

Комментарии к теме «Слабая сходимость распределений»

Практические занятия по теории вероятностей, 322 гр., СМ, 2013 г.

В. В. Некруткин

1 Определение и основные свойства

1.1 Общая постановка задачи

Пусть \mathcal{P}_1 и \mathcal{P}_2 — два распределения в \mathbb{R}^d с борелевской σ -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 ε равномерное распределение $U(-\varepsilon, \varepsilon)$ на отрезке $[-\varepsilon, \varepsilon]$ «похоже» на распределение δ_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 метрику¹

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

которая эквивалентна ρ 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 ρ .

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 усложняется 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 Π_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.

¹Докажите, что это метрика.

1.2 Расстояние по вариации

Определение 1.1. (*Расстояние по вариации*). Пусть Π_d — множество распределений, определенных на борелевской σ -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 вариации² 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 $\mathcal{P}_n \xrightarrow{\text{var}} \mathcal{P}$.

Замечание 1.1. 1. Формула (1.1) задает метрику в множестве Π_d .³ 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.⁴

Теорема 1.1. (*Теорема Шеффе*).

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

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

1б. 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

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

2а. 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 $\mathcal{P}_n \xrightarrow{\text{var}} \mathcal{P}$.

2б. Если $\mathcal{P}_n, \mathcal{P}$ — дискретные распределения⁵ с вероятностями $p_\ell^{(n)}$ и p_ℓ соответственно. Если $p_\ell^{(n)} \rightarrow p_\ell$ при всех ℓ , 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 положительных⁶ ε равно $\varepsilon/(1 + \varepsilon)$, в то время как $\rho_{\text{var}}(U(0, 1), \text{Exp}(\lambda)) = e^{-\lambda}$ при $\lambda \leq 1$.⁷

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

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

²В другой терминологии — *сильно сходится*.

³Докажите это.

⁴Здесь она формулируется отдельно для абсолютно непрерывных и дискретных распределений в \mathbb{R}^d . Более общий случай см. в [1, с. 306] или в [2, гл. III §9 лемма 2]

⁵Сосредоточенные на одном множестве и том же не более чем счетном множестве.

⁶А при отрицательных?

⁷А при $\lambda > 1$?

распределения Стьюдента S_n с параметром n (так называемым «числом степеней свободы») поточно стремится⁸ при $n \rightarrow \infty$ к плотности распределения стандартного нормального закона сразу же следует, что $S_n \xrightarrow{\text{var}} N(0, 1)$.⁹

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 ε . Действительно, взяв $B = \{0\}$, получим, что $\mathcal{P}_\varepsilon(B) = 0$, а $\mathcal{Q}(B) = 1$. Это выглядит странно: при $\varepsilon \downarrow 0$ распределения $U(-\varepsilon, \varepsilon)$ все больше и больше концентрируются около нуля, а расстояние между этими распределениями и δ_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 $\xi_\varepsilon \equiv \varepsilon$ сходятся¹⁰ при $\varepsilon \rightarrow 0$ к случайной величине $\xi \equiv 0$, а их распределения не сходятся.¹¹

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

Оказывается, имеет место следующее утверждение ([2, 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 1.¹²

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

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

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

⁸Проверьте!

⁹В 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?

¹⁰В любом разумном смысле.

¹¹В этом примере расстояние по вариации аналогично дискретной метрике в \mathbb{R} .

¹²Исходя 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 $2 \sup_{B \in \mathcal{B}_d} |\mathcal{P}(B) - \mathcal{Q}(B)|$.

Замечание 1.2. 1. Легко видеть, что формула (1.5) действительно определяет метрику.
2. Поскольку

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

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

К сожалению, для BL-метрики не существует конструктивного представления в стиле теоремы Шеффе. Однако, если рассматривать не саму метрику $\rho_{\text{var}}(\mathcal{P}, \mathcal{Q})$, 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 $\rho_{\text{BL}}(\mathcal{P}_n, \mathcal{P}) \rightarrow 0$ 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 сложно¹³ (в отличие от Теоремы 1.1 и Предложения 1.1), и поэтому саму сходимость (1.6) обычно берут в качестве определения, а метрику ρ_{BL} рассматривают только при доказательстве теоретических результатов. Начиная с этого места мы будем поступать именно таким образом.

1.3 Слабая сходимость

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 $\mathcal{P}_n \Rightarrow \mathcal{P}$.¹⁴

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 σ -алгебре \mathcal{B}_d подмножеств \mathbb{R}^d . Обозначим соответствующие характеристические функции φ_n, φ 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 ∂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.

¹³С ним можно ознакомиться в [2, гл. III §7].

¹⁴Иногда пишут $\mathcal{P}_n \xrightarrow{w} \mathcal{P}$. Используются и другие обозначения.

Действительно, согласно определению сходимости по вариации сходимость $\mathcal{P}_n(B) \rightarrow \mathcal{P}(B)$ должна выполняться¹⁵ 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 ψ — непрерывная в нуле функция, то ψ 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 совпадет 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 поточечно.¹⁶

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 $B = \langle \alpha, \beta \rangle$,¹⁷ 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 ∂B

¹⁵Да еще и равномерно по B .

¹⁶Более того, в этом случае сходимость равномерна, то есть $\sup_x |F_n(x) - F(x)| \rightarrow 0$. Докажите это.

¹⁷Значки “ \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 $(\Omega, \mathcal{F}, \mathbb{P})$ и пусть ξ_n, ξ — 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 ω , 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

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

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 $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$.¹⁸ Конечно, так и есть. Проиллюстрируем это с помощью следующей диаграммы.

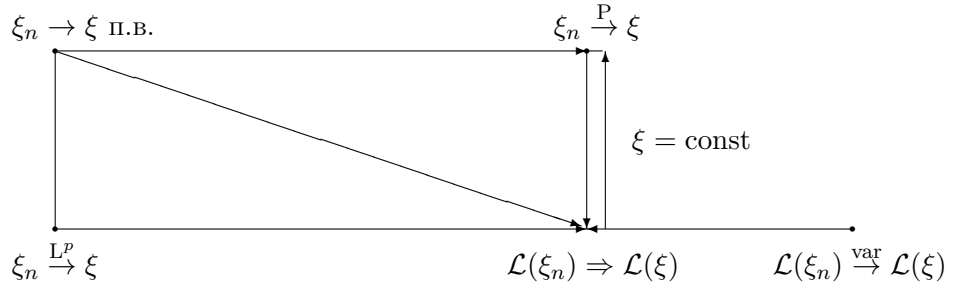


Рис. 1: Соотношения между разными видами сходимости

Рассматриваются 3 вида сходимости случайных величин: сходимость почти всюду¹⁹, по вероятности²⁰ и в L^p .²¹ Диаграмма показывает, что каждая из этих сходимостей влечет за собой слабую сходимость соответствующих распределений.²² 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 ξ_n, ξ , что $\xi_n \rightarrow \xi$ в одном из трех 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.

¹⁸Иногда вместо $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$ пишут $\xi_n \xrightarrow{d} \xi$ и говорят, что ξ_n сходится к ξ по распределению. При использовании такой терминологии нужно четко понимать, что речь идет не о сходимости случайных величин ξ_n к случайной величине ξ , а о сходимости их распределений. Например, в записи (1.7) ничего не мешает случайным величинам ξ_n и ξ быть заданным на разных вероятностных пространствах.

¹⁹Вероятность того, что $\xi_n \rightarrow \xi$, равна 1. Часто для такой сходимости применяю термин «почти наверное».

²⁰То есть $\mathbb{P}(|\xi_n - \xi| > \varepsilon) \rightarrow 0$ для любого $\varepsilon > 0$.

²¹При $d = 1$ это означает, что $\mathbb{E}|\xi_n - \xi|^p \rightarrow 0$. Здесь, конечно, предполагается, что $\xi_n, \xi \in L^p$. При $d > 1$ все аналогично.

²²При желании это можно считать еще одним основанием называть такую сходимость «слабой».

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, три²³ 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.²⁴

2. Если мы пользуемся вторым пунктом Предложения 1.2, то, взяв $x \in [0,1]$,²⁵ получаем, что

$$F_n(x) = \frac{\lceil nx \rceil}{n} \rightarrow x = F(x).$$

3. Если использовать для доказательства слабой сходимости характеристические функции, то при любом фиксированном t и достаточно большом n ²⁶

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

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 $\xi_n = \lfloor n\xi \rfloor / n$.²⁷ Тогда при $k = 0, \dots, n-1$

$$\mathbb{P}(\xi_n = k/n) = \mathbb{P}(\lfloor n\xi \rfloor = k) = 1/n$$

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

$$\frac{\lfloor n\xi \rfloor}{n} = \frac{n\xi - \{n\xi\}}{n} = \xi - \frac{\{n\xi\}}{n} \rightarrow \xi$$

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

²³Как уже говорилось, четвертый пункт Предложения 1.2 мало пригоден для решения подобных задач.

²⁴Здесь мы столкнулись с тем редким случаем, когда факт слабой сходимости может быть проверен прямо по определению.

²⁵При других x сходимость функций распределения очевидна.

²⁶Чтобы t/n не было целым числом при фиксированном $t \neq 0$.

²⁷Тем самым ξ_n и ξ определены на одном вероятностном пространстве.

2 О законе больших чисел и центральной предельной теореме

Среди многих утверждений о слабой сходимости распределений исторически выделяются две группы утверждений, традиционно называемые Слабым Законом Больших Чисел и Центральной Предельной Теоремой. Дадим соответствующие определения.

Определение 2.1. Говорят, что последовательность случайных величин $\xi_1, \dots, \xi_n, \dots$, обладающих математическими ожиданиями $\mathbb{E}\xi_k = a_k$, удовлетворяет слабому²⁸ закону больших чисел (стандартная аббревиатура — СЗБЧ), если

$$\frac{\xi_1 + \dots + \xi_n - (a_1 + \dots + a_n)}{n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0. \quad (2.1)$$

Замечание 2.1. Поскольку (2.1) может быть переписано в виде

$$\mathcal{L}\left(\frac{\xi_1 + \dots + \xi_n - (a_1 + \dots + a_n)}{n}\right) \implies \delta_0,$$

то утверждение, что какая-то последовательность случайных величин подчиняется СЗБЧ, является утверждением о слабой сходимости распределений. Но предельное распределение здесь является вырожденным.

Если все математические ожидания a_k одинаковы (обозначим их a), то (2.1) переписывается как

$$\frac{\xi_1 + \dots + \xi_n}{n} \xrightarrow[n \rightarrow \infty]{\mathbb{P}} a. \quad (2.2)$$

Это выражение лучше поддается интерпретации: если последовательность $\{\xi_n, n \geq 1\}$ с $\mathbb{E}\xi_n = a$ удовлетворяет слабому закону больших чисел, то при больших n среднее арифметическое ξ_k с большой вероятностью близко к a .²⁹

В случае, когда у случайных величин ξ_n имеются конечные вторые моменты, достаточным условием выполнения слабого закона больших чисел является сходимость $\mathbb{D}(\xi_1 + \dots + \xi_n)/n^2 \rightarrow 0$. Этот факт легко выводится из неравенства Чебышева, которое утверждает, что для любого $\varepsilon > 0$ и любой случайной величины ξ с конечной дисперсией

$$\mathbb{P}(|\xi - \mathbb{E}\xi| > \varepsilon) \leq \frac{\mathbb{D}\xi}{\varepsilon^2}.$$

В частности, если случайные величины ξ_i независимы, то для выполнения (2.1) достаточно, чтобы

$$\sum_{i=0}^n \mathbb{D}\xi_i = o(n^2)$$

при $n \rightarrow \infty$. Если же случайные величины ξ_i не только независимы, но и одинаково распределены с математическим ожиданием $\mathbb{E}\xi_i = a$, то, согласно теореме Хинчина, Слабый Закон Больших Чисел (в традиционной форме (2.2)) выполняется автоматически, причем конечность дисперсии здесь не предполагается.

²⁸В так называемом усиленном законе больших чисел рассматривается сходимость почти наверное, а не по вероятности.

²⁹Физики описывают это свойство такими словами: среднее по времени сходится к среднему по пространству, имея в виду, что $a = \int_{\mathbb{R}} x \mathcal{P}_{\xi_k}(dx)$.

Определение 2.2. Говорят, что последовательность случайных величин $\xi_1, \dots, \xi_n, \dots$, обладающих математическими ожиданиями $\mathbb{E}\xi_k = a_k$ и конечными вторыми моментами, удовлетворяет центральной предельной теореме (стандартная аббревиатура — ЦПТ), если

$$\mathcal{L}\left(\frac{\xi_1 + \dots + \xi_n - (a_1 + \dots + a_n)}{B_n}\right) \implies N(0, 1), \quad (2.3)$$

где $B_n^2 = \mathbb{D}(\xi_1 + \dots + \xi_n)$.³⁰

Структура формулы (2.3) очень проста. Из случайных величин ξ_k образуются последовательные суммы $S_n = \xi_1 + \dots + \xi_n$, которые потом центрируются вычитанием математического $\mathbb{E}S_n$ и нормируются делением на $\sqrt{\mathbb{D}(S_n)}$. В результате полученная случайная величина $(S_n - \mathbb{E}S_n)/\sqrt{\mathbb{D}(S_n)}$ имеет нулевое среднее и единичную дисперсию, что совпадает с параметрами 0 и 1 нормального распределения в правой части (2.3).

Тем самым то, что в пределе получается именно *стандартное* нормальное распределение, можно считать естественным. А вот сама нормальность, конечно, не является очевидной.

Замечание 2.2. Легко видеть, что предельные соотношения (2.1) и (2.3) останутся в силе, если случайные величины ξ_k заменить на $\xi_k + c_k$, где c_k — некоторые постоянные. Поэтому при теоретических исследованиях обычно (и без ущерба общности) считают, что $\mathbb{E}\xi_k = 0$.

Что касается ЦПТ, то наиболее популярной является теорема Поля Леви, утверждающая, что любая последовательность независимых одинаково распределенных величин, имеющих конечную ненулевую дисперсию, подчиняется Центральной Предельной Теореме. Сходимость (2.3) в этом случае приобретает вид

$$\mathcal{L}\left(\frac{\xi_1 + \dots + \xi_n - na}{\sigma\sqrt{n}}\right) \implies N(0, 1),$$

где $a = \mathbb{E}\xi_i$ и $\sigma^2 = \mathbb{D}\xi_i$.

Список литературы

- [1] П. Биллингсли (1977), Сходимость вероятностных мер. М., Наука.
- [2] А.Н. Ширяев (2004), Вероятность-1, М., Изд-во МЦНМО.

³⁰При этом подразумевается, что $B_n^2 > 0$ при всех достаточно больших n .