Лекции, не прочитанные во вторник 20.05.2008 г. на 1-й и 2-й парах

Статистическое моделирование

322 гр., специализация СМ

1 Комментарий к теореме непрерывности (прошлое занятие)

Теорема непрерывности, доказанная на предыдущей лекции.

Теорема 1.1. *Пусть $\mathcal{P}_n$, $\mathcal{P}$ — распределения в $\mathbb{R}^d$, φ_n — характеристические функции распределений $\mathcal{P}_n$.*
1. Для того, что бы последовательность $\mathcal{P}_n$ слабо сходилась к $\mathcal{P}$, н. и д., чтобы функции φ_n поточечно сходились к характеристической функции φ распределения $\mathcal{P}$.
2. Для того, что бы последовательность $\mathcal{P}_n$ имела некоторый слабый предел $\mathcal{P}$, н. и д., чтобы функции φ_n поточечно сходились к некоторой функции φ, непрерывной в нуле.

Замечание 1.1. 1. Первая часть теоремы 1.1 доказывается с помощью теоремы о классе функций, определяющих слабую сходимость, доказательство второй, кроме того, использует результат теоремы Прохорова.
2. Во второй части теоремы функция φ оказывается в конечном итоге характеристической функцией распределения $\mathcal{P}$.
3. Из поточечной сходимости φ_n к φ автоматически следует равномерна сходимость $\varphi_n \to \varphi$ на любом компакте.

Из теоремы 1.1 можно вывести следующее следствие. Пусть $\mathcal{P}_n$, $\mathcal{P}$ — распределения в $\mathbb{R}^d$ и $k \geq 1$ — целое.

Следствие 1.1. *Для того, что бы последовательность $\mathcal{P}_n$ слабо сходилась к $\mathcal{P}$, н. и д., чтобы при $n \to \infty$*

$$\int\limits_{\mathbb{R}^d} f d\mathcal{P}_n \to \int\limits_{\mathbb{R}^d} f d\mathcal{P}$$

для любой ограниченной k раз непрерывно дифференцируемой функции.

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. $\square$

2 Слабая сходимость и отображения

2.1 Основная теорема

Рассмотрим два измеримых пространства $(X_1, \mathcal{B}_1)$ и $(X_2, \mathcal{B}_2)$ и некоторое измеримое отображение $f : (X_1, \mathcal{B}_1) \mapsto (X_2, \mathcal{B}_2)$. Если в $(X_1, \mathcal{B}_1)$ задано распределение $\mathcal{P}$, то отображение f

порождает в $(X_2, \mathcal{B}_2)$ распределение $\mathcal{P}f^{-1}$ равенством

$$\mathcal{P}f^{-1}(B) = \mathcal{P}(f^{-1}(B)) = \mathcal{P}(x \in X_1 : f(x) \in B), \quad B \in \mathcal{B}_2. \tag{1}$$

Замечание 2.1. Замечания к определению.
1. Конечно, в этом определении совсем не обязательно, чтобы $\mathcal{B}_1$ и $\mathcal{B}_2$ были борелевскими σ-алгебрами — никаких метрик или топологий нет.
2. Запись $\mathcal{P}(f^{-1}(B))$ служит основанием для обозначения $\mathcal{P}f^{-1}$.
3. На вероятностном языке распределение $\mathcal{P}f^{-1}$ можно назвать распределением отображения f относительно $\mathcal{P}$.

Действительно, рассмотрим вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ и случайную величину $\xi : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}, \mathcal{B})$ (то, что ξ принимает значения на прямой — только для наглядности). Тогда распределение $\mathcal{P}_\xi$ случайной величины ξ — это распределение в $(\mathbb{R}, \mathcal{B})$, определяемое равенством

$$\mathcal{P}_\xi(B) = \mathbb{P}(\omega \in \Omega : \xi(\omega) \in B), \quad B \in \mathcal{B}.$$

Конечно, это то же самое определение, что в (1). Основная разница состоит в том, что в теории вероятностей вероятность $\mathbb{P}$ как правило (но не всегда) считается фиксированной, а случайных величин может быть много. Здесь же наоборот — можно рассматривать много мер $\mathcal{P}$ (например, целую последовательность) с одним и тем же отображением f. Отсюда и разница в обозначениях.
4. Если $g : (X_2, \mathcal{B}_1 \mapsto (\mathbb{R}, \mathcal{B})$ — подходящая функция (скажем, ограниченная), то имеет место равенство

$$\int\limits_{X_2} g \, d\mathcal{P}f^{-1} = \int\limits_{X_1} (g \circ f) \, d\mathcal{P} \left(= \int\limits_{X_1} g(f) \, d\mathcal{P} \right). \tag{2}$$

На самом деле это равенство — следствие из теоремы о замене переменной в интеграле Лебега. Если обратиться к теоретико-вероятностным ассоциациям, то (2) — это то же самое, что равенство

$$\int\limits_{\mathbb{R}} g(x) \, \mathcal{P}_\xi(dx) = \int\limits_{\Omega} (g \circ \xi) \; d\mathbb{P} \left(= \int\limits_{\Omega} g(\xi) \; d\mathbb{P} = \mathbb{E}g(\xi) \right),$$

стандартное в теории вероятностей.

Перейдем теперь в слабой сходимости. Пусть $(X_1, \rho_1, \mathcal{C}_1, \mathcal{B}_1)$ и $(X_2, \rho_2, \mathcal{C}_2, \mathcal{B}_2)$ — измеримые метрические пространства и $f : (X_1, \mathcal{B}_1) \mapsto (X_2, \mathcal{B}_2)$ — некоторое измеримое отображение. Пусть $\mathcal{P}_n$ — последовательность распределений в $(X_1, \mathcal{B}_1)$, слабо сходящаяся к распределению $\mathcal{P}$. Ставится вопрос: какими свойствами должно обладать отображение f, чтобы выполнялось соотношение $\mathcal{P}_n f^{-1} \Rightarrow \mathcal{P}f^{-1}$?

На вероятностном языке это можно записать так. Пусть ξ_n и ξ — случайные величины со значениями в X_1 (иначе говоря, измеримые отображения $(\Omega, \mathcal{F}) \mapsto (X_1, \mathcal{B}_1)$) и пусть $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$. Какими свойствами должно обладать отображение $f : (X_1, \mathcal{B}_1) \mapsto (X_2, \mathcal{B}_2)$, чтобы имела место сходимость $\mathcal{L}(\eta_n) \Rightarrow \mathcal{L}(\eta)$, где $\eta_n = f(\xi_n)$ и $\eta = f(\xi)$?

Простейший ответ на этот вопрос — достаточно, чтобы отображение было непрерывным.

Лемма 2.1. *Если f — непрерывное отображение (т.е., $f : (X_1, \mathcal{C}_1) \mapsto (X_2, \mathcal{C}_2)$), то $\mathcal{P}_n f^{-1} \Rightarrow \mathcal{P}f^{-1}$.*

Доказательство. Обсуждалось на занятиях по теории вероятностей. См. также [1, гл. 1, §5, стр. 47]. □

Это простое утверждение имеет много применений в теории вероятностей. Например, такое.

Следствие 2.1. *Пусть* ξ_n, η_n *— случайные величины,* $\zeta_n = (\xi_n, \eta_n)$ *и* $\zeta = (\xi, \eta)$*. Пусть, кроме того,* $\mathcal{L}(\zeta_n) \Rightarrow \mathcal{L}(\zeta)$*. Тогда, если* f *— непрерывная функция двух переменных,* $\beta_n = f(\xi_n, \eta_n)$ *и* $\beta = f(\xi, \eta)$*, то* $\mathcal{L}(\beta_n) \Rightarrow \mathcal{L}(\beta)$*.*

На самом деле таких отображений гораздо больше. Обозначим $D_f \subset X_1$ множество точек разрыва отображения f. Оказывается (доказательство можно найти в [1, добавление II, стр. 309], но мы его опускаем), что это множество измеримо (то есть принадлежит борелевской σ-алгебре $\mathcal{B}_1$) даже если само отображение f не является измеримым.

Имеет место следующее утверждение.

Теорема 2.1. *Если* $\mathcal{P}(D_f) = 0$ *и* $\mathcal{P}_n \Rightarrow \mathcal{P}$*, то* $\mathcal{P}_n f^{-1} \Rightarrow \mathcal{P} f^{-1}$*.*

Доказательство. См. [1, гл. 1, §5, теорема 5.1 стр. 49]. □

В качестве простого примера применения это теоремы можно взять пример следствия 2.1, но потребовать, чтобы случайные величины ξ и η имели совместную плотность, а также заменить непрерывную функцию f на кусочно-непрерывную (то есть имеющую конечное число точек разрыва). Проверьте это.

Есть, однако, гораздо более интересное следствие из теоремы 2.1.

Следствие 2.2. *Пусть* $(X, \rho, \mathcal{C}, \mathcal{B})$ *— измеримое метрическое пространство, распределения* $\mathcal{P}_n, \mathcal{P}$ *определены на* $\mathcal{B}$ *и* $\mathcal{P}_n \Rightarrow \mathcal{P}$*. Если* f *— ограниченная функция, множество точек разрыва которой имеет* $\mathcal{P}$*-меру ноль, то*

$$\int\limits_X f d\mathcal{P}_n \to \int\limits_X f d\mathcal{P}. \tag{3}$$

Доказательство. См. [1, гл. 1, §5, теорема 5.2, п. 3, стр. 49]. □

Замечание 2.2. О чем говорит следствие 2.2? По определению, для слабой сходимости необходимо и достаточно выполнения сходимости (3) для всех непрерывных ограниченных функций.

Обсуждая классы функций, определяющие слабую сходимость, мы решали такую задачу: как сузить класс непрерывных ограниченный функций, для которых имеет место сходимость (3), чтобы все равно отсюда вытекала сходимость $\mathcal{P}_n \Rightarrow \mathcal{P}$.

Теперь мы решаем обратную задачу: мы знаем, что (3) имеет место для любых непрерывных ограниченных функций и задаемся вопросом: а для каких еще функций выполняется (3). Оказывается, можно ослабить условие непрерывности функции f до условия ее почти непрерывности относительно меры $\mathcal{P}$ (при этом условие ограниченности остается). На самом деле условие ограниченности тоже можно существенно ослабить (об этом написано в [1, гл. 1, §5]), но эти вопросы мы здесь опускаем.

2.2 Интеграл Римана и слабая сходимость

Проиллюстрируем, насколько существенен результат следствия 2.2, на примере интеграла Римана.

Рассмотрим $X = [0,1]^d$. Пусть Ξ_n — последовательность разбиений (полуоткрытого) множества $(0,1]^d$ на (попарно непересекающиеся) полуоткрытые слева ячейки $\Lambda_j^{(n)}$ $j = 1, \ldots, m_n$, $n \to \infty$. Обозначим $\Delta_j^{(n)}$ замыкания этих ячеек.

Пусть $\Im$ — множество последовательностей разбиений Ξ_n таких, что $\max_{1 \le j \le m_n} \operatorname{diam} \Delta_j^{(n)} \to 0$ при $n \to \infty$.

Обозначим $\left|\Delta_j^{(n)}\right|$ объем ячейки $\Delta_j^{(n)}$. По определению, функция f, определенная на X, называется интегрируемой по Риману, если для любой последовательности разбиений $\Xi_n \in \Im$ и любом выборе точек $x_j^{(n)} \in \Delta_j^{(n)}$ римановы суммы

$$S_n = \sum_{1\le j\le m_n} f\big(x_j^{(n)}\big)\big|\,\Delta_j^{(n)}\big|$$

имеет один и то же конечный предел J, который называется интегралом Римана функции f.

Мы будем использовать два известных факта из математического анализа.

1. Любая функция $f \in \mathbf{C}_b(X)$ является интегрируемой по Риману;

2. Интеграл Римана совпадает с интегралом Лебега по мере mes_d, являющейся на самом деле равномерным распределением в X.

Мы сейчас на языке слабой сходимости легко получим результат, совсем непросто доказываемый на языке математического анализа.

Предложение 2.1. *Пусть f — непрерывная почти всюду по мере mes_d ограниченная функция. Тогда f интегрируема по Риману.*

Доказательство. Рассмотрим некоторую последовательность разбиений $\Xi_n \in \Im$ и некоторые точки $x_j^{(n)} \in \Delta_j^{(n)}$. Если $f \in \mathbf{C}_b(X)$, то при $n \to \infty$

$$\sum_{1\le j\le m_n} f\big(x_j^{(n)}\big)\big|\,\Delta_j^{(n)}\big| \to \int\limits_X f(x)\,\mathrm{mes}_d(dx) \tag{4}$$

(это следствие того, что непрерывные ограниченные функции интегрируемы по Риману и того, что интеграл Римана совпадает с интегралом Лебега).

Введем в $(X, \mathcal{B}_X)$ последовательность дискретных распределений $\mathcal{P}_n$ с помощью таблиц

$$\mathcal{P}_n: \quad \begin{pmatrix} x_1^{(n)} & \dots & x_j^{(n)} & \dots & x_{m_n}^{(n)} \\ \big|\,\Delta_1^{(n)}\big| & \dots & \big|\,\Delta_j^{(n)}\big| & \dots & \big|\,\Delta_{m_n}^{(n)}\big| \end{pmatrix}.$$

Тогда. очевидно,

$$\sum_{1\le j\le m_n} f\big(x_j^{(n)}\big)\big|\,\Delta_j^{(n)}\big| = \int\limits_X f\,d\mathcal{P}_n,$$

и (4) попросту означает слабую сходимость $\mathcal{P}_n \Rightarrow \mathcal{P} = \mathrm{mes}_d$. Поэтому по следствию 2.2 сходимость (4) имеет место и для ограниченных п.в. непрерывных по мере mes_d функций f. $\square$

Замечание 2.3. *1. Конечно, простота доказательства предложения 2.1 — это следствие того, что мы уже достаточно глубоко "вошли" в теорию слабой сходимости распределений. 2. Несложно доказать от противного и обратное утверждение: если множество D_f разрывов функции f имеет положительную меру Лебега mes_d, или есть f не является ограниченной, то f не интегрируема по Риману.*

То есть на самом деле множество интрегрируемых по Риману функций совпадает с множеством ограниченных непрерывных почти всюду по мере Лебега функций.

3 Метризация сходимости распределений

Когда мы говорим о том, что последовательности каких-то объектов Λ_n (точек, функций, распределений и пр.) сходится к другому объекту Λ, возникает естественное желание узнать, насколько близок объект Λ_{n_0} с большим номером n_0 к предельному объекту Λ. Иначе говоря, всегда интересен вопрос о скорости сходимости Λ_n к Λ.

Стандартный способ решения этого вопроса — введение метрики ρ на множестве объектов, которому принадлежат Λ_n и Λ.

В случае, когда это множество — числовая прямая $\mathbb{R}$ (то есть когда Λ_n и Λ — числа), большой проблемы здесь не возникает: нужно всего лишь определить $\rho(\Lambda_n, \Lambda) = |\Lambda_n - \Lambda|$. (Хотя и здесь возникает некоторая неопределенность: почему бы, скажем, не умножить это ρ на 2?)

В $\mathbb{R}^d$ ситуация более сложная. Если сходимость векторов в $\mathbb{R}^d$ понимается как поточечная сходимость их координат, то существует много топологически эквивалентных метрик, порождающих эту сходимость (например, евклидова, равномерная, эллиптическая и пр.). И в практических задачах нужно, в принципе, задуматься, какая из этих метрик более адекватна решаемой проблеме.

Коснемся этого вопроса в случае, когда Λ_n и Λ — распределения (конечно, мы вернемся при этом к старым обозначениям $\mathcal{P}_n$ и $\mathcal{P}$).

3.1 О метризации слабой сходимости

Определение слабой сходимости распределений имеет ту же структуру, что и определение сходимости векторов через поточечную сходимость их координат. Иначе говоря, в определении слабой сходимости не присутствует никакая метрика.

Возникает вопрос о метризации слабой сходимости: можно ли задать на множестве распределений такую метрику, чтобы сходимость по ней была эквивалентна слабой сходимости?

Ответ, безусловно, положительный. Более того, таких метрик можно ввести много (это не удивительно). Беда состоит в том, что вид этих расстояний, как правило, далек от наглядности, поэтому они в основном используются только в теоретических построениях. Приведем одну из таких метрик — ту, определение которой тесно связано с классами функций, определяющими слабую сходимость.

Пусть $(X, \rho, \mathcal{C}, \mathcal{B})$ — измеримое метрическое пространство. Будем говорить, что функция $f : X \mapsto \mathbb{R}$ удовлетворяет условию Липшица с постоянной L, если $|f(x) - f(y)| \le L\rho(x, y)$ при любых $x, y \in X$. Ясно, что липшицевские функции непрерывны (и равномерно непрерывны).

Обозначим $\mathbf{BL} = \mathbf{BL}(X)$ множество вещественнозначных функций, определенных на X и таких, что a) $|f(x)| \le 1$ при любом x и b) f удовлетворяет условию Липшица с постоянной 1.

Конечно, $\mathbf{BL}(X) \subset \mathbf{C}_b(X)$. (Аббревиатура **BL** расшифровывается очень просто — "Bounded Lipschitz".) То, что функции по модулю ограничиваются именно единицей (и постоянная Липшица равна 1) — это просто удобные условия нормировки.

Пусть $\mathcal{P}$ и $\mathcal{Q}$ — распределения в $(X, \mathcal{B})$. Положим

$$\mathrm{dist}\,(\mathcal{P}, \mathcal{Q}) = \sup_{f \in \mathbf{BL}} \left| \int\limits_X f\, d\mathcal{P} - \int\limits_X f\, d\mathcal{Q} \right|. \tag{5}$$

Теорема 3.1. *Равенство* (5) *определяет метрику, порождающую слабую сходимость распределений в* $(X, \rho, \mathcal{C}, \mathcal{B})$.

Мы не будем приводить доказательство теоремы 3.1 (его краткое изложение можно найти в [2, гл.3 §7 п.8 стр. 450], где используется чуть-чуть другое определение множества **BL**, но на доказательстве это не сказывается), а только немного обсудим его ход.

Для доказательства теоремы 3.1 нужно доказать три утверждения:
1. *Правая часть* (5) *задает некоторую метрику.* Здесь, конечно, основное — проверка неравенства треугольника.
2. *Если* $\mathrm{dist}\,(\mathcal{P}_n, \mathcal{P}) \to 0$, *то* $\mathcal{P}_n \Rightarrow \mathcal{P}$. Здесь главная идея состоит в том, что, как оказывается (т.е. доказывается), множество $\mathbf{BL}$ образует класс функций, определяющих слабую сходимость. После этого утверждение становится очевидным (почему?).
3. *Если* $\mathcal{P}_n \Rightarrow \mathcal{P}$, *то* $\mathrm{dist}\,(\mathcal{P}_n, \mathcal{P}) \to 0$. Так как $\mathbf{BL}(X) \subset \mathbf{C}_b(X)$, то все проблема состоит в том, чтобы доказать, что сходимость (3) для всех $f \in \mathbf{BL}$ (она очевидна для каждой фиксированной f) является равномерной по $f \in \mathbf{BL}$.

Приведем простейший пример использования метрики (5). Пусть на одном вероятностном пространстве $(\Omega, \mathcal{F}, \mathbb{P})$ заданы случайные величины ξ_n и ξ со значениями в метрическом пространстве (X, ρ) и распределениями $\mathcal{P}_n$ и $\mathcal{P}$. Тогда (5) записывается как

$$\mathrm{dist}\,(\mathcal{P}_n, \mathcal{P}) = \sup_{f \in \mathbf{BL}} \left|\mathbb{E}f(\xi_n) - \mathbb{E}f(\xi)\right| = \sup_{f \in \mathbf{BL}} \left|\mathbb{E}\big(f(\xi_n) - f(\xi)\big)\right|$$

(сама по себе эта формула тоже проясняет суть дела). Отсюда, поскольку f — липшицевская функция с постоянной 1, получаем, что

$$\mathrm{dist}\,(\mathcal{P}_n, \mathcal{P}) \le \sup_{f \in \mathbf{BL}} \mathbb{E}\left|f(\xi_n) - f(\xi)\right| \le \mathbb{E}\rho(\xi_n, \xi).$$

Таким образом, из сходимости $\mathbb{E}\rho(\xi_n, \xi) \to 0$ следует сходимость $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$. Если $X = \mathbb{R}$ со стандартной метрикой $\rho(x, y) = |x - y|$, то это хорошо известный факт: из сходимости случайных величин в $\mathbb{L}^1$ следует слабая сходимость их распределений. Здесь же получается, что аналог этого факта верен и в произвольном метрическом пространстве (например, для случайных векторов).

Приведем еще один способ метризации слабой сходимости — на этот раз для распределений в $\mathbb{R}^d$. Пусть $\mathcal{P}$ и $\mathcal{Q}$ — распределения в $\mathbb{R}^d$ с характеристическими функциями $\varphi^{(\mathcal{P})}$ и $\varphi^{(\mathcal{Q})}$. Зафиксируем непрерывную положительную функцию $w : \mathbb{R}^d \mapsto \mathbb{R}$, сходящуюся к нулю при $\mathbb{R}^d \ni t \to \infty$, и положим

$$\mathrm{dist}_w\,(\mathcal{P}, \mathcal{Q}) = \sup_{t \in \mathbb{R}^d} \left|\varphi^{(\mathcal{P})}(t) - \varphi^{(\mathcal{Q})}(t)\right| w(t). \tag{6}$$

Теорема 3.2. *Равенство* (6) *определяет метрику, порождающую слабую сходимость распределений в* $\mathbb{R}^d$.

Снова прокомментируем формулу (6) и доказательство теоремы 3.2. В целом оно такое же, как у теоремы 3.1, и тоже основано на классе функций, определяющих слабую сходимость, только не на классе ограниченных липшицевских функций, а на классе тригонометрических полиномов (иначе говоря, на теореме непрерывности, которую, конечно, уже не нужно специально доказывать). Эта одна из причин, почему теорема 3.2 проще, чем теорема 3.1.

Вторая причина состоит в том, что супремум в правой части (6) достигается на некотором конкретном векторе t_0 (из-за того, что функция $\left|\varphi^{(\mathcal{P})}(t) - \varphi^{(\mathcal{Q})}(t)\right| w(t)$ непрерывна и стремится к нулю на бесконечности). Отсюда, в частности, легко выводится неравенство треугольника для $\mathrm{dist}_w\,(\mathcal{P}, \mathcal{Q})$.

Кроме этого факта, роль весовой функции $\omega(t)$ состоит еще и в следующем. Слабая сходимость распределений эквивалентна поточечной сходимости характеристических функций, а из этой поточечной сходимости следует их равномерная сходимость только на (любых) конечных интервалах, а не на всем $\mathbb{R}^d$ (заодно — придумайте пример, когда при $d = 1$ есть слабая сходимость распределений, но нет равномерной сходимости в $\mathbb{R}$ их характеристических функций).

Поскольку расстояние между распределениями не должно зависеть от точек, в которых рассматриваются характеристические функции, приходится вводить весовую функцию ω, “убивающую” характеристические функции на бесконечности.

Кроме того, здесь мы получаем целый класс метрик, зависящих от функции ω, и можем в конкретных ситуациях выбирать наиболее удобную из них.

3.2 Сходимость по вариации. Теорема Шеффе

Существует еще одна популярная метрика, заданная на множестве распределений. Эта метрика порождает так называемую "сходимость по вариации", которая отличается от слабой сходимости. Дадим соответствующие определения.

На измеримом пространстве $(X, \mathcal{X})$ рассмотрим множество $\mathcal{S}$ конечных зарядов — знакопеременных конечных мер. (Сразу же отметим, что в X не предполагается существование никакой метрики или топологии, и поэтому σ-алгебра $\mathcal{X}$ совсем не обязана быть борелевской.)

Очевидно, множество $\mathcal{S}$ является линейным пространством (так как линейная комбинация конечных зарядов — конечный заряд). Положим для $S \in \mathcal{S}$

$$\mathrm{Var}(S) = \sup_{A \in \mathcal{X}} |S(A)|.$$

Величина $\mathrm{Var}(S)$ называется полной вариацией заряда S.

Легко доказать, что $(\mathcal{S}, \mathrm{Var})$ является нормированным пространством, то есть что $\mathrm{Var}(\,\cdot\,)$ задает норму в линейном пространстве $\mathcal{S}$. Поэтому иногда используют обозначение $\|S\|_{\mathrm{var}}$ вместо $\mathrm{Var}(S)$.

Как обычно, любая норма порождает метрику, и на множестве зарядов возникает расстояние $\mathrm{dist}_{\mathrm{var}}(S_1, S_2) = \mathrm{Var}(S_1 - S_2)$. Оно называется расстоянием по вариации, а сходимость по этой метрике — сходимостью по вариации.

Поскольку множество Π распределений на измеримом пространстве $(X, \mathcal{X})$ является подмножеством множества $\mathcal{S}$ конечных зарядов, то метрика $\mathrm{dist}_{\mathrm{var}}$ автоматически переносится на Π (отметим, что Π является выпуклым множеством, но не линейным пространством).

Таким образом, мы приходим к следующему определению.

Определение 3.1. 1. *Расстоянием по вариации между распределениями $\mathcal{P}$ и $\mathcal{Q}$, заданными на измеримом пространстве $(X, \mathcal{X})$ называется*

$$\mathrm{dist}_{\mathrm{var}}(\mathcal{P}, \mathcal{Q}) = \sup_{A \in \mathcal{X}} \big|\mathcal{P}(A) - \mathcal{Q}(A)\big|.$$

2. *Последовательность распределений $\mathcal{P}_n$ сходится по вариации к распределению $\mathcal{P}$* (пишется $\mathcal{P}_n \overset{\mathrm{var}}{\to} \mathcal{P}$), если

$$\sup_{A \in \mathcal{X}} \big|\mathcal{P}_n(A) - \mathcal{P}(A)\big| \to 0.$$

Замечание 3.1. *1. Ясно, что $0 \le \mathrm{dist}_{\mathrm{var}}(\mathcal{P}, \mathcal{Q}) \le 1$ для любых распределений $\mathcal{P}$ и $\mathcal{Q}$.*
2. Пусть σ-алгебра $\mathcal{X}$ содержит все одноточечные множества. Если $\mathcal{P}(\{x\}) = 0$ для любого $x \in X$, а $\mathcal{Q}$ сосредоточено на не более чем счетном множестве X_0 (иными словами, $\mathcal{P}$ является непрерывным распределением, а $\mathcal{Q}$ — дискретным распределением), то $\mathrm{dist}_{\mathrm{var}}(\mathcal{P}, \mathcal{Q}) = 1$ (для проверки этого достаточно рассмотреть $\mathcal{P}(X_0)$ и $\mathcal{Q}(X_0)$). Отсюда следует, что последовательность непрерывных распределений не может сходиться по вариации к дискретному распределению и наоборот.

Следующее утверждение дает описание сходимости по вариации в терминах интегралов (а не самих распределений).

Пусть S — конечный заряд. Хорошо известно, что $S = S^+ - S^-$, где S^+ и S^- — (положительные) меры, имеющие дизъюнктные носители. Иначе говоря, существует такое $E \in \mathcal{X}$ (носитель меры S^+), что $S^-(E) = 0$ и $S^+(E^{\mathrm{c}}) = 0$.

Обозначим $|S| = S^+ + S^-$. Ясно, что $|S|$ является (положительной) мерой, которую иногда называют вариацией заряда S (не нужно путать меру $|S|$ с полной вариацией $\mathrm{Var}(S)$, которая является числом).

Обозначим $\mathbb{F}_M$ множество измеримых функций, ограниченных по модулю постоянной M.

Предложение 3.1. *Последовательность распределений $\mathcal{P}_n$ сходится по вариации к распределению $\mathcal{P}$ тогда и только тогда, когда для любого $M > 0$*

$$\sup_{f\in\mathbb{F}_M}\left|\int\limits_X f\,d\mathcal{P}_n - \int\limits_X f\,d\mathcal{P}\right| \to 0. \tag{7}$$

Доказательство. Достаточность условия (7) очевидна (нужно взять $M = 1$ и рассмотреть в качестве функций f индикаторы всевозможных измеримых множеств). Докажем необходимость этого условия.

Рассмотрим распределения $\mathcal{P}$ и $\mathcal{Q}$ и обозначим $S = \mathcal{P} - \mathcal{Q}$. Тогда $S(X) = 0$ и, обозначая E носитель меры S^+, получим, что $S^+(E) = S^-(E^c)$, $|S|(X) = S^+(E) + S^-(E^c) = 2S^+(E)$ и

$$\begin{aligned}\mathrm{Var}(S) = \sup_A |S(A)| = \sup_A |S^+(A) - S^-(A)| = \sup_A |S^+(A\cap E) - S^-(A\cap E^c)| \le \\ \le \sup_A \max(S^+(A\cap E), S^-(A\cap E^c)) \le \max(S^+(E), S^-(E^c)) = S^+(E) = S(E).\end{aligned}$$

Таким образом, $2\mathrm{Var}(S) = |S|(X)$.

Разложим теперь функцию $f \in \mathbb{F}_M$ на положительную и отрицательную составляющую: $f = f^+ - f^-$. Тогда $|f| = f^+ + f^-$ и

$$\begin{aligned}\left|\int f dS\right| = \left|\int f^+dS^+ - \int f^-dS^+ - \int f^+dS^- + \int f^-dS^-\right| \le \\ \le \int f^+dS^+ + \int f^-dS^+ + \int f^+dS^- + \int f^-dS^- = \int |f|dS^+ + \int |f|dS^- = \\ = \int |f|d|S| \le M\,|S|(X) = 2M\,\mathrm{Var}(S).\end{aligned}$$

Отсюда сразу же следует, что

$$\sup_{f\in\mathbb{F}_M}\left|\int fd\mathcal{P}_n - \int fd\mathcal{P}\right| = \sup_{f\in\mathbb{F}_M}\left|\int fd(\mathcal{P}_n - \mathcal{P})\right| \le 2M\,\mathrm{Var}(\mathcal{P}_n - \mathcal{P}) \to 0$$

при $n \to \infty$. □

Смысл расстояния по вариации на множестве распределений проясняет следующее утверждение, носящее название теоремы Шеффе. Рассмотрим λ — (положительную) меру на $(X, \mathcal{X})$.

Теорема 3.3. *1. Пусть $\mathcal{P}$ и $\mathcal{Q}$ — распределения, заданные на $(X, \mathcal{X})$ и предположим, что у распределений $\mathcal{P}$ и $\mathcal{Q}$ есть плотности p и q относительно меры λ. Тогда*

$$\mathrm{dist}_{\mathrm{var}}(\mathcal{P}, \mathcal{Q}) = \frac{1}{2}\int\limits_X |p - q|\,d\lambda. \tag{8}$$

2. Пусть распределения $\mathcal{P}_n$ и $\mathcal{P}$ имеют плотности p_n и p относительно меры λ. Если $p_n \to p$ λ-п.в., то $\mathcal{P}_n \stackrel{\mathrm{var}}{\to} \mathcal{P}$.

Доказательство. См. [1, стр. 306]. □

Прокоммментируем утверждение теоремы Шеффе.

Замечание 3.2. *1. Слова "у распределений $\mathcal{P}$ и $\mathcal{Q}$ есть плотности p и q относительно меры λ" — это просто несколько жаргонный эквивалент более точному "у распределений $\mathcal{P}$ и $\mathcal{Q}$ существуют производные Радона-Никодима p и q относительно меры λ".*
2. Равенство (8) *означает, что расстояние по вариации между распределениями, абсолютно*

непрерывными относительно одной и той же меры λ — это просто (с точностью до множителя $1/2$) расстояние в $\mathbb{L}^1(d\lambda)$ между соответствующими плотностями.
3. Существуют два стандартных частных случая применения теоремы Шеффе. Первый — когда $X = \mathbb{R}^d$ и λ — d-мерная мера Лебега mes_d. Тогда p_n и p — обычные плотности распределения и теорема Шеффе утверждает, что из сходимости p_n к p (почти всюду по мере Лебега) следует сходимость по вариации распределений.

Второй — случай дискретных распределений. Пусть (например) $X = \{0, 1, \dots\}$, а σ-алгебра $\mathcal{X}$ — σ-алгебра всех подмножеств множества X. Рассмотрим дискретные распределения $\mathcal{P}_n$ и $\mathcal{P}$, задаваемые вероятностями $\mathcal{P}_n(\{i\}) = p_n(i)$ и $\mathcal{P}(\{i\}) = p(i)$, $i = 0, 1, \dots$. Пусть λ — считающая мера на X, то есть $\lambda(\{i\}) = 1$ для любого i. Тогда, очевидно, для любого $A \subset X$

$$\mathcal{P}(A) = \sum_{i\in A} p(i) = \int_X p\, d\lambda \quad и \quad \mathcal{P}_n(A) = \sum_{i\in A} p_n(i) = \int_X p_n d\lambda,$$

где функции p и p_n принимают соответсвенно значения $p(i)$ и $p_n(i)$ в точке i. Поэтому, если $p_n(i)$ сходится к $p(i)$ для любого i (слова "почти всюду" здесь можно опустить, так как $\lambda(A) = 0$ только если $A = \varnothing$), то $\mathcal{P}_n$ сходится по вариации к $\mathcal{P}$.

Закончим эту тему замечанием о связи между слабой сходимостью и сходимостью по вариации. При определении сходимости по вариации предполагалось, что распределения $\mathcal{P}_n$ и $\mathcal{P}$ заданы на абстрактном измеримом пространстве $(X, \mathcal{X})$. Поэтому в общем случае нельзя даже ставить вопрос о слабой сходимости $\mathcal{P}_n$ к $\mathcal{P}$.

Пусть теперь множество X оснащено метрикой ρ, а $\mathcal{X} = \mathcal{B}_X$ — σ-алгебра борелевских подмножеств X. Тогда из сходимости $\mathcal{P}_n \overset{\mathrm{var}}{\to} \mathcal{P}$ будет следовать слабая сходимость $\mathcal{P}_n \Rightarrow \mathcal{P}$. Действительно, согласно предложению 3.1 из сходимости по вариации следует сходимость

$$\int_X f d\mathcal{P}_n \to \int_X f d\mathcal{P}$$

для любой ограниченной измеримой функции f, в то время как по определению слабой сходимости такие соотношения обязаны иметь место только непрерывных ограниченных функций.

Можно привести и другой вариант доказательства. По основной теореме о слабой сходимости условие $\mathcal{P}_n(A) \to \mathcal{P}(A)$, выполненное для всех множествх непрерывности A, эквивалентно сходимости $\mathcal{P}_n \Rightarrow \mathcal{P}$. Конечно, это условие непосредственно следует из определения сходимости по вариации, где сходимость $\mathcal{P}_n(A) \to \mathcal{P}(A)$ должна выполняться для любого измеримого A, да еще и раномерно по A.

В то же время из слабой сходимости распределений сходимость по вариации, вообще говоря, не следует (см. второй пункт замечания 3.1).

Список литературы

[1] П. Биллингсли, Сходимость вероятностных мер, М., Наука, 1977.

[2] А.Н. Ширяев, Вероятность. М., МЦНМО, 2004.