = (z_1, z_2 - z_1^2)^T$$

и положим  $\bar{a} = (m_1, m_2)$ . Тогда  $f(\bar{a}) = f(m_1, m_2) = (Ex_1, D(x_1 - Ex_1)^2)^T$  и  $f(\hat{m}_1^{(n)}, \hat{m}_2^{(n)}) = (\bar{x}_n, \bar{s}_n^2)^T$ .

Поскольку  $\partial f_1(z_1, z_2)/dz_1 = 1$ ,  $\partial f_1(z_1, z_2)/dz_2 = 0$ ,  $\partial f_2(z_1, z_2)/dz_1 = -2z_1$  и  $\partial f_2(z_1, z_2)/dz_2 = 1$ , то для ковариационной матрицы  $\Sigma$ , определенной в (9.32),

$$\begin{aligned} \Sigma_f &= \Delta_f^T(\bar{a}) \Sigma \Delta_f(\bar{a}) = \begin{pmatrix} 1 & 0 \\ -2m_1 & 1 \end{pmatrix} \begin{pmatrix} m_2 - m_1^2 & m_3 - m_1m_2 \\ m_3 - m_1m_2 & m_4 - m_2^2 \end{pmatrix} \begin{pmatrix} 1 & -2m_1 \\ 0 & 1 \end{pmatrix} = \\ &= \begin{pmatrix} 1 & 0 \\ -2m_1 & 1 \end{pmatrix} \begin{pmatrix} m_2 - m_1^2 & -2m_1(m_2 - m_1^2) + m_3 - m_1m_2 \\ m_3 - m_1m_2 & -2m_1(m_3 - m_1m_2) + m_4 - m_2^2 \end{pmatrix} = \\ &= \begin{pmatrix} m_2 - m_1^2 & m_3 - 3m_1m_2 + 2m_1^3 \\ m_3 - 3m_1m_2 + 2m_1^3 & m_4 - m_2^2 + 8m_2m_1^2 - 4m_1m_3 - 4m_1^4 \end{pmatrix}. \end{aligned}$$

Поскольку  $D(x_1 - Ex_1) = m_2 - m_1^2$ ,<sup>147</sup>  $D(x_1 - Ex_1)^2 = m_4 - m_2^2 + 8m_2m_1^2 - 4m_1m_3 - 4m_1^4$ <sup>148</sup> и<sup>149</sup>

$$\text{cov}(x_1 - Ex_1, (x_1 - Ex_1)^2) = E(x_1 - Ex_1)(x_1 - Ex_1)^2 = m_3 - 3m_1m_2 + 2m_1^3,$$

то ссылка на общую Теорему 9.7 обеспечивает сходимость (9.35).

<sup>144</sup>Подсчитайте!

<sup>145</sup>Проверьте!

<sup>146</sup>Постройте!

<sup>147</sup>Это очевидно.

<sup>148</sup>Это уже установлено при нахождении предельного распределения выборочной дисперсии.

<sup>149</sup>Проверьте!

**Замечание 9.13.** Хорошо известно<sup>150</sup> что для выборки из нормального распределения выборочное среднее и выборочная дисперсия независимы. Отсюда сразу же следует, что в этом случае предельная ковариационная матрица  $\Sigma_f$  будет диагональной. А для каких еще распределений это свойство будет выполняться?

Иначе говоря, при каких условиях на моменты случайной величины  $x_1$  координаты вектора  $\sqrt{n}(\bar{x}_n - Ex_1, \bar{s}_n^2 - Dx_1)^T$  будут асимптотически независимы?

Следующие примеры относятся к двумерным выборкам. Они оформлены в виде задач для самостоятельного решения.

**Пример 9.10.** *Выборочная ковариация.*

Пусть  $(x, y)^T$  — двумерный случайный вектор, обладающий конечными четвертыми моментами и ковариацией  $\text{Cov}(x, y)$ , а  $(x_1, y_1), \dots, (x_n, y_n)$  — соответствующая повторная выборка. Обозначим  $\bar{x}_n$  и  $\bar{y}_n$  выборочные средние, соответствующие случайным величинам  $x$  и  $y$ .

Тогда выборочная ковариация имеет вид

$$\text{cov}_n = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x}_n \bar{y}_n.$$

Нужно доказать, что

$$\mathcal{L}\left(\sqrt{n}(\text{cov}_n - \text{Cov}(x, y))\right) \Rightarrow N\left(0, D((x - Ex)(y - Ey))\right).$$

В гауссовском случае асимптотическая дисперсия выборочной ковариации равна  $\sigma_x^2 \sigma_y^2 (1 + \rho^2)$ , где  $\rho$  — коэффициент корреляции между  $x$  и  $y$ , а  $\sigma_x^2, \sigma_y^2$  — соответствующие дисперсии.

**Пример 9.11.** *Выборочный коэффициент корреляции.*

В условиях Примера 9.10 положим  $\rho = \text{corr}(x, y) = \text{Cov}(x, y) / (\sigma_x \sigma_y)$  и

$$r_n = \frac{\text{cov}_n}{\bar{s}_n(x) \bar{s}_n(y)},$$

где  $\bar{s}_n(x)$  и  $\bar{s}_n(y)$  — выборочные стандарты, относящиеся к переменным  $x$  и  $y$ . Кроме того, положим

$$\Sigma_f = D(\rho(z^2 + u^2) - 2zu)/4, \quad (9.36)$$

где  $z = (x - Ex)/\sigma_x$  и  $u = (y - Ey)/\sigma_y$ . Тогда

$$\mathcal{L}(\sqrt{n}(r_n - \rho) \mid \bar{s}_n(x) \bar{s}_n(y) > 0) \Rightarrow N(0, \Sigma_f).$$

Если исходная выборка гауссовская, то  $\Sigma_f = (1 - \rho)^2$ .

Заметим, что при проверке гипотезы о значимости коэффициента корреляции  $\rho$ <sup>151</sup> естественной<sup>152</sup> оценкой предельной дисперсии  $\Sigma_f$  является

$$\hat{\Sigma}_f^2 := \frac{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}{n \bar{s}_n^2(x) \bar{s}_n^2(y)},$$

так как заменять  $\rho$ , входящий в (9.36) на его состоятельную оценку в этом случае нет нужды.<sup>153</sup>

**Пример 9.12.** *Выборочный коэффициент линейной регрессии.*

Если  $(x, y)^T$  — случайный вектор, обладающий конечными вторыми моментами, то квадрат расстояния  $E(y - ax - b)^2$  между  $y$  и  $ax + b$  достигает своего минимума при  $a = \text{Cov}(x, y)/\sigma_x^2$  и  $b = Ey - aEx$ . Здесь мы будем рассматривать только коэффициент регрессии  $a$ .

<sup>150</sup> Действительно?

<sup>151</sup> А как это делается?

<sup>152</sup> А почему она состоятельная?

<sup>153</sup> А почему?

Рассмотрим повторную выборку  $(x_1, y_1), \dots, (x_n, y_n)$ , соответствующую вектору  $(x, y)^T$ , предполагая выполненными условия Примера 9.10. Выборочный вариант  $\tilde{a}_n$  коэффициента регрессии  $a$  имеет вид

$$\tilde{a}_n = \frac{\text{cov}_n}{\bar{s}_n^2(x)} = \frac{\sum_{i=1}^n x_i y_i / n - \bar{x}_n \bar{y}_n}{\sum_{i=1}^n x_i^2 / n - \bar{x}_n^2}.$$

Если положить  $x_* = x - \mathbb{E}x$ ,  $y_* = y - \mathbb{E}y$  и  $\Sigma_f = \mathbb{E}(x_*^2(y_* - ax_*)^2)/\sigma_x^4$ , то

$$\mathcal{L}(\sqrt{n}(\tilde{a}_n - a) \mid \bar{s}_n(x) > 0) \Rightarrow N(0, \Sigma_f).$$

В гауссовском случае  $\Sigma_f = (1 - \rho^2)\sigma_y^2/\sigma_x^2$ .

**Пример 9.13.** *Выборочные параметры линейной регрессии.*

Линейная регрессия Примера 9.12 обычно описывается тремя параметрами: коэффициентом регрессии  $a$ , свободным членом  $b$  и остаточной дисперсией  $\sigma^2 = \mathbb{E}(y - ax - b)^2$ . Постройте выборочные варианты этих параметров (для  $a$  это уже сделано) и докажите соответствующую трехмерную предельную теорему. Проанализируйте предельную ковариационную матрицу в гауссовском случае.

## Список литературы

- [1] П. Биллингсли, Сходимость вероятностных мер. М., Наука, 1977.
- [2] П.А. Мейер, Вероятность и потенциалы, М., Мир, 1973.
- [3] А.Н. Ширяев, Вероятность-1, М., Изд-во МЦНМО, 2004.
- [4] А.А. Боровков, Математическая статистика. М., Физматлит, 2007.
- [5] С.Р. Рао, Линейные статистические методы и их применения, М., Наука, 1968.
- [6] М. Лозв, Вероятность, М., ИЛ, 1962.
- [7] П. Халмош, Теория меры, М., ИЛ, 1953.
- [8] Ю.А. Розанов, Теория вероятностей, случайные процессы и математическая статистика, М., Наука, 1985.
- [9] A. Pires, C. Amado, Interval Estimators for a Binomial Proportion, REVSTAT – Statistical Journal, Volume 6, Number 2, June 2008, 165–197.
- [10] Jun Shao, Mathematical statistics, Springer texts in statistics, Springer-Verlag, New-York, 1999.
- [11] R. J. Serfling, Approximation Theorems of Mathematical Statistics, Wiley series in propability and statistics, Wiley-Interscience, 2001.