

Центральная предельная теорема: отклонения от условий независимости и одинаковой распределенности (материалы к специальному семинару)

В.В. Некруткин

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

Содержание

1	ЦПТ для независимых случайных величин	2
1.1	Схема серий и теорема Леви	2
1.2	Отказ от одинаковой распределенности: теорема Линдеберга-Феллера	4
2	ЦПТ для стационарных последовательностей	13
2.1	Стационарные вещественные последовательности и их свойства	13
2.1.1	Стационарные в широком смысле процессы	13
2.1.2	Стационарные в узком смысле последовательности.	18
2.1.3	Предельная дисперсия выборочного среднего стационарной последовательности	19
2.2	ЦПТ для стационарных последовательностей	22
2.2.1	ЦПТ для линейного процесса	22
2.2.2	ЦПТ для конечно-зависимых стационарных последовательностей	24

1 ЦПТ для независимых случайных величин

1.1 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 функций¹ $\varphi_n : \mathbb{R}^n \mapsto \mathbb{R}$ 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*²) 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 разницы³ 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 1.1. (*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 $B_n^2 \stackrel{\text{def}}{=} \text{D}S_n > 0$ при достаточно больших n . Если при

$$\mathcal{L}\left(\frac{S_n - \text{E}S_n}{B_n}\right) \Rightarrow N(0, 1), \quad (1.1)$$

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.

¹Вообще 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.

²В англоязычной литературе — «triangular array».

³В этом примере и вообще.

Определение 1.2. (ЦПТ для последовательностей случайных величин.)

Говорят, что последовательность случайных величин $\xi_1, \dots, \xi_n, \dots$, обладающих математическими ожиданиями $E\xi_k = a_k$ и конечными вторыми моментами, удовлетворяет центральной предельной теореме, если

$$\mathcal{L}\left(\frac{\xi_1 + \dots + \xi_n - (a_1 + \dots + a_n)}{B_n}\right) \Rightarrow N(0, 1), \quad (1.2)$$

где $B_n^2 = D(\xi_1 + \dots + \xi_n)$.⁴

Сформулируем и докажем теорему Поля Леви в схеме серий.

Теорема 1.1. (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. □

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 1.1.

Следствие 1.1. (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 $a = E\xi_i$.

Доказательство. Для доказательства достаточно положить $\xi_{in} = \xi_i$ при $n \geq i$ и воспользоваться Теоремой 1.1. □

⁴Здесь и далее в аналогичных случаях подразумевается, что $B_n^2 > 0$ при всех достаточно больших n .

1.2 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 (1.1) и (1.2).

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 1.2. (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 $\eta_{in} = \xi_{in} - E\eta_{in}$.

1. Если для любого $\delta > 0$

$$\frac{1}{B_n^2} \sum_{i=1}^n E(\eta_{in}^2, |\eta_{in}| > \delta B_n) \rightarrow 0 \quad (1.3)$$

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,

$$\frac{1}{B_n^2} \max_{1 \leq i \leq n} \sigma_{in}^2 \rightarrow 0, \quad (1.4)$$

то для любого ε имеет место сходимость (1.3).

Конечно, из Теоремы 1.2 вытекает ее аналог для последовательностей случайных величин.

Следствие 1.2. (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 $\eta_i = \xi_i - E\eta_i$. Кроме того, будем считать, что $B_n > 0$ при достаточно больших n .⁵

1. Если для любого $\delta > 0$

$$\frac{1}{B_n^2} \sum_{i=1}^n E(\eta_i^2, |\eta_i| > \delta B_n) \rightarrow 0 \quad (1.5)$$

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,

$$\frac{1}{B_n^2} \max_{1 \leq i \leq n} \sigma_i^2 \rightarrow 0, \quad (1.6)$$

то для любого ε имеет место сходимость (1.5).

Условия (1.3) и (1.5)⁶ называются *условиями Линдеберга*, условия (1.4) и (1.6) — условиями *асимптотической пренебрежимости слагаемых*⁷. Последний термин понятен: дисперсия каждого слагаемого в сумме S_n гораздо меньше дисперсии всей суммы.

Мы не будем заниматься доказательством теоремы Линдеберга-Феллера,⁸ его можно найти в любом достаточно продвинутом учебнике по теории вероятностей.⁹ Это доказательство достаточно громоздко, хотя его идея точно такая же, как идея доказательства теоремы Леви — нужно рассматривать характеристические функции и аккуратно считать их предел.

⁵Иначе говоря, не все случайные величины ξ_{in} постоянны.

⁶Подразумевается, что они выполнены для любого $\varepsilon > 0$.

⁷В сумме S_n .

⁸Тем не менее — выведите Следствие 1.2 из Теоремы 1.2!

⁹Например, в [1, т.1, гл. III, §4] или в [2, гл. XIV §6]. Первый пункт теоремы был доказан Я. Линдебергом в 1922 году, второй — В. Феллером в 1935 году.

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 (1.2) налагается на «хвосты» дисперсий, то есть на то, насколько быстро выражения $E(\eta_i^2, |\eta_i| > \delta B_n)$ стремятся к нулю при $n \rightarrow \infty$.¹⁰

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 .¹¹

Далее, ЦПТ может выполняться и без условия (1.6). Действительно, если взять случайные величины $\xi_i \in N(a_i, \sigma_i^2)$ независимыми, то левая часть (1.2) будет иметь стандартное нормальное распределение при любом n и любых σ_i^2 . Нетрудно подобрать эти дисперсии так,¹² чтобы (1.6) не выполнялось.

Для прояснения роли условия асимптотической пренебрежимости слагаемых докажем следующее утверждение.

Лемма 1.1. Пусть выполнено условие Линдеберга (1.5). Тогда

1. Имеет место сходимость (1.6);
2. Для любого $\delta > 0$

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

3. $B_n \rightarrow \infty$;

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

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

а B_n^2 должно стремиться к бесконечности, так как не все дисперсии σ_i^2 в левой части (1.6) нулевые.

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

$$\begin{aligned} \frac{\sigma_i^2}{B_n^2} &= \frac{1}{B_n^2} \left(E(\xi_i^2, |\xi_i| > \delta B_n) + E(\xi_i^2, |\xi_i| \leq \delta B_n) \right) \leq \\ &\leq \frac{1}{B_n^2} \left(\sum_{i=1}^n E(\xi_i^2, |\xi_i| > \delta B_n) + \delta^2 B_n^2 \right) = \frac{1}{B_n^2} \left(\sum_{i=1}^n E(\xi_i^2, |\xi_i| > \delta B_n) \right) + \delta^2. \end{aligned} \quad (1.7)$$

Поскольку правая часть (1.7) 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. $\square$

Замечание 1.1. Отметим, что для схемы серий пункты 1 и 2 Леммы 1.1 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.¹³

¹⁰ Попробуйте построить пример двух последовательностей независимых случайных величин с попарно одинаковыми дисперсиями, первая из которых удовлетворяла бы ЦПТ, а вторая — нет.

¹¹ Так и будем делать, если не оговорено противное.

¹² А как?

¹³ Приведите пример!

Условие асимптотической пренебрежимости имеет свою интерпретацию, но ее удобнее проводить в терминах схемы серий. Более того, перепишем условия (1.3) и (1.4) 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 (1.3) переписывается как

$$\sum_{i=1}^n E(\beta_{in}^2, |\beta_{in}| > \delta) \rightarrow 0, \quad (1.8)$$

а (1.4) — как $\max_{1 \leq i \leq n} D\beta_{in} \rightarrow 0$, Сама же ЦПТ приобретает вид

$$\mathcal{L}(\beta_{1n} + \dots + \beta_{nn}) \Rightarrow N(0, 1). \quad (1.9)$$

Именно такая запись принята, например, в [1].

Пусть теперь мы имеем «результат измерения»¹⁴ 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.¹⁵

Часто предполагают, что ε имеет нормальное распределение $N(0, \sigma^2)$ с какой-то дисперсией σ^2 .¹⁶ Словесно это предположение можно объяснить следующим образом.

Естественно считать, что происхождение ошибки ε связано с большим числом каких-то независимых воздействий на эксперимент.¹⁷ 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 $\sum_{1 \leq i \leq n} D\beta_{in} = 1$.¹⁸

Отсюда сразу становится ясно, что предположение о нормальности ε тесно связано с ЦПТ в форме (1.9). Конечно, формально эту ЦПТ нужно доказывать, например, проверять условие Линдеберга в форме (1.8). 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 двух¹⁹ 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,²⁰ 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 1.2 следует, что ЦПТ выполняется для любых независимых одинаково распределенных случайных величин с конечной положительной дисперсией σ^2 .²¹ Действительно, рассмотрим для определенности условие Линдеберга (1.5) для последовательностей. В нашем случае

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

а последнее выражение стремится к нулю при $n \rightarrow \infty$ ввиду²² конечности дисперсии случайной величины η_1 .

¹⁴То есть случайную величину.

¹⁵Если $E\gamma \neq a$, то это означает, что мы измеряем не a , а что-то другое.

¹⁶Конечно, это очень удобное предположение. Например, оно позволяет на практике строить доверительные интервалы для a даже при небольшом числе измерений. А как?

¹⁷Ну, например, что-то куда-то случайно сдвинулось и т.п.

¹⁸Конечно, 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 ссылаться на ЦПТ. Случайные величины β_{in} считаются «элементарными», и поэтому не делается никаких предположений об их распределении — кроме существования вторых моментов.

¹⁹Например, $\beta_1 = \beta_{1n}$, а $\beta_2 = \beta_{2n} + \dots + \beta_{nn}$.

²⁰А как это обосновать? Подсказка: используйте аппарат слабой сходимости распределений.

²¹Иначе говоря, из теоремы Линдеберга следует утверждение теоремы Леви.

²²А подробнее?

Гораздо интереснее и полезнее другое следствие из теоремы Линдеберга, которое мы для разнообразия сформулируем и докажем в терминах схемы серий.²³

Теорема 1.3. (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

$$\frac{1}{B_n^\alpha} \sum_{i=1}^n E|\eta_{in}|^\alpha \rightarrow 0 \quad \text{при } n \rightarrow \infty, \quad (1.10)$$

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 (1.10)

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

что и заканчивает доказательство. $\square$

Правую часть (1.10) обычно называют *дробью Ляпунова*, а саму сходимость (1.10) — *условием Ляпунова*.

Замечание 1.2. 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).²⁴

Приведем несколько примеров использования условий Линдеберга и Ляпунова.

Пример 1.1. 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.

²³А отсюда она (вернее, ее надлежащий вариант) будет вытекать и для последовательностей. Не правда ли?

²⁴А почему «более того»?

Пример 1.2. Пусть $\xi_1, \dots, \xi_n, \dots$ — последовательность независимых случайных величин таких, что они все являются ограниченными²⁵ 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 (1.5) при больших n просто обнуляется и условие Линдеберга оказывается выполненным.

Конечно, эти рассуждения верны и для схемы серий. □

Для третьего примера нам понадобится следующее утверждение.

Предложение 1.1. Пусть $\varepsilon_1, \dots, \varepsilon_n, \dots$ последовательность независимых одинаково распределенных случайных величин с нулевым средним и единичной дисперсией. Обозначим $\xi_i = a_i \varepsilon_i$, где a_i — 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 последовательность $\{\xi_n, n \geq 1\}$ удовлетворяет ЦПТ.

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

$$\frac{1}{B_n^2} \sum_{i=1}^n E(\xi_i^2, |\xi_i| > \delta B_n) \leq \frac{c^2}{B_n^2} \sum_{i=1}^n E(\varepsilon_i^2, |\varepsilon_i| > \delta B_n/c) = \frac{c^2 n}{B_n^2} E(\varepsilon_1^2, |\varepsilon_1| > \delta B_n/c) \rightarrow 0$$

что очевидным²⁶ образом стремится к нулю. □

Рассмотрим частный случай Предложения 1.1, имеющий отношение к статистике случайных процессов.

Пример 1.3. Пусть $\omega \in (0, 1/2)$ и $\{\varepsilon_n, n \geq 1\}$ — последовательность независимых 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, что²⁷

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

Тем самым мы попадаем в условия Предложения 1.1, так как $|\sin(2\pi\omega n)| \leq 1$ и $B_n^2 \rightarrow \infty$. 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. □

Следующий пример относится к методу Монте-Карло и связан со стратифицированными выборками в статистике.

²⁵С вероятностью 1.

²⁶Действительно?

²⁷Докажите 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 $\xi_n = \cos(2\pi\omega n)\varepsilon_n$.

Пример 1.4. Рассмотрим интеграл $J = \int_0^1 f(x)dx$, где функция f непрерывна и имеет непрерывную ограниченную производную на отрезке $[0, 1]$. Простейший метод Монте-Карло для вычисления интеграла J состоит в моделировании независимых случайных величин $\alpha_1, \dots, \alpha_n$, равномерно распределенных на $[0, 1]$ и использовании случайной величины

$$\hat{J}_n = \frac{1}{n} (f(\alpha_1) + \dots + f(\alpha_n)) \quad (1.11)$$

в качестве оценки J .

В этих общих построениях, однако, не учитывается гладкость функции f . Оказывается, что при учете этой гладкости можно построить оценку J , быстрее сходящуюся к J , чем стандартная.

Для этого выберем большое n и разделим отрезок $(0, 1]$ на n отрезков вида $D_{i,n} = ((i-1)/n, i/n]$, $i = 1, \dots, n$.

На каждом из таких отрезков независимо промоделируем случайную величину $\beta_{in} \in U(D_{in})$ и рассмотрим случайную величину

$$\hat{I}_n = \frac{1}{n} (f(\beta_{1n}) + \dots + f(\beta_{in})) \quad (1.12)$$

как новую оценку интеграла J . Так как

$$J_{in} \stackrel{\text{def}}{=} E f(\beta_{in}) = n \int_{(i-1)/n}^{i/n} f(x)dx,$$

то новая оценка является несмещенной.

Докажем, что

$$\mathcal{L} \left(n^{3/2} (\hat{I}_n - J) \right) \Rightarrow N(0, \sigma_f^2), \quad (1.13)$$

где

$$\sigma_f^2 = \frac{1}{12} \int_0^1 (f'(x))^2 dx. \quad (1.14)$$

Начнем с проверки условия Ляпунова.

Начнем с изучения дисперсии оценки $\hat{I}_n$. Поскольку все β_{in} независимы, то

$$D\hat{I}_n = \frac{1}{n^2} \sum_{i=1}^n Df(\beta_{in}). \quad (1.15)$$

Сначала приведем немного неточные (но на самом деле правильные) соображения показывающие, откуда берется правая часть равенства (1.14).

Обозначив $x_i = i/n$, получим, что на отрезке $((i-1)/n, i/n]$

$$f(x) = f(x_{i-1}) + f'(x_{i-1})(x - x_{i-1}) + o(1/n), \quad (1.16)$$

причем остаточный член стремится к нулю равномерно²⁸ по $1 \leq i \leq n$. Пренебрегая этим остаточным членом, введем на отрезке $((i-1)/n, i/n]$ линейную функцию

$$g_{in}(x) = f(x_{i-1}) + f'(x_{i-1})(x - x_{i-1})$$

и будем вместо дисперсии (1.15) оценки (1.12) изучать дисперсию

$$D\hat{I}_n^* = \frac{1}{n^2} \sum_{i=1}^n Dg_{in}(\beta_{in})$$

²⁸ А почему равномерно? А как эта равномерность проявляется при строгих построениях?

«испорченной» оценки

$$\widehat{I}_n^* = \frac{1}{n} (g_1(\beta_{1n}) + \dots + g_{in}(\beta_{in})),$$

справедливо предполагая, что она будет не слишком отличаться от искомой дисперсии (1.15).

Поскольку

$$g_{in}(\beta_{in}) = f(x_{i-1}) + f'(x_{i-1})(\beta_{in} - x_{i-1}),$$

а случайная величина $\beta_{in} - x_{i-1}$ равномерно распределена на $[0, 1/n]$,²⁹ то $Dg_{in}(\beta_{in}) = n^{-2}(f'(x_{i-1}))^2/12$ и

$$D\widehat{I}_n^* = \frac{1}{12n^4} \sum_{i=1}^n (f'(x_{i-1}))^2 \sim \frac{1}{12n^3} \int_0^1 (f'(x))^2 dx \quad (1.17)$$

при $n \rightarrow \infty$, где в последнем переходе использованы свойства сумм Римана.

Таким образом, у нас есть основания предположить, что дисперсия оценки (1.12) определяется линейными членами аппроксимации функции f на отрезках $[x_{i-1}, x_i]$ и главный член ее дисперсии при больших n имеет вид правой части соотношения (1.17).

Заметим кстати, что

$$E\widehat{I}_n^* = \frac{1}{n} \sum_{i=1}^n E g_{in}(\beta_{in}) = \frac{1}{n} \sum_{i=0}^{n-1} f(i/n) + \frac{1}{2n^2} \sum_{i=0}^{n-1} f'(i/n) = \int_0^1 f(x) dx + O(n^{-2}).$$

Займемся теперь Центральной Предельной Теоремой в схеме серий, обозначив $\xi_{in} = g_{in}(\beta_{in})/n$. Как уже доказано, $E(\xi_{1n} + \dots + \xi_{nn}) = J$ и

$$B_n^2 = D(\xi_{1n} + \dots + \xi_{nn}) = D\widehat{I}_n^* \sim \frac{\sigma_f^2}{12n^3}.$$

Таким образом, нам осталось изучить поведение выражения

$$\sum_{i=1}^n E \xi_{in}^4 = \frac{1}{n^4} \sum_{i=1}^n E (g_{in}(\beta_{in}) - J_{in}^*)^4 = \frac{1}{n^3} \sum_{i=1}^n \int_{(i-1)/n}^{i/n} (g_{in}(x) - J_{in}^*)^4 dx, \quad (1.18)$$

где $J_{in}^* = f((i-1)/n)$.

Поскольку $|g_{in}(x) - J_{in}^*| \leq c/n$, где c не зависит от n , то левая часть (1.18) имеет порядок убывания $O(n^{-7})$, в то время как B_n^4 убывает как n^{-6} . Тем самым дробь Ляпунова стремится к нулю и сходимость (1.1) доказана с заменой $\widehat{I}_n$ на $\widehat{I}_n^*$.

Теперь перейдем к точным оценкам. Легко видеть³⁰, что

$$D\widehat{I}_n = \frac{1}{n^2} \sum_{i=1}^n n \int_{(i-1)/n}^{i/n} (f(x) - J_{in})^2 dx.$$

Далее, используя (1.16), получим, что

$$J_{in} = f(x_{i-1}) + \frac{1}{2n} f'(x_{i-1}) + o(n^{-1}) \quad (1.19)$$

и, проводя несложные вычисления с учетом (1.16) и (1.19),

$$\begin{aligned} \int_{(i-1)/n}^{i/n} (f(x) - J_{in})^2 dx &= \int_{(i-1)/n}^{i/n} \left(f(x) - f(x_{i-1}) - \frac{1}{2n} f'(x_{i-1}) + o(n^{-1}) \right)^2 dx = \\ &= \frac{1}{12n^3} (f'(x_{i-1}))^2 + o(n^{-3}). \end{aligned} \quad (1.20)$$

²⁹И, следовательно, ее дисперсия равна $n^{-2}/12$.

³⁰Действительно?

Следовательно, используя непрерывность производной f' ,³¹ получим, что

$$D\hat{I}_n = \frac{1}{12n^4} \sum_{i=1}^n (f'(x_{i-1}))^2 + o(n^{-3}) = \frac{\sigma_f^2}{n^3} + o(n^{-3}).$$

Тем самым дисперсия новой оценки имеет вид $O(n^{-3})$, в то время как традиционная оценка (1.11) имеет дисперсию вида $O(n^{-1})$.

Проверка условия Ляпунова (с тем же 4-м моментом) происходит точно так же, как и раньше. Поэтому, заменяя B_n на ее асимптотику, приходим к (1.13).

Замечание 1.3. 1. Таким образом, скорость сходимости (например, длина доверительного интервала) в рассматриваемой задаче оказалась равной $n^{-3/2}$ вместо обычных $n^{-1/2}$. Нужно отметить, что эта быстрая скорость сходимости не является универсальной. Если рассмотреть аналогичную задачу не на отрезке прямой, а в многомерном единичном кубе,³² то преимущество стратифицированного метода Монте-Карло над стандартным останется, но будет уменьшаться при росте размерности куба.

Кроме того, здесь существенно используются свойства гладкости подинтегральной функции, неприемлемые при общей постановке задачи о вычислении зарядов методом Монте-Карло.

2. Если мы умеем вычислять значение в произвольной точке не только функции f , но и ее производной, то нет проблем³³ с построением состоятельной оценки асимптотической дисперсии σ_f^2 .

Если же производная f' нам недоступна, то все равно ее состоятельную оценку можно построить, используя приближения типа $f'(x_i) \approx n(f(x_{i+1}) - f(x_i))$.

Если мы рассматриваем независимые и одинаково распределенные случайные вектора, переход от одномерной к многомерной ЦПТ осуществляется очень просто, соответствующее утверждение называется многомерной теоремой Леви. В общем случае этот переход не является таким очевидным.³⁴

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

Лемма 1.2. 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. □

Пример 1.5. В условиях Примера 1.3 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.

³¹А как?

³²Это делается аналогично.

³³Действительно?

³⁴Например, многомерный аналог теоремы Линдеберга можно найти в [3, гл.8 §7].

Доказательство. Прежде всего заметим, что³⁵

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

при $n \rightarrow \infty$.

Далее будем использовать Лемму 1.2. 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 1.1 и Лемму 1.2, получаем нужный нам результат. □

³⁵Убедитесь!

2 ЦПТ для стационарных последовательностей

В предыдущих разделах обсуждались условия, при которых ЦПТ имеет место для независимых, но не обязательно одинаково распределенных случайных величин. Здесь мы рассмотрим ситуацию, когда исходные случайные величины ξ_i не обязаны быть независимыми.

Такая постановка задачи является гораздо более неопределенной, чем предыдущая, так как отклонения от независимости у случайных величин могут быть чрезвычайно многообразными. Поэтому мы ограничимся³⁶ рассмотрением так называемых стационарных последовательностей. При этом, имея в виду статистические применения, будем писать x_i вместо ξ_i .

2.1 Стационарные вещественные последовательности и их свойства

2.1.1 Стационарные в широком смысле процессы

Определение 2.1. Вещественная³⁷ случайная последовательность $\{x_n, n \geq 1\}$ с конечными вторыми моментами называется *стационарной в широком смысле*³⁸, если

1. математические ожидания случайных величин x_n одинаковы, то есть $E x_n = a$;
2. для любых $m \geq 1$ и $n \geq 0$ математическое ожидание $E x_m x_{m+n}$ 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 2.1 сведется к двум пунктам: $E x_n = a$ и $\text{Cov}(x_m, x_n) = R(m - n)$.

Функция R называется *ковариационной функцией* последовательности $\{x_n, n \geq 1\}$. Ясно, что $R(0) = D x_n$ для любого $n \geq 1$.³⁹ Конечно, ковариационная функция не зависит от сдвига последовательности x_n на любую постоянную, поэтому при изучении ее свойств обычно полагают $E x_n = 0$.

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 $z_1, \dots, z_n$

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

Действительно, полагая $E x_n = 0$, получим, что

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

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

Теорема 2.1. (Теорема Герглота).

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

$$R(n) = \int_{(-\pi, \pi]} e^{in\lambda} \mu(d\lambda). \quad (2.1)$$

³⁶И то не в полной мере.

³⁷В общей теории рассматриваются комплекснозначные случайные последовательности.

³⁸Или *слабо стационарной*. Такое же определение дается и для двусторонней бесконечной последовательности.

³⁹Везде далее мы предполагаем, что $R(0) > 0$.

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-Никодима⁴⁰ f по мере Лебега на интервале $(-\pi, \pi]$. Тогда f называется *спектральной плотностью*, а (2.1) приобретает вид

$$R(n) = \int_{-\pi}^{\pi} e^{in\lambda} f(\lambda) d\lambda = \int_{-\pi}^{\pi} \cos(n\lambda) f(\lambda) d\lambda. \quad (2.2)$$

Из (2.2) сразу же следует, что спектральную плотность можно считать⁴¹ 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-Лебега⁴² сразу же следует, что 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 2.1. Если $\sum_m |R(m)| < \infty$, то

$$f(\lambda) = \frac{1}{2\pi} \sum_m e^{-im\lambda} R(m) = \frac{1}{2\pi} \sum_m e^{im\lambda} R(m) = \frac{1}{2\pi} \left(R(0) + 2 \sum_{m=1}^{\infty} \cos(m\lambda) R(m) \right), \quad (2.3)$$

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

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

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 2.1. 1. Из равенства (2.3) сразу же следует, что

$$R(0) + 2 \sum_{m=1}^{\infty} R(m) = 2\pi f(0). \quad (2.4)$$

⁴⁰Конечно, эта производная является неотрицательной почти всюду.

⁴¹А почему только «можно считать»? И почему это «можно»?

⁴²А что это такое?

2. Рассуждениями, аналогичными приведенным в Предложении 2.1, можно доказать,⁴³ 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.⁴⁴ Таким образом, гладкость спектральной плотности оказывается тесно связанной со скоростью убывания ковариационной функции.

Приведем несколько примеров стационарных в широком смысле последовательностей.

Пример 2.1. 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 мера на этот раз имеет вид⁴⁵ $\sigma^2 \delta_\pi(d\lambda)$.

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

$$R(n) = \begin{cases} \sigma^2 & \text{при } n = 0, \\ 0 & \text{иначе,} \end{cases} \quad (2.5)$$

где $\sigma^2 = Dx_n$. Спектральной плотностью такой последовательность 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 (2.5).

Стационарную последовательность с ковариационной функцией (2.5) обычно называют *белым шумом*, добавляя слова *с дисперсией σ^2* . Если $\sigma^2 = 1$, то говорят о *стандартном белом шуме*, часто опуская слово *стандартный*.⁴⁶ 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

$$x_n = \rho x_{n-1} + \sigma \sqrt{1 - \rho^2} \varepsilon_n, \quad n \geq 1. \quad (2.6)$$

Тогда (это легко проверяется по индукции) $E x_n = 0$. Сосчитаем ковариацию $E x_n x_m$, считая, что $n > m$. Для этого заметим,⁴⁷ 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 (2.6) на x_m и беря математическое ожидание от обеих частей),

$$E x_n x_m = \rho E x_{n-1} x_m + \sigma \sqrt{1 - \rho^2} E x_m \varepsilon_n = \rho E x_{n-1} x_m = \rho^{n-m} E x_m^2 = \rho^{n-m} \sigma^2.$$

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

⁴³ Попробуйте!

⁴⁴ А как эта производная будет выражаться через ковариационную функцию?

⁴⁵ Проверьте!

⁴⁶ А если при этом $\varepsilon_n \in N(0, 1)$ (и, следовательно, ε_n независимы), то о *гауссовском белом шуме*.

⁴⁷ Проверьте!

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 (2.6) 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 2.2. Пусть $\{\varepsilon_n\}_{n=-\infty}^{\infty}$ — двусторонняя последовательность случайных величин, причем $E\varepsilon_n = 0$ и $E\varepsilon_n \varepsilon_m = \delta_{nm}$.

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

$$x_n = \sum_m c_{n-m} \varepsilon_m \quad (2.7)$$

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

$$R(n) = \sum_{\ell=-\infty}^{\infty} c_{\ell} c_{\ell+n} \quad (2.8)$$

и спектральной плотностью

$$f(\lambda) = \frac{1}{2\pi} \left| \sum_{m=-\infty}^{\infty} c_m e^{-i\lambda m} \right|^2. \quad (2.9)$$

Доказательство. Заметим, что из сходимости ряда $\sum_m c_m^2$ следует⁴⁸ (абсолютная) сходимость рядов вида $\sum_{\ell} c_{\ell} c_{\ell+n}$.

Далее, ряд в правой части (2.7) сходится в среднеквадратическом.⁴⁹ 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

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

⁴⁸ Действительно?

⁴⁹ А почему?

Делая в последнем выражении замену $\ell = t - j - m$, приходим к (2.8) и заодно доказываем, что x_n является стационарной последовательностью.

Наконец,

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

что и доказывает, что функция (2.9) действительно определяет спектральную плотность. $\square$

Легко написать в терминах коэффициентов c_j достаточные условия гладкости спектральной плотности для последовательности (2.7). 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 (2.9) будет непрерывной.

Более того, имеет место следующее утверждение.

Лемма 2.1. 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. $\square$

Переход к свойствам спектральной плотности происходит с помощью Следствия 2.1.

2.1.2 Стационарные в узком смысле последовательности.

Определение 2.2. 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.⁵⁰

Конечно, стационарная в узком смысле последовательность не обязана быть стационарной в широком смысле, так как в Определении 2.2 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.⁵¹

Приведем примеры стационарных в узком смысле последовательностей, в основном основываясь на Примере 2.1, относящемся к стационарным в широком смысле последовательностям.

Пример 2.2. 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.⁵²

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

3. Рассмотрим процесс авторегрессии (2.6) 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 (2.6) 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 (2.6) будет сильно стационарной, если белый шум ε_n является гауссовским.⁵³

4. Еще один пример — однородная марковская цепь, начальное распределение которой совпадает со стационарным распределением этой марковской цепи.⁵⁴

5. Другой пример, не связанный с Примером 2.1 — это последовательность $x_n = \xi_n \xi_{n+1}$, где ξ_i — независимые одинаково распределенные случайные величины.⁵⁵

Следующий пример будет в дальнейшем для нас основным.

Теорема 2.2. Пусть $\varepsilon_i, i \in \mathbb{Z}$ 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

$$x_n = \sum_m c_{n-m} \varepsilon_m. \quad (2.10)$$

Тогда для любого n ряд (2.10) сходится в среднеквадратическом, причем $E x_n = 0$, а последовательность $x_1, \dots, x_n, \dots$ является сильно стационарной.

⁵⁰ Иногда в этом случае употребляют термин *сильно стационарная последовательность*. Понятие сильно стационарной последовательности автоматически переносится на случайные величины, принимающие значения в произвольном измеримом пространстве.

⁵¹ А почему?

⁵² А какие у них конечномерные распределения?

⁵³ А почему?

⁵⁴ Действительно?

⁵⁵ Действительно? А при каких условиях такая последовательность будет слабо стационарной?

Доказательство. Сходимость ряда (2.10) в среднеквадратическом (как и равенство $Ex_n = 0$) уже доказано при рассмотрении в Предложении 2.2. Для доказательства сильной стационарности перепишем (2.10) 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.⁵⁶ $\square$

Замечание 2.1. 1. Если дополнительно потребовать, чтобы $\sum_{i \in \mathbb{Z}} |c_i| < \infty$, то последовательность будет обладать ковариационной функцией, определенной в (2.8) и спектральной плотностью (2.9). 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.

2.1.3 Предельная дисперсия выборочного среднего стационарной последовательности

Для нас предельные теоремы в первую очередь интересны с точки зрения математической статистики. Пусть у нас есть некоторая выборка⁵⁷ $x_1, \dots, x_n$ 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) несмещенных⁵⁸ 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,⁵⁹ 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

⁵⁶ А если точнее?

⁵⁷ Вообще говоря, не обязательно независимая и не обязательно повторная.

⁵⁸ А если они не такие, то вообще разговаривать не о чем.

⁵⁹ Пока не важно, в широком или узком смысле.

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 хватит⁶⁰ 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.

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

$$\mathcal{L}\left(\frac{S_n - na}{B_n}\right) \Rightarrow N(0, 1), \quad (2.11)$$

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 (2.11) 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

$$\mathcal{L}(\sqrt{n} (\bar{x}_n - a)) \Rightarrow N(0, \sigma^2). \quad (2.12)$$

Поскольку $DS_n = B_n^2/n \rightarrow \sigma^2$, то число σ^2 можно называть *асимптотической*⁶¹ дисперсией выборочного среднего $\bar{x}_n$.

Заметим, что (2.12) следует из (2.11) 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 2.2. *Дисперсия $D\bar{x}_n$ может быть представлена в виде*

$$D\bar{x}_n = \frac{1}{n} \sum_{j=-(n-1)}^{n-1} (1 - |j|/n) R(j). \quad (2.13)$$

Доказательство. Действительно,

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

что и заканчивает доказательство. □

⁶⁰ А почему?

⁶¹ Или *предельной*.

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

Лемма 2.3. *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. $\square$

Теперь не представляет труда найти асимптотическую дисперсию для выборочного среднего.

Предложение 2.3. *Если $\sum_m |R(m)| < \infty$, то при $n \rightarrow \infty$*

$$nD \bar{x}_n \rightarrow R(0) + 2 \sum_{m \geq 1} R(m) = 2\pi f(0). \quad (2.14)$$

Доказательство. Утверждение вытекает из равенства (2.13), Леммы 2.3 и Следствия 2.1. $\square$

Замечание 2.2. 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.

2.2 ЦПТ для стационарных последовательностей

В разделе 1.2 мы обсуждали ЦПТ, которая отличается от теоремы Леви тем, что там допускается отклонение от условия одинаковой распределенности случайных величин.⁶² Теперь мы откажемся от условия независимости, но зато сохраним условие одинаковой распределенности. Иначе говоря, мы будем рассматривать стационарные в узком смысле последовательности.

Из предыдущих рассуждений мы можем ожидать, что при выполнении некоторых дополнительных условий ЦПТ в форме (2.12) будет выполняться для стационарных в узком смысле последовательностей, обладающих конечными вторыми моментами, где σ^2 определяется правой частью (2.14).

2.2.1 ЦПТ для линейного процесса

Лемма 2.4. *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

$$J_{n,x} = \left(\sum_m c_m \right) J_{n,\varepsilon} + V_n \quad (2.15)$$

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

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

$$U_{n,m} = - \sum_{\ell=1}^n \varepsilon_\ell + \sum_{\ell=1-m}^{n-m} \varepsilon_\ell \quad (2.16)$$

и

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

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

$$\begin{aligned} J_{n,x} &= \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\sum_m c_m \varepsilon_{j-m} \right) = \frac{1}{\sqrt{n}} \sum_m c_m \sum_{j=1}^n \varepsilon_{j-m} = \\ &= \frac{1}{\sqrt{n}} \sum_m c_m \sum_{\ell=1-m}^{n-m} \varepsilon_\ell = \sum_m c_m \frac{1}{\sqrt{n}} \sum_{\ell=1-m}^{n-m} \varepsilon_\ell = \\ &= \left(\sum_m c_m \right) \left(\frac{1}{\sqrt{n}} \sum_{\ell=1}^n \varepsilon_\ell \right) + \frac{1}{\sqrt{n}} \sum_m c_m U_{n,m}(\lambda) = \left(\sum_m c_m \right) J_{n,\varepsilon} + V_n. \end{aligned} \quad (2.17)$$

⁶²Но сохраняется условие их независимости.

Нетрудно видеть, что после сокращения подобных членов в сумме (2.16) 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 2.3. В условиях Леммы 2.4 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 (2.15) и сходимости V_n к нулю по вероятности. $\square$

Замечание 2.3. Из утверждения Теоремы 2.3 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 2.3. 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 2.3. 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,⁶³ 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 несчастный случай, когда предельная дисперсия считается аналитически.

2.2.2 ЦПТ для конечно-зависимых стационарных последовательностей

Теоремы Леви (одномерная и многомерная) говорят о слабой сходимости к нормальному распределению суммы (центрированной и нормированной) независимых одинаково распределенных случайных величин. Теорема Линдберга осуществляет отход от одинаковой распределенности, но сохраняет условие независимости. Что делать, если случайные величины зависимы?

Простым отклонением от независимости является конечно-зависимость.

Определение 2.3. Случайные величины $x_1, \dots, x_n, \dots$ называются *конечно-зависимыми*, если существует такое $N \geq 0$, что для любого $l \geq 1$, любых $k_1, \dots, k_l$ и любых $j_1, \dots, j_l$ таких, что $\min j_i - \max k_i > N$ случайные вектора $(x_{k_1}, \dots, x_{k_l})$ и $(x_{j_1}, \dots, x_{j_l})$ независимы.

Если число M — минимальное из всех таких N , то величины называются M -зависимыми. Допуская некоторую вольность, мы будем также говорить о $(\leq M)$ -зависимых случайных величинах.

Приведем несколько примеров конечно-зависимых случайных величин.

Пример 2.4.

1. Обычные независимые случайные величины — ноль-зависимые.
2. Пусть $\xi_1, \dots, \xi_n, \dots$ — независимые случайные величины. Обозначим $x_i = \xi_i \xi_{i+1}$. Тогда x_i являются (≤ 1) -зависимыми.
3. Положим в условиях предыдущего примера

$$x_k = \sum_{l=1}^M c_l \xi_{k+l}.$$

Такие случайные величины x_k конечно-зависимы. Если при этом $E \xi_i = \text{const}_1$ и $D \xi_i = \text{const}_2 < \infty$, то x_i — стационарная в широком смысле линейная последовательность. Если ξ_i еще и одинаково распределены, то она является стационарной в узком смысле.

⁶³ А почему? А почему это важно?

4. Пусть ξ_i — независимые случайные величины. Положим

$$x_n = \begin{cases} 1 & \text{при } \xi_n > \xi_{n+1}, \\ 0 & \text{иначе.} \end{cases}$$

Тогда, очевидно, x_n являются конечно-зависимыми. Если при этом ξ_n одинаково распределены, последовательность x_n — стационарная в узком (и широком) смысле.

Аналогичными свойствами обладают последовательности

$$y_n = \begin{cases} 1 & \text{при } \xi_n \xi_{n+1} < 0, \\ 0 & \text{иначе.} \end{cases}$$

и

$$z_n = \begin{cases} 1 & \text{при } \xi_{n+1} > \max(\xi_n, \xi_{n+2}) \text{ или } \xi_{n+1} < \min(\xi_n, \xi_{n+2}), \\ 0 & \text{иначе.} \end{cases}$$

Покажем, что при естественных условиях стационарные конечно-зависимые случайные последовательности удовлетворяют ЦПТ. Для это будем использовать следующую лемму.

Лемма 2.5. Пусть $\xi_n = \eta_{kn} + \beta_{kn}$ при $n \geq n_0(k)$, где η_{kn} и β_{kn} принимают значения в $\mathbb{R}$. Предположим, что выполнены следующие условия.

1. Последовательность β_{kn} равномерно по n сходится по вероятности к нулю при $k \rightarrow \infty$ в том смысле, что $\sup_n P(|\beta_{kn}| > \varepsilon) \rightarrow 0$ при $k \rightarrow \infty$.
2. Для любого k $\mathcal{L}(\eta_{kn}) \Rightarrow \mathcal{P}_k$ при $n \rightarrow \infty$.
3. $\mathcal{P}_k \Rightarrow \mathcal{P}$ при $k \rightarrow \infty$.

Тогда $\mathcal{L}(\xi_n) \Rightarrow \mathcal{P}$ при $n \rightarrow \infty$.

Доказательство. Возьмем $f \in \mathbf{C}_{BL}$, $0 \leq f \leq M$. Тогда

$$\begin{aligned} Ef(\xi_n) &= Ef(\eta_{kn} + \beta_{kn}) = Ef((\eta_{kn} + \beta_{kn}), |\beta_{kn}| > \varepsilon) + Ef(f(\eta_{kn} + \beta_{kn}), |\beta_{kn}| \leq \varepsilon) = \\ &= Ef(f(\eta_{kn} + \beta_{kn}), |\beta_{kn}| > \varepsilon) + Ef(f(\eta_{kn} + \beta_{kn}) - f(\eta_{kn}), |\beta_{kn}| \leq \varepsilon) + Ef(f(\eta_{kn}), |\beta_{kn}| \leq \varepsilon) = \\ &= J_1(k, n) + J_2(k, n) + J_3(k, n). \end{aligned}$$

Далее для любого $\varepsilon > 0$

$$\sup_n J_1(k, n) = \sup_n Ef(f(\eta_{kn} + \beta_{kn}), |\beta_{kn}| > \varepsilon) \leq M \sup_n P(|\beta_{kn}| > \varepsilon) \rightarrow 0$$

при $k \rightarrow \infty$. Значит, выбором k можно сделать J_1 сколь угодно малым равномерно по n . Далее,

$$|J_2(k, n)| \leq LE(|\beta_{kn}|, |\beta_{kn}| \leq \varepsilon) \leq L\varepsilon.$$

То есть выбором ε слагаемое $J_2(k, n)$ можно сделать сколь угодно малым равномерно по k, n . Затем,

$$J_3(k, n) = Ef(f(\eta_{kn}), |\beta_{kn}| \leq \varepsilon) = Ef(\eta_{kn}) - Ef(f(\eta_{kn}), |\beta_{kn}| > \varepsilon) = I_1(k, n) + I_2(k, n).$$

Слагаемое $I_2(k, n)$ будет равномерно по n мало при большом k . Что касается $I_1(k, n)$, то

$$I_2(k, n) - \int f d\mathcal{P} = \left(Ef(\eta_{kn}) - \int f d\mathcal{P}_k \right) + \left(\int f d\mathcal{P}_k - \int f d\mathcal{P} \right) = I_3(k, n) + I_4(k).$$

Выбором k делаем I_4 малым, а затем при этом фиксированном k выбираем такое n , что мало I_3 . Тем самым утверждение доказано. $\square$

Прежде чем доказывать предельную теорему, остановимся на следующем простом факте.

Лемма 2.6. Пусть случайные величины $\xi_1, \dots, \xi_L$ имеют одинаковые дисперсии, равные σ^2 . Тогда $D(\xi_1 + \dots + \xi_L) \leq L^2 \sigma^2$.

Доказательство. Действительно,

$$\begin{aligned} D(\xi_1 + \dots + \xi_L) &= \sum_{i=1}^L D(\xi_i) + 2 \sum_{1 \leq i < j \leq L} \text{Cov}(\xi_i, \xi_j) = \\ &= L\sigma^2 + 2\sigma^2 \sum_{1 \leq i < j \leq L} \rho(x_i, x_j) \leq L\sigma^2 + L(L-1)\sigma^2 = L^2\sigma^2, \end{aligned}$$

что и требовалось доказать. □

Теорема 2.4. (ЦПТ для конечно-зависимых стационарных последовательностей)

Пусть $x_1, \dots, x_n, \dots$ — сильно стационарная ($\leq M$)-зависимая последовательность со средним a , ковариационной функцией $R(m)$ и спектральной плотностью f . Тогда

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

где

$$\sigma^2 = \sum_{l=-\infty}^{\infty} R(l) = R(0) + 2 \sum_{l=1}^M R(l) = 2\pi f(0). \quad (2.18)$$

Доказательство. Прежде всего заметим, что $R(l) = 0$ при $l > M$ и положим $a = 0$.

Зафиксируем $k > 2M$ и положим $N = \lfloor n/k \rfloor$. Таким образом, $n = kN + r$, где $0 \leq r < k$. Определим следующие случайные величины.

$$\eta'_{kn} = \left((x_1 + \dots + x_{k-M}) + (x_{k+1} + \dots + x_{2k-M}) + \dots + (x_{k(N-1)+1} + \dots + x_{kN-M}) \right) / \sqrt{n} \quad (2.19)$$

(в каждой из N сумм по $k - M$ слагаемых, суммы независимы и одинаково распределены),

$$\beta_{kn} = \left((x_{k-M+1} + \dots + x_k) + (x_{2k-M+1} + \dots + x_{2k}) + \dots + (x_{kN-M+1} + \dots + x_{kN}) \right) / \sqrt{n}$$

(в каждой из N сумм по M слагаемых, суммы независимы и одинаково распределены),

$$\eta''_{kn} = (x_{kN+1} + \dots + x_n) / \sqrt{n}$$

(оставшиеся иксы, их число не больше $r < k$).

Заметим, что $\sqrt{n} \bar{x}_n = \eta'_{kn} + \beta_{kn} + \eta''_{kn}$. Рассмотрим сумму (2.19), считая, что k фиксировано, а $n \rightarrow \infty$. Обозначим

$$B_k^2 \stackrel{\text{def}}{=} D(x_1 + \dots + x_{k-M}).$$

Если $n \rightarrow \infty$, то $N \rightarrow \infty$ и поэтому по теореме П. Леви

$$\mathcal{L}(\sqrt{n/N} \eta'_{kn}) \Rightarrow N(0, D(x_1 + \dots + x_{k-M})) = N(0, B_k^2).$$

Так как одновременно $n/N \rightarrow k$, то $\mathcal{L}(\sqrt{k} \eta'_{kn}) \Rightarrow N(0, B_k^2)$, и следовательно,

$$\mathcal{L}(\eta'_{kn}) \Rightarrow \mathcal{P}_k = N(0, \sigma_k^2),$$

где $\sigma_k^2 = B_k^2/k$. Далее, согласно Предложению 2.3, при $k \rightarrow \infty$

$$\sigma_k^2 = \frac{k-M}{k} \frac{D(x_1 + \dots + x_{k-M})}{k-M} = \left(\frac{k-M}{k} \right) (k-M) D(\bar{x}_{k-M}) \rightarrow \sigma^2,$$

где σ^2 определено в (2.18). Поэтому $\mathcal{P}_k \Rightarrow \mathcal{P} = N(0, \sigma^2)$.

Обозначая $r = n - kN$, получим из Леммы 2.6, что

$$D\eta_{kn}'' = \frac{1}{n} D(x_{kN+1} + \dots + x_n) \leq \frac{r^2}{n} R(0) \leq \frac{k^2}{n} R(0) \rightarrow 0$$

при фиксированном k и $n \rightarrow \infty$. Поэтому $\eta_{kn}'' \xrightarrow{\mathbb{P}} 0$ и

$$\mathcal{L}(\eta_{kn}' + \eta_{kn}'') \Rightarrow \mathcal{P}_k = N(0, \sigma_k^2).$$

Наконец, $N/n \leq 1/k$ и

$$\sup_n D\beta_{kn} = \sup_n \frac{N}{n} D(x_1 + \dots + x_M) \leq \sup_n \frac{M^2 N}{n} R(0) \leq \frac{M^2}{k} R(0) \rightarrow 0$$

при $k \rightarrow \infty$. Применяя Лемму 2.5, мы видим, что утверждение доказано. $\square$

Пример 2.5. 1. Рассмотрим независимые одинаково распределенные случайные величины $\{\varepsilon_n\}_{n \geq 1}$ такие, что $E\varepsilon_i = 0$ и $E\varepsilon_i^2 = 1$. Положим $x_k = \varepsilon_k \varepsilon_{k+1}$. Как уже говорилось, эти случайные величины образуют сильно стационарную последовательность, которая является конечно-зависимой с $M = 1$. Поэтому $R(m) = Ex_k x_{k+m} = 0$ при $m \geq 2$.

Поскольку $R(0) = E\varepsilon_1^2 \varepsilon_2^2 = 1$ и $R(1) = Ex_1 x_2 = E(\varepsilon_1 \varepsilon_2^2 \varepsilon_3) = 0$, то по теореме 2.4

$$\mathcal{L}\left((x_1 + \dots + x_n)/\sqrt{n}\right) \Rightarrow N(0, 1).$$

2. В условиях пункта 4 Примера 2.4 дополнительно потребуем, чтобы случайные величины ξ_i были одинаково распределены. Тогда, конечно, для них будет выполняться ЦПТ. Если распределение $\mathcal{L}(\xi_i)$ является непрерывным, то предельная дисперсия в ЦПТ для последовательностей x_n и y_n будет не зависеть от $\mathcal{L}(\xi_i)$.

То же самое будет верно и для последовательности z_n , если вместо непрерывности $\mathcal{L}(\xi_n)$ потребовать, чтобы $P(\xi_i < 0) = P(\xi_i > 0) = 1/2$. Докажите эти факты и сосчитайте предельные дисперсии.⁶⁴

Список литературы

- [1] А.Н. Ширяев, Вероятность 1,2. М., Изд-во МЦНМО, 2004.
- [2] В. Феллер, Введение в теорию вероятностей и ее приложения, Т.2. М., Наука, 1967.
- [3] А.А. Боровков, Теория вероятностей. М., Эдиториал УРСС, 1999.

⁶⁴Эти варианты ЦПТ используются в статистике случайных процессов для проверки гипотезы о том, что случайные величины ξ_i являются независимыми и одинаково распределенными. Догадайтесь, как устроены эти критерии и против каких альтернатив они хорошо работают?