

Предварительные сведения. Материал к курсу «Статистика случайных процессов»

В.В. Некруткин

кафедра статистического моделирования <http://statmod.ru>, матмех СПбГУ
Материал к специальному курсу, 2016 г.

Содержание

1	Многомерные гауссовские распределения	3
1.1	Общий случай	3
1.2	Невырожденный случай	5
1.3	Вырожденный случай	6
2	Слабая сходимость распределений	7
2.1	О сходимости распределений	7
2.2	О расстоянии по вариации	7
2.3	О слабой сходимости	10
2.4	Слабая сходимость и отображения	13
2.5	О сходимости по вероятности к константе	14
3	Центральная предельная теорема для независимых слагаемых	16
3.1	Схема серий и теорема Леви	16
3.2	Отказ от одинаковой распределенности: теорема Линдеберга-Феллера	18
3.3	Многомерная ЦПТ	23
4	О сходимости к гауссовскому распределению	25
4.1	Критерий сходимости	25
4.2	Сохранение предельной гауссовости при отображениях	26
5	Вещественные стационарные последовательности	28
5.1	Стационарные в широком смысле процессы	28
5.2	Стационарные в узком смысле последовательности.	33
5.3	Предельная дисперсия выборочного среднего стационарной последовательности	34
5.4	ЦПТ для линейного процесса	37
6	Процессы авторегрессии	39
6.1	Основные понятия и свойства.	39
6.2	Реализуемые процессы авторегрессии.	40
6.3	Авторегрессия первого порядка и марковские гауссовские стационарные последовательности.	42
7	Авторегрессионные продолжения	44
7.1	Уравнения Юла-Уолкера	44
7.2	Теорема об авторегрессионном продолжении	46
7.3	Спектральные плотности авторегрессионных продолжений.	47
	Список литературы	50

В этом файле собран материал, на который будет опираться курс «Статистика случайных процессов», читаемый в 10-м семестре 5-го курса (магистры) на кафедре статистического моделирования. Весь этот материал уже прочитан на 3-4 курсах (бакалавриат) специализации СМ-СМ этой кафедры и приводится здесь для удобства слушателей.

В принципе, можно было бы ограничиться только определениями и формулировками соответствующих утверждений, но, тем не менее, сюда включены некоторые (не слишком громоздкие) доказательства и примеры.

1 Многомерные гауссовские распределения

Определение 1.1. Говорят, что случайный вектор $\bar{\xi} = (\xi_1, \dots, \xi_d)^T \in \mathbb{R}^d$ имеет (d -мерное) гауссовское распределение, если для любого $\bar{t} \in \mathbb{R}^d$ случайная величина $(\bar{\xi}, \bar{t})$ имеет нормальное распределение¹ с некоторыми параметрами a_t и σ_t^2 .

Нужные нам свойства многомерных гауссовских распределений объединены в три леммы.

1.1 Общий случай

Лемма 1. (Характеристическая функция гауссовского вектора)

Пусть $\bar{\xi} = (\xi_1, \dots, \xi_d)^T \in \mathbb{R}^d$ имеет гауссовское распределение. Тогда характеристическая функция $\varphi_{\bar{\xi}}(\bar{t})$ случайного вектора $\bar{\xi}$ имеет вид

$$\varphi_{\bar{\xi}}(\bar{t}) = \exp(i(\bar{a}, \bar{t})) \exp(-(\Sigma \bar{t}, \bar{t})/2), \quad (1.1)$$

где $\bar{a} = \mathbb{E}\bar{\xi}$ и Σ — ковариационная матрица вектора $\bar{\xi}$.

Доказательство. Прежде всего заметим, что компоненты гауссовского вектора обладают всеми моментами. Действительно, взяв j -ый орт в качестве $\bar{t}$, мы получим, что ξ_j имеет нормальное распределение. Поэтому мы можем говорить о среднем $\bar{a}$ и ковариационной матрице Σ вектора $\bar{\xi}$.

Сосчитаем теперь среднее a_t и дисперсию σ_t^2 скалярного произведения $(\bar{\xi}, \bar{t}) = \bar{t}^T \bar{\xi}$. Ясно, что $a_t = (\bar{a}, \bar{t})$. Для вычисления дисперсии достаточно рассмотреть случай $\bar{a} = \mathbf{0}$. Тогда

$$\sigma_t^2 = \mathbb{E}(\bar{\xi}, \bar{t})^2 = \mathbb{E}(\bar{t}^T \bar{\xi} \bar{\xi}^T \bar{t}) = \bar{t}^T \mathbb{E}(\bar{\xi} \bar{\xi}^T) \bar{t} = \bar{t}^T \Sigma \bar{t} = (\Sigma \bar{t}, \bar{t}).$$

Следовательно, характеристическая функция $\psi_t(\tau)$ скалярного произведения $(\bar{\xi}, \bar{t})$ равна

$$\psi_t(\tau) = \mathbb{E} \exp(i(\bar{\xi}, \bar{t})\tau) = \exp(i(\bar{a}, \bar{t})\tau) \exp(-(\Sigma \bar{t}, \bar{t})\tau^2/2) \quad (1.2)$$

Полагая $\tau = 1$ в (1.2), приходим к нужному нам результату. $\square$

Таким образом, распределение d -мерного гауссовского случайного вектора $\bar{\xi}$ полностью определяется его средним $\bar{a}$ и ковариационной матрицей Σ . Поэтому мы будем использовать запись $\bar{\xi} \in N_d(\bar{a}, \Sigma)$. Отметим, что невырожденность ковариационной матрица при этом не предполагается.

Следствие 1. Для того, чтобы компоненты гауссовского вектора $\bar{\xi} \in N_d(\bar{a}, \Sigma)$ были независимы, необходимо и достаточно, чтобы они были попарно некоррелированы.²

Доказательство. Случайный вектор $\bar{\xi} = (\xi_1, \dots, \xi_d)^T$ имеет независимые компоненты тогда и только тогда, когда его характеристическая функция $\varphi_{\bar{\xi}}(t_1, \dots, t_d)$ распадается в произведение некоторых функций вида $\psi_j(t_j)$. Поскольку попарная некоррелированность компонент этого вектора эквивалентна диагональности его ковариационной матрицы, то утверждение следует из формулы (1.1). $\square$

¹Напомним, что по нашему соглашению постоянная a имеет нормальное распределение $N(a, 0)$.

²Здесь нам удобно считать, что случайные величины ξ и η некоррелированы, если $\text{Cov}(\xi, \eta) = 0$.

Лемма 2. (Сохранение гауссовости при аффинных преобразованиях. Существование.)

1. Если $\bar{\xi} \in \mathbb{R}^d$, $\bar{\xi} \in N_d(\bar{a}, \Sigma)$, $\bar{b} \in \mathbb{R}^k$ и матрица A имеет d столбцов и k строк, то $\bar{\eta} \stackrel{\text{def}}{=} A\bar{\xi} + \bar{b} \in N_k(\bar{b} + A\bar{a}, A\Sigma A^T)$
2. Каковы бы ни были вектор $\bar{a} \in \mathbb{R}^d$ и $d \times d$ неотрицательно определенная матрица Σ , существует такое вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ и такая случайная величина $\bar{\xi} : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}^d, \mathcal{B}_d)$, что $\bar{\xi} \in N_d(\bar{a}, \Sigma)$.

Доказательство. 1. Гауссовость случайного вектора $\bar{\eta}$ следует из цепочки равенств

$$(\bar{\eta}, \bar{s}) = (A\bar{\xi} + \bar{b}, \bar{s}) = (\bar{b}, \bar{s}) + (\bar{\xi}, A^T \bar{s})$$

и Определения 1.1. Вычисляя среднее и ковариационную матрицу вектора $\bar{\eta}$, получаем окончательный результат.

2. Прежде всего, нам достаточно рассматривать случай $\bar{a} = \mathbf{0}$. Пусть $\mu_1 \geq \dots \mu_d \geq 0$ — собственные числа матрицы Σ с учетом их кратности. Обозначим $U_1, \dots, U_d$ соответствующую ортонормированную систему собственных векторов и положим $U = [U_1 : \dots : U_d]$. Ясно, что матрица U является ортонормированной, причем $\Sigma = U\Lambda U^T$, где матрица Λ диагональна с элементами $\lambda_{jj} = \mu_j$.

Рассмотрим теперь случайный вектор $\bar{\eta} = (\eta_1, \dots, \eta_d)^T$ такой, что $\eta_j \in N(0, \mu_j)$ и потребуем, чтобы эти компоненты были независимы. Такие случайные величины существуют.³ Более того, так как линейная комбинация независимых нормальных случайных величин имеет нормальное распределение, то вектор $\bar{\eta}$ имеет гауссовское распределение. Точнее, $\bar{\eta} \in N_d(\mathbf{0}, \Lambda)$.

Теперь, положив $\bar{\xi} = U\bar{\eta}$, получим требуемое из первого утверждения леммы. $\square$

Замечание 1. Как уже говорилось, каждая из координат ξ_i многомерного гауссовского вектора $\bar{\xi} = (\xi_1, \dots, \xi_d)^T$ имеет нормальное (возможно, вырожденное) распределение. Покажем, что обратное, вообще говоря, неверно. Для этого рассмотрим функцию

$$p(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} + \frac{1}{2\pi} e^{-(x^2+y^2)/2} \varepsilon(x) \varepsilon(y),$$

где $\varepsilon(z)$ — нечетная измеримая функция, по модулю не превосходящая единицы.

Ясно, что при этих условиях $p(x, y) \geq 0$ и

$$\int_{\mathbb{R}^2} p(x, y) dx dy = \int_{\mathbb{R}^2} \frac{1}{2\pi} e^{-(x^2+y^2)/2} dx dy = 1,$$

так что p является плотностью распределения некоторого случайного вектора $(\xi_1, \xi_2)^T$. Конечно, это распределение не является гауссовским.⁴ В то же самое время,

$$p_{\xi_1}(x) \stackrel{\text{def}}{=} \int_{\mathbb{R}} p(x, y) dy = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

и аналогичное равенство выполняется для плотности распределения случайной величины ξ_2 . Тем самым $\xi_1, \xi_2 \in N(0, 1)$.

³ Действительно, приведенное описание полностью определяет распределение случайного вектора $\bar{\eta}$ как меры на борелевской σ -алгебре в $\mathbb{R}^d$. А для любой вероятностной тройки $(\mathbb{R}^d, \mathcal{B}_d, \mathcal{P})$ существует такое вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$ и такой случайный вектор $\bar{\xi} : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}^d, \mathcal{B}_d)$, что $\mathcal{L}(\bar{\xi}) = \mathcal{P}$. Для доказательства последнего факта достаточно взять $\Omega = \mathbb{R}^d$, $\mathcal{F} = \mathcal{B}_d$, $\mathbb{P} = \mathcal{P}$ и $\bar{\xi}(\omega) = \omega$ для любого $\omega \in \Omega$.

⁴ А почему?

1.2 Невырожденный случай

Предложение 1. Пусть $\bar{\xi} \in N_d(\bar{a}, \Sigma)$. Если матрица Σ невырождена, то распределение вектора $\bar{\xi}$ абсолютно непрерывно с плотностью

$$p_{\bar{\xi}}(X) = \frac{\det \Sigma^{-1/2}}{(2\pi)^{d/2}} e^{-(\Sigma^{-1}(X - \bar{a}), X - \bar{a})/2}, \quad X \in \mathbb{R}^d. \quad (1.3)$$

Доказательство. Нам снова достаточно рассматривать случай $\bar{a} = \mathbf{0}$. Действуя так же, как в Лемме 2, рассмотрим собственные числа μ_j матрицы Σ и соответствующие матрицы U и Λ . Так как матрица Σ невырождена, то $\mu_j > 0$ при всех j .

Введем гауссовский случайный вектор $\bar{\eta} = (\eta_1, \dots, \eta_d)^T$ с независимыми компонентами и такой, что $\eta_j \in N(0, \mu_j)$. Нетрудно видеть, что плотность распределения этого вектора имеет вид

$$p_{\bar{\eta}}(Y) = \frac{\det \Lambda^{-1/2}}{(2\pi)^{d/2}} e^{-(\Lambda^{-1}Y, Y)/2}, \quad Y \in \mathbb{R}^d.$$

Положим теперь

$$\bar{\xi} = U\bar{\eta} = \sum_{j=1}^d \eta_j U_j. \quad (1.4)$$

Согласно Лемме 2, $\bar{\xi} \in N_d(\mathbf{0}, \Sigma)$. С другой стороны, так как $U^{-1} = U^T$, $\det \Sigma = \det \Lambda$ и $\Sigma^{-1} = U\Lambda^{-1}U^T$, то из [1, разд. 7.2.1, формула (7.2.1)] мы получаем, что

$$\begin{aligned} p_{\bar{\xi}}(X) &= |\det U|^{-1} p_{\bar{\eta}}(U^{-1}X) = \frac{\det \Lambda^{-1/2}}{(2\pi)^{d/2}} e^{-(\Lambda^{-1}U^{-1}X, U^{-1}X)/2} = \\ &= \frac{\det \Sigma^{-1/2}}{(2\pi)^{d/2}} e^{-(U\Lambda^{-1}U^T X, X)/2} = \frac{\det \Sigma^{-1/2}}{(2\pi)^{d/2}} e^{-(\Sigma^{-1}X, X)/2}. \end{aligned}$$

Утверждение доказано. $\square$

Рассмотрим теперь частный случай $d = 2$. Тогда Σ (в терминах дисперсий σ_1^2, σ_2^2 и коэффициента корреляции ρ) имеет вид

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Для того, чтобы эта матрица была невырождена, необходимо и достаточно, чтобы $|\rho| < 1$, причем $\det \Sigma = \sigma_1^2\sigma_2^2(1 - \rho^2)$, а

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1 - \rho^2)} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix},$$

так что при $Y = (y_1, y_2)^T$

$$(\Sigma^{-1}Y, Y) = \frac{\sigma_2^2 y_1^2 + \sigma_1^2 y_2^2 - 2\rho\sigma_1\sigma_2 y_1 y_2}{\sigma_1^2\sigma_2^2(1 - \rho^2)}.$$

Таким образом, для случая нулевого среднего

$$\begin{aligned} p_{\xi_1 \xi_2}(y_1, y_2) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} e^{-\frac{1}{2} \frac{\sigma_2^2 y_1^2 + \sigma_1^2 y_2^2 - 2\rho\sigma_1\sigma_2 y_1 y_2}{\sigma_1^2\sigma_2^2(1 - \rho^2)}} = \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} e^{-\frac{1}{2(1 - \rho^2)} \left(\frac{y_1^2}{\sigma_1^2} + \frac{y_2^2}{\sigma_2^2} - 2\rho \frac{y_1 y_2}{\sigma_1\sigma_2} \right)}. \end{aligned} \quad (1.5)$$

Если же среднее равно $(a_1, a_2)^T$, то в правой части формулы (1.5) нужно заменить y_1, y_2 на $y_1 - a_1, y_2 - a_2$.

1.3 Вырожденный случай

Как только что доказано, при условии $\det \Sigma > 0$ многомерное гауссовское распределение является абсолютно непрерывным. Пусть теперь $\text{rank } \Sigma = k < d$.

Рассмотрим снова разложение (1.4) и заметим, что при доказательстве того, что $\mathcal{L}(\bar{\xi}) = N(\mathbf{0}, \Sigma)$, никак не использовалось невырожденность Σ .⁵ Для удобства будем считать, что собственные числа матрицы Σ упорядочены по убыванию. Тогда $\mu_1 \geq \dots \geq \mu_k > 0$ и $\mu_j = 0$ при $j > k$. При этих условиях (1.4) приобретает вид

$$\bar{\xi} = \sum_{j=1}^k \eta_j U_j$$

Следовательно, случайный вектор $\bar{\xi}$, имеющий распределение $N(\mathbf{0}, \Sigma)$, оказывается сосредоточенным на k -мерном линейном подпространстве $\mathcal{L}_k$ с базисом $U_1, \dots, U_k$.⁶ Так как эта подпространство имеет нулевую d -мерную меру Лебега, то в вырожденном случае многомерное гауссовское распределение плотности не имеет.

⁵Невырожденность нужна была только для получения вида плотности этого распределения.

⁶А для распределения $N(\bar{a}, \Sigma)$ — на гиперплоскости $\mathcal{L}_k + \bar{a}$.

2 Слабая сходимость распределений

2.1 О сходимости распределений

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 метрику⁷

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

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.

2.2 О расстоянии по вариации

Определение 2.1. (*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

$$\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) = \sup_{B \in \mathcal{B}_d} |\mathcal{P}(B) - \mathcal{Q}(B)|. \quad (2.1)$$

Говорят, что последовательность $\mathcal{P}_n \in \Pi_d$ сходится по вариации⁸ 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 $\mathcal{P}_n \xrightarrow{\text{var}} \mathcal{P}$.

⁷Докажите, что это метрика.

⁸В другой терминологии — *сильно сходится*.

Замечание 2. 1. Формула (2.1) задает метрику в множестве Π_d .⁹ 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.¹⁰

Теорема 1. (*Теорема Шеффе*).

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

$$\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) = \frac{1}{2} \int_{\mathbb{R}^d} |p(x) - q(x)| dx. \quad (2.2)$$

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

$$\rho_{\text{var}}(\mathcal{P}, \mathcal{Q}) = \frac{1}{2} \sum_{\ell} |p_\ell - q_\ell|. \quad (2.3)$$

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 распределения¹¹ 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 (2.2) дает возможность вычислять расстояние по вариации. Например, нетрудно посчитать, что расстояние по вариации между распределениями $U(0, 1 + \varepsilon)$ и $U(0, 1)$ при положительных¹² ε равно $\varepsilon/(1 + \varepsilon)$, в то время как $\rho_{\text{var}}(U(0, 1), \text{Exp}(\lambda)) = e^{-\lambda}$ при $\lambda \leq 1$.¹³

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 стремится¹⁴ при $n \rightarrow \infty$ к плотности распределения стандартного нормального закона сразу же следует, что $S_n \xrightarrow{\text{var}} N(0, 1)$.¹⁵

Все это, несомненно, весьма привлекательные свойства. Рассмотрим, однако, расстояние по вариации (2.1) 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

⁹Докажите это.

¹⁰Здесь она формулируется отдельно для абсолютно непрерывных и дискретных распределений в $\mathbb{R}^d$. Более общий случай с доказательством см. в [1, разд. 8], а также в [5, с. 306] или в [2, гл. III §9 лемма 2]

¹¹Сосредоточенные на одном множестве и том же не более чем счетном множестве.

¹²А при отрицательных?

¹³А при $\lambda > 1$?

¹⁴Проверьте!

¹⁵В 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 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 $\xi_\varepsilon \equiv \varepsilon$ сходятся¹⁶ при $\varepsilon \rightarrow 0$ к случайной величине $\xi \equiv 0$, а их распределения не сходятся.¹⁷

Это представляется неестественным и, следовательно, метрика Определения 2.1 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 (доказательство см. [1, разд. 8], а также в [2, гл. III §9 лемма 1]).

Предложение 2.

$$\sup_{B \in \mathcal{B}_d} |\mathcal{P}(B) - \mathcal{Q}(B)| = \frac{1}{2} \sup_{|f| \leq 1} \left| \int_{\mathbb{R}^d} f d\mathcal{P} - \int_{\mathbb{R}^d} f d\mathcal{Q} \right|, \quad (2.4)$$

где супремум берется по всем $\mathcal{B}_d$ -измеримым функциям f , по модулю меньшим 1.¹⁸

Равенство (2.4) 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 (2.4) расстояния $\rho_{\text{var}}(\mathcal{P}, \mathcal{Q})$.

Как это естественно сделать? В множество рассматриваемых в (2.4) 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 2.2. (BL-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 BL-расстоянием между $\mathcal{P}$ и $\mathcal{Q}$ называется число

$$\rho_{\text{BL}}(\mathcal{P}, \mathcal{Q}) = \sup_{f \in \text{BL}(1,1)} \left| \int_{\mathbb{R}^d} f d\mathcal{P} - \int_{\mathbb{R}^d} f d\mathcal{Q} \right|, \quad (2.5)$$

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

Замечание 3. 1. Легко видеть, что формула (2.5) действительно определяет метрику.¹⁹

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.

¹⁶В любом разумном смысле.

¹⁷В этом примере расстояние по вариации аналогично дискретной метрике в $\mathbb{R}$.

¹⁸Исходя из равенства (2.4), 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 (2.4) приведено в [1, разд. 8]

¹⁹Покажите 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 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

$$\int_{\mathbb{R}^d} f d\mathcal{P}_n \rightarrow \int_{\mathbb{R}^d} f d\mathcal{P} \quad (2.6)$$

для любой $f \in C_b(\mathbb{R}^d)$.

Доказательство Теоремы 2 достаточно сложно²⁰ (в отличие от Теоремы 1 и Предложения 2), и поэтому саму сходимость (2.6) 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.

2.3 О слабой сходимости

Определение 2.3. (*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 (2.6) выполняется для любой функции $f \in C_b(\mathbb{R}^d)$.

Факт слабой сходимости $\mathcal{P}_n$ к $\mathcal{P}$ записывается в виде $\mathcal{P}_n \Rightarrow \mathcal{P}$.²¹

Замечание 4. Из Теоремы 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 2.3.

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 3. 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 5. 1. Второй и третий пункт Предложения 3 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.

Действительно, согласно определению сходимости по вариации сходимость $\mathcal{P}_n(B) \rightarrow \mathcal{P}(B)$ должна выполняться²² 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, $\mathcal{P}_n \Rightarrow \mathcal{Q}$.

²⁰С ним можно ознакомиться в [2, гл. III §7].

²¹Иногда пишут $\mathcal{P}_n \xrightarrow{w} \mathcal{P}$. Используются и другие обозначения.

²²Да еще и равномерно по B .

Пример 1. Обсудим результат Предложения 3 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 поточечно.²³

4. Чем же отличается точка a от других точек на прямой? Перекидывая мостик к четвертому пункту Предложения 3, мы видим, что a — это единственная точка $x \in \mathbb{R}$, для которой $\mathcal{P}(\{x\}) > 0$.

Более того, если рассмотреть всевозможные интервалы $B = \langle \alpha, \beta \rangle$,²⁴ 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 множества B . Таким образом мы приходим к иллюстрации (конечно, на простейшем примере и только для $B = \{\alpha, \beta\}$) четвертого пункта Предложения 3. 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 $(\Omega, \mathcal{F}, \mathcal{P})$ и пусть ξ_n, ξ — случайные величины, определенные на этом вероятностном пространстве. Обозначим $\mathcal{P}_n = \mathcal{L}(\xi_n)$, $\mathcal{P} = \mathcal{L}(\xi)$.

²³Более того, в этом случае сходимость равномерна, то есть $\sup_x |F_n(x) - F(x)| \rightarrow 0$. Докажите это.

²⁴Значки “ $\langle$ ” и “ $\rangle$ ” означают, что нам все равно, принадлежат концы интервала этому множеству или нет.

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 (2.6) может быть переписано как

$$Ef(\xi_n) \rightarrow Ef(\xi), \quad f \in C_b(\mathbb{R}). \quad (2.7)$$

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 $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$.²⁵ Конечно, так и есть. Проиллюстрируем это с помощью следующей диаграммы.²⁶

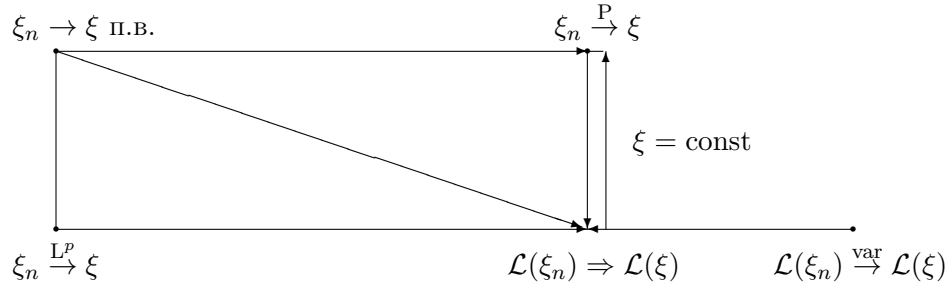


Рис. 1: Соотношения между разными видами сходимости

Рассматриваются 3 вида сходимости случайных величин: сходимость почти всюду²⁷, по вероятности²⁸ и в L^p .²⁹ Диаграмма показывает, что каждая из этих сходимостей влечет за собой слабую сходимость соответствующих распределений.³⁰ 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.

Пример 2. Путь распределения $\mathcal{P}_n$ определяются своими таблицами

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

²⁵Иногда 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 (2.7) ничего не мешает случайным величинам ξ_n и ξ быть заданным на разных вероятностных пространствах.

²⁶Доказательство этих утверждений можно найти, например, в [2].

²⁷Вероятность того, что $\xi_n \rightarrow \xi$, равна 1. Часто для такой сходимости применяю термин «почти на верное».

²⁸То есть $P(|\xi_n - \xi| > \varepsilon) \rightarrow 0$ для любого $\varepsilon > 0$.

²⁹При $d = 1$ 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.

³⁰При желании это можно считать еще одним основанием называть такую сходимость «слабой».

Докажем, что $\mathcal{P}_n \Rightarrow U(0, 1)$ четырьмя способами, три³¹ из которых связаны с Предложением 3, 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.³²

2. Если мы пользуемся вторым пунктом Предложения 3, то, взяв $x \in [0, 1]$,³³ получаем, что

$$F_n(x) = \frac{\lfloor nx \rfloor}{n} \rightarrow x = F(x).$$

3. Если использовать для доказательства слабой сходимости характеристические функции, то при любом фиксированном t и достаточно большом n ³⁴

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

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 $\xi_n = \lfloor n\xi \rfloor / n$.³⁵ Тогда при $k = 0, \dots, n-1$

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

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

2.4 Слабая сходимость и отображения

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.

³¹Как уже говорилось, четвертый пункт Предложения 3 мало пригоден для решения подобных задач.

³²Здесь мы столкнулись с тем редким случаем, когда факт слабой сходимости может быть проверен прямо по определению.

³³При других x сходимость функций распределения очевидна.

³⁴Чтобы t/n не было целым числом при фиксированном $t \neq 0$.

³⁵Тем самым ξ_n и ξ определены на одном вероятностном пространстве.

Предложение 4. Пусть $\bar{\xi}_n$ и $\bar{\xi}$ — случайные вектора со значениями в $\mathbb{R}^d$, причем

$$\mathcal{P}_n \stackrel{\text{def}}{=} \mathcal{L}(\xi_n) \Rightarrow \mathcal{P} \stackrel{\text{def}}{=} \mathcal{L}(\xi).$$

Если $f : \mathbb{R}^d \mapsto \mathbb{R}^k$ — непрерывное отображение, то

$$\mathcal{L}(f(\xi_n)) \Rightarrow \mathcal{L}(f(\xi)). \quad (2.8)$$

Доказательство. По определению нам нужно доказать, что для любой функции $g \in C_b(\mathbb{R}^k)$ имеет место сходимость

$$\mathbb{E}(g(f(\xi_n))) \rightarrow \mathbb{E}(g(f(\xi))).$$

Эти сходимость можно переписать в виде

$$\mathbb{E} g \circ f(\xi_n) \rightarrow \mathbb{E} g \circ f(\xi).$$

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

Сейчас мы сформулируем гораздо более интересное утверждение о сохранении слабой сходимости при отображениях.

Предложение 5. Обозначим G_f множество точек разрыва отображения f . Если $\mathbb{P}(\bar{\xi} \in G_f) = 0$, то имеет место слабая сходимость (2.8).

Таким образом, слабая сходимость сохраняется не только при непрерывных, но и при непрерывных почти всюду (относительно предельной меры $\mathbb{P}$) отображениях.

2.5 О сходимости по вероятности к константе

Как уже говорилось, сходимость по вероятности последовательности случайных величин ξ_n к случайной величине ξ означает, что для любого $\varepsilon > 0$ вероятность $\mathbb{P}(|\xi_n - \xi| > \varepsilon)$ сходится к нулю при $n \rightarrow \infty$. Обозначается это факт как $\xi_n \xrightarrow{\mathbb{P}} \xi$.

Аналогичным образом дается определение сходимости по вероятности $\bar{\xi}_n \xrightarrow{\mathbb{P}} \bar{\xi}$ для случайных векторов $\bar{\xi}_n$ и $\bar{\xi}$, только вместо модуля $|\xi_n - \xi|$ берется евклидова норма $\|\bar{\xi}_n - \bar{\xi}\|$.

Теорема 3. Пусть $\xi_n, \xi : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}^d, \mathcal{B}_d)$, $\eta_n : (\Omega, \mathcal{F}) \mapsto (\mathbb{R}^k, \mathcal{B}_k)$ и $\theta \in \mathbb{R}^k$. Если $\mathcal{L}(\xi_n) \Rightarrow \mathcal{L}(\xi)$ и $\eta_n \xrightarrow{\mathbb{P}} \theta$, то а) $\mathcal{L}((\xi_n, \eta_n)) \Rightarrow \mathcal{L}((\xi, \theta))$;
б) для любой непрерывного отображения $f : \mathbb{R}^d \times \mathbb{R}^k \mapsto \mathbb{R}^\ell$ имеет место сходимость $\mathcal{L}(f(\xi_n, \eta_n)) \Rightarrow \mathcal{L}(f(\xi, \theta))$.

Следствие 2. Вот простые (и используемые наиболее часто) примеры применения Теоремы 3.

1. Аддитивность.

Если в условиях Теоремы 3 θ — 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.³⁶

Скажем, пусть $\beta_1, \dots, \beta_n$ — независимые одинаково распределенные случайные величины со средним θ и единичной дисперсией. Обозначим (как принято в статистике) $\bar{x}_n$ среднее арифметическое β_i и

³⁶Обратите внимание, что здесь ничего не говорится о зависимости или независимости случайных векторов ξ_n и η_n ! Если бы они были независимы, то техника характеристических функция позволила бы доказать этот факт очень просто. Действительно?

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 η имеет распределение Коши, то слабый предел последовательности распределений $\mathcal{L}(\xi_n + \eta_n)$ не изменится,³⁷ несмотря на то, что у случайной величины ξ_n среднее и дисперсия совпадают с параметрами предельного распределения $N(0, 1)$, а математического ожидания у случайной величины $\xi_n + \eta_n$ не существует.

2. Мультипликативность.

Если в условиях Теоремы 3 $k = 1$, 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 β_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 $\bar{x}_n^2 \xrightarrow{P} 0$.

³⁷Да?

3 Центральная предельная теорема для независимых слагаемых

3.1 Схема серий и теорема Леви

Предельные теоремы несколько условно можно разделить на две категории. Одна из них имеет дело с последовательностью случайных величин $\xi_1, \dots, \xi_n, \dots$. Затем рассматривается последовательность измеримых функций³⁸ $\varphi_n : \mathbb{R}^n \mapsto \mathbb{R}$ и изучается предельное поведение случайных величин $\varphi_n(\xi_1, \dots, \xi_n)$. Именно так выглядит, например, предельная теорема для минимума из n независимых равномерно распределенных на $[0, 1]$ случайных величин. А именно, если $\xi_1, \dots, \xi_n$ независимы, причем $\xi_i \in U(0, 1)$, и если $\varphi_n(x_1, \dots, x_n) = n \min(x_1, \dots, x_n)$, то при $n \rightarrow \infty$

$$\mathcal{L}(\varphi_n(\xi_1, \dots, \xi_n)) \Rightarrow \text{EXP}(1).$$

Другая конструкция (так называемая *схема серий*³⁹) устроена несколько иначе. При любом $n \geq 1$ рассматриваются случайные величины $\xi_{1n}, \dots, \xi_{nn}$ и снова некоторые функции $\varphi_n : \mathbb{R}^n \mapsto \mathbb{R}$. А изучается предельное поведение случайных величин $\varphi_n(\xi_{1n}, \dots, \xi_{nn})$.

Наиболее известным примером предельной теоремы в схеме серий является классическая теорема Пуассона, где в n -й серии случайные величины $\xi_{1n}, \dots, \xi_{nn}$ считаются независимыми и имеющими одинаковое распределение $\text{Ber}(p_n)$, причем $np_n \rightarrow \lambda > 0$, а $\varphi_n(x_1, \dots, x_n) = x_1 + \dots + x_n$. Тогда

$$\mathcal{L}(\varphi_n(\xi_{1n}, \dots, \xi_{nn})) \xrightarrow{\text{var}} \Pi(\lambda).$$

Конечно, предельные теоремы для последовательности случайных величин являются частными случаями предельных теорем в схеме серий: для того, чтобы убедиться в этом, достаточно положить в схеме серий $\xi_{in} = \xi_i$ для любого $n \geq i$. Предельная теорема Пуассона показывает, что обратное, вообще говоря, неверно. Суть разницы⁴⁰ состоит в том, что здесь не делаются никаких предположений о связи случайных величин между сериями (то есть при различных n), в теореме Пуассона независимость случайных величин необходима только внутри каждой серии.

Схемы серий (в их детерминированных вариантах) характерны для многих задач вычислительной математики, когда переход от параметра n вычислительной схемы к параметру $n + 1$ требует совершенно новых расчетов. Простейшим примером является метод прямоугольников с равноотстоящими узлами при вычислении интеграла, заданного на конечном промежутке, если в этом примере считать n числом узлов квадратурной формулы.

Перейдем теперь к центральной предельной теореме.

Определение 3.1. (*ЦПТ в схеме серий.*)

Пусть для любого $n \geq 1$ имеются случайные величины $\xi_{1n}, \dots, \xi_{nn}$, имеющие конечные вторые моменты. Обозначим $S_n = \xi_{1n} + \dots + \xi_{nn}$ и предположим, что $B_n^2 \stackrel{\text{def}}{=} \text{DS}_n > 0$ при достаточно больших n . Если при

$$\mathcal{L}\left(\frac{S_n - \text{ES}_n}{B_n}\right) \Rightarrow N(0, 1), \quad (3.1)$$

то говорят, что схема серий $\{\xi_{n1}, \dots, \xi_{nn}, n \geq 1\}$ удовлетворяет *центральной предельной теореме* (стандартная аббревиатура — ЦПТ).

Пусть у нас есть последовательность случайных величин $\xi_1, \dots, \xi_n, \dots$. Положим $\xi_{in} = \xi_i$. Тогда определение ЦПТ в схеме серий автоматически переходит в определение «обычной» ЦПТ.

³⁸ Вообще говоря, функция φ_n может быть определена на всем $\mathbb{R}^n$, а на некотором его измеримом подмножестве. Здесь для простоты мы игнорируем это обстоятельство.

³⁹ В англоязычной литературе — «triangular array».

⁴⁰ В этом примере и вообще.

Определение 3.2. (ЦПТ для последовательностей случайных величин.)

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, если

$$\mathcal{L}\left(\frac{\xi_1 + \dots + \xi_n - (a_1 + \dots + a_n)}{B_n}\right) \Rightarrow N(0, 1), \quad (3.2)$$

где $B_n^2 = D(\xi_1 + \dots + \xi_n)$.⁴¹

Структура формулы (3.2) (как и формулы (3.1)) 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 (3.2).

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

Замечание 6. Легко видеть, что предельные соотношения (3.1) и (3.2) 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 $\mathbb{E}\xi_k = 0$.

Сформулируем и докажем теорему Поля Леви в схеме серий.

Теорема 4. (Теорема П. Леви в схеме серий).

Пусть для каждого $n \geq 1$ случайные величины $\xi_{1n}, \dots, \xi_{nn}$ независимы. Если все ξ_{in} имеют одинаковое распределение с математическим ожиданием a и ненулевой дисперсией σ^2 , то

$$\mathcal{L}\left(\frac{\xi_{1n} + \dots + \xi_{nn} - na}{\sigma\sqrt{n}}\right) \Rightarrow N(0, 1).$$

Доказательство. Как всегда, не умаляя общности, можно считать, что $a = 0$. Пусть $\varphi(t)$ — характеристическая функция ξ_{in} , обозначим

$$\eta_n = \frac{\xi_{1n} + \dots + \xi_{nn}}{\sigma\sqrt{n}}$$

и φ_n — характеристическую функцию η_n . Тогда $\varphi(t) = 1 - \sigma^2 t^2/2 + o(t^2)$ при $t \rightarrow \infty$ и поэтому для фиксированного t и $n \rightarrow \infty$

$$\varphi_n(t) = \varphi^n(t/\sigma\sqrt{n}) = (1 - t^2/2n + o(n^{-1}))^n \rightarrow e^{-t^2/2},$$

что и доказывает утверждение. $\square$

Конечно, это доказательство ничуть не отличается от доказательства «стандартной» теоремы П. Леви для последовательности случайных величин. Тем не менее мы получим последнюю как следствие Теоремы 4.

Следствие 3. (Теорема П. Леви для последовательности случайных величин).

Пусть случайные величины $\xi_1, \dots, \xi_n, \dots$ независимы и одинаково распределены. Если они обладают конечной ненулевой дисперсией σ^2 , то

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

где $a = \mathbb{E}\xi_i$.

Доказательство. Для доказательства достаточно положить $\xi_{in} = \xi_i$ при $n \geq i$ и воспользоваться Теоремой 4. $\square$

⁴¹Здесь и далее в аналогичных случаях подразумевается, что $B_n^2 > 0$ при всех достаточно больших n .

3.2 Отказ от одинаковой распределенности: теорема Линдеберга-Феллера

Теорема П. Леви (не важно, в какой форме) имеет дело с независимыми одинаково распределенными случайными величинами. На самом деле можно ослабить как условие одинаковой распределенности, так и условие независимости с сохранением предельных соотношений (3.1) и (3.2).

Мы начнем с отказа от одинаковой распределенности независимых случайных величин ξ_{in} (или ξ_i). Оказывается, в такой постановке имеет место результат, близкий к необходимому и достаточному условию выполнения ЦПТ.

Теорема 5. (Теорема Линдеберга-Феллера для схемы серий).

Рассмотрим схему серий $\{\xi_{1n}, \dots, \xi_{nn}, n \geq 1\}$, в каждой из которых случайные величины ξ_{in} независимы при $1 \leq i \leq n$ и имеют конечные дисперсии σ_{in}^2 . Обозначим $B_n^2 = \sigma_{1n}^2 + \dots + \sigma_{nn}^2$ и $\eta_{in} = \xi_{in} - E\xi_{in}$.

1. Если для любого $\delta > 0$

$$\frac{1}{B_n^2} \sum_{i=1}^n E(\eta_{in}^2, |\eta_{in}| > \delta B_n) \rightarrow 0 \quad (3.3)$$

при $n \rightarrow \infty$, то схема серий $\{\xi_{1n}, \dots, \xi_{nn}, n \geq 1\}$ удовлетворяет ЦПТ.

2. Если схема серий $\{\xi_{1n}, \dots, \xi_{nn}, n \geq 1\}$ удовлетворяет ЦПТ и, кроме этого,

$$\frac{1}{B_n^2} \max_{1 \leq i \leq n} \sigma_{in}^2 \rightarrow 0, \quad (3.4)$$

то для любого ε имеет место сходимость (3.3).

Конечно, из Теоремы 5 вытекает ее аналог для последовательностей случайных величин.

Следствие 4. (Теорема Линдеберга-Феллера для последовательностей).

Пусть случайные величины $\xi_1, \dots, \xi_n, \dots$ независимы и обладают конечными дисперсиями σ_i^2 . Обозначим $B_n^2 = \sigma_1^2 + \dots + \sigma_n^2$ и $\eta_i = \xi_i - E\xi_i$. Кроме того, будем считать, что $B_n > 0$ при достаточно больших n .⁴²

1. Если для любого $\delta > 0$

$$\frac{1}{B_n^2} \sum_{i=1}^n E(\eta_i^2, |\eta_i| > \delta B_n) \rightarrow 0 \quad (3.5)$$

при $n \rightarrow \infty$, то последовательность $\{\xi_n, n \geq 1\}$ удовлетворяет ЦПТ.

2. Если последовательность $\{\xi_n, n \geq 1\}$ удовлетворяет ЦПТ и, кроме этого,

$$\frac{1}{B_n^2} \max_{1 \leq i \leq n} \sigma_i^2 \rightarrow 0, \quad (3.6)$$

то для любого ε имеет место сходимость (3.5).

Условия (3.3) и (3.5)⁴³ называются *условиями Линдеберга*, условия (3.4) и (3.6) условиями *асимптотической пренебрежимости слагаемых*⁴⁴. Последний термин понятен: дисперсия каждого слагаемого в сумме S_n гораздо меньше дисперсии всей суммы.

⁴²Иначе говоря, не все случайные величины ξ_{in} постоянны.

⁴³Подразумевается, что они выполнены для любого $\varepsilon > 0$.

⁴⁴В сумме S_n .

Мы не будем заниматься доказательством теоремы Линдеберга-Феллера,⁴⁵ его можно найти в любом достаточно продвинутом учебнике по теории вероятностей.⁴⁶ Это доказательство достаточно громоздко, хотя его идея точно такая же, как идея доказательства теоремы Леви — нужно рассматривать характеристические функции и аккуратно считать их предел.

Вместо этого мы обсудим сами условия теоремы Линдеберга-Феллера и приведем несколько примеров. Для удобства изложения будем делать это в терминах одной последовательности случайных величин, а не в терминах схемы серий.

Теорема Линдеберга-Феллера показывает, что для произвольной последовательности независимых случайных величин свойство «удовлетворять ЦПТ» не может быть выражено только в терминах дисперсий: условие (4) налагается на «хвосты» дисперсий, то есть на то, насколько быстро выражения $E(\eta_i^2, |\eta_i| > \delta B_n)$ стремятся к нулю при $n \rightarrow \infty$.⁴⁷

Прежде всего заметим, что в теории ЦПТ мы всегда можем считать, что $E\xi_i = 0$ для всех i , и поэтому в условии Линдеберга можно писать ξ_i вместо η_i .⁴⁸

Далее, ЦПТ может выполняться и без условия (3.6). Действительно, если взять случайные величины $\xi_i \in N(a_i, \sigma_i^2)$ независимыми, то левая часть (3.2) будет иметь стандартное нормальное распределение при любом n и любых σ_i^2 . Нетрудно подобрать эти дисперсии так,⁴⁹ чтобы (3.6) не выполнялось.

Для прояснения роли условия асимптотической пренебрежимости слагаемых докажем следующее утверждение.

Лемма 3. Пусть выполнено условие Линдеберга (3.5). Тогда

1. Имеет место сходимост (3.6);
2. Для любого $\delta > 0$

$$\max_{1 \leq i \leq n} P(|\xi_n| > \delta B_n) \rightarrow 0, \quad n \rightarrow \infty;$$

3. $B_n \rightarrow \infty$;

Доказательство. Заметим, что второе и третье утверждение леммы следуют из первого. Действительно,

$$\max_{1 \leq i \leq n} P(|\xi_n| > \delta B_n) \leq \frac{1}{\delta^2 B_n^2} \max_{1 \leq i \leq n} \sigma_i^2,$$

а B_n^2 должно стремиться к бесконечности, так как не все дисперсии σ_i^2 в левой части (3.6) нулевые.

Докажем, что условие асимптотической пренебрежимости следует из условия Линдеберга. Это делается просто: для любых i и $\delta > 0$

$$\begin{aligned} \frac{\sigma_i^2}{B_n^2} &= \frac{1}{B_n^2} \left(E(\xi_i^2, |\xi_i| > \delta B_n) + E(\xi_i^2, |\xi_i| \leq \delta B_n) \right) \leq \\ &\leq \frac{1}{B_n^2} \left(\sum_{i=1}^n E(\xi_i^2, |\xi_i| > \delta B_n) + \delta^2 B_n^2 \right) = \frac{1}{B_n^2} \left(\sum_{i=1}^n E(\xi_i^2, |\xi_i| > \delta B_n) \right) + \delta^2. \end{aligned} \quad (3.7)$$

Поскольку правая часть (3.7) не зависит от i (а левая — от δ), то осталось заметить, что при любом i отношение σ_i^2/B_n^2 может быть сделано равномерно по i сколь угодно малым выбором (сначала) δ , а потом n — ввиду выполнения условия Линдеберга. $\square$

⁴⁵Тем не менее — выведите Следствие 4 из Теоремы 5!

⁴⁶Например, в [2, т.1, гл. III, §4] или в [3, гл. XIV §6]. Первый пункт теоремы был доказан Я. Линдебергом в 1922 году, второй — В. Феллером в 1935 году.

⁴⁷Попробуйте построить пример двух последовательностей независимых случайных величин с попарно одинаковыми дисперсиями, первая из которых удовлетворяла бы ЦПТ, а вторая — нет.

⁴⁸Так и будем делать, если не оговорено противное.

⁴⁹А как?

Замечание 7. Отметим, что для схемы серий пункты 1 и 2 Леммы 3 по-прежнему имеют место, в то время как условие $B_n \rightarrow \infty$ выполняться не обязано. Например, вполне допустим вариант, когда $\max_i \sigma_{in}^2 \rightarrow 0$, а B_n^2 стремится к положительной постоянной.⁵⁰

Условие асимптотической пренебрежимости имеет свою интерпретацию, но ее удобнее проводить в терминах схемы серий. Более того, перепишем условия (3.3) и (3.4) немного иначе. А именно, вместо случайных величин ξ_{in} (или η_{in}) введем случайные величины $\beta_{in} = (\xi_{in} - E\xi_{in})/B_n$. Ясно, что $E\beta_{in} = 0$, $D\beta_{in} = \sigma_{in}^2/B_n^2$ и $D(\beta_{1n} + \dots + \beta_{nn}) = 1$. Тогда (3.3) переписывается как

$$\sum_{i=1}^n E(\beta_{in}^2, |\beta_{in}| > \delta) \rightarrow 0, \quad (3.8)$$

а (3.4) — как $\max_{1 \leq i \leq n} D\beta_{in} \rightarrow 0$, Сама же ЦПТ приобретает вид

$$\mathcal{L}(\beta_{in} + \dots + \beta_{nn}) \Rightarrow N(0, 1). \quad (3.9)$$

Именно такая запись принята, например, в [2].

Пусть теперь мы имеем «результат измерения»⁵¹ γ некоторой величины a . Ошибка измерения — это разность $\varepsilon = \gamma - a$. Если $E\gamma = a$ (то есть $E\varepsilon = 0$), то это означает, что наше измерение не имеет систематической погрешности.⁵²

Часто предполагают, что ε имеет нормальное распределение $N(0, \sigma^2)$ с какой-то дисперсией σ^2 .⁵³ Словесно это предположение можно объяснить следующим образом.

Естественно считать, что происхождение ошибки ε связано с большим числом каких-то независимых воздействий на эксперимент.⁵⁴ Если предположить, что эти воздействия аддитивны, то получится, что для большого n должно выполняться равенство $\varepsilon = \sigma(\beta_{1n} + \dots + \beta_{nn})$, где β_{in} независимы и $\sum_{1 \leq i \leq n} D\beta_{in} = 1$.⁵⁵

Отсюда сразу становится ясно, что предположение о нормальности ε тесно связано с ЦПТ в форме (3.9). Конечно, формально эту ЦПТ нужно доказывать, например, проверять условие Линдеберга в форме (3.8). Но нас сейчас интересует не это, а естественность условия асимптотической пренебрежимости слагаемых β_{in} в сумме $\beta_{1n} + \dots + \beta_{nn}$.

Представим себе, что это не так. Для простоты рассмотрим сумму всего двух⁵⁶ независимых слагаемых $\beta_1 + \beta_2$, где $E\beta_1 = E\beta_2 = 0$ и $D\beta_1 \gg D\beta_2$, причем β_1 имеет распределение, сильно отличающееся от нормального. Тогда $D\beta_1/D(\beta_1 + \beta_2) \approx 1$.

В этом случае естественно ожидать,⁵⁷ что распределение суммы $\beta_1 + \beta_2$ будет близко к распределению слагаемого β_1 . Поскольку это последнее распределение не похоже на нормальное, то и распределение суммы тоже будет сильно отличаться от нормального.

А формализация этих рассуждений и приводит нас к требованию асимптотической пренебрежимости.

Посмотрим теперь, какие следствия можно вывести из условия Линдеберга. Прежде всего, из первого пункта Теоремы 4 следует, что ЦПТ выполняется для любых независимых одинаково распределенных случайных величин с конечной положительной дисперсией σ^2 .⁵⁸ Действительно, рассмотрим для определенности условие Линдеберга (3.5) для последовательностей. В нашем случае

$$\frac{1}{B_n^2} \sum_{i=1}^n E(\eta_i^2, |\eta_i| > \delta B_n) = \frac{1}{\sigma^2} E(\eta_1^2, |\eta_1| > \delta \sigma \sqrt{n}),$$

⁵⁰Приведите пример!

⁵¹То есть случайную величину.

⁵²Если $E\gamma \neq a$, то это означает, что мы измеряем не a , а что-то другое.

⁵³Конечно, это очень удобное предположение. Например, оно позволяет на практике строить доверительные интервалы для a даже при небольшом числе измерений. А как?

⁵⁴Ну, например, что-то куда-то случайно сдвинулось и т.п.

⁵⁵Конечно, n мы не знаем — кроме того, что оно большое. Но обозначение β_{in} , включающее n , необходимо, так как у нас сумма дисперсий всех β_{in} равно 1 и мы хотим сослаться на ЦПТ. Случайные величины β_{in} считаются «элементарными», и поэтому не делается никаких предположений об их распределении — кроме существования вторых моментов.

⁵⁶Например, $\beta_1 = \beta_{1n}$, а $\beta_2 = \beta_{2n} + \dots + \beta_{nn}$.

⁵⁷А как это обосновать? Подсказка: используйте аппарат слабой сходимости распределений.

⁵⁸Иначе говоря, из теоремы Линдеберга следует утверждение теоремы Леви.

а последнее выражение стремится к нулю при $n \rightarrow \infty$ ввиду⁵⁹ конечности дисперсии случайной величины η_1 .

Гораздо интереснее и полезнее другое следствие из теоремы Линдеберга, которое мы для разнообразия сформулируем и докажем в терминах схемы серий.⁶⁰

Теорема 6. (Теорема Ляпунова.)

Пусть в схеме серий ξ_{in} — независимые случайные величины, обладающие при некотором $\alpha > 2$ конечными моментами порядка α . Снова обозначим $\eta_{in} = \xi_{in} - E\xi_{in}$ и $B_n^2 = D(\eta_{1n} + \dots + \eta_{nn})$. Как и раньше, предполагается, что $B_n > 0$ при достаточно больших n .

Если

$$\frac{1}{B_n^\alpha} \sum_{i=1}^n E|\eta_{in}|^\alpha \rightarrow 0 \quad \text{при } n \rightarrow \infty, \quad (3.10)$$

то схема серий $\{\xi_{1n}, \dots, \xi_{nn}, n \geq 1\}$ удовлетворяет ЦПТ.

Доказательство. Прежде всего заметим, что из выполнения неравенства $|\eta_{in}| > \delta B_n$ следует, что

$$\eta_{in}^2 = |\eta_{in}|^\alpha |\eta_{in}|^{2-\alpha} \leq |\eta_{in}|^\alpha \left(\frac{1}{\delta B_n} \right)^{\alpha-2}.$$

Поэтому при выполнении условия (3.10)

$$\frac{1}{B_n^2} \sum_{i=1}^n E(\eta_{in}^2, |\eta_{in}| > \delta B_n) \leq \frac{1}{B_n^2} \left(\frac{1}{\delta B_n} \right)^{\alpha-2} \sum_{i=1}^n E(\eta_{in}^\alpha, |\eta_{in}| > \delta B_n) \leq \delta^{2-\alpha} \frac{1}{B_n^\alpha} \sum_{i=1}^n E|\eta_{in}|^\alpha,$$

что и заканчивает доказательство. $\square$

Правую часть (3.10) обычно называют *дробью Ляпунова*, а саму сходимост (3.10) — *условием Ляпунова*.

Замечание 8. Условие Ляпунова, конечно, слабее условия Линдеберга хотя бы потому, что в первом из них требуется существование моментов порядка больше 2, а во втором — ровно 2. Тем не менее, если эти моментные условия выполнены, условие Ляпунова часто оказывается более удобным для проверки, чем условие Линдеберга, особенно если мы можем взять число α целым (и более того — четным).⁶¹

Приведем несколько примеров использования условий Линдеберга и Ляпунова.

Пример 3. Если у независимых случайных величин ξ_{in} есть равномерно ограниченные моменты порядка $\alpha > 2$ (то есть $E\xi_{in}^\alpha < c$ при любых i, n), а $n/B_n^\alpha \rightarrow 0$, то такая схема серий удовлетворяет ЦПТ.

Доказательство. Для доказательства будем использовать условие Ляпунова. Действительно,

$$\frac{1}{B_n^\alpha} \sum_{i=1}^n E|\eta_{in}|^\alpha \leq \frac{c^\alpha n}{B_n^\alpha} \rightarrow 0,$$

откуда и следует требуемое. $\square$

Таким образом, здесь доказательство ЦПТ сводится к проверке равномерной ограниченности моментов случайных величин ξ_{in} и изучению скорости роста дисперсии их суммы.

⁵⁹А подробнее?

⁶⁰А отсюда она (вернее, ее надлежащий вариант) будет вытекать и для последовательностей. Не правда ли?

⁶¹А почему «более того»?

Пример 4. Пусть $\xi_1, \dots, \xi_n, \dots$ — последовательность независимых случайных величин таких, что они все являются ограниченными⁶² одной и той же постоянной c .

Если при этом $D(\xi_1 + \dots + \xi_n) \rightarrow \infty$, то эта последовательность удовлетворяет ЦПТ.

Доказательство. Действительно, в этом случае все случайные величины η_i тоже оказываются равномерно ограниченными и поэтому (так как $B_n \rightarrow \infty$) при достаточно большом n все события вида $\{|\eta_i| > \delta B_n\}$ будут иметь нулевую вероятность. Следовательно, левая часть (3.5) при больших n просто обнуляется и условие Линдеберга оказывается выполненным.

Конечно, эти рассуждения верны и для схемы серий. $\square$

Для третьего примера нам понадобится следующее утверждение.

Предложение 6. Пусть $\varepsilon_1, \dots, \varepsilon_n, \dots$ последовательность независимых одинаково распределенных случайных величин с нулевым средним и единичной дисперсией. Обозначим $\xi_{in} = a_n \varepsilon_i$, где a_n — некоторые постоянные, и положим $B_n^2 = a_1^2 + \dots + a_n^2$.

Если $|a_n| < c$ и $n/B_n^2 = O(1)$, то схема последовательность $\{\xi_n, n \geq 1\}$ удовлетворяет ЦПТ.

Доказательство. Для доказательства используем условие Линдеберга (3.3). Так как $|\xi_i| \leq c|\varepsilon_i|$, то

$$\frac{1}{B_n^2} \sum_{i=1}^n E(\xi_i^2, |\xi_i| > \delta B_n) \leq \frac{c^2}{B_n^2} \sum_{i=1}^n E(\varepsilon_i^2, |\varepsilon_i| > \delta B_n/c) = \frac{c^2 n}{B_n^2} E(\varepsilon_1^2, |\varepsilon_1| > \delta B_n/c),$$

что очевидным⁶³ образом стремится к нулю. $\square$

Рассмотрим частный случай Предложения 6, имеющий отношение к статистике случайных процессов.

Пример 5. Пусть $\omega \in (0, 1/2)$ и $\{\varepsilon_n, n \geq 1\}$ — последовательность независимых одинаково распределенных случайных величин с $E\xi_n = 0$ и $D\xi_n = 1$. Обозначим $\xi_n = \sin(2\pi\omega n)\varepsilon_n$. Тогда

$$\mathcal{L}\left(\frac{\xi_1 + \dots + \xi_n}{\sqrt{n}}\right) \Rightarrow N(0, 1/4),$$

Доказательство. Прежде всего заметим, что⁶⁴

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \sin^2(2\pi\omega k) = 1/2.$$

Тем самым мы попадаем в условия Предложения 6, так как $|\sin(2\pi\omega n)| \leq 1$ и $n/B_n^2 \rightarrow 2$. Поэтому

$$\mathcal{L}\left(\frac{\xi_1 + \dots + \xi_n}{B_n}\right) \Rightarrow N(0, 1).$$

Поскольку $B_n \sim \sqrt{n}/2$, результат доказан. $\square$

⁶²С вероятностью 1.

⁶³Действительно?

⁶⁴Докажите этот факт. Из него, кстати, будет сразу же следовать, что

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \cos^2(2\pi\omega k) = 1/2.$$

Поэтому наш пример автоматически распространяется и на последовательность $\xi_n = \cos(2\pi\omega n)\varepsilon_n$.

3.3 Многомерная ЦПТ

Если мы рассматриваем независимые и одинаково распределенные случайные вектора, переход от одномерной к многомерной ЦПТ осуществляется очень просто, соответствующее утверждение называется многомерной теоремой Леви.

Теорема 7. Пусть $\bar{\xi}_1, \dots, \bar{\xi}_n, \dots$ — последовательность d -мерных независимых одинаково распределенных векторов со средним $\bar{a}$ и ковариационной матрицей Σ . Тогда

$$\mathcal{L}((\bar{\xi}_1 + \dots + \bar{\xi}_n - n\bar{a})/\sqrt{n}) \Rightarrow N(\mathbf{0}, \Sigma). \quad (3.11)$$

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

$$\varphi(\bar{t}) = 1 - (\Sigma \bar{t}, \bar{t})/2 + o(|\bar{t}|^2). \quad (3.12)$$

Пусть φ_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 (3.12) 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. $\square$

Приведем два содержательных примера использования Теоремы 7.

Пример 6. *Предельная теорема для независимых испытаний с m исходами.*

Пусть $x_1, \dots, x_n$ — повторная независимая выборка из распределения

$$\mathcal{P} : \begin{pmatrix} 1 & 2 & \dots & m \\ p_1 & p_2 & \dots & p_m \end{pmatrix}, \quad (3.13)$$

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,⁶⁵ которая утверждает, что $\mathcal{L}((k - np)/\sqrt{np}) \Rightarrow N(0, p(1-p))$, 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.⁶⁶ Положим

$$\bar{\xi}_i = (\xi_i^{(1)}, \dots, \xi_i^{(m)})^T = e_k \quad \text{при} \quad x_i = k, \quad 1 \leq k \leq m.$$

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

$$\sigma_{jk} \stackrel{\text{def}}{=} \text{Cov}(\xi_i^{(j)}, \xi_i^{(k)}) = E\xi_i^{(j)} \xi_i^{(k)} - p_j p_k = \begin{cases} p_j(1-p_j) & \text{при } k = j, \\ -p_j p_k & \text{при } k \neq j. \end{cases} \quad (3.14)$$

Поэтому многомерная теорема П. Леви выглядит в данном случае следующим образом:

$$\mathcal{L}\left(\left((k_1, \dots, k_m) - n(p_1, \dots, p_m)\right)/\sqrt{n}\right) \Rightarrow N(\mathbf{0}, \Sigma_m), \quad (3.15)$$

где Σ_m — матрица размера $m \times m$ с элементами σ_{jk} .

⁶⁵Фактически — частный случай одномерной теоремы П. Леви.

⁶⁶То есть $e_1 = (1, 0, \dots, 0)^T$, $e_2 = (0, 1, \dots, 0)^T$, ..., $e_m = (0, 0, \dots, 1)^T$.

Замечание 9. Поскольку $k_1 + \dots + k_m = n$, то ковариационная матрица Σ_m является вырожденной, но, если убрать из нее последний столбец и последнюю строку, то вырожденность пропадет.⁶⁷ 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 Σ_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 7. *Асимптотическое распределение начальных моментов.*

Рассмотрим независимые одинаково распределенные случайные величины $x_1, \dots, x_n$ в предположении, что $Ex_i^{2k} < \infty$ 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 $m_j = Ex_1^j$. Столь же очевидно, что элемент σ_{st} ковариационной матрицы Σ 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 выборочный⁶⁸ 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 7 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.

В общем случае этот переход не является таким очевидным.⁶⁹

⁶⁷Убедитесь как в первом, так и во втором. А как будет выглядеть предельная теорема, соответствующая этой операции?

⁶⁸Слово «выборочный» здесь используется в следующем смысле. Пусть $x_1, \dots, x_n$ повторная выборка из распределения $\mathcal{P}$ и θ — 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.

⁶⁹Например, многомерный аналог теоремы Линдберга можно найти в [4, гл.8 §7].

4 О сходимости к гауссовскому распределению

4.1 Критерий сходимости

Здесь мы опишем один из приемов, полезных при доказательстве многомерных ЦПТ и проиллюстрируем его на двух примерах.

Лемма 4. Пусть $\bar{\xi}_n = (\xi_{1,n}, \dots, \xi_{d,n})^T$ — последовательность d -мерных случайных векторов.

Для того, чтобы последовательность $\mathcal{L}(\bar{\xi}_n)$ слабо сходилась к d -мерному гауссовскому распределению $N(\bar{a}, \Sigma)$, необходимо и достаточно, чтобы для любых $\bar{t} = (t_1, \dots, t_d)^T$ последовательность $\mathcal{L}(\sum_{k=1}^d \xi_{k,n} t_k)$ слабо сходилась к одномерному нормальному распределению $N((\bar{a}, \bar{t}), (\Sigma \bar{t}, \bar{t}))$.

Доказательство. 1. Необходимость. Обозначим $\bar{t} = (t_1, \dots, t_d)^T$. Если $\mathcal{L}(\bar{\xi}_n) \Rightarrow N(\bar{a}, \Sigma)$, то по свойствам многомерного гауссовского распределения

$$\mathcal{L}\left(\sum_{k=1}^d \xi_{k,n} t_k\right) \Rightarrow N((\bar{a}, \bar{t}), (\Sigma \bar{t}, \bar{t})).$$

2. Достаточность. Из условий следует, что для любого $\tau \in \mathbb{R}$

$$\mathbb{E} e^{i\tau(\bar{\xi}_n, \bar{t})} \rightarrow e^{i\tau(\bar{a}, \bar{t})} e^{-\tau^2(\Sigma \bar{t}, \bar{t})/2}.$$

Полагая $\tau = 1$, получаем нужный результат. $\square$

Пример 8. В условиях Примера 5 рассмотрим не одну, а две последовательности случайных величин

$$\xi_n^{(1)} = \sin(2\pi\omega n)\varepsilon_n, \quad \xi_n^{(2)} = \cos(2\pi\omega n)\varepsilon_n$$

и положим $\bar{\xi}_n = (\xi_n^{(1)}, \xi_n^{(2)})^T$. Тогда

$$\mathcal{L}(\bar{\xi}_1 + \dots + \bar{\xi}_n)/\sqrt{n} \Rightarrow N(\mathbf{0}, \Sigma),$$

где $\mathbf{0} = (0, 0)^T$ и Σ — диагональная 2×2 матрица с числами $1/2$ на диагонали.

Доказательство. Прежде всего заметим, что⁷⁰

$$\frac{1}{n} \sum_{k=1}^n \sin(2\pi\omega k) \cos(2\pi\omega k) \rightarrow 0$$

при $n \rightarrow \infty$.

Далее будем использовать Лемму 4. Для этого при произвольных $s, t \in \mathbb{R}$ рассмотрим последовательность случайных величин

$$\eta_n = s\xi_n^{(1)} + t\xi_n^{(2)} = (s \sin(2\pi\omega n) + t \cos(2\pi\omega n))\varepsilon_n.$$

Поскольку $|s \sin(2\pi\omega n) + t \cos(2\pi\omega n)| < |s + t|$ и при $n \rightarrow \infty$

$$\begin{aligned} B_n^2/n &= \frac{1}{n} \sum_{k=1}^n (s^2 \sin^2(2\pi\omega k) + t^2 \cos^2(2\pi\omega k) + 2st \sin(2\pi\omega k) \cos(2\pi\omega k)) = \\ &= s^2 \frac{1}{n} \sum_{k=1}^n \sin^2(2\pi\omega k) + t^2 \frac{1}{n} \sum_{k=1}^n \cos^2(2\pi\omega k) + 2ts \frac{1}{n} \sum_{k=1}^n \sin(2\pi\omega k) \cos(2\pi\omega k) \rightarrow (s^2 + t^2)/2, \end{aligned}$$

то, применяя Предложение 6 и Лемму 4, получаем нужный нам результат. $\square$

⁷⁰Убедитесь!

4.2 Сохранение предельной гауссовости при отображениях

Линейные отображения

Предложение 7. Пусть $\bar{\xi}_n \in \mathbb{R}^d$ — случайные вектора, причем $\mathcal{L}(\bar{\xi}_n) \Rightarrow N(\bar{a}, \Sigma)$. 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 $\mathcal{L}(\bar{\eta}_n) \Rightarrow N(\bar{b} + A\bar{a}, A\Sigma A^T)$.

Доказательство. Обозначим φ_n и ψ_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. $\square$

Гладкие отображения. Теорема о сохранении асимптотической нормальности довольно широко распространена, наглядна и часто используется как в теоретических построениях, так и на практике.

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 σ^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 γ для $f(a)$ можно взять интервал⁷¹

$$(f(\bar{x}_n) - |f'(\bar{x}_n)| \bar{s}_n z_\gamma / \sqrt{n}, f(\bar{x}_n) + |f'(\bar{x}_n)| \bar{s}_n z_\gamma / \sqrt{n}),$$

где z_γ — решение уравнения $\Phi(z) - \Phi(-z) = \gamma$.

Конечно, эти нестрогие соображения можно сделать точными, в том числе и для многомерного случая (заметим, что для линейного отображения f у нас уже есть этот результат, доказанный в Предложении 7). Нам здесь, однако, достаточно одномерного случая.

Предложение 8. Рассмотрим последовательность случайных величин ξ_n и предположим, что

$$\mathcal{L}(c_n(\xi_n - a)) \Rightarrow N(0, \sigma^2),$$

где $c_n \rightarrow +\infty$. Если функция $f : \mathbb{R} \mapsto \mathbb{R}$ дифференцируема в точке a , то

$$\mathcal{L}(c_n(f(\xi_n) - f(a))) \Rightarrow N(0, (\sigma^2 f'(a))^2). \quad (4.1)$$

Приведем несколько простых примеров.

⁷¹Здесь, конечно, предполагается, что $\bar{s}_n > 0$ и $f'(\bar{x}_n) \neq 0$.

Пример 9. Из (4.1) следует, что $\mathcal{L}(c_n(f(\eta_n) - f(a))) \Rightarrow \delta_0$ при $f'(a) = 0$.

Например, пусть $x_1, \dots, x_n$ — независимая повторная выборка со средним a и положительной дисперсией σ^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 $\mathcal{L}(\sqrt{n}\sin(\bar{x}_n)) \Rightarrow N(0, \sigma^2)$.

Пример 10. *Стабилизация предельной дисперсии.*

Как уже говорилось, проблемы с построением доверительных интервалов возникают тогда, когда предельное распределение, возникающее, например, в ЦПТ, 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:

$$\mathcal{L}(\sqrt{n}(\bar{x}_n - p)) \Rightarrow N(0, p(1 - p)), \quad (4.2)$$

и например, один из вариантов границ такого интервала имеет вид

$$\bar{x}_n \pm z_\gamma \sqrt{\bar{x}_n(1 - \bar{x}_n)/n}, \quad (4.3)$$

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 параметра.⁷² Поскольку в доверительном интервале (4.3) $\sqrt{|\bar{x}_n|} < |\bar{x}_n|$, 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*.

А именно, взяв $f(x) = \arcsin \sqrt{x}$, мы получим, что в нашем случае (4.2) 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

$$\left(\sin \left(\arcsin \sqrt{\bar{x}_n} + 0.5z_\gamma/\sqrt{n} \right) \right)^2 < p < \left(\sin \left(\arcsin \sqrt{\bar{x}_n} - 0.5z_\gamma/\sqrt{n} \right) \right)^2,$$

с вероятностью, близкой к γ .

Как видно отсюда, «стабилизация» дисперсии заключается в выборе такого преобразования, после которого предельная дисперсия становится независимой от неизвестных параметров.⁷³

⁷²Вряд ли имеет смысл оценивать число 0.01 с точностью 0.1.

⁷³Пусть $x_1, \dots, x_n$ — повторная выборка из показательного распределения $\text{EXP}(\lambda)$, где λ — 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 λ и как его «стабилизировать»? Тот же вопрос — для выборки из распределения Пуассона.

5 Вещественные стационарные последовательности

5.1 Стационарные в широком смысле процессы

Определение 5.1. Вещественная⁷⁴ случайная последовательность $\{x_n, n \geq 1\}$ с конечными вторыми моментами называется *стационарной в широком смысле*⁷⁵, если

1. математические ожидания случайных величин x_n одинаковы, то есть $Ex_n = a$;
2. для любых $m \geq 1$ и $n \geq 0$ математическое ожидание $Ex_m x_{m+n}$ зависит только от n .

Из этого определения сразу же следует, что для любых $m, n \geq 1$ ковариация $\text{Cov}(x_m, x_n)$ зависит только от $|n - m|$. Поэтому, если ввести при $n \geq 0$ функцию $R(n) = \text{Cov}(x_{n+1}, x_1)$ и доопределить ее при отрицательных значениях n равенством $R(n) = R(-n)$, то Определение 5.1 сведется к двум пунктам: $Ex_n = a$ и $\text{Cov}(x_m, x_n) = R(m - n)$.

Функция R называется *ковариационной функцией* последовательности $\{x_n, n \geq 1\}$. Ясно, что $R(0) = \text{D}x_n$ для любого $n \geq 1$.⁷⁶ Конечно, ковариационная функция не зависит от сдвига последовательности x_n на любую постоянную, поэтому при изучении ее свойств обычно полагают $Ex_n = 0$.

Легко видеть, что ковариационная функция является неотрицательно определенной, то есть для любого $n \geq 1$, любых целых $k_1, \dots, k_n \geq 1$ и любых комплексных $z_1, \dots, z_n$

$$\sum_{j, \ell=1}^n R(k_j - k_m) z_j \bar{z}_\ell \geq 0.$$

Действительно, полагая $Ex_n = 0$, получим, что

$$\sum_{j, m=1}^n R(k_j - k_m) z_j \bar{z}_m = \sum_{j, m=1}^n Ex_{k_j} x_{k_m} z_j \bar{z}_m = E \left(\sum_{j, \ell=1}^n x_{k_j} z_j x_{k_m} \bar{z}_m \right) = \left| \sum_{j=1}^n x_{k_j} z_j \right|^2 \geq 0.$$

Имеет место следующее утверждение, которое мы оставим без доказательства (его можно найти в [2, гл. VI, §1]).

Теорема 8. (Теорема Герглотца).

Пусть $R(n)$ — некоторая неотрицательно определенная функция. Тогда существует единственная мера m , определенная на борелевской σ -алгебре подмножеств интервала $(-\pi, \pi]$ такая, то

$$R(n) = \int_{(-\pi, \pi]} e^{in\lambda} m(d\lambda). \quad (5.1)$$

В случае, когда R является ковариационной функцией последовательности x_n , мера m называется *спектральной мерой* последовательности x_n .

Нас будет интересовать только тот случай, когда мера m обладает производной Радона-Никодима⁷⁷ f по мере Лебега на интервале $(-\pi, \pi]$. Тогда f называется *спектральной плотностью*, а (5.1) приобретает вид

$$R(n) = \int_{-\pi}^{\pi} e^{in\lambda} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(n\lambda) f(\lambda) d\lambda. \quad (5.2)$$

⁷⁴В общей теории рассматриваются комплекснозначные случайные последовательности.

⁷⁵Или *слабо стационарной*. Такое же определение дается и для двусторонней бесконечной последовательности.

⁷⁶Везде далее мы предполагаем, что $R(0) > 0$.

⁷⁷Конечно, эта производная является неотрицательной почти всюду.

Из (5.2) сразу же следует, что спектральную плотность можно считать⁷⁸ четной функцией, то есть что $f(-\lambda) = f(\lambda)$. Поэтому

$$R(n) = 2 \int_0^\pi \cos(n\lambda) f(\lambda) d\lambda.$$

Из теоремы Римана-Лебега⁷⁹ сразу же следует, то при наличии спектральной плотности $R(n) \rightarrow 0$ при $n \rightarrow \infty$.

Если наложить условия на скорость сходимости ковариационной функции к нулю, то можно получить достаточное простое выражение спектральной плотности через ковариационную функцию.

Предложение 9. Если $\sum_m |R(m)| < \infty$, то

$$f(\lambda) = \frac{1}{2\pi} \sum_m e^{-im\lambda} R(m) = \frac{1}{2\pi} \sum_m e^{im\lambda} R(m) = \frac{1}{2\pi} \left(R(0) + 2 \sum_{m=1}^{\infty} \cos(m\lambda) R(m) \right), \quad (5.3)$$

причем f непрерывна и ограничена.

Доказательство. Обозначим

$$g(\lambda) = \frac{1}{2\pi} \sum_m e^{-im\lambda} R(m).$$

Ряд в правой части последнего равенства сходится, причем

$$\begin{aligned} |g(\lambda_1) - g(\lambda_2)| &\leq \frac{1}{2\pi} \sum_m |e^{-im\lambda_1} - e^{-im\lambda_2}| |R(m)| \leq \\ &\leq \frac{1}{2\pi} \sum_{|m| \leq m_0} |1 - e^{-im(\lambda_2 - \lambda_1)}| |R(m)| + \frac{1}{\pi} \sum_{m > m_0} |R(m)|. \end{aligned}$$

Следовательно, функция g непрерывна. Далее,

$$\int_{-\pi}^{\pi} e^{ik\lambda} g(\lambda) d\lambda = \sum_m R(m) \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(m+k)\lambda} d\lambda = R(-k) = R(k),$$

что и заканчивает доказательство. □

Следствие 5. 1. Из равенства (5.3) сразу же следует, что

$$R(0) + 2 \sum_{m=1}^{\infty} R(m) = 2\pi f(0). \quad (5.4)$$

2. Рассуждениями, аналогичными приведенным в Предложении 9, можно доказать,⁸⁰ что для любого целого $k \geq 1$ из того, что $\sum_m |m|^k |R(m)| < \infty$, то следует, что f k раз непрерывно дифференцируема.⁸¹ Таким образом, гладкость спектральной плотности оказывается тесно связанной со скоростью убывания ковариационной функции.

Приведем несколько примеров стационарных в широком смысле последовательностей.

Пример 11. 1. Пусть $E\xi = 0$, $D\xi = \sigma^2$ и $x_n \equiv \xi$. Тогда $E\xi_n \xi_m = \sigma^2$ и последовательность x_n является стационарной. Но спектральной плотности здесь не существует, так как ковариационная функция не стремится к нулю. Нетрудно видеть, что спектральная мера этой последовательности равна $\sigma^2 \delta_0(d\lambda)$.

2. Если в предыдущем примере положить $x_n = (-1)^n \xi$, то, очевидно, $E\xi_n \xi_m = (-1)^{n-m} \sigma^2$ и по-прежнему мы получаем стационарную последовательность, не имеющую спектральной плотности. А спектральная

⁷⁸ А почему только «можно считать»? И почему это «можно»?

⁷⁹ А что это такое?

⁸⁰ Попробуйте!

⁸¹ А как эта производная будет выражаться через ковариационную функцию?

мера на этот раз имеет вид⁸² $\sigma^2 \delta_\pi(d\lambda)$.

3. Важным примером слабо стационарной последовательности является последовательность попарно некоррелированных случайных величин x_i с одинаковыми средними и дисперсиями. В этом случае

$$R(n) = \begin{cases} \sigma^2 & \text{при } n = 0, \\ 0 & \text{иначе,} \end{cases} \quad (5.5)$$

где $\sigma^2 = Dx_n$. Спектральной плотностью такой последовательности является функция $f(\lambda) = \sigma^2/2\pi$. Действительно, для этой функции

$$\int_{-\pi}^{\pi} e^{in\lambda} f(\lambda) d\lambda = \frac{\sigma^2}{2\pi} \int_{-\pi}^{\pi} e^{in\lambda} d\lambda,$$

что совпадает с правой частью (5.5).

Стационарную последовательность с ковариационной функцией (5.5) обычно называют *белым шумом*, добавляя слова *с дисперсией* σ^2 . Если $\sigma^2 = 1$, то говорят о *стандартном белом шуме*, часто опуская слово *стандартный*.⁸³ Обычное обозначение стандартного белого шума — ε_n .

4. Пусть ε_n — (стандартный) белый шум, $\sigma > 0$ и $|\rho| \leq 1$. Положим $x_0 = \sigma\varepsilon_0$ и

$$x_n = \rho x_{n-1} + \sigma \sqrt{1 - \rho^2} \varepsilon_n, \quad n \geq 1. \quad (5.6)$$

Тогда (это легко проверяется по индукции) $Ex_n = 0$. Сосчитаем ковариацию $Ex_n x_m$, считая, что $n > m$. Для этого заметим,⁸⁴ что x_n является линейной комбинацией $\varepsilon_0, \dots, \varepsilon_{n-1}$ и, следовательно, x_n ортогональна ε_k при $k > n - 1$. Отсюда сразу же следует, что

$$Dx_n = \rho^2 Dx_{n-1} + \sigma^2(1 - \rho^2),$$

откуда по индукции заключаем, что $Dx_n = \sigma^2$.

Далее (домножая равенство (5.6) на x_m и беря математическое ожидание от обеих частей),

$$Ex_n x_m = \rho Ex_{n-1} x_m + \sigma \sqrt{1 - \rho^2} Ex_m \varepsilon_n = \rho Ex_{n-1} x_m = \rho^{n-m} Ex_m^2 = \rho^{n-m} \sigma^2.$$

Следовательно, последовательность x_n является стационарной (в широком смысле) с ковариационной функцией $R(n) = \rho^n \sigma^2$.

Ясно, что при $\rho = 1$ мы приходим к первому примеру из этой серии, а при $\rho = -1$ — ко второму. Если же $\rho = 0$, то $x_n = \sigma\varepsilon_n$, то есть является белым шумом с дисперсией σ^2 .

При $|\rho| < 1$ последовательность (5.6) обычно называют *процессом авторегрессии первого порядка*. Поскольку в этом случае $\sum_m |R(m)| < \infty$, то у таких последовательностей есть спектральная плотность. Непосредственными вычислениями можно убедиться, что она имеет вид

$$f(\lambda) = \frac{1}{2\pi} \frac{\sigma^2}{|1 - \rho e^{-i\lambda}|^2} = \frac{1}{2\pi} \frac{\sigma^2}{1 - 2\rho \cos(\lambda) + \rho^2}.$$

Приведем еще одну конструкцию, приводящую к построению слабо стационарной последовательности из белого шума.

Предложение 10. Пусть $\{\varepsilon_n\}_{n=-\infty}^{\infty}$ — двусторонняя последовательность случайных величин, причем $Ex_n = 0$ и $Ex_n \varepsilon_m = \delta_{nm}$.

Кроме того, пусть $\{c_n\}_{n=-\infty}^{\infty}$ — некоторая вещественная числовая последовательность, удовлетворяющая условию $\sum_n c_n^2 < \infty$. Положим

$$x_n = \sum_m c_{n-m} \varepsilon_m \quad (5.7)$$

⁸²Проверьте!

⁸³А если при этом $\varepsilon_n \in N(0, 1)$ (и, следовательно, ε_n независимы), то о *гауссовском белом шуме*.

⁸⁴Проверьте!

Тогда последовательность $\{x_n\}_{n=-\infty}^{\infty}$ является слабо стационарной с нулевым средним, ковариационной функцией

$$R(n) = \sum_{\ell=-\infty}^{\infty} c_{\ell} c_{\ell+n} \quad (5.8)$$

и спектральной плотностью

$$f(\lambda) = \frac{1}{2\pi} \left| \sum_{m=-\infty}^{\infty} c_m e^{-i\lambda m} \right|^2. \quad (5.9)$$

Доказательство. Заметим, что из сходимости ряда $\sum_m c_m^2$ следует⁸⁵ (абсолютная) сходимость рядов вида $\sum_{\ell} c_{\ell} c_{\ell+n}$.

Далее, ряд в правой части (5.7) сходится в среднеквадратическом.⁸⁶ Поэтому $E x_n^2 < \infty$ и, следовательно, $E|x_n| < \infty$. Отсюда

$$E x_n = \sum_{|m| \leq N} c_{n-m} E \varepsilon_m + E y_n^{(N)} = E y_n^{(N)}, \quad \text{где} \quad y_n^{(N)} = \sum_{|m| > N} c_{n-m} \varepsilon_m.$$

Так как

$$\left(E y_n^{(N)} \right)^2 \leq E \left(y_n^{(N)} \right)^2 = \sum_{|m|, |\ell| > N} c_{n-m} c_{n-\ell} E \varepsilon_m \varepsilon_{\ell} = \sum_{|m| > N} c_{n-m}^2 \rightarrow 0$$

при $N \rightarrow \infty$, то $E x_n = 0$.

Совершенно аналогично

$$\begin{aligned} E x_n^2 &= \sum_{m, \ell} c_{n-m} c_{n-\ell} E \varepsilon_m \varepsilon_{\ell} = \sum_m c_{n-m}^2 = \sum_m c_m^2 \quad \text{и} \\ E x_j x_{n+j} &= E \left(\sum_s c_{s-j} \varepsilon_s \right) \left(\sum_t c_{t-n-j} \varepsilon_t \right) = \sum_{s, t} c_{s-j} c_{t-n-j} E \varepsilon_s \varepsilon_t = \\ &= \sum_t c_{t-j} c_{t-j-n}. \end{aligned}$$

Делая в последнем выражении замену $\ell = t - j - m$, приходим к (5.8) и заодно доказываем, что x_n является стационарной последовательностью.

Наконец,

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\lambda n} \left| \sum_m c_m e^{-i\lambda m} \right|^2 d\lambda &= \frac{1}{2\pi} \sum_{\ell, m} c_{\ell} c_m \int_{-\pi}^{\pi} e^{i\lambda n} e^{-i\lambda(\ell-m)} d\lambda = \\ &= \frac{1}{2\pi} \sum_{\ell, m} c_{\ell} c_m \int_{-\pi}^{\pi} e^{-i\lambda(\ell-m+n)} d\lambda = \frac{2\pi}{2\pi} \sum_{\ell} c_{\ell} c_{\ell+n} = \sum_{\ell} c_{\ell} c_{\ell+n} = R(n), \end{aligned}$$

что и доказывает, что функция (5.9) действительно определяет спектральную плотность. $\square$

Легко написать в терминах коэффициентов c_j достаточные условия гладкости спектральной плотности для последовательности (5.7). Действительно, если $\sum_j |c_j| < \infty$, то

$$\sum_n |R(n)| = \sum_n \left| \sum_j c_j c_{j+n} \right| \leq \sum_j |c_j| \sum_n |c_{j+n}| = \left(\sum_n |c_n| \right)^2 < \infty.$$

Следовательно, в этом случае спектральная плотность (5.9) будет непрерывной.

Более того, имеет место следующее утверждение.

⁸⁵ Действительно?

⁸⁶ А почему?

Лемма 5. Пусть $\beta \geq 0$. $\sum_j |j|^\beta |c_j| < \infty$, то $\sum_j |j|^\beta |R(j)| < \infty$.

Доказательство. Действительно, так как $|k + j|^\beta \leq 2^\beta \max(|k|^\beta, |j|^\beta) \leq 2^\beta (|k|^\beta + |j|^\beta)$, то

$$\begin{aligned} \sum_j |j|^\beta |R(j)| &\leq \sum_j |j|^\beta \sum_k |c_k| |c_{k+j}| = \sum_k |c_k| \sum_j |j|^\beta |c_{k+j}| = \sum_k |c_k| \sum_m |m - k|^\beta |c_m| \leq \\ &2^\beta \sum_k |c_k| \sum_m (|m|^\beta + |k|^\beta) |c_m| = 2^{\beta+1} \left(\sum_k |k|^\beta |c_k| \right) \left(\sum_k |c_k| \right) < \infty. \end{aligned}$$

Утверждение доказано. □

Переход к свойствам спектральной плотности происходит с помощью Следствия 5.

5.2 Стационарные в узком смысле последовательности.

Определение 5.2. Случайная последовательность $x_1, \dots, x_n, \dots$ называется *стационарной в узком смысле*, если для любых n и k распределения случайных векторов $(x_n, \dots, x_{n+k})$ и $(x_1, \dots, x_{k+1})$ совпадают.⁸⁷

Конечно, стационарная в узком смысле последовательность не обязана быть стационарной в широком смысле, так как в Определении 5.2 никакие моментные условия не накладываются. Верно и обратное — стационарность в широком смысле определяется только через средние и ковариации, а не через совместные распределения.

В то же время, если стационарная в узком смысле последовательность обладает конечными вторыми моментами, то она является стационарной в широком смысле. Что же касается гауссовских последовательностей, то для понятия стационарности в широком и узком смысле совпадают.⁸⁸

Приведем примеры стационарных в узком смысле последовательностей, в основном основываясь на Примере 11, относящемся к стационарным в широком смысле последовательностям.

Пример 12. 1. Если ξ — некоторая случайная величина (не обязательно имеющая какие-то конечные моменты), то последовательности $x_n = \xi$ и $x_n = (-1)^n \xi$ являются сильно стационарными.⁸⁹

2. Если ε_n — независимые и одинаково распределенные случайные величины, то последовательность $x_n = \varepsilon_n$ является сильно стационарной.

3. Рассмотрим процесс авторегрессии (5.6) и поставим вопрос, при каких условиях на белый шум ε_n последовательность x_n будет стационарной не только в широком смысле, но и в узком.

Естественно потребовать, что бы этот белый шум состоял из независимых и одинаково распределенных случайных величин (предположения о нулевом среднем и единичной дисперсии, естественно, сохраняются). Этого, однако, не достаточно. Действительно (касаясь только одномерных распределений), с какой стати распределение правой части (5.6) совпадает с распределением x_{n-1} (и, следовательно, с распределением $x_0 = \varepsilon_0$), даже если x_{n-1} и ε_n независимы?

Не пытаясь разрешить эту проблему во всей ее полноте, заметим, что последовательность (5.6) будет сильно стационарной, если белый шум ε_n является гауссовским.⁹⁰

4. Еще один пример — однородная марковская цепь, начальное распределение которой совпадает со стационарным распределением этой марковской цепи.⁹¹

5. Другой пример, не связанный с Примером 11 — это последовательность $x_n = \xi_n \xi_{n+1}$, где ξ_i — независимые одинаково распределенные случайные величины.⁹²

Следующий пример будет в дальнейшем для нас основным.

Теорема 9. Пусть ε_i , $i \in \mathbb{Z}$ независимые одинаково распределенные случайные величины с $E\varepsilon_n = 0$ и $D\varepsilon_n = 1$. и $\sum_{i \in \mathbb{Z}} c_i^2 < \infty$. Положим

$$x_n = \sum_m c_{n-m} \varepsilon_m. \quad (5.10)$$

Тогда для любого n ряд (5.10) сходится в среднеквадратическом, причем $E x_n = 0$, а последовательность $x_1, \dots, x_n, \dots$ является сильно стационарной.

Доказательство. Сходимость ряда (5.10) в среднеквадратическом (как и равенство $E x_n = 0$) уже доказано при рассмотрении в Предложении 10. Для доказательства сильной стационарности перепишем (5.10) в виде

$$x_n = \sum_{\ell} c_{\ell} \varepsilon_{n-\ell}.$$

Отсюда сразу же видно, что вектор $(x_n, \dots, x_{n+k})$ получается из вектора $(x_1, \dots, x_{k+1})$ сдвигом последовательности $\{\varepsilon_i\}$ на $n-1$ шаг вправо. Поскольку свойства этой последовательности от такого сдвига не меняются, распределения рассматриваемых векторов совпадают.⁹³ $\square$

⁸⁷ Иногда в этом случае употребляют термин *сильно стационарная последовательность*. Понятие сильно стационарной последовательности автоматически переносится на случайные величины, принимающие значения в произвольном измеримом пространстве.

⁸⁸ А почему?

⁸⁹ А какие у них конечномерные распределения?

⁹⁰ А почему?

⁹¹ Действительно?

⁹² Действительно? А при каких условиях такая последовательность будет слабо стационарной?

⁹³ А если точнее?

Замечание 10. Если же в теореме 9 ввести дополнительное условие $E\varepsilon_i^2 = 1$ (а это все время будет предполагаться в дальнейшем), то из утверждения [2, Т.2, гл. IV, §2, теор. 2] сразу будет следовать, что ряд (5.10) сходится с вероятностью 1 при выполнении условия $\sum_{i \in \mathbb{Z}} c_i^2 < \infty$.

У нас, однако, условие $\sum_{i \in \mathbb{Z}} |c_i| < \infty$ (более сильное, чем это необходимо для сходимости ряда (5.10)) будет использоваться и в дальнейшем, поэтому здесь мы опираемся именно на него.

Определение 5.3. Пусть последовательность $\{x_i\}_{n \geq 1}$ удовлетворяет условиям Теоремы 5.10. Если дополнительно $E\varepsilon_n^2 = 1$, то такая последовательность будет называться *линейной стационарной*.

Замечание 11. 1. Тем самым линейная стационарная последовательность обладает ковариационной функцией, определенной в (5.8) и спектральной плотностью (5.9).

2. Далее мы всегда будем предполагать, что $\sum_i |c_i| > 0$, то есть будем исключать тот тривиальный случай, когда линейная последовательность вырождается до нулевой.

5.3 Предельная дисперсия выборочного среднего стационарной последовательности

Для нас предельные теоремы в первую очередь интересны с точки зрения математической статистики. Пусть у нас есть некоторая выборка⁹⁴ $x_1, \dots, x_n$ и нас интересует какой-то неизвестный параметр $\theta \in \mathbb{R}$, связанный с распределением этой выборки.

Если мы оцениваем этот параметр с помощью случайной величины $\hat{\theta}_n = \hat{\theta}_n(x_1, \dots, x_n)$, то естественная иерархия статистических задач, связанных с этим оцениванием, выглядит следующим образом.

1. Проверка несмещенности (или асимптотической несмещенности) оценки: верно ли, что $E\hat{\theta}_n = \theta$ или $E\hat{\theta}_n \rightarrow \theta$ при $n \rightarrow \infty$?
2. Проверка состоятельности оценки: верно ли что $P(|\hat{\theta}_n - \theta| > \delta) \rightarrow 0$ для любого $\delta > 0$ при $n \rightarrow \infty$?
Для (асимптотически) несмещенных⁹⁵ оценок естественным способом доказательства состоятельности является проверка того, что $D\hat{\theta}_n \rightarrow 0$ при $n \rightarrow \infty$.
3. Наконец, последними (и самыми сложными) являются задачи построения доверительных интервалов для параметра θ и проверка статистических гипотез для этого параметра. Если эти задачи решаются на основе оценки $\hat{\theta}_n$, то так или иначе их решения следуют и какой-то предельной теореме для этой оценки.

Перенесем эту общую схему на интересующий нас случай стационарной последовательности. Итак, пусть $x_1, \dots, x_n, \dots$ — некоторая стационарная последовательность,⁹⁶ обладающая неизвестным нам математическим ожиданием $Ex_n = \theta$.

В качестве оценки $\hat{\theta}_n$ величины θ естественно выбрать среднее арифметическое

$$\bar{x}_n = (x_1 + \dots + x_n)/n.$$

Очевидно, $E\bar{x}_n = \theta$, так что оценка является несмещенной, и для установления этого факта (необходимо) и достаточно существования математических ожиданий Ex_n .

Если мы будем изучать проблему состоятельности оценки $\bar{x}_n$ через ее дисперсию, то, очевидно, нам дополнительно нам потребуется существование вторых моментов у x_n . Это означает, что мы должны потребовать, чтобы последовательность $x_1, \dots, x_n, \dots$ была слабо стационарной. При этом, как нетрудно видеть, ответ (то есть условие сходимости дисперсии оценки $\bar{x}_n$) будет выражаться в терминах ковариационной функции этой последовательности.

Для построения (асимптотического) доверительного интервала математического ожидания θ с помощью $\bar{x}_n$ нам, конечно, понадобится какая-то предельная теорема. Поэтому условия слабой стационарности последовательности нам не хватит⁹⁷ и произойдет естественный переход к условию сильной стационарности.

Естественно ожидать, что эта предельная теорема будет (при некоторых условиях) Центральной Предельной Теоремой для исходной последовательности $x_1, \dots, x_n, \dots$. Попробуем сообразить, какой именно вид будет иметь эта ЦПТ.

⁹⁴Вообще говоря, не обязательно независимая и не обязательно повторная.

⁹⁵А если они не такие, то вообще разговаривать не о чем.

⁹⁶Пока не важно, в широком или узком смысле.

⁹⁷А почему?

Стандартная форма ЦПТ для некоторой последовательности $\{x_n\}_{n \geq 1}$ с одинаковыми средними, равными a , и конечными вторыми моментами может быть записана как

$$\mathcal{L}\left(\frac{S_n - na}{B_n}\right) \Rightarrow N(0, 1), \quad (5.11)$$

где $S_n = x_1 + \dots + x_n$, $B_n^2 = DS_n$ и $n \rightarrow \infty$.

Если при этом $B_n^2 \sim n\sigma^2$, где $\sigma^2 = \text{const} > 0$, то (5.11) можно переписать в виде

$$\mathcal{L}\left(\frac{S_n - na}{\sigma\sqrt{n}}\right) \Rightarrow N(0, 1),$$

или, вводя обозначение $\bar{x}_n = S_n/n$, в виде

$$\mathcal{L}\left(\sqrt{n} \frac{\bar{x}_n - a}{\sigma}\right) \Rightarrow N(0, 1),$$

что эквивалентно

$$\mathcal{L}(\sqrt{n} (\bar{x}_n - a)) \Rightarrow N(0, \sigma^2). \quad (5.12)$$

Поскольку $D\bar{x}_n = B_n^2/n \rightarrow \sigma^2$, то число σ^2 можно называть *асимптотической*⁹⁸ *дисперсией* выборочного среднего $\bar{x}_n$.

Заметим, что (5.12) следует из (5.11) и в том вырожденном случае, когда $B_n^2/n \rightarrow \sigma^2 = 0$, если понимать под $N(b, 0)$ распределение, сосредоточенное в точке b .

Таким образом, если сначала мы покажем, что $B_n^2 \sim n\sigma^2$, то мы не только докажем состоятельность оценки $\bar{x}_n$, но и приобретем правдоподобную гипотезу, о том, какой вид может иметь ЦПТ для рассматриваемой последовательности.

Итак, займемся стационарной в широком смысле последовательностью $\{x_n\}_{n \geq 1}$ и найдем теперь выражение для дисперсии $\bar{x}_n$. Как обычно, для этого мы можем рассматривать только случай $a = 0$.

Лемма 6. Дисперсия $D\bar{x}_n$ может быть представлена в виде

$$D\bar{x}_n = \frac{1}{n} \sum_{j=-(n-1)}^{n-1} (1 - |j|/n) R(j). \quad (5.13)$$

Доказательство. Действительно,

$$D\bar{x}_n = \frac{1}{n^2} \mathbb{E}(x_1 + \dots + x_n)^2 = \frac{1}{n^2} \sum_{k,l=1}^n \mathbb{E}x_k x_l = \frac{1}{n^2} \sum_{k,l=1}^n R(k-l) = \frac{1}{n} \sum_{j=-(n-1)}^{n-1} (1 - |j|/n) R(j),$$

что и заканчивает доказательство. $\square$

Перейдем теперь к предельное поведение дисперсии выборочного среднего. Сначала докажем следующую элементарную лемму.

Лемма 7. Если ряд $\sum_{k \geq 1} a_k$ сходится абсолютно, то

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{n-1} (1 - k/n) a_k = \sum_{k \geq 1} a_k.$$

Доказательство. Найдем n_0 такое, что $\sum_{k \geq n_0} |a_k| < \varepsilon$. Тогда при $n > n_0$

$$\begin{aligned} \left| \sum_{k \geq 1} a_k - \sum_{k=0}^{n-1} (1 - k/n) a_k \right| &\leq \sum_{k \geq n} |a_k| + \frac{1}{n} \sum_{k=1}^{n_0-1} k |a_k| + \sum_{k=n_0}^{n-1} (k/n) |a_k| \leq \\ &\leq \sum_{k \geq n} |a_k| + \frac{1}{n} \sum_{k=1}^{n_0-1} k |a_k| + \sum_{k \geq n_0} |a_k|. \end{aligned}$$

Первое слагаемое стремится к нулю при $n \rightarrow \infty$, последнее по условию меньше ε , а среднее стремится к нулю при $n \rightarrow \infty$ и фиксированном ε . Утверждение доказано. $\square$

⁹⁸Или *предельной*.

Теперь не представляет труда найти асимптотическую дисперсию для выборочного среднего.

Предложение 11. Если $\sum_m |R(m)| < \infty$, то при $n \rightarrow \infty$

$$nD\bar{x}_n \rightarrow R(0) + 2 \sum_{m \geq 1} R(m) = 2\pi f(0). \quad (5.14)$$

Доказательство. Утверждение вытекает из равенства (5.13), Леммы 7 и Следствия 5. □

Замечание 12. 1. Таким образом, если $f(0) \neq 0$, то $D\bar{x}_n \sim 2\pi f(0)/n$ и имеет порядок $O(1/n)$ при $n \rightarrow \infty$. Если же $f(0) = 0$, то дисперсия $\bar{x}_n$ стремится к нулю быстрее, чем n^{-1} .

2. На самом деле можно доказать, что оценка $\bar{x}_n$ является состоятельной, если у спектральной меры m нет нагрузки в нуле (например, если существует спектральная плотность). Сейчас мы не будем этим заниматься.

5.4 ЦПТ для линейного процесса

Лемма 8. Рассмотрим линейный процесс x_n , записанный в виде

$$x_n = \sum_m c_m \varepsilon_{n-m},$$

где $\sum_m |c_m| < \infty$, и ε_n — последовательность независимых одинаково распределенных случайных величин такая, что $E\varepsilon_n = 0$, $E\varepsilon_n^2 = 1$. Обозначим

$$J_{n,x} = \frac{1}{\sqrt{n}} \sum_{\ell=1}^n x_\ell \quad \text{и} \quad J_{n,\varepsilon} = \frac{1}{\sqrt{n}} \sum_{\ell=1}^n \varepsilon_\ell.$$

Тогда

$$J_{n,x} = \left(\sum_m c_m \right) J_{n,\varepsilon} + V_n \quad (5.15)$$

где V_n стремится по вероятности к нулю при $n \rightarrow \infty$.

Доказательство. Положим при $m = 1, \dots, n$

$$U_{n,m} = - \sum_{\ell=1}^n \varepsilon_\ell + \sum_{\ell=1-m}^{n-m} \varepsilon_\ell \quad (5.16)$$

и

$$V_n = \frac{1}{\sqrt{n}} \sum_m c_m U_{n,m}.$$

Тогда, делая замену $\ell = j - m$, получим, что

$$\begin{aligned} J_{n,x} &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\sum_m c_m \varepsilon_{j-m} \right) = \frac{1}{\sqrt{n}} \sum_m c_m \sum_{j=1}^n \varepsilon_{j-m} = \\ &= \frac{1}{\sqrt{n}} \sum_m c_m \sum_{\ell=1-m}^{n-m} \varepsilon_\ell = \sum_m c_m \frac{1}{\sqrt{n}} \sum_{\ell=1-m}^{n-m} \varepsilon_\ell = \\ &= \left(\sum_m c_m e^{-im\lambda} \right) \left(\frac{1}{\sqrt{n}} \sum_{\ell=1}^n \varepsilon_\ell \right) + \frac{1}{\sqrt{n}} \sum_m c_m U_{n,m}(\lambda) = \left(\sum_m c_m \right) J_{n,\varepsilon} + V_n. \end{aligned} \quad (5.17)$$

Нетрудно видеть, что после сокращения подобных членов в сумме (5.16) останется ровно $2|m|$ независимых слагаемых при $|m| < n$ и ровно $2n$ независимых слагаемых при $|m| \geq n$. Поэтому

$$\mathbb{E}|U_{n,m}|^2 \leq 2 \min(|m|, n).$$

Далее, зафиксировав некоторое n_0 и взяв $n > n_0$, получим, что

$$\begin{aligned} \mathbb{E}|V_n| &\leq \frac{1}{\sqrt{n}} \sum_m |c_m| \mathbb{E}|U_{n,m}| \leq \frac{1}{\sqrt{n}} \sum_m |c_m| \left(\mathbb{E}|U_{n,m}|^2 \right)^{1/2} \leq \\ &\leq \frac{\sqrt{2}}{\sqrt{n}} \sum_m |c_m| (\min(|m|, n))^{1/2} \leq \frac{\sqrt{2}}{\sqrt{n}} \sum_{|m| < n_0} |c_m| |m|^{1/2} + \sqrt{2} \sum_{|m| \geq n_0} |c_m|. \end{aligned}$$

Поэтому при фиксированном n_0

$$\limsup_n \mathbb{E}|V_n| \leq \sqrt{2} \sum_{|m| \geq n_0} |c_m|.$$

Так как n_0 произвольно, то $\mathbb{E}|V_n| \rightarrow 0$. Следовательно, V_n стремится к нулю по вероятности. $\square$

Отсюда уже легко получается ЦПТ для линейных процессов.

Теорема 10. В условиях Леммы 8 последовательность $\{x_n\}_{n \geq 1}$ удовлетворяет Центральной Предельной Теореме. А именно,

$$\mathcal{L}((x_1 + \dots + x_n)/\sqrt{n}) \Rightarrow N(0, \sigma^2),$$

где

$$\sigma^2 = \left(\sum_n c_n \right)^2 = 2\pi f(0).$$

Доказательство. Утверждение сразу же следует из того, что $\mathcal{L}(J_{n,\varepsilon}) \Rightarrow N(0, 1)$, представления (5.15) и сходимости V_n к нулю по вероятности. $\square$

Замечание 13. Из утверждения Теоремы 10 следует, что если нам удастся построить состоятельную оценку $\hat{f}_n(0)$ числа $f(0)$, то мы стандартным образом получим возможность строить асимптотические доверительные интервалы для среднего стационарного процесса. Но это уже относится к статистике случайных процессов.

Пример 13. Пусть ε_n , $n = 0, \pm 1, \dots$ — независимые одинаково распределенные случайные величины с нулевым средним и единичной дисперсией. Возьмем числа ρ, σ с $|\rho| < 1$ и $\sigma > 0$ и положим

$$x_n = \sigma \sqrt{1 - \rho^2} \sum_{k=0}^{\infty} \rho^k \varepsilon_{n-k}.$$

Легко видеть, что для этой последовательности выполняются условия Теоремы 10. Поскольку

$$\begin{aligned} x_n &= \sqrt{1 - \rho^2} \varepsilon_n + \sqrt{1 - \rho^2} \sum_{k=1}^{\infty} \rho^k \varepsilon_{n-k} = \sqrt{1 - \rho^2} \varepsilon_n + \rho \sqrt{1 - \rho^2} \sum_{k=1}^{\infty} \rho^{k-1} \varepsilon_{n-k} = \\ &= \sqrt{1 - \rho^2} \varepsilon_n + \rho \sqrt{1 - \rho^2} \sum_{\ell=0}^{\infty} \rho^{\ell} \varepsilon_{n-1-\ell} = \sqrt{1 - \rho^2} \varepsilon_n + \rho x_{n-1}, \end{aligned}$$

причем x_{n-1} и ε_n независимы,⁹⁹ то, как мы уже выясняли, ковариационная функция этой последовательности имеет вид $R(m) = \sigma^2 \rho^{|m|}$.

Впрочем, это можно проверить и непосредственно: при $m \geq 0$

$$E x_n x_{m+n} = \sigma^2 (1 - \rho^2) \sum_{k,j=1}^{\infty} \rho^{k+j} E \varepsilon_{n-k} \varepsilon_{n+m-j} = \sigma^2 (1 - \rho^2) \sum_{k=1}^{\infty} \rho^{2k+m} = \frac{\sigma^2 (1 - \rho^2) \rho^m}{1 - \rho^2} = \sigma^2 \rho^m.$$

Так как

$$\begin{aligned} \sum_m R(m) &= R(0) + 2 \sum_{m \geq 1} R(m) = \sigma^2 \left(1 + 2 \sum_{m \geq 1} \rho^m \right) = \sigma^2 \left(2 \sum_{m \geq 0} \rho^m - 1 \right) = \\ &= \sigma^2 \left(\frac{2}{1 - \rho} - 1 \right) = \sigma^2 \frac{1 + \rho}{1 - \rho}. \end{aligned}$$

Таким образом,

$$\mathcal{L}((x_1 + \dots + x_n)/\sqrt{n}) \Rightarrow N(0, \sigma_1^2), \text{ где } \sigma_1^2 = \sigma^2 \frac{1 + \rho}{1 - \rho}.$$

Это, конечно, тот нечастый случай, когда предельная дисперсия считается аналитически.

⁹⁹А почему? А почему это важно?

6 Процессы авторегрессии

6.1 Основные понятия и свойства.

Пусть ξ_n , $n = 0, \pm 1, \pm 2, \dots$ — вещественная стационарная в широком смысле последовательность, имеющая нулевое среднее, ковариационную функцию $R_\xi(n)$ и спектральную меру m_ξ .

Если существуют такие натуральные $q \geq 1$ и $p \geq 0$, (вещественные) числа $a_0, \dots, a_p$, $b_0, \dots, b_q$, удовлетворяющие условию $a_0 b_0 a_p b_q \neq 0$, а также такой белый шум ε_n , что для любого n

$$b_0 \xi_n + b_1 \xi_{n-1} + \dots + b_q \xi_{n-q} = a_0 \varepsilon_n + a_1 \varepsilon_{n-1} + \dots + a_p \varepsilon_{n-p}, \quad (6.1)$$

то процесс ξ называется *процессом авторегрессии-скользящего суммирования порядка (q, p)* , сокращенно — АРСС(q, p). Процесс (6.1) называется реализуемым относительно белого шума $\{\varepsilon_n\}$, если ε_n ортогонально ξ_m при $m < n$.

В случае, когда $p = 0$, ξ_n называется процессом (чистой) авторегрессии и обозначается АР(q).

С коэффициентами $(a_0, \dots, a_p)$ и $(b_0, \dots, b_q)$ связывают следующие полиномы комплексной переменной. Введем обозначения

$$Q(z) = b_0 z^q + b_1 z^{q-1} + \dots + b_{q-1} z + b_q \quad \text{и} \quad Q^*(z) = b_q z^q + b_{q-1} z^{q-1} + \dots + b_1 z + b_0$$

и аналогичные обозначения P и P^* для полиномов с коэффициентами $(a_0, \dots, a_p)$. Поскольку полиномы Q, Q^* не обращаются в ноль, то их корни взаимно обратны. Действительно, так как

$$Q(z) = \sum_{k=0}^q b_k z^{q-k} = z^q \sum_{k=0}^q b_k z^{-k} = z^q Q^*(z^{-1}), \quad (6.2)$$

то из $Q(z_0) = 0$ следует $Q^*(1/z_0) = 0$ и наоборот. В частности, из (6.2) следует, что полиномы Q и Q^* одновременно имеют корни или не имеют корней на единичной окружности $|z| = 1$.

Предложение 12. Пусть $Q(z) \neq 0$ при $|z| = 1$. Тогда процесс (6.1) имеет спектральную плотность f_ξ , выражающуюся формулой

$$f_\xi(\lambda) = \frac{1}{2\pi} \frac{|P^*(e^{-i\lambda})|^2}{|Q^*(e^{-i\lambda})|^2}, \quad \lambda \in (-\pi, \pi]. \quad (6.3)$$

Доказательство. Обозначим η_n левую (и правую) часть равенства (6.1). Так как η является линейным преобразованием ξ , то мы можем стандартным способом вычислить спектральную меру этого процесса. Действительно,

$$\eta_n = \sum_{k=0}^q b_k \xi_{n-k} = \sum_{k=0}^q b_k \int_{-\pi}^{\pi} e^{i\lambda(n-k)} \mu_\xi(d\lambda) = \int_{-\pi}^{\pi} e^{i\lambda n} \left(\sum_{k=0}^q b_k e^{-i\lambda k} \right) \mu_\xi(d\lambda) = \int_{-\pi}^{\pi} e^{i\lambda n} Q^*(e^{-i\lambda}) \mu_\xi(d\lambda),$$

так что

$$m_\eta(d\lambda) = |Q^*(e^{-i\lambda})|^2 m_\xi(d\lambda).$$

Аналогично из правой части (6.1) получаем, что

$$m_\eta(d\lambda) = |Q^*(e^{-i\lambda})|^2 m_\epsilon(d\lambda) = \frac{1}{2\pi} |P^*(e^{-i\lambda})|^2 d\lambda.$$

Поэтому

$$|Q^*(e^{-i\lambda})|^2 m_\xi(d\lambda) = |P^*(e^{-i\lambda})|^2 d\lambda$$

и, если $Q^*(e^{-i\lambda}) \neq 0$ ни при каком λ , то

$$m_\xi(d\lambda) = \frac{1}{2\pi} \frac{|P^*(e^{-i\lambda})|^2}{|Q^*(e^{-i\lambda})|^2} d\lambda.$$

Утверждение доказано. $\square$

В дальнейшем, если не оговорено противное, мы всегда предполагаем, что полином Q не имеет корней на единичной окружности.

Замечание 14. Так как $|Q(z)| = |z^q Q^*(1/z)| = |Q^*(1/z)|$ при $|z| = 1$, то спектральная плотность f_ξ может быть представлена также в виде

$$f_\xi(\lambda) = \frac{1}{2\pi} \frac{|P(e^{i\lambda})|^2}{|Q(e^{i\lambda})|^2}.$$

Покажем теперь, что изучение процесса АРСС (6.1) можно свести к изучению процесса чистой авторегрессии с теми же коэффициентами b_k .

Предложение 13. Пусть η — процесс чистой авторегрессии, имеющий вид

$$b_0\eta_n + b_1\eta_{n-1} + \dots + b_q\eta_{n-q} = \varepsilon_n \quad (6.4)$$

и

$$\xi_n = a_0\eta_n + a_1\eta_{n-1} + \dots + a_p\eta_{n-p}. \quad (6.5)$$

Тогда 1. Процесс ξ удовлетворяет (6.1). 2. Если процесс η является реализуемым относительно ε , то таким же свойством обладает и ξ .

Доказательство. Действительно,

$$\sum_{k=0}^q b_k \xi_{n-k} = \sum_{k=0}^q b_k \left(\sum_{\ell=0}^p a_\ell \eta_{n-k-\ell} \right) = \sum_{\ell=0}^p a_\ell \left(\sum_{k=0}^q b_k \eta_{n-k-\ell} \right) = \sum_{\ell=0}^p a_\ell \varepsilon_{n-k}.$$

Второе утверждение очевидным образом следует из (6.4) и (6.5). $\square$

Ввиду Предложения 13 дальнейшие рассуждения проводятся для процессов чистой авторегрессии

$$b_0\xi_n + b_1\xi_{n-1} + \dots + b_q\xi_{n-q} = a_0\varepsilon_n, \quad (6.6)$$

имеющей спектральную плотность

$$f_\xi(\lambda) = \frac{1}{2\pi} \frac{|a_0|^2}{|Q^*(e^{-i\lambda})|^2}, \quad \lambda \in (-\pi, \pi].$$

Заметим, что эта спектральная плотность не обращается в ноль, поэтому (согласно общему результату теории стационарных в широком смысле последовательностей) процесс ξ_n является линейным преобразованием некоторого белого шума, принадлежащего замыканию в $\mathbb{L}^2(dP)$ всевозможных линейных комбинаций случайных величин ξ_n . Для процессов авторегрессии этот результат можно уточнить.

Предложение 14. Для любого белого шума ε_n и для любых чисел $(b_0, \dots, b_q)$ таких, что $b_0 b_q \neq 0$ и полином Q^* не имеет корней на единичной окружности существует процесс авторегрессии ξ_n , удовлетворяющий (6.6).

6.2 Реализуемые процессы авторегрессии.

Теорема 11. Пусть $Q^*(z) \neq 0$ при $|z| = 1$. Для того, чтобы процесс авторегрессии (6.6) был реализуемым, необходимо и достаточно, чтобы все корни полинома $Q^*(z)$ лежали вне единичного круга.

Доказательство. Запишем условие реализуемости в явном виде. Скалярно умножив (6.6) на ξ_m при $m < n$, получим, что реализуемость эквивалентна бесконечной системе

$$b_0 R_\xi(n-m) + b_1 R_\xi(n-1-m) + \dots + b_q R_\xi(n-q-m) = 0, \quad n > m.$$

Обозначая $k = n - m$ получим, что процесс ξ_n реализуем тогда и только тогда, когда при любом $k \geq 1$

$$b_0 R_\xi(k) + b_1 R_\xi(k-1) + \dots + b_q R_\xi(k-q) = 0. \quad (6.7)$$

Поскольку

$$R_\xi(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\lambda} \frac{|a_0|^2}{|Q^*(e^{-i\lambda})|^2} d\lambda,$$

то (6.7) эквивалентно тому, что при любом $k \geq 1$

$$\begin{aligned} 0 &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\sum_{\ell=0}^q b_\ell e^{i(k-\ell)\lambda} \right) \frac{1}{|Q^*(e^{-i\lambda})|^2} d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\lambda} \left(\sum_{\ell=0}^q b_\ell e^{-i\ell\lambda} \right) \frac{1}{|Q^*(e^{-i\lambda})|^2} d\lambda = \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\lambda} \frac{Q^*(e^{-i\lambda})}{|Q^*(e^{-i\lambda})|^2} d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\lambda} \frac{1}{\overline{Q^*(e^{-i\lambda})}} d\lambda = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\lambda} \frac{1}{Q^*(e^{i\lambda})} d\lambda = \\ &= \frac{1}{2\pi i} \int_{|z|=1} z^{k-1} \frac{1}{Q^*(z)} dz. \end{aligned}$$

Как хорошо известно из теории функций комплексной переменной, интеграл от мероморфной функции по замкнутому контуру равен нулю, если это все нули этой функции расположены вне контура. Таким образом (так как $k \geq 1$), процесс авторегрессии будет реализуемым, если функция $Q^*(z)$ не имеет корней внутри единичного круга.

Доказательство необходимости для простоты проведем в случае, когда все корни $z_1, \dots, z_m$ ($m > 0$) полинома $Q^*(z)$, лежащие внутри единичного круга, являются простыми. Поскольку среди чисел z_ℓ нет нуля, то при любом $k \geq 1$

$$\frac{1}{2\pi i} \int_{|z|=1} z^{k-1} \frac{1}{Q^*(z)} dz = \sum_{\ell=1}^m \text{Res}_{v_\ell} \left(\frac{z^{k-1}}{Q^*(z)} \right) = \sum_{\ell=1}^m \frac{z_\ell^{k-1}}{H(z_\ell)} = 0,$$

где $H(z) = dQ^*(z)/dz$. Этого, конечно, быть не может, так как вектора $(1, z_\ell, \dots, z_\ell^s)^T$ линейно независимы при достаточно больших s . В случае кратных корней доказательство сохраняется, но линейно независимые вектора выглядят более сложно. $\square$

Замечание 15. В терминах полинома Q условия Теоремы 11 означают, что все корни этого полинома лежат внутри единичного круга.

Предложение 15. Пусть $b_0, \dots, b_q$ и $a_0, \dots, a_p$ — некоторые числа, причем $b_0 b_q a_0 a_p \neq 0$. Обозначим $Q^*(z) = \sum_{k=0}^q b_k z^k$ и $P^*(z) = \sum_{k=0}^p a_k z^k$. Если $Q^*(z) \neq 0$ при $|z| = 1$, Если корни полинома $Q^*(z) = \sum_{\ell=0}^q b_\ell z^\ell$ лежат вне единичного круга, то для любого белого шума ε_n существует стационарная последовательность АРСС(q,p), реализуемая относительно ε_n и имеющая спектральную плотность (6.3).

Доказательство. Во первых, нам достаточно доказывать результат лишь для процесса чистой регрессии, то есть в случае $P^*(z) = 1$. Далее, если все корни полинома $Q^*(z)$ лежат вне единичного круга, то мы можем воспользоваться Предложением 14.

Рассмотрим случай, когда $Q^*(z) = \left(\prod_{j=1}^d (z - z_j) \right) R(z)$, где все нули функции $R(z)$ лежат вне единичного круга, а $|z_j| < 1$ при $j = 1, \dots, d$. Далее,

$$|Q^*(z)|^2 = Q^*(z) \overline{Q^*(z)} = |R(z)|^2 \prod_{j=1}^d (z - z_j)(\bar{z} - \bar{z}_j).$$

Заметим, что в формуле для спектральной плотности полином $Q^*(z)$ нужен только при $|z| = 1$. Поэтому при $|z| = 1$

$$(z - z_j)(\bar{z} - \bar{z}_j) = |z|^2 |z_j|^2 (1/z_j - 1/z)(1/\bar{z}_j - 1/\bar{z}) = |z_j|^2 (\bar{z} - 1/z_j)(z - 1/\bar{z}_j),$$

и мы приходим к полиному

$$Q_1^*(z) = \prod_{j=1}^d z_j \left(\prod_{j=1}^d (z - 1/\bar{z}_j) \right) R(z),$$

который совпадает с $Q^*(z)$ при $|z| = 1$ и все корни которого лежат вне единичного круга. Осталось снова сослаться на Предложение 14. $\square$

6.3 Авторегрессия первого порядка и марковские гауссовские стационарные последовательности.

Рассмотрим вещественный реализуемый процесс чистой авторегрессии первого порядка, который можно записать в форме

$$\xi_n = a\xi_{n-1} + b\varepsilon_n, \quad (6.8)$$

где $a, b \in \mathbb{R}$ и $\mathbb{E}(\xi_m \varepsilon_n) = 0$ при $m < n$. Из этих условий сразу же следует, что $R_\xi(n-m) = aR_\xi(n-m-1)$ при любых $n > m$, то есть $R_\xi(k) = a^k R_\xi(0)$, $k > 0$. Тем самым $\rho_\xi(k) = a^k$ и, следовательно, $|a| \leq 1$. Для наглядности записи будем вместо a употреблять букву ρ . Также положим $R_\xi(0) = \sigma^2$ и выразим через ρ и σ число b .

Домножив обе части (6.8) на ξ_n и взяв математическое ожидание, получим, что

$$\sigma^2 = \rho R_\xi(1) + b\mathbb{E}\xi_n \varepsilon_n = \rho^2 \sigma^2 + b\mathbb{E}(\xi_n \varepsilon_n).$$

С другой стороны, домножив (6.8) на ε_n , получим после осреднения, что $\mathbb{E}(\xi_n \varepsilon_n) = b$. Тем самым приходим к равенству $b^2 = \sigma^2(1 - \rho^2)$. Выбирая знак “плюс” перед квадратным корнем, приходим к представлению

$$\xi_n = \rho\xi_{n-1} + \sigma\sqrt{1 - \rho^2} \varepsilon_n, \quad |\rho| \leq 1, \quad \sigma > 0, \quad (6.9)$$

которым и будем пользоваться в дальнейшем. Разберем сначала случай $|\rho| = 1$. Если $\rho = 1$, то $\xi_n = \xi_{n-1}$ и спектральная мера m_ξ процесса ξ_n сосредоточена в нуле. При $\rho = -1$, получаем, что $\xi_n = -\xi_{n-1}$ и, следовательно, m_ξ сосредоточена в точке π . В обоих случаях спектральной плотности не существует.

Для проверки условий существования спектральной плотности обратимся к Предложению 12. В нашем случае полином Q^* имеет вид $Q^*(z) = 1 - az$, то есть условие Предложения 12 приобретает вид $|a| \neq 1$. С учетом того, что мы исследуем реализуемые процессы, приходим к выводу, что необходимое и достаточное условие существования спектральной плотности у процесса авторегрессии (6.9) является неравенство $|\rho| < 1$, и при этом

$$f_\xi(\lambda) = \frac{1}{2\pi} \frac{\sigma^2(1 - \rho^2)}{|1 - \rho e^{i\lambda}|^2} = \frac{1}{2\pi} \frac{\sigma^2(1 - \rho^2)}{1 + \rho^2 - 2\rho \cos(\lambda)}. \quad (6.10)$$

Легко видеть, что спектральная плотность (6.10) монотонно убывает при $\rho > 0$ и монотонно возрастает при $\rho < 0$ на отрезке $[0, \pi]$ (а при $\rho = 0$ последовательность ξ_n просто является белым шумом). При этом

$$f_\xi(0) = \frac{1}{2\pi} \frac{\sigma^2(1 - \rho^2)}{(1 - \rho)^2} = \frac{\sigma^2}{2\pi} \frac{1 + \rho}{1 - \rho} \quad \text{и} \quad f_\xi(\pi) = \frac{\sigma^2}{2\pi} \frac{1 - \rho}{1 + \rho}.$$

Пусть $\rho \rightarrow 1$. Нетрудно видеть, что тогда $f_\xi(\lambda) \rightarrow 0$ при $\lambda > 0$ в то время как $f_\xi(0) \rightarrow \infty$. Это означает, что в спектральном разложении процесса (6.9) при $\rho \approx 1$ будут доминировать низкие частоты (а на уровне реализаций процесса — низкочастотные колебания). Поскольку в оптике такая ситуация характерна для красного цвета, то процесс авторегрессии с ρ , близким к $+1$, называют *красным шумом*.

Если $\rho \approx -1$, то ситуация аналогичная, только реализации процесса будут походить на последовательность с периодом 2.

Легко найти явное представление процесса авторегрессии (6.9) как линейного преобразования белого шума ε_n . Действительно,

$$\begin{aligned} \xi_n &= \sigma\sqrt{1 - \rho^2} \sum_{k=0}^{\infty} \rho^k \varepsilon_{n-k} = \sqrt{1 - \rho^2} \varepsilon_n + \sqrt{1 - \rho^2} \sum_{k=1}^{\infty} \rho^k \varepsilon_{n-k} = \\ &= \sqrt{1 - \rho^2} \varepsilon_n + \rho\sqrt{1 - \rho^2} \sum_{k=1}^{\infty} \rho^{k-1} \varepsilon_{n-k} = \sqrt{1 - \rho^2} \varepsilon_n + \rho\sqrt{1 - \rho^2} \sum_{\ell=0}^{\infty} \rho^\ell \varepsilon_{n-1-\ell} = \sqrt{1 - \rho^2} \varepsilon_n + \rho\xi_{n-1}. \end{aligned}$$

Наконец, обратимся к связи между реализуемыми процессами авторегрессии первого порядка и марковскими гауссовскими стационарными последовательностями.

Предложение 16. Пусть ξ_n , $n = 0, 1, \dots$, — вещественная гауссовская стационарная последовательность с нулевым средним. Для того, чтобы она была марковской, необходимо и достаточно, ξ_n было реализуемой авторегрессией первого порядка относительно некоторого гауссовского белого шума.

Доказательство. Пусть последовательность ξ_n марковская. Рассмотрим разложение

$$\xi_n = \mathbb{E}(\xi_n | \xi_{n-1}, \dots, \xi_0) + \sigma_n \varepsilon_n,$$

где $\mathbb{E}\varepsilon_n = 0$, $\mathbb{D}\varepsilon_n = 1$ и ε_n ортогональна любой случайной величине вида $\varphi(\xi_0, \dots, \xi_{n-1})$, имеющей конечный второй момент.

Поскольку вектор $(\xi_0, \dots, \xi_n)$ — гауссовский, то $\varepsilon_n \in N(0, 1)$, $\mathbb{E}(\xi_n | \xi_{n-1}, \dots, \xi_0) = \sum_{j=0}^{n-1} c_j \xi_j$ и случайная величина ε_n не зависит от $(\xi_0, \dots, \xi_n)$.

Марковость последовательности ξ_n означает, что $\mathbb{P}(\xi_n \in A | \xi_{n-1}, \dots, \xi_0) = \mathbb{P}(\xi_n \in A | \xi_{n-1})$ и поэтому $\mathbb{E}(\xi_n | \xi_{n-1}, \dots, \xi_0) = \mathbb{E}(\xi_n | \xi_{n-1}) = c_{n-1} \xi_{n-1}$, так как вектор (ξ_{n-1}, ξ_n) гауссовский. Наконец, из стационарности ξ_n следует, что числа σ_n и c_n не зависят от n .

Пусть теперь гауссовская стационарная последовательность ξ_n имеет вид

$$\xi_n = a\xi_{n-1} + b\varepsilon_n, \tag{6.11}$$

где ε_n ортогонален ξ_m при $m = 0, \dots, n-1$ и $\mathbb{E}\varepsilon_n^2 = 1$. Если в рекуррентной формуле (6.11) эту ортогональность можно заменить на независимость, то марковость последовательности становится очевидной по построению. Но, поскольку вектор (ξ_{n-1}, ξ_n) — гауссовский, то $\varepsilon_n \in N(0, 1)$ и ε_n не зависит от ξ_m . Доказательство завершено. $\square$

7 Авторегрессионные продолжения

7.1 Уравнения Юла-Уолкера

Пусть $R(0), \dots, R(p)$ — некоторые вещественные числа и $R(-k) = R(k)$. При $t \leq p+1$ обозначим $\Sigma_t = \{r_{kl}\}_{k,l=1}^t$ с $r_{kl} = R(k-l)$.

Лемма 9. Если матрица Σ_{p+1} неотрицательно определена, а матрица Σ_p положительно определена, то существует единственное решение $A = (a_1, \dots, a_p)^T$ системы

$$R(m) + \sum_{k=1}^p a_k R(m-k) = 0, \quad m = 1, \dots, p, \quad (7.1)$$

причем это решение удовлетворяет неравенству

$$\sigma^2 \stackrel{\text{def}}{=} R(0) + \sum_{k=1}^p a_k R(k) \geq 0. \quad (7.2)$$

Если к тому же матрица Σ_{p+1} положительно определена, то $\sigma^2 > 0$.

Доказательство. Обозначим $B = (R(1), \dots, R(p))^T$ и $A = (a_1, \dots, a_p)^T$. Тогда система (7.1) перепишется в виде $B + \Sigma_p A = \mathbf{0}$. Поскольку матрица Σ_p — невырожденная, то единственность вектора $A = -\Sigma_p^{-1}B$ очевидна.

Нам будет удобно при доказательстве (7.2) перейти на вероятностный язык. Прежде всего заметим, что

$$\sum_{k=1}^p a_k R(k) = B^T A,$$

и поэтому равенство (7.2) переписывается в виде

$$\sigma^2 = R(0) - B^T \Sigma_p^{-1} B.$$

Возьмем случайный вектор $(\eta_p, \dots, \eta_0)^T$ с распределением $\mathbb{N}(\mathbf{0}, \Sigma_{p+1})$ и заметим, что по определению $\mathbb{D}\eta_m = R(0)$ и $\mathbb{E}\eta_j \eta_{m-j} = R(m)$.

Рассмотрим условное математическое ожидание $\mathbb{E}(\eta_p | \bar{\eta}_{p-1})$ с $\bar{\eta}_{p-1} = (\eta_{p-1}, \dots, \eta_0)^T$, которое, как хорошо известно, является линейной функцией случайных величин $\eta_0, \dots, \eta_{p-1}$. Запишем этот факт в виде равенства

$$\mathbb{E}(\eta_p | \eta_{p-1}, \dots, \eta_0) = \sum_{k=1}^p c_k \eta_{p-k} = C^T \bar{\eta}_{p-1}.$$

Поскольку ковариационная матрица вектора $\bar{\eta}_{p-1}$ есть Σ_p , а $\mathbb{E}\eta_p \eta_{p-m} = R(m)$, то $C = \Sigma_p^{-1} B = -A$.

Как хорошо известно, дисперсия случайной величины $e_p \stackrel{\text{def}}{=} \eta_p - C^T \bar{\eta}_{p-1}$ (так называемая «остаточная дисперсия») имеет вид

$$0 \leq \mathbb{D}e_p = \mathbb{D}\eta_p - B^T \Sigma_p^{-1} B = R(0) + B^T A,$$

что в точности совпадает с (7.2).

Для окончания доказательства осталось заметить, что, если остаточная дисперсия равна нулю, то $e_p = 0$, η_p является линейной комбинацией $\eta_0, \dots, \eta_{p-1}$ и, следовательно, матрица Σ_{p+1} вырождена. $\square$

Лемма 10. Пусть $R(n)$ — ненулевая вещественная неотрицательно определенная функция и m — соответствующая спектральная мера. Если m обладает отделенной от нуля спектральной плотностью f , то для любого $p \geq 1$ матрица $\Sigma_p = \{R(l-k)\}_{k,l=1}^p$ является положительно определенной.

Доказательство. Рассмотрим комплекснозначные $z_0, \dots, z_{p-1}$. Так как

$$R(n) = \int_{-\pi}^{\pi} e^{in\lambda} f(\lambda) d\lambda,$$

то

$$\sum_{k,l=1}^p R(k-l) z_k \bar{z}_l = \int_{-\pi}^{\pi} \left(\sum_{k,l=1}^p e^{i(k-l)\lambda} z_k \bar{z}_l \right) f(\lambda) d\lambda = \int_{-\pi}^{\pi} \left| \sum_{k=1}^p e^{ik\lambda} z_k \right|^2 f(\lambda) d\lambda.$$

Поскольку функция

$$g(\lambda) = \left| \sum_{k=1}^p e^{ik\lambda} z_k \right|^2$$

представляет собой тригонометрический полином, то она (кроме тривиального случая $z_0 = \dots = z_{p-1} = 0$) имеет на отрезке $(-\pi, \pi)$ лишь конечное число корней. Следовательно,

$$\sum_{k,l=1}^p R(k-l) z_k \bar{z}_l = 0$$

лишь в случае $z_0 = \dots = z_{p-1} = 0$, и утверждение доказано. $\square$

Рассмотрим реализуемый вещественный авторегрессионный процесс порядка $p+1$ (или меньшего порядка, если $a_p = 0$):

$$\xi_n + \sum_{k=1}^p a_k \xi_{n-k} = \sigma \varepsilon_n, \quad (7.3)$$

где $\sigma > 0$ и $\varepsilon_1, \dots, \varepsilon_n, \dots$ — (нормированный и центрированный) белый шум. Как всегда, предполагается, что $\mathbb{E}\xi_n = 0$. Реализуемость ξ_n (относительно белого шума ε_n) означает, что $\mathbb{E}\xi_n \varepsilon_m = 0$ при $m > n$. Коэффициенты a_m предполагаются вещественными.

Спектральная плотность процесса авторегрессии (7.3) имеет вид

$$f_{\xi}(\lambda) = \frac{1}{2\pi} \frac{\sigma^2}{\left| 1 + \sum_{k=1}^p a_k e^{-ik\lambda} \right|^2}, \quad |\lambda| \leq \pi.$$

Предложение 17 (Уравнения Юла-Уолкера). Пусть $R_{\xi}(n)$ — ковариационная функция процесса (7.3). Тогда

1. При $m \geq 1$

$$R_{\xi}(m) + \sum_{k=1}^p a_k R_{\xi}(m-k) = 0. \quad (7.4)$$

2. Числа $a_1, \dots, a_p$ однозначно определяются системой уравнений (7.4) при $m = 1, \dots, p$ и удовлетворяют неравенству

$$R_{\xi}(0) + \sum_{k=1}^p a_k R_{\xi}(k) = \sigma^2 > 0. \quad (7.5)$$

Доказательство. Первое утверждение уже доказано в Теореме 11 о реализуемых процессах авторегрессии. Докажем второе утверждение. Умножим обе части (7.3) на ξ_n и возьмем математическое ожидание. Получим, что

$$R_{\xi}(0) + \sum_{k=1}^p a_k R_{\xi}(k) = \sigma \mathbb{E}\xi_n \varepsilon_n.$$

С другой стороны, домножая (7.3) на ε_n и снова беря математическое ожидание приходим к равенству

$$\mathbb{E}\xi_n\varepsilon_n = -\sum_{k=1}^p a_k \mathbb{E}\xi_{n-k}\varepsilon_n + \sigma \mathbb{E}\varepsilon_n^2 = \sigma.$$

Так как $\sigma^2 > 0$, то утверждение доказано. $\square$

Замечание 16. 1. Положительность левой части (7.5) непосредственно следует из лемм 9, 10.
2. Система уравнений (7.4) при $1 \leq m \leq p$ называется *системой уравнений Юла-Уолкера*.

7.2 Теорема об авторегрессионном продолжении

Предложение 18. Пусть η_n ($n \geq 0$) — вещественный стационарный в широком смысле процесс с ковариационной функцией $R_\eta(m)$ и ненулевой спектральной плотностью $f_\eta(\lambda)$. Тогда для любого $p \geq 0$ существует реализуемый процесс авторегрессии ξ_n порядка, не большего $p+1$, такой, что $R_\xi(m) = R_\eta(m)$ при $m = 0, \dots, p$.

Доказательство. Конечно, можно считать, что $\mathbb{E}\eta_n = 0$. При $p = 0$ результат очевиден: достаточно в качестве ξ_n взять белый шум с дисперсией $R_\xi(0)$. Рассмотрим случай $p > 0$.

Согласно Лемме 10, для любого p матрица $\Sigma_{p+1} = \{R_\eta(l-k)\}_{k,l=1}^{p+1}$ является положительно определенной. Поэтому (см. Лемму 9) ковариации $R_\eta(n)$ удовлетворяют равенствам

$$R_\eta(m) + \sum_{k=1}^p a_k R_\eta(m-k) = 0, \quad m = 1, \dots, p, \quad (7.6)$$

причем при фиксированном p числа $a_1, \dots, a_p$ определяются однозначно, а

$$\sigma_p^2 \stackrel{\text{def}}{=} R_\eta(0) + \sum_{k=1}^p a_k R_\eta(k) > 0. \quad (7.7)$$

Построим процесс ξ_n следующим образом. При $n = 0, \dots, p-1$ положим $\xi_n = \eta_n$. При $n \geq p$ определим

$$\xi_n = -\sum_{k=1}^p a_k \xi_{n-k} + \sigma_p \varepsilon_n, \quad (7.8)$$

где (центрированный и нормированный) белый шум ε_n ортогонален ξ_j при $0 \leq j < n$.

Рассмотрим сначала случай $n = p$, то есть равенство

$$\xi_p = -\sum_{k=1}^p a_k \xi_{p-k} + \sigma_p \varepsilon_p. \quad (7.9)$$

Домножив обе части (7.9) скалярно на ξ_ℓ при $0 \leq \ell < p$ и воспользовавшись тем, что $\xi_\ell = \eta_\ell$ и $\xi_{p-k} = \eta_{p-k}$, получим равенство

$$\mathbb{E}\xi_p \xi_\ell = -\sum_{k=1}^p a_k \mathbb{E}\eta_{p-k} \eta_\ell = -\sum_{k=1}^p a_k R_\eta(p-k-\ell). \quad (7.10)$$

Используя равенство (7.6) с $m = p-\ell$ получаем, что $\mathbb{E}\xi_p \xi_\ell = R_\eta(p-\ell)$, так как $1 \leq p-\ell \leq p$.

Теперь перейдем к случаю $\ell = p$. Тогда, так как $0 \leq p-k < p$, то по уже доказанному

$$\mathbb{E}\xi_p^2 = -\sum_{k=1}^p a_k \mathbb{E}\xi_{p-k} \xi_p + \sigma_p \mathbb{E}\xi_p \varepsilon_p = -\sum_{k=1}^p a_k R_\eta(k) + \sigma_p \mathbb{E}\xi_p \varepsilon_p,$$

Как следует из (7.9), $\mathbb{E}\eta_p \varepsilon_p = \sigma_p$ и поэтому окончательно получаем, что

$$\mathbb{E}\xi_p^2 = -\sum_{k=1}^p a_k \mathbb{E}\xi_{p-k} \xi_p + \sigma_p \mathbb{E}\xi_p \varepsilon_p = -\sum_{k=1}^p a_k R_\eta(k) + \sigma_p^2,$$

то есть, согласно (7.7), $\mathbb{E}\xi_p^2 = R_\eta(0)$.

Далее все по индукции. Предположим, что для любых ℓ, m таких, что $0 \leq \ell \leq m < n$ ковариация $\mathbb{E}\xi_n \xi_\ell$ зависит только от разности $m - \ell$ (обозначим эту ковариацию $R_\xi(m - \ell)$), а при $0 \leq m - \ell \leq p$ эта ковариация совпадает с $R_\eta(m - \ell)$. Нужно доказать, что это же свойство имеет место при $m = n$ и $0 \leq \ell \leq m$.

Обозначим лишь ход рассуждений. Снова рассматриваем (7.7), скалярно умножаем это равенство на ξ_ℓ при $\ell < n$ и приходим к равенству

$$\mathbb{E}\xi_n \xi_\ell = - \sum_{k=1}^p a_k \mathbb{E}\xi_{n-k} \xi_\ell.$$

Согласно индуктивному предположению, $\mathbb{E}\xi_{n-k} \xi_\ell$ зависит только от модуля разности $|n - \ell - k|$ и равна $R_\xi(|n - \ell - k|)$, так что $\mathbb{E}\xi_{n-k} \xi_\ell$ действительно зависит только разности $n - \ell$.

Далее, если $n - \ell \leq p$, то, как нетрудно видеть, $0 \leq |n - k - \ell| < p$ при $1 \leq k \leq p$ и $0 \leq \ell \leq n$. Поэтому при этих условиях $\mathbb{E}\xi_{n-k} \xi_\ell = R_\eta(n - k - \ell)$ и мы снова используем равенство (7.6).

Наконец, доказательство того, что $\mathbb{E}\xi_n^2 = R_\eta(0)$, повторяет уже приведенные рассуждения для $n = p$. Утверждение доказано. $\square$

Замечание 17. Как доказано в Предложении 18, процесс ξ_n является реализуемым процессом авторегрессии. Порядок этой авторегрессии может оказаться меньше p , если $a_p = 0$.

Последовательность $R_\xi(m)$ можно называть *авторегрессионным продолжением* отрезка $R_\eta(i)$, $0 \leq i \leq p$, ковариационной функции процесса η_n . Очевидно, ковариации $R_\xi(m)$ при $m > p$ определяются рекуррентными равенствами Юла-Уолкера (7.4).

Замечание 18. Предположим, что процесс η_n Предложения 18 является реализуемым процессом авторегрессии порядка d :

$$\eta_n = - \sum_{k=1}^d b_k \eta_{n-k} + \sigma \varepsilon'_n.$$

Если в Предложении 18 выбрать $p = d$, то авторегрессионным продолжением ряда $R_\eta(0), \dots, R_\eta(d)$ будет сама ковариационная функция $R_\eta(n)$. Действительно, в этом случае коэффициенты a_k в (7.6) будут равны b_k , а число σ_p^2 в (7.7) равно σ^2 . Поэтому последовательность (7.8) можно расценивать как другую (отличающуюся лишь выбором белого шума) реализацию исходного процесса η_n .

Если $p < d$, то это уже не так просто потому, что по определению $b_d \neq 0$, а процесс авторегрессии (7.8) имеет порядок, меньший d .

Рассмотрим случай $p > d$. Согласно (7.4),

$$R_\eta(m) = \sum_{k=1}^d b_k R_\eta(m - k)$$

при $m \geq 1$. В то же время выполнено (7.6), причем коэффициенты a_k однозначно определены. Отсюда сразу же следует, что $a_k = b_k$ при $k \leq d$ и $a_k = 0$ при $k > d$. Поэтому $\sigma_p^2 = \sigma$ и (как и в случае $p = d$) авторегрессионным продолжением отрезка $R_\eta(0), \dots, R_\eta(d)$ будет ковариационная функция $R_\eta(n)$.

7.3 Спектральные плотности авторегрессионных продолжений.

Пусть $p \rightarrow \infty$. Рассмотрим последовательность $\{R_\xi^{(p)}(m), m \geq 0\}$ авторегрессионных продолжений отрезков $R_\eta(0), \dots, R_\eta(p - 1)$ ковариационной функции процесса η_n . Естественнo ожидать, что спектральные плотности $f_\xi^{(p)}(\lambda)$, соответствующие ковариационным функциям $R_\xi^{(p)}$, будут сходиться к спектральной плотности $f_\eta(\lambda)$ исходного процесса η_n .

Легко доказать, что это имеет место в некотором слабом смысле. Для этого нам понадобится следующая лемма.

Лемма 11. Пусть $R_n(k)$ — последовательность ковариационных функций со спектральными мерами m_n , сосредоточенными на $(-\pi, \pi]$. Если для любого k имеет место сходимостъ $R_n(k) \rightarrow R(k)$, где $R(k)$ — ковариационная функция со спектральной мерой m , сосредоточенной в $(-\pi, \pi)$, то $m_n \Rightarrow m$ в том смысле, что $\int_{(-\pi, \pi]} g dm_n \rightarrow \int_{(-\pi, \pi]} g dm$ для любой ограниченной непрерывной в $(-\pi, \pi]$ функции g .

Доказательство. Перейдем от ковариационных функций к корреляционным. А именно, положим $\rho_n(k) = R_n(k)/R_n(0)$ и $\rho(k) = R(k)/R(0)$, а также $\mathcal{P}_n = m_n/R_n(0)$ и $\mathcal{P} = m/R(0)$. Ясно, что $\mathcal{P}_n$ и $\mathcal{P}$ являются распределениями и одновременно — спектральными мерами последовательностей ρ_n и ρ соответственно. Кроме того, сходимость $R_n(k) \rightarrow R(k)$ эквивалентна сходимости $\rho_n(k) \rightarrow \rho(k)$, а слабая сходимость $m_n \Rightarrow m$ — слабой сходимости $\mathcal{P}_n \Rightarrow \mathcal{P}$. Таким образом, мы имеем

$$\rho_n(k) = \int_{(-\pi, \pi]} e^{ik\lambda} \mathcal{P}_n(d\lambda) \rightarrow \rho(k) = \int_{(-\pi, \pi]} e^{ik\lambda} \mathcal{P}(d\lambda).$$

Пусть ξ_n — случайная величина, имеющая распределение $\mathcal{P}_n$, а $\mathcal{L}(\xi) = \mathcal{P}$. Тем самым $\mathbb{E}e^{ik\xi_n} \rightarrow \mathbb{E}e^{ik\xi}$. Введем случайные величины $\eta_n = (\xi_n + \pi)/2\pi$ и $\eta = (\xi + \pi)/2\pi$. Ясно, что распределение $\mathcal{Q}_n = \mathcal{L}(\xi_n)$ сосредоточено на $(0, 1]$, а распределение $\mathcal{Q} = \mathcal{L}(\xi)$ — на $(0, 1)$. Кроме того,

$$\rho_n(k) = \mathbb{E}e^{ik\xi_n} = \mathbb{E}e^{ik(2\pi\eta_n - \pi)} = e^{-\pi k} \mathbb{E}e^{i2\pi k\eta_n} \rightarrow \rho(k) = \mathbb{E}e^{ik\xi} = \mathbb{E}e^{ik(2\pi\eta - \pi)} = e^{-\pi k} \mathbb{E}e^{i2\pi k\eta}.$$

Поэтому $\mathbb{E}e^{i2\pi k\eta_n} \rightarrow \mathbb{E}e^{i2\pi k\eta}$. Из теоремы о характеристизации сходимости на единичном торе ([5, гл. 7 §7]) следует, что $\mathcal{Q}_n \Rightarrow \mathcal{Q}$. Поэтому (так как линейное преобразование непрерывно) $\mathcal{P}_n \Rightarrow \mathcal{P}$ и утверждение доказано. $\square$

Следствие 6. В случае, когда у предельной меры m нет нагрузок на всем промежутке $(-\pi, \pi]$, из этого утверждения следует в частности, что для любого отрезка $A = [a, b] \subset (-\pi, \pi]$ имеет место сходимость $m_n(A) \rightarrow m(A)$.

Кроме того, легко доказать, что любая непрерывная отделенная от нуля спектральная плотность может быть равномерно приближена спектральной плотностью некоторого процесса авторегрессии.

Предложение 19. Пусть $f(\lambda)$ — спектральная плотность некоторой вещественной стационарной последовательности. Если f непрерывна и отделена от нуля, то существует такой реализуемый процесс авторегрессии, что его спектральная плотность отличается от f сколь угодно мало в равномерной метрике.

Доказательство. Пусть $0 < m^2 \leq f(\lambda) \leq M^2$. Обозначим $h(\lambda) = 1/\sqrt{f(\lambda)}$. Тогда $1/M \leq h(\lambda) \leq 1/m$. Функция h , очевидно, непрерывна и может быть продолжена на всю вещественную ось по периодичности, причем продолжение тоже будет непрерывной функцией.

По теореме Вейерштрасса о тригонометрических полиномах функцию h можно равномерно приблизить с точностью до сколь угодно малого ε с помощью тригонометрического полинома Q периода 2π . Ясно, что $1/M - \varepsilon \leq Q(\lambda) \leq 1/m + \varepsilon$.

Поскольку всегда можно представить положительный полином Q в виде

$$Q(\lambda) = \sum_{k=0}^q c_k e^{-ik\lambda} = \left| \sum_{k=0}^q c_k e^{-ik\lambda} \right|,$$

то в обозначениях

$$\hat{f}(\lambda) = \frac{1}{|Q(\lambda)|^2}$$

получим, что

$$\begin{aligned} |\hat{f}(\lambda) - f(\lambda)| &\leq \left| \frac{1}{Q(\lambda)} - \frac{1}{h(\lambda)} \right| \left| \frac{1}{Q(\lambda)} + \frac{1}{h(\lambda)} \right| \leq \max_{\lambda} (1/h(\lambda) + 1/Q(\lambda)) \left| \frac{h(\lambda) - Q(\lambda)}{h(\lambda)Q(\lambda)} \right| \leq \\ &\leq \frac{2}{1/M - \varepsilon} \frac{\varepsilon M}{1/M - \varepsilon}. \end{aligned}$$

Тем самым $\hat{f}$ равномерно аппроксимирует f . $\square$

Предложение 19 говорит о том, что непрерывная строго положительная спектральная плотность может быть приближена сколь угодно точно с помощью спектральной плотности некоторого процесса авторегрессии. В то же время остается непонятным, можно ли это сделать с помощью обсуждаемой специальной процедуры.

Лемма 11, напротив, рассматривает (в том числе) эту специальную процедуру, но гарантирует лишь слабую сходимость спектральных мер вместо равномерной сходимости плотностей.

На самом деле имеет место следующий результат [6, разд. 3.5].

Теорема 12. Пусть $R_\eta(k)$ — ковариационная функция некоторого вещественного стационарного процесса, имеющего непрерывную ограниченную и отделенную от нуля спектральную плотность $f_\eta(\lambda)$, удовлетворяющую условию Гельдера с показателем γ . Обозначим $f_\xi^{(p)}(\lambda)$ спектральную плотность реализуемого процесса авторегрессии, первые p значений ковариационной функции которого совпадают со значениями R_η . Тогда существует такая положительная постоянная C , зависящая от f_η , что

$$|f_\xi^{(p)}(\lambda) - f_\eta(\lambda)| \leq Cp^{-\gamma}.$$

Эта теорема объясняет популярность обсуждаемого метода. Но доказательство теоремы довольно сложно и поэтому опускается.

Список литературы

- [1] В.В. Некруткин, Моделирование распределений, <http://statmod.ru>, 2014.
- [2] А.Н. Ширяев, Вероятность 1,2. М., Изд-во МЦНМО, 2004.
- [3] В. Феллер, Введение в теорию вероятностей и ее приложения, Т.2. М., Наука, 1967.
- [4] А.А. Боровков, Теория вероятностей. М., Эдиториал УРСС, 1999.
- [5] П. Биллингсли, Сходимость вероятностных мер. М., Наука, 1977.
- [6] У. Гренандер, Г. Сегё (1961), Теплицевы формы и их приложения, М., Изд. иностр. лит.