

Эконометрический ликбез: асимптотика временных рядов

Зависимость случайных процессов и теория стохастических пределов*

Джонатан Хилл[†]

Университет Северной Каролины, Чепел Хилл, США

В настоящем эссе представлена базовая асимптотическая теория, лежащая в основе оценивания временных рядов. Рассматриваются типы зависимостей, используемые в стохастической теории пределов, в том числе перемешивание, миксингалы и свойство почти эпохальной зависимости. Затем излагаются наиболее общие предельные теоремы (как для сходимости по вероятности, так и сходимости по распределению) для процессов, обладающих вышеперечисленными свойствами и часто используемых в теории временных рядов и ее приложениях.

1 Введение

Настоящее эссе представляет собой краткую справку по предельной теории случайных процессов. Эта теория позволит вычислить пределы по вероятности и по распределению для оценок нелинейного МНК, квази-ММП, ОММ и информационно-теоретических оценок, и в том числе оценок обобщенного метода эмпирического правдоподобия. Таким образом, настоящее эссе будет полезно студентам старших курсов университетов и аспирантам младших курсов по экономике, финансам, эконометрике и статистике для вывода асимптотических свойств параметрических и непараметрических оценок в регрессионных моделях.

Множество концепций и аргументов подчерпнуты в замечательных монографиях Doob (1953), Davidson (1994) и Billingsley (1999). Основная часть теории зависимых процессов, изложенной в данном эссе, обязана прогрессу 1970–1990-х годов в теории мартингалов, перемешивания и измеримых функций от перемешивающих последовательностей. Я особо заостряю внимание на свойствах миксингалов и почти эпохальной зависимости как простых, но мощных инструментах описания серийной зависимости и нестационарности временных рядов. Эти идеи принадлежат, среди прочих, авторам Ibragimov (1962), McLeish (1974), Bierens (1981), а также Gallant & White (1988).

Из-за краткости изложения множество концепций зависимости и соответствующих предельных теорем не вошли в настоящее эссе. Важные ссылки, включающие работы по изучению зависимостей Поля Духана и Вей Бяо Ву, даны в работе Dedecker, Doukhan, Lang, Leon, Louhichi & Prieur (2007).

В разделе 2 я напомню читателю базовые определения вероятностных пространств и случайных процессов. В разделе 3 рассматриваются стационарность и эргодичность. Концепции зависимости вводятся в разделе 4, а в разделах 5 и 6 представлены законы больших чисел и центральные предельные теоремы. Наконец, в разделе 7 я определю броуновское движение и слабую сходимость. Некоторые доказательства вынесены в Приложение.

*Перевод Г. Франгуриди. Цитировать как: Хилл, Джонатан (2012) «Зависимость случайных процессов и теория стохастических пределов», Квантиль, №10, стр. 1–31. Citation: Hill, Jonathan (2012) “Dependence and stochastic limit theory,” *Quantile*, No.10, pp. 1–31.

[†]Адрес: Department of Economics, Gardner Hall 208B, University of North Carolina–Chapel Hill, NC 27599-3305, USA. Электронная почта: jbhill@email.unc.edu

Отныне будем считать, что $K > 0$ — конечная константа, значение которой может меняться по ходу изложения. L_p -норма для $m \times n$ матриц определяется как

$$\|x_t\|_p = \left(\sum_{i=1}^m \sum_{j=1}^n \mathbb{E}|x_{t,i,j}|^p \right)^{1/p}.$$

2 Вероятностные пространства, борелевские функции, случайные процессы

2.1 σ -алгебры, борелевские функции, измеримость по Борелю

Пусть заданы пространство элементарных исходов Ω , σ -алгебра \mathfrak{F} исходов $\omega \in \Omega$ и вероятностная мера \mathbb{P} . *Вероятностным пространством* называется тройка $(\Omega, \mathfrak{F}, \mathbb{P})$.

Сигма-алгеброй \mathfrak{F} (σ -алгеброй, или σ -полем) называется множество подмножеств Ω , содержащее \emptyset и Ω и замкнутое относительно операций взятия дополнений и счетных объединений. В частности, \mathfrak{F} является σ -алгеброй *тогда и только тогда, когда*

$$\begin{aligned} \mathfrak{F} &\neq \emptyset, \emptyset \in \mathfrak{F}, \Omega \in \mathfrak{F}, \\ \text{если } A &\subseteq \mathfrak{F}, \text{ то } A^c \subseteq \mathfrak{F}, \\ \text{если } \{A_i\}_{i=1}^\infty, &\text{ где } A_i \subseteq \mathfrak{F} \forall i, \text{ то } \bigcup_{i=1}^\infty A_i \subseteq \mathfrak{F}. \end{aligned}$$

Случайная величина $x : \Omega \rightarrow \mathbb{R}$ *\mathfrak{F} -измерима* (является \mathfrak{F} -измеримой функцией или \mathfrak{F}/\mathbb{R} -измерима) *тогда и только тогда, когда* $x^{-1}(A) \in \mathfrak{F}$ для любого $A \in \mathbb{R}$. Таким образом, если реализации x на вещественной прямой связаны с подмножествами исходов $\omega \in \mathfrak{F}$, то вероятности событий, связанных с x , могут быть вычислены на основе шансов реализации исходов $\omega \in \mathfrak{F}$. Более кратко: реализации x ассоциированы с событиями в \mathfrak{F} .

σ -алгеброй, порожденной x и обозначаемой как $\sigma(x)$, называется пересечение всех σ -алгебр \mathfrak{F}_i , относительно которых x измерима. То есть если x измерима только на элементах последовательности $\{\mathfrak{F}_i\}_{i=1}^\infty$ σ -алгебр, то

$$\sigma(x) = \bigcap_{i=1}^\infty \mathfrak{F}_i.$$

Это наименьшее множество событий, ассоциированных со всеми реализациями x .

Пример 2.1 Пусть $\Omega = \{\emptyset, 0, 1, 2\}$. Пусть $x(0) = x(1) = 5$ и $x(2) = 10$. Зададим

$$\mathfrak{F} = \{\{\emptyset\}, \{\Omega\}, \{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}, \{0, 1, 2\}\}.$$

Легко проверить, что \mathfrak{F} замкнуто относительно дополнений и счетных пересечений. Например, $\{0, 1\}^c = \{2\} \in \mathfrak{F}$. Тогда

$$x^{-1}(5) = \{0, 1\} \in \mathfrak{F} \text{ и } x^{-1}(10) = \{2\} \in \mathfrak{F}.$$

Значит $x(\omega)$ является $\mathfrak{F}/\{0, 1, 2\}$ -измеримой. Более того, $\sigma(x) = \mathfrak{F}$, так как любая другая σ -алгебра, относительно которой x измерима, будет «больше» чем \mathfrak{F} . Другими словами, \mathfrak{F} есть наименьшая σ -алгебра, относительно которой x измерима.

Борелевской σ -алгеброй β называется σ -алгебра, порожденная подмножествами вида $(-\infty, x]$ числовой прямой $(-\infty, \infty)$:

$$\begin{aligned} \emptyset &\in \beta, \mathbb{R} \subseteq \beta, \\ (-\infty, x] &\subseteq \beta \text{ и } (-\infty, x]^c \subseteq \beta \forall x \in \mathbb{R}, \\ \bigcup_{q \in \mathbb{Q}} &(-\infty, q] \subseteq \beta. \end{aligned}$$

Объединение берется по рациональным концам отрезков, так как рациональные числа суть счетное множество (Doob 1953).

2.2 Борелевские функции

Борелевская функция $g : \Omega \rightarrow \mathbb{R}$, или функция, измеримая по Борелю (или β/\mathbb{R} -измеримая функция) удовлетворяет условию

$$x^{-1}(A) \in \beta \text{ для каждого } A \in \mathbb{R}.$$

Очень трудно построить пример функции, не измеримой по Борелю. Вообще говоря, мы просто соглашаемся с тем, что можем ассоциировать вещественные интервалы $(-\infty, w]$ с абстрактными событиями ω , и «измерить» шансы $x(\omega)$, «измеряя» $(-\infty, w]$.

Изложение на таком уровне абстракции бессмысленно. Множество эконометрических результатов получено в очень слабых предположениях относительно участвующих в них функций. Хотя есть и такие, которые предполагают непрерывность функций, для других достаточно измеримости по Борелю. На самом деле, свойство непрерывности строго сильнее свойства измеримости по Борелю.

Лемма 2.1 *Любая функция $g : \mathbb{R} \rightarrow \mathbb{R}$, непрерывная почти всюду, измерима по Борелю.*

2.3 Случайные процессы

2.3.1 Определение

Случайный процесс есть обобщение измеримого отображения $\{x\} : \Omega \rightarrow \mathbb{R}^{\mathbb{Z}}$:

$$\{x(\omega)\} = \{x_t(\omega)\}_{t=-\infty}^{\infty} = \{x_t(\omega) : t \in \mathbb{Z}\}.$$

Вся последовательность $\{\dots, x_t, x_{t+1}, \dots\}$ должна быть совместно измерима: должно быть возможно приписать функцию совместного распределения подпоследовательности $\{x_{t_1}, \dots, x_{t_l}\}$ для любого непустого l -набора $\{t_1, \dots, t_l\} \in \mathbb{Z}^l$.

Заметим, что вся последовательность $\{x_t(\omega)\}_{t=-\infty}^{\infty}$ есть функция одного случайного исхода $\omega \in \Omega$. Мы должны представлять последовательные во времени события как одну возможную реализацию случайного эксперимента $\omega \in \Omega$.

Мы говорим об «обобщении», потому что индекс не обязан быть ни целочисленным, ни сохранять естественный линейный порядок. На практике координата t обозначает время, а $x_{t+1}(\omega)$ реализуется после $x_t(\omega)$ и никогда в обратную сторону.

Траекторией называется наблюдаемая последовательность, зависящая от конкретного (не обязательно наблюдаемого) случайного исхода $\omega \in \Omega$:

$$\{x_t(\omega)\}_{t=1}^n.$$

2.3.2 Свойства

Процесс $\{x(\omega)\}$ равномерно ограничен по вероятности, если для любого $\varepsilon > 0$ существует $K_\varepsilon < \infty$, такое что

$$\sup_{t \in \mathbb{Z}} \mathbb{P}\{|x_t| > K_\varepsilon\} \leq \varepsilon.$$

Равномерная ограниченность исключительно важна, так как многие процессы зависят от времени и могут неограниченно расти или падать. Рассмотрим простой линейный тренд

$$x_t = \beta_0 + \beta_1 t + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2),$$

где ϵ_t обладает строго положительной и непрерывной плотностью (например, гауссовской). Ясно, что процесс x_t не является равномерно ограниченным по вероятности, так как для любого конечного K_ϵ

$$\sup_{t \in \mathbb{Z}} \mathbb{P} \{ |\beta_0 + \beta_1 t + \epsilon_t| > K_\epsilon \} = 1.$$

Случайный процесс *равномерно L_p -ограничен*, $p > 0$, если

$$\sup_{t \in \mathbb{Z}} \|x_t\|_p < \infty.$$

Заметим, что равномерная L_p -ограниченность влечет равномерную ограниченность по вероятности в силу неравенства Маркова:

$$\sup_{t \in \mathbb{Z}} \mathbb{P} \{ |x_t| > K_\epsilon \} \leq \frac{1}{K_\epsilon^p} \sup_{t \in \mathbb{Z}} \|x_t\|_p^p < \infty.$$

Поэтому для любого $\epsilon > 0$ выберем

$$\frac{1}{\epsilon^{1/p}} \sup_{t \in \mathbb{Z}} \|x_t\|_p = K_\epsilon.$$

Если $p = \infty$, то L_p -норма превращается в *равномерную норму*:

$$\sup_{t \in \mathbb{Z}} \|x_t\|_p = \sup_{t \in \mathbb{Z}} |x_t| < \infty \text{ a.s.},$$

а значит x_t *почти наверное равномерно ограничен* (например, $\mathbb{P}\{a \leq x_t \leq b\} = 1$ для любого конечного $a \leq b$: это верно, если данные усечены (скажем, $x_t \times \mathbb{I}\{|x_t| \leq a\} \in [-a, a]$ п.н. равномерно по t для любого конечного $a > 0$).

Случайная величина x_t *равномерно интегрируема*, если

$$\lim_{M \rightarrow \infty} \mathbb{E} [|x_t| \times \mathbb{I}\{|x_t| \geq M\}] = 0.$$

Это значит, что экстремальные значения (т.е. хвосты распределения) не доминируют, так как реализуются слишком редко. Процесс $\{x(\omega)\}$ *равномерно интегрируем*, если

$$\lim_{M \rightarrow \infty} \sup_{t \in \mathbb{Z}} \mathbb{E} [|x_t| \times \mathbb{I}\{|x_t| \geq M\}] = 0.$$

Теорема 2.1 $L_{1+\delta}$ -ограниченность для любого $\delta > 0$ влечет равномерную интегрируемость.

3 Стационарность и эргодичность

3.1 Стационарность

Процесс $\{x_t(\omega)\}$ называется *строго стационарным*, если *преобразования сдвига сохраняют меру*: совместное распределение $\{x_{t-m_1}, x_{t-m_2}, \dots, x_{t-m_k}\}$ зависит только от $\{m_1, \dots, m_k\}$, но не от t . Это означает, что *конечномерные распределения* (совместные распределения векторов $\{x_{t-m_1}, x_{t-m_2}, \dots, x_{t-m_k}\}$) зависят только от относительного смещения, но не от конкретных моментов времени.

Процесс $\{x_t(\omega)\}$ *слабо стационарен*, если

$$\begin{aligned} \mathbb{E} [x_t] &= \mu \quad \forall t, \\ \mathbb{E} [(x_t - \mathbb{E} [x_t])^2] &= \sigma^2 < \infty \quad \forall t, \\ \gamma(s, t) &:= \mathbb{E} [(x_s - \mathbb{E} [x_s]) (x_t - \mathbb{E} [x_t])] = \gamma(h), \text{ где } h = |s - t|. \end{aligned}$$

Иначе говоря, стационарными характеристиками процесса являются только среднее, дисперсия и автоковариации.

Если процесс $\{x_t(\omega)\}$ строго стационарен и L_p -ограничен, то $\forall r \leq p$

$$\mathbb{E}|x_t|^r = \mathbb{E}|x_1|^r.$$

3.2 Эргодичность: эргодическая теорема как ЗБЧ

3.2.1 Эргодичность: определение и пример

Пусть процесс $\{x_t(\omega)\}$ задан на $(\Omega, \mathfrak{F}, \mathbb{P})$. Рассмотрим траекторию этого процесса для некоторого исхода ω_0 , $\{x_t(\omega_0)\}_{t=1}^n$, и последовательность сечений этого процесса для некоторого момента времени t_0 , $\{x_{t_0}(\omega_i)\}_{i=1}^N$. Мы хотим узнать, при каких условиях среднее по реализациям для любого момента t_0

$$\frac{1}{N} \sum_{i=1}^N x_{t_0}(\omega_i)$$

совпадает в пределе со средним по времени при фиксированном ω_0

$$\frac{1}{n} \sum_{t=1}^n x_t(\omega_0).$$

Если процесс $\{x_t(\omega)\}$ стационарен, то $\mathbb{E}[x_t] = \mathbb{E}[x_s] \forall s, t$, в случае чего нам хотелось бы узнать, когда

$$\frac{1}{n} \sum_{t=1}^n x_t(\omega_0) \rightarrow \mathbb{E}[x_{t_0}] \text{ для произвольного момента } t_0.$$

Пример 3.1 Синоптик A собирает данные об осадках $x_{t_0}(\omega_i)$ из N точек $\{\omega_i\}_{i=1}^N$ в Северной Каролине в момент времени t_1 . Синоптик B собирает данные об осадках $x_t(\omega_1)$ в моменты времени $t \in \{1, \dots, n\}$ в Северной Каролине в точке ω_1 . Ясно, что данные, собранные двумя синоптиками, могут не иметь одинаковых популяционных характеристик. Точка ω_1 может быть очень засушливым местом, а в момент времени t_1 погода может быть очень дождливая, так что

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_{t_1}(\omega_i) > \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n x_t(\omega_1).$$

Стационарный процесс $\{x_t(\omega)\}$ называется *эргодичным для среднего*, если

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_{t_0}(\omega_i) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n x_t(\omega) = \mathbb{E}[x_t] \quad \forall t.$$

В пространстве элементарных исходов нет «систематического сдвига» (то есть доминирующих засушливых регионов в Северной Каролине: флуктуации во времени в точке ω в среднем не отличаются от флуктуаций в пространстве).

3.2.2 Эргодичность: память в σ -алгебрах

Пусть процесс $\{x_t(\omega)\}$ стационарен. На практике стационарность обычно не является реалистичным предположением в силу наличия временных трендов в среднем или дисперсии или сезонных эффектов. Тем не менее, эргодичность стационарного процесса допускает значительную персистентность и поэтому интересна, так как представляет собой уровень, с которым можно сравнивать персистентность других концепций зависимости.

Пусть $\mathfrak{F} = \sigma(x)$ и $A, B \in \mathfrak{F}$. Пусть также T есть преобразование сдвига, сохраняющее меру: $\mathbb{P}\{TB\} = \mathbb{P}\{B\}$, и TB сдвигает исходы, составляющие событие A . Таким образом, $T^k B$ производит k операций сдвига.

Говорят, что *сохраняющее меру преобразование сдвига T эргодично*, если средняя память исчезает.

Теорема 3.1 *Сохраняющее меру преобразование сдвига T эргодично тогда и только тогда, когда $\forall A, B \in \mathfrak{F}$*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [\mathbb{P}\{A \cap T^i B\} - \mathbb{P}\{A\} \mathbb{P}\{B\}] = 0. \quad (1)$$

То есть в среднем события независимы.

Процесс $\{x_t(\omega)\}$ *эргодичен*, если для любого момента времени t

$$x_t(T^k \omega) = x_{t+k}(\omega),$$

где преобразование T сохраняет меру и эргодично. Это ключевая идея (хотя и абстрактная). Процесс *эргодичен*, когда сдвиг в пространстве исходов эквивалентен сдвигу по времени. Таким образом, случайная выборка из сечения процесса в конкретный момент времени даст (асимптотически) ту же информацию, что и наблюдения во времени. Сечение момента t (которого мы обычно не имеем) будет иметь ту же вероятностную структуру, что и последовательность во времени (которую мы имеем).

Пример 3.2 Пусть $\Omega = \{0, 1\}$, так что $\omega = 0$ или 1 , и $\mathfrak{F} = \{\{\emptyset\}, \{\Omega\}, \{0\}, \{1\}, \{0, 1\}\}$, $\mathbb{P}\{0\} = \mathbb{P}\{1\} = 1/2$. Пусть $T0 = 1$ и $T1 = 0$. Ясно, что T сохраняет меру: $\mathbb{P}\{T\omega\} = \mathbb{P}\{\omega\} = 1/2$. Если $A = \{0, 1\}$ и $B = \{0\}$, то для любого четного k

$$\mathbb{P}\{A \cap T^k B\} - \mathbb{P}\{A\} \mathbb{P}\{B\} = \mathbb{P}\{0\} - \mathbb{P}\{\{0, 1\}\} \mathbb{P}\{0\} = 0,$$

а для любого нечетного k

$$\mathbb{P}\{A \cap T^k B\} - \mathbb{P}\{A\} \mathbb{P}\{B\} = \mathbb{P}\{1\} - \mathbb{P}\{\{0, 1\}\} \mathbb{P}\{0\} = 0.$$

Аналогично, если $A = \{0\}$ и $B = \{1\}$, то для любого четного k

$$\mathbb{P}\{A \cap T^k B\} - \mathbb{P}\{A\} \mathbb{P}\{B\} = \mathbb{P}\{\emptyset\} - \mathbb{P}\{0\} \mathbb{P}\{1\} = -1/4,$$

а для любого нечетного k

$$\mathbb{P}\{A \cap T^k B\} - \mathbb{P}\{A\} \mathbb{P}\{B\} = \mathbb{P}\{0\} - \mathbb{P}\{0\} \mathbb{P}\{1\} = 1/4,$$

что в среднем дает ноль. И так далее. Преобразование T эргодично. Определим теперь $x_1(\omega) := \omega$ и $x_t(\omega) := x_1(T^{t-1}\omega)$ для $t = 2, 3, \dots$. Тогда последовательность случайно выходит из 0 или 1, а после этого чередуется, скажем, как 0, 1, 0, 1, Среднее по времени равно, очевидно, $1/2$, и по построению среднее по реализациям равно $1/2$: $\mathbb{E}[x_t(\omega)] = 0 \times .5 + 1 \times .5 = .5$. Таким образом, процесс $\{x_t(\omega)\}$ эргодичен.

Пример 3.3 Пусть $\mathfrak{S} = \{\{\emptyset\}, \{\Omega\}, \omega_1, \dots, \omega_6\}$, где $\mathbb{P}\{\omega_i\} = 1/6$. Рассмотрим подмножества $A = \{\omega_1, \omega_2, \omega_3\}$, $B = \{\omega_1, \omega_2, \omega_4\}$. Преобразование сдвига $T^k B = \{\omega_{1+2k}, \omega_{2+2k}, \omega_{4+2k}\}$ сохраняет меру, так как $\mathbb{P}\{T^k B\} = \mathbb{P}\{B\} = 1/2$, где $\omega_b = \omega_{b-6}$ для любого $b > 6$. Тогда

$$\begin{aligned}\mathbb{P}\{A \cap TB\} - \mathbb{P}\{A\}\mathbb{P}\{B\} &= \mathbb{P}\{\omega_3\} - 1/4 = 1/6 - 1/4 = -1/12 \\ \mathbb{P}\{A \cap T^2 B\} - \mathbb{P}\{A\}\mathbb{P}\{B\} &= \mathbb{P}\{\omega_2\} - 1/4 = 1/6 - 1/4 = -1/12 \\ \mathbb{P}\{A \cap T^3 B\} - 1/4 &= \mathbb{P}\{\omega_1, \omega_2\} - 1/4 = 1/3 - 1/4 = 1/12 \\ \mathbb{P}\{A \cap T^4 B\} - 1/4 &= \mathbb{P}\{\omega_3\} - 1/4 = 1/6 - 1/4 = -1/12,\end{aligned}$$

и так далее. Таким образом,

$$\frac{1}{n} \sum_{i=1}^n [\mathbb{P}\{A \cap T^i B\} - \mathbb{P}\{A\}\mathbb{P}\{B\}] = \frac{1}{n} [-1/12 - 1/12 + 1/12 + \dots] \rightarrow -1/3,$$

и значит, преобразование T неэргодично.

Есть два способа описать память эргодичного процесса. Первый способ просто переформулирует (1) в терминах расстояния по времени, в противовес абстрактному «сдвигу».

Теорема 3.2 Пусть \mathfrak{S}_t обозначает σ -алгебру, порожденную $\{x_t, x_{t-1}, \dots\}$. Процесс $\{x_t\}$ эргодичен тогда и только тогда, когда для любых $A_t, B_t \subseteq \mathfrak{S}_t$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [\mathbb{P}\{A_t \cap B_{t-i}\} - \mathbb{P}\{A_t\}\mathbb{P}\{B_{t-i}\}] = 0. \quad (2)$$

Теорема утверждает, что процесс x_t эргодичен, если его серийная память (то есть зависимости между x_t и x_{t-i}) в среднем исчезает. Это не значит, что далекие по времени события независимы, то есть эргодичность допускает слишком сильную персистентность, чтобы найти хорошее применение в теории (предельная теория требует более сильных ограничений). Если наложить дополнительные структурные ограничения, из эргодичности будет следовать, что средние по времени имеют корректно определенный предел по вероятности.

Теорема 3.3 Если $\{x_t(\omega)\}$ стационарен, эргодичен и L_2 -ограничен, то

$$\frac{1}{n} \sum_{i=1}^n \mathbb{C}(x_i, x_1) \rightarrow 0.$$

Следующий результат представляет собой простое достаточное условие эргодичности для среднего.

Теорема 3.4 Если процесс $\{x_t(\omega)\}$ слабо стационарен и $\sum_{i=0}^{\infty} |\mathbb{C}(x_{i+1}, x_1)| < \infty$, то он эргодичен для среднего:

$$\frac{1}{n} \sum_{t=1}^n x_t(\omega) \rightarrow \mathbb{E}[x_t(\omega)].$$

Пример 3.4 Пусть $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$, $\sigma^2 < \infty$. Тогда процесс $MA(1)$, определяемый как

$$x_t = \theta \epsilon_{t-1} + \epsilon_t, \quad |\theta| < \theta,$$

имеет очень короткую память. Ясно, что для всех $A_t, B_t \subseteq \mathfrak{S}_t = \sigma(x_\tau : \tau \leq t)$ и всех $i \geq 2$, $\mathbb{P}\{A_t \cap B_{t-i}\} = \mathbb{P}\{A_t\}\mathbb{P}\{B_{t-i}\}$. Следовательно, $1/n \sum_{i=1}^n [\mathbb{P}\{A_t \cap B_{t-i}\} - \mathbb{P}\{A_t\}\mathbb{P}\{B_{t-i}\}] \rightarrow 0$, и $1/n \sum_{t=1}^n x_t(\omega) \rightarrow 0$. Ясно, что $\sum_{i=0}^{\infty} |\mathbb{C}(x_{i+1}, x_1)| = \sigma^2(1 + \theta^2) + \sigma^2|\theta| < \infty$.

Пример 3.5 Пусть $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$ и

$$x_t = a + bt + \epsilon_t.$$

Так как $\mathbb{E}[x_t] = a + bt$ не стационарно (и не слабо стационарно), теорема 3.4 неприменима.

Лемма 3.1 Рассмотрим процесс $AR(1)$

$$x_t = \phi x_{t-1} + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2) \text{ и } |\phi| < 1.$$

Тогда $\sum_{i=0}^{\infty} |\mathbb{C}(x_{i+1}, x_1)| = \sigma^2 / [(1 - \phi^2)(1 - |\phi|)] < \infty$. Следовательно, x_t эргодичен для среднего.

4 Концепции зависимости и неоднородности

4.1 Перемешивание

4.1.1 Преобразование перемешивания

Пусть $\{x_t(\omega)\}$ — случайный процесс. Эргодичность — довольно слабое свойство памяти, применимое к последовательностям по времени в среднем, но не означающее, что далекие события асимптотически независимы. В экономике обычно предполагается, что бесконечно далекие события (x_t и x_{t-N} при $N \rightarrow \infty$) независимы, или этого требует подходящая центральная предельная теорема. Это следует также из условия перемешивания.

Пусть $\mathfrak{F} = \sigma(x)$ и $A, B \in \mathfrak{F}$. Эргодичное преобразование T , сохраняющее меру, называется перемешивающим, если для любых $A, B \in \mathfrak{F}$

$$\lim_{k \rightarrow \infty} \mathbb{P} \left\{ T^k A \cap B \right\} = \mathbb{P} \{A\} \mathbb{P} \{B\}.$$

Процесс $\{x_t(\omega)\}$ является перемешивающим, если

$$x_t(T^k \omega) = x_{t+k}(\omega),$$

где T — эргодичное сохраняющее меру перемешивающее преобразование.

4.1.2 Сильное и равномерное перемешивание

Рассмотрим более конкретные случаи перемешивания. Пусть $\{x_t\}$ есть случайный процесс с σ -алгеброй

$$\mathfrak{F}_a^b := \sigma(x_t : a \leq t \leq b).$$

Определим последовательности α_m и ϕ_m следующим образом:

$$\alpha_m = \sup_{t \in \mathbb{Z}} \sup_{G \in \mathfrak{F}_{-\infty}^t, H \in \mathfrak{F}_{t+m}^{\infty}} |\mathbb{P} \{G \cap H\} - \mathbb{P} \{G\} \mathbb{P} \{H\}|$$

и

$$\phi_m = \sup_{t \in \mathbb{Z}} \sup_{G \in \mathfrak{F}_{-\infty}^t, H \in \mathfrak{F}_{t+m}^{\infty}} |\mathbb{P} \{H|G\} - \mathbb{P} \{H\}|.$$

Процесс $\{x_t\}$ является равномерно перемешивающим, если

$$\phi_m \rightarrow 0 \text{ при } m \rightarrow \infty,$$

и *сильно перемешивающим*, если

$$\alpha_m \rightarrow 0 \text{ при } m \rightarrow \infty.$$

Перемешивание просто означает, что асимптотически далекие события $G \in \mathfrak{S}_{-\infty}^t$ и $H \in \mathfrak{S}_{t+m}^{\infty}$ независимы. Поэтому случайные величины x_{t-m} и x_t независимы при $m \rightarrow \infty$ для любого t .

Нетрудно доказать, что

$$\alpha_m \leq \phi_m,$$

то есть «сильное перемешивание» на самом деле слабее «равномерного перемешивания», а значит представляет больший теоретический интерес.

4.1.3 Размер перемешивания

Скорость, с которой x_t и x_{t-k} становятся независимыми при $k \rightarrow \infty$, измеряется «размером». Есть несколько способов определить размер, однако наиболее популярным является гиперболическая ограниченность.

Говорят, что процесс x_t *сильно* или *равномерно перемешивающий* размера $\lambda > 0$, если

$$\alpha_m = o(m^{-\lambda}) \text{ или } \phi_m = o(m^{-\lambda}).$$

Гиперболическая ограниченность — наиболее слабое условие, применяемое в теории, так как оно представляет собой «*длинную*» память: x_t и x_{t-k} становятся независимыми очень медленно.

Обсудим, что на самом деле означает размер перемешивания. Если x_t — *сильно перемешивающий* процесс размера 2, то $\forall \delta \leq 2$

$$m^\delta \times \sup_{t \in \mathbb{Z}} \sup_{G \in \mathfrak{S}_{-\infty}^t, H \in \mathfrak{S}_{t+m}^{\infty}} |\mathbb{P}\{G \cap H\} - \mathbb{P}\{G\}\mathbb{P}\{H\}| \rightarrow 0$$

при $m \rightarrow \infty$. Таким образом, x_t и x_{t-k} становятся независимыми со скоростью m^2 . Чем больше размер, тем больше скорость сходимости, и тем слабее зависимость (меньше персистентность).

С другой стороны, говорят, что процесс x_t *геометрически сильно* или *равномерно перемешивающий*, если существует число $\rho \in (0, 1)$, такое что

$$\alpha_m = o(\rho^m) \text{ или } \phi_m = (\rho^m).$$

Легко доказать, что геометрическое перемешивание влечет перемешивание любого размера $\lambda > 0$, то есть геометрическое перемешивание — *сильное* условие, и представляет собой «*короткую*» память: x_t и x_{t-k} становятся независимыми относительно быстро.

Из перемешивания размера $\lambda > 1$ (а значит и геометрического перемешивания) следует суммируемость коэффициентов перемешивания:

$$\sum_{m=0}^{\infty} \alpha_m < \infty \text{ или } \sum_{m=0}^{\infty} \phi_m < \infty.$$

4.1.4 Перемешивание во временных рядах

Для того чтобы понять, что собой представляет перемешивание, рассмотрим следующий пример.

Пример 4.1 Рассмотрим бросание одного красного и одного синего плавающих мячей в бесконечно большой бассейн, и пусть x_t обозначает евклидово расстояние между ними. Рассмотрим также последовательность шок по времени: мы опускаем в воду большую палку и шевелим ее. Если мы пошевелим палкой однократно, расстояние между мячами станет x_{t+1} . Ясно, что x_{t+1} будет зависеть от x_t (если $x_t = 10$, то x_{t+1} будет равно скорее 12, чем, скажем, 20000). Однако после k -кратного шевеления или *перемешивания* бассейна, расстояние x_{t+k} при $k \rightarrow \infty$ не будет зависеть от x_t , потому что в долгосрочной перспективе x_{t+k} может принять любое значение в \mathbb{R}_+ , независимо от исходного расстояния x_t . Таким образом, x_t называется «перемешивающим».

Этот пример не абстрактен. Если случайные шоки ϵ_t входят в процесс, порождающий, например, дневные данные по некоторому обменному курсу x_t , в соответствии с формулой

$$x_t = f(x_{t-1}, \dots, x_{t-p}; \phi) + \epsilon_t,$$

где $f : \mathbb{R}^p \times \Phi \rightarrow \mathbb{R}$ неизвестна, $\Phi \subseteq \mathbb{R}^p$, то как текущий курс x_t зависит от прошлых курсов x_{t-k} при $k \rightarrow \infty$? В примере с бассейном прошлые шоки ϵ_{t-i} (шевеления) оказывали все уменьшающийся эффект на текущее расстояние x_t , а недавние шоки (шевеления) имели больший эффект. Вообще говоря, очень трудно ответить на поставленный вопрос с точки зрения перемешиваний, и только недавно полученные результаты прояснили, какие виды нелинейных процессов перемешиваются.

Рассмотрим частный случай гораздо более общего доказательства для нелинейных авторегрессий. См. An & Huang (1996) и Leibscher (2005).

Лемма 4.1 *Рассмотрим процесс $AR(1)$*

$$x_t = \phi x_{t-1} + \epsilon_t,$$

где $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$ обладает строго положительной и непрерывной функцией плотности, и $|\phi| < \infty$. Тогда процесс x_t геометрически сильно перемешивающий.

Условие непрерывности и положительности плотности ошибок ϵ_t на всем носителе очень ограничительно, хотя плотности нормального, экспоненциального и многих других распределений ему удовлетворяют. Несложно найти простой стационарный эргодичный процесс $AR(1)$, не являющийся сильно перемешивающим, потому что плотность ошибок ϵ_t разрывна (например, распределение Бернулли). См. Andrews (1984).

4.1.5 Свойства перемешивания

Приводимые ниже результаты полезны для расширения понятий эргодичности и законов больших чисел на случайные величины с перемешиванием.

Лемма 4.2 *Для $p > 1$ и $r \geq p/(p-1)$*

$$|\mathbb{C}(x_t, x_{t-m})| \leq 2 \left(2^{1/p} + 1 \right) \alpha_m^{1-1/p-1/r} \|x_t\|_p \|x_t\|_r,$$

где процесс x_t является $L_{\max\{p,r\}}$ -ограниченным.

Лемма 4.3 *Для $r \geq 1$*

$$|\mathbb{C}(x_t, x_{t-m})| \leq 2 \phi_m^{1/r} \|x_t\|_r \|x_t\|_{r/(r-1)},$$

где процесс x_t является $L_{\max\{r,r/(r-1)\}}$ -ограниченным.

Оба результата мгновенно влекут эргодичность для среднего в процессах с перемешиванием. Случай геометрического перемешивания тривиален.

Теорема 4.1 *Если процесс x_t слабо стационарен, $L_{2+\delta}$ -ограничен и является геометрически сильно перемешивающим, то он эргодичен для среднего: $1/n \sum_{t=1}^n x_t \xrightarrow{P} \mathbb{E}[x_t]$.*

Одна из наиболее полезных характеристик перемешивания случайных величин — это результат о том, что свойство перемешивания распространяется на функции конечного числа лагов от перемешивающихся случайных величин.

Лемма 4.4 *Пусть $y_t := g(x_t, x_{t-1}, \dots, x_{t-k})$ есть измеримая функция для конечного $k > 0$. Если x_t сильно или равномерно перемешивающий размера $\lambda > 0$, то и y_t тоже.*

Пример 4.2 Если процесс x_t сильно перемешивающий размера $\lambda > 0$, то сильно перемешивающий и процесс

$$\sum_{i=0}^k \psi_i x_{t-i}$$

для любого конечного $k > 0$ и любого $\psi_i \in \mathbb{R}$.

Пример 4.3 Если ϵ_t сильно перемешивающий размера $\lambda > 0$, а $\{x_t\}$ — случайный процесс, то $\mathbb{E}[x_t | \epsilon_t, \dots, \epsilon_{t-k}]$ сильно перемешивающий размера $\lambda > 0$.

4.2 Мартингалы и мартингал-разности

Рассмотрим процесс $\{x_t\}$ с σ -алгеброй $\mathfrak{F}_t := \sigma(x_\tau : \tau \leq t)$. Мартингал $\{x_t, \mathfrak{F}_t\}_{-\infty}^{\infty}$ удовлетворяет условию

$$\mathbb{E}[x_t | \mathfrak{F}_{t-1}] = x_{t-1}.$$

Слово «мартингал» изначально относилось к игровой стратегии, а как концепция в теории вероятностей появилось по крайней мере в работах Lévy (1925, 1954) и Doob (1953).

Пример 4.4 Рассмотрим процесс AR(1) с единичным наклоном:

$$x_t = x_{t-1} + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2).$$

Автоковариации не суммируются, и x_t слишком персистентен, чтобы быть перемешивающим. Однако такой процесс является мартингалом, так как из независимости шоков следует, что $\mathbb{E}[x_t | \mathfrak{F}_{t-1}] = x_{t-1}$. Гипотеза эффективного рынка утверждает, что цены активов являются мартингалами: цена актива в момент времени $t - 1$ на эффективном рынке учитывает всю информацию, так что лучший (в среднеквадратичном смысле) прогноз на будущее и есть цена в момент t .

Хотя в силу своих математических свойств мартингалы широко применяются в экономике и финансах, они допускают гораздо более сильную память, чем требуют законы больших чисел и центральные предельные теоремы.

Напротив, первая разность мартингала x_t ,

$$\epsilon_t := x_t - x_{t-1},$$

обладает такой простой структурой, что для нее верны ЦПТ и ЗБЧ. Процесс $\{\epsilon_t, \mathfrak{F}_t\}_{-\infty}^{\infty}$ называется *мартингал-разностью* с характерными свойствами

$$\mathbb{E}[\epsilon_t | \mathfrak{F}_{t-1}] = 0 \quad (3)$$

и

$$\mathbb{E}[\epsilon_s \epsilon_t] = \mathbb{E}(\epsilon_s \mathbb{E}[\epsilon_t | \mathfrak{F}_{t-1}]) = \mathbb{E}(\epsilon_s \times 0) = 0. \quad (4)$$

В общем случае мартингал-разность определяется через свойство (3) без привязки к какому-либо мартингалу. Ясно, что последовательность независимых одинаково распределенных центрированных величин $\{\epsilon_t, \mathfrak{F}_t\}_{-\infty}^{\infty}$, $\mathfrak{F}_t := \sigma(\epsilon_\tau : \tau \leq t)$ является мартингал-разностью $\{\epsilon_t, \mathfrak{F}_t\}_{-\infty}^{\infty}$, так как по построению $\mathbb{E}[\epsilon_t | \mathfrak{F}_{t-1}] = \mathbb{E}[\epsilon_t] = 0$.

В силу того, что все ковариации равны нулю, *мартингал-разность* $\{\epsilon_t\}$ эргодична для среднего (см. теорему 3.4).

4.2.1 Линейная модель распределенных лагов

Линейная модель распределенных лагов *мартингал-разности* — полезная модель временных рядов, которая проявляется в различных постановках (например слабо стационарные модели ARIMA или случайные блуждания).

Пусть $\{\epsilon_t, \mathfrak{F}_t\}_{-\infty}^{\infty}$ — мартингал-разность, $\mathbb{E}[\epsilon_t] = 0$, $\mathbb{E}[\epsilon_t^2] = \sigma^2$, и положим

$$x_t := \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}.$$

Тогда автоковариации $\gamma(h) := \mathbb{E}[x_t x_{t-h}]$ равны (используем (4))

$$\gamma(h) = \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h}.$$

Автоковариации $\{\gamma(h)\}_{h \in \mathbb{N}}$ абсолютно суммируемы, если $\{\psi_i\}_{i=0}^{\infty}$ абсолютно суммируемы:

$$\sum_{h=0}^{\infty} |\gamma(h)| = \sigma^2 \sum_{h=0}^{\infty} \left| \sum_{i=0}^{\infty} \psi_i \psi_{i+h} \right| \leq \sigma^2 \sum_{h=0}^{\infty} \sum_{i=0}^{\infty} (|\psi_i| \times |\psi_{i+h}|) \leq \sigma^2 \left(\sum_{i=0}^{\infty} |\psi_i| \right)^2 < \infty.$$

Лемма 4.5 Пусть $\{\epsilon_t, \mathfrak{F}_t\}_{-\infty}^{\infty}$ является мартингал-разностью, $\mathbb{E}[\epsilon_t] = 0$, $\mathbb{E}[\epsilon_t^2] = \sigma^2 < \infty$, и положим $x_t := \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$. Если числа $\{\psi_i\}$ абсолютно суммируемы, то процесс x_t эргодичен для среднего: $1/n \sum_{t=1}^n x_t \xrightarrow{p} \mathbb{E}[x_t]$.

4.3 Почти эпохальная зависимость

Мартингалы обладают слишком сильной персистентностью для применения ЗБЧ или ЦПТ в общем случае (без серьезных ограничений), а мартингал-разности структурно слишком бедны для моделирования памяти, встречающейся в экономических и финансовых временных рядах. Тем не менее, многие значительно персистентные и неоднородные процессы можно приблизить мартингал-разностями, что делает последние важнейшими компонентами ЗБЧ и ЦПТ.

Одним из решений проблемы является наложение ограничений типа перемешивания, но их очень сложно проверить, и это можно сделать только при достаточно абстрактных или ограничительных условиях (например, положительная и непрерывная на всем своем носителе плотность f).

Свойство *почти эпохальной зависимости* (ПЭЗ), напротив, тривиально содержит перемешивание, и характеризует огромное количество моделей временных рядов. В частности, это свойство не требует явно положительности и непрерывности плотности, и формулируется в терминах математических ожиданий, что делает его не столь трудным для проверки на практике.

4.3.1 Определение почти эпохальной зависимости

Пусть $\{F_t\}$ есть последовательность σ -алгебр, порожденная последовательностью случайных величин $\{v_t\}$,

$$F_t := \sigma(v_\tau : \tau \leq t).$$

Отметим, что v_t может быть вектором (например, шоки в модели векторной авторегрессии VAR).

Говорят, что $\{x_t\}$ обладает L_p -ПЭЗ-свойством на $\{F_t\}$ (или $\{v_t\}$) размера $\lambda > 0$, если

$$\|x_t - \mathbb{E}[x_t | F_{t-m}^{t+m}]\|_p \leq d_t \times \varphi_m,$$

где $d_t > 0$ и $\varphi_m = o(m^{-\lambda})$. Константы $\{d_t\}$ позволяют вычлнить тренды моментов, и возможно стремление $d_t \rightarrow \infty$ при $t \rightarrow \infty$ (например, гетероскедастичность типа линейного тренда $\mathbb{E}[x_t^2] = a + bt$, $a, b > 0$). Коэффициенты позволяют моделировать персистентность. Случайная величина $\{v_t\}$ называется *базой* ПЭЗ.

Двусторонность F_{t-m}^{t+m} позволяет включить двусторонние временные ряды, такие как линейные лаги, распределенные в обе стороны:

$$x_t = \sum_{i=-\infty}^{\infty} \psi_i \epsilon_{t-i}. \quad (5)$$

Вообще говоря, двусторонние ряды редко встречаются в экономических задачах; в этом случае мы заменяем F_{t-m}^t на F_{t-m}^{t+m} .

Словесно, $\{x_t\}$ обладает L_p -ПЭЗ-свойством на $\{F_t\}$, когда последовательность случайных величин $\{v_\tau\}_{\tau=t-m}^{t+m}$ можно использовать для предсказания x_t с нулевой ошибкой в L_p -норме при $m \rightarrow \infty$. Как и в случае перемешивания, $\{x_t\}$ обладает L_p -ПЭЗ-свойством на $\{F_t\}$ размера 2, если, например, $\forall \delta \leq 2$

$$m^\delta \times \|x_t - \mathbb{E}[x_t | F_{t-m}^{t+m}]\|_p \rightarrow 0 \text{ при } m \rightarrow \infty.$$

4.3.2 Примеры почти эпохальной зависимости

Очевидно, iid-последовательность является ПЭЗ на себе: если $\mathfrak{F}_t := \sigma(x_\tau : \tau \leq t)$, то $\mathbb{E}[x_t | \mathfrak{F}_{t-m}^{t+m}] = \mathbb{E}[x_t | \mathfrak{F}_{t-m}^t] = x_t \forall m \geq 1$, и значит $\varphi_m = 0 \forall m \geq 1$.

ПЭЗ-свойство обобщает перемешивание, так как процесс v_t может быть любым. Если $x_t = v_t$ сильно перемешивающий, то x_t , очевидно, ПЭЗ на себе (то есть на $\{F_t\} = \{\mathfrak{F}_t\}$) с $\varphi_m = 0 \forall m \geq 1$.

Проверка ПЭЗ-свойства обычно очень проста. Не все слабостационарные и эргодичные процессы с линейными распределенными лагами сильно перемешивающие (см., например, Andrews 1984, Guegan & Ladoucette 2001), но все обладают ПЭЗ-свойством.

Лемма 4.6 Пусть процесс ϵ_t имеет нулевое среднее и равномерно L_p -ограничен. Положим $x_t := \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}$, где $\sum_{i=0}^{\infty} |\psi_i| < \infty$. Тогда $\{x_t\}$ обладает L_p -ПЭЗ-свойством на $\{\epsilon_t\}$. Размер равен $\lambda > 0$, если $\psi_i = O(i^{-1-\lambda-\iota})$ для бесконечно малого $\iota > 0$.

Любой временной ряд, представимый в виде распределенных лагов бесконечного порядка, строящийся на L_p -ограниченных центрированных инновациях, и с абсолютно суммируемыми коэффициентами, стало быть, обладает L_p -ПЭЗ-свойством. Этот класс процессов включает все стационарные эргодичные процессы AR(p), MA(q) и ARMA(p, q) для конечных p, q .

4.3.3 Свойства почти эпохальной зависимости

Ясно, что не любая функция ПЭЗ-случайных величин обязательно обладает ПЭЗ-свойством. Сравним это наблюдение с тем, что мы знаем про перемешивание: измеримые функции конечного числа лагов перемешивающих случайных величин сами всегда перемешивающие (лемма 4.4). Тем не менее, некоторые простые свойства облегчают получение асимптотических распределений выборочных средних и выборочных ковариаций ПЭЗ-векторов.

Лемма 4.7 (Линейные комбинации сохраняют ПЭЗ) Если $\{x_{i,t} : i = 1, \dots, k\}$ являются L_p -ограниченными и L_p -ПЭЗ на некотором процессе $\{F_t\}$ с константами $\{d_{i,t}\}$ и коэффициентами $\{\varphi_{i,m}\}$ размера $\{\lambda_i\}$, то $\sum_{i=1}^k \psi_i x_{i,t}$ является L_p -ограниченным и L_p -ПЭЗ на $\{F_t\}$ с константами $d_t := \sum_{i=1}^k |\psi_i| d_{i,t}$ и коэффициентами $\varphi_m := \sum_{i=1}^k \varphi_{i,m}$ размера $\min_{1 \leq i \leq k} \{\lambda_i\}$.

Лемма 4.8 (Произведения сохраняют ПЭЗ) Если $\{x_t, y_t\}$ являются L_p -ограниченными и L_p -ПЭЗ на некотором $\{F_t\}$ с константами $\{d_{x,t}, d_{y,t}\}$ и коэффициентами $\{\varphi_{x,m}, \varphi_{y,m}\}$ размера $\{\lambda_x, \lambda_y\}$, то $x_t \times y_t$ является $L_{p/2}$ -ПЭЗ на $\{F_t\}$ с константами $d_t := d_{x,t} d_{y,t}$ и коэффициентами $\varphi_m := \varphi_{x,m} \varphi_{y,m}$ размера $\min_{1 \leq i \leq k} \{\lambda_i\}$.

5 Законы больших чисел

Пусть случайный процесс $\{x_t\}$ имеет среднее $\mu_t := \mathbb{E}(x_t)$, дисперсию $\sigma_t^2 := \mathbb{E}(x_t - \mu_t)^2$ и автоковариации $\gamma(|s - t|) = \mathbb{E}(x_s - \mu_s)(x_t - \mu_t)$. Мы неявно позволяем процессу быть нестационарным (например иметь тренд в среднем или быть гетероскедастичным), но для ясности предполагаем, что автоковариации зависят только от смещений $h = |s - t|$. Например, если x_t есть просто зашумленный линейный тренд

$$x_t = \beta_0 + \beta_1 t + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2),$$

то

$$\begin{aligned} \mu_t &= \beta_0 + \beta_1 t, \\ \sigma_t^2 &= \sigma^2, \\ \mathbb{E}(x_s - \mu_s)(x_t - \mu_t) &= 0 \quad \forall s \neq t. \end{aligned}$$

Разложим выборочную дисперсию следующим образом:

$$\mathbb{E} \left(\frac{1}{n} \sum_{t=1}^n (x_t - \mu_t) \right)^2 = \frac{1}{n^2} \sum_{t=1}^n \sigma_t^2 + \frac{2}{n} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n} \right) \gamma(i). \quad (6)$$

Если мы хотим использовать неравенство Чебышева

$$\mathbb{P} \left\{ \left| \frac{1}{n} \sum_{t=1}^n (x_t - \mu_t) \right| > \varepsilon \right\} \leq \varepsilon^{-2} \mathbb{E} \left(\frac{1}{n} \sum_{t=1}^n (x_t - \mu_t) \right)^2$$

для доказательства ЗБЧ, мы должны убедиться в том, что

$$\frac{1}{n^2} \sum_{t=1}^n \sigma_t^2 + \frac{2}{n} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \gamma(i) \rightarrow 0. \quad (7)$$

Безусловно, существуют и другие методы доказательства различных ЗБЧ. Однако для краткости, а также для того, чтобы интуитивно понять, что говорят различные типы зависимости о (7), мы сфокусируем внимание на предложенном нехитром приеме.

Случай мартингал-разностей наиболее прост и, конечно, включает в себя iid-случай.

5.1 Мартингал-разности

Теорема 5.1 Если $\{x_t, \mathfrak{F}_t\}$ есть мартингал-разность с нулевым средним и дисперсией $\{\sigma_t^2\}$, $1/n \sum_{t=1}^n \sigma_t^2 = o(n)$, то $1/n \sum_{t=1}^n x_t \xrightarrow{p} 0$.

Замечание 1. Если дисперсии равномерно ограничены, $\sup_{t \in \mathbb{N}} \sigma_t^2 \leq K < \infty$, то ясно, что $1/n \sum_{t=1}^n \sigma_t^2 \leq K$, и значит $1/n^2 \sum_{t=1}^n \sigma_t^2 = o(1)$ в силу $K/n \rightarrow 0$. Предположение о мартингал-разности влечет $\sum_{i=1}^{n-1} (1 - i/n) \gamma(i) = 0$.

Замечание 2. Заметим, что масштабирование на $1/n$ всего лишь простой частный случай. Для большого количества интересных нестационарных процессов необходимое масштабирование отличается от $1/n$.

5.2 Неоднородные мартингал-разности

Справедливы следующие еще более интересные результаты.

Лемма 5.1 Если $\{S_n, \mathfrak{F}_n\}$ мартингал, то для любого $p > 1$

$$\mathbb{P} \left\{ \max_{1 \leq k \leq n} |S_k| > \varepsilon \right\} \leq \frac{\mathbb{E} |S_n|^p}{\varepsilon^p}.$$

Замечание. Это достаточно тонкий результат, так как из неравенства Чебышева следует лишь, что

$$\mathbb{P} \left\{ \max_{1 \leq k \leq n} |S_k| > \varepsilon \right\} \leq \frac{\mathbb{E} [\max_{1 \leq k \leq n} |S_k|]^p}{\varepsilon^p}.$$

В следующей лемме в правой части неравенства убирается операция взятия максимума, что делает результат исключительно полезным.

Лемма 5.2 Пусть $\{x_t\}$ есть случайный процесс и $S_n = \sum_{t=1}^n x_t$. Предположим, что существует последовательность положительных чисел $\{c_t\}$, такая что для некоторого $p > 0$, каждого $t \geq 0$ и $n > t$, и каждого $\varepsilon > 0$

$$\mathbb{P} \left\{ \max_{m < j \leq n} |S_j - S_m| > \varepsilon \right\} \leq \frac{K}{\varepsilon^2} \sum_{t=m+1}^n c_t^p.$$

Если $\sum_{t=1}^{\infty} c_t^p < \infty$, то $S_n \xrightarrow{a.s.} S$ (последовательность сходится).

Теорема 5.2 Если $\{x_t, \mathfrak{F}_t\}$ есть мартингал-разность с нулевым средним и последовательностью дисперсий $\{\sigma_t^2\}$, удовлетворяющей $\sum_{t=1}^{\infty} \sigma_t^2 / a_t^2 < \infty$ для последовательности положительных чисел $\{a_t\}$, $a_t \uparrow \infty$, то $1/a_n \sum_{t=1}^n x_t \xrightarrow{p} 0$.

Легко привести простые примеры. Если $\{x_t\}$ — слабостационарная мартингал-разность, то $\sigma^2 = \sigma_t^2 < \infty$, и $\sum_{t=1}^{\infty} 1/t^2 < \infty$ выполнено тривиально, и значит $1/n \sum_{t=1}^n x_t \xrightarrow{P} 0$. Вообще, если $\sup_{t \in \mathbb{N}} \sigma_t^2 \leq K < \infty$, то $\sum_{t=1}^{\infty} \sigma_t^2/t^2 \leq K \sum_{t=1}^{\infty} 1/t^2 < \infty$.

Предположение о конечных дисперсиях можно отбросить.

Теорема 5.3 Если $\{x_t, \mathfrak{F}_t\}$ имеет нулевое среднее и удовлетворяет $\sum_{t=1}^{\infty} \mathbb{E}|x_t|^p/a_t < \infty$ для некоторой последовательности положительных чисел $\{a_t\}$, $a_t \uparrow \infty$, и $1 \leq p \leq 2$, то $1/a_n \sum_{t=1}^n x_t \xrightarrow{a.s.} 0$.

5.3 Процессы с сильным перемешиванием

Предположим, что дисперсии σ_t^2 и ковариации $\gamma(t, t-h)$ равномерно ограничены,

$$B := \sup_{t \in \mathbb{N}} \sigma_t^2 < K \text{ и } B_h := \sup_{t \in \mathbb{N}} |\gamma(t, t-h)| < K,$$

и запишем

$$\mathbb{E} \left(\frac{1}{n} \sum_{t=1}^n (x_t - \mu_t) \right)^2 = \frac{1}{n^2} \sum_{t=1}^n \sigma_t^2 + \frac{1}{n^2} \sum_{i=1}^{n-1} (n-i) \gamma(i) \leq \frac{B}{n} + \frac{1}{n} \sum_{i=1}^n B_i.$$

Лемма 5.3 Если $\{x_t\}$ равномерно L_2 -ограничен и $\sum_{i=1}^{\infty} i^{-1} B_i < \infty$, то $1/n \sum_{t=1}^n x_t \xrightarrow{P} 0$.

Приложения леммы 5.3 очевидны. Необходимо лишь установить ограниченную память в том смысле, что $\sum_{i=1}^{\infty} i^{-1} B_i < \infty$. Хотя это условие не вполне интуитивно, легко проверить, что оно выполняется для обширного класса случайных процессов. Рассмотрим следующий простой результат.

Теорема 5.4 Пусть $\{x_t\}$ имеет нулевое среднее, равномерно $L_{2+\delta}$ -ограничен, и сильно перемешивающий с коэффициентами $\sum_{m=1}^{\infty} m^{-1} \alpha_m^{\delta/(2+\delta)} < \infty$. Тогда $1/n \sum_{t=1}^n x_t \xrightarrow{P} 0$.

5.3.1 Почти эпохально зависимые процессы

Красота ПЭЗ-случайных величин x_t состоит в том, что они являются миксингалами, а миксингалы при минимальных дополнительных предположениях имеют $O(1/n)$ -ограниченные выборочные дисперсии $\mathbb{E}(1/n \sum_{t=1}^n (x_t - \mu_t))^2$. Иными словами, при подходящих условиях

$$\mathbb{E} \left(\frac{1}{n} \sum_{t=1}^n (x_t - \mu_t) \right)^2 \rightarrow 0,$$

и установления неравенства Чебышева достаточно для справедливости ЗБЧ. Формальное определение миксингала см. в работах McLeish (1974, 1975) и Davidson (1994). Следующие рассуждения частично адаптированы из работы de Jong (1997).

Необходимо ввести понятие двойного (или стохастического) массива $\{x_{n,t}\}$. Далее по тексту $\{x_{n,t}\}$ нужно представлять себе как треугольный массив: $\{\{x_{n,t}\}_{t=1}^n\}_{n=1}^{\infty}$. Классический пример такого массива:

$$x_{n,t} := \frac{1}{\sqrt{n}} \left(\frac{x_t - \mathbb{E}[x_t]}{\sqrt{\mathbb{E}(x_t - \mathbb{E}[x_t])^2}} \right), \text{ где } x_t \sim iid.$$

Определим σ -алгебру, порожденную базой ПЭЗ: $F_t := \sigma(\epsilon_\tau : \tau \leq t)$.

Теорема 5.5 *Предположим следующее:*

- a. $\{x_{n,t}\}$ является центрированным, L_r -ограниченным, $r > 2$, стохастическим массивом.
- b. $\{x_{n,t}\}$ является L_2 -ПЭЗ размера $1/2$ с константами $\{d_{n,t}\}$ и коэффициентами $\{\varphi_m\}$. База ϵ_t равномерно перемешивающая размера $r/[2(r-1)]$, или сильно перемешивающая размера $r/(r-2)$.
- c. Константы $\{d_{n,t}\}$ удовлетворяют

$$\sum_{t=1}^n (\max\{\|x_{n,t}\|_r, d_{n,t}\})^2 = O(1).$$

Тогда существует последовательность положительных постоянных $\{a_n\}$, $a_n \nearrow \infty$, удовлетворяющих $1/a_n \sum_{t=1}^n x_{n,t} \xrightarrow{P} 0$.

Случай строгой стационарности формулируется в терминах $\{x_t\}$ вместо массива $\{x_{n,t}\}$, что позволяет избежать громоздких обозначений.

Следствие 5.1 *Пусть $\{x_t\}$ строго стационарен, имеет среднее $\mathbb{E}[x_t]$, L_r -ограничен для некоторого $r > 2$. Также предположим, что $\{x_t\}$ обладает L_2 -ПЭЗ-свойством размера $1/2$ с константами $d_t = d$ и коэффициентами φ_m и ПЭЗ-базой $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$. Тогда $\{x_t\}$ эргодичен для среднего: $1/n \sum_{t=1}^n x_t \xrightarrow{P} \mathbb{E}[x_t]$.*

6 Центральные предельные теоремы

В настоящем разделе мы преследуем цель получить достаточно ограничительные условия, которые гарантировали бы, что некоторый стандартизованный треугольный массив $\{x_{n,t}\}$, удовлетворяющий $\|\sum_{t=1}^n x_{n,t}\|_2 = 1$, подчинялся бы

$$\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, 1).$$

6.1 Мартингал-разности

Несмотря на то, что мартингал-разности обладают тривиальной структурой, в отличие от ЗБЧ для мартингал-разностей необходимо будет наложить некоторые условия, касающиеся неоднородности. Следующий результат, представленный в работе Davidson (1994, теорема 24.3), использует более общее утверждение, доказанное в работе McLeish (1974).

Теорема 6.1 *Пусть $\{x_{n,t}, \mathfrak{F}_{n,t}\}$ есть последовательность мартингал-разностей с конечной дисперсией $\{\sigma_{n,t}^2\}$, и $\sum_{t=1}^n \sigma_{n,t}^2 = 1$. Если $\sum_{t=1}^n x_{n,t}^2 \xrightarrow{P} 1$ и $\max_{1 \leq t \leq n} |x_{n,t}| \xrightarrow{P} 0$, то $\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, 1)$.*

Рассмотрим стандартный случай

$$x_{n,t} = \frac{1}{\sqrt{n}} \left(\frac{x_t - \mathbb{E}[x_t]}{\sqrt{\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n (x_t - \mathbb{E}[x_t])^2)}} \right)$$

где $\{x_t, \mathfrak{F}_t\}$ есть мартингал-разность с дисперсиями $\{\sigma_t^2\}$, $\sigma_t^2 > 0$ равномерно по t . Заметим, что $\mathfrak{S}_{n,t} = \mathfrak{F}_t$, так как неслучайное n не рафинирует σ -алгебру. Предположение о мартингал-разности влечет

$$\mathbb{E} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n (x_t - \mathbb{E}[x_t]) \right)^2 = \frac{1}{n} \sum_{t=1}^n \sigma_t^2,$$

и значит

$$\sigma_{n,t}^2 = \frac{1}{n} \frac{\sigma_t^2}{\frac{1}{n} \sum_{t=1}^n \sigma_t^2} = \frac{\sigma_t^2}{\sum_{t=1}^n \sigma_t^2},$$

поэтому, очевидно, $\sum_{t=1}^n \sigma_{n,t}^2 = 1$.

Условие $\max_{1 \leq t \leq n} |x_{n,t}| \xrightarrow{p} 0$ влечет

$$\max_{1 \leq t \leq n} \left| \frac{x_t - \mathbb{E}[x_t]}{\sqrt{\sum_{t=1}^n \sigma_t^2}} \right| \xrightarrow{p} 0,$$

что почти тривиально, так как $\sum_{t=1}^n \sigma_t^2 \rightarrow \infty$ из-за того, что $\inf_t \sigma_t^2 > 0$. Необходимо лишь предположить, что ни одно $|x_t - \mathbb{E}[x_t]|$ не доминирует над всеми остальными.

Наконец, используем свойство мартингал-разности, чтобы получить

$$\sum_{t=1}^n x_{n,t}^2 = \frac{\sum_{t=1}^n (x_t - \mathbb{E}[x_t])^2}{\mathbb{E}(\sum_{t=1}^n (x_t - \mathbb{E}[x_t])^2)} = \frac{\sum_{t=1}^n (x_t - \mathbb{E}[x_t])^2}{\sum_{t=1}^n \mathbb{E}(x_t - \mathbb{E}[x_t])^2}.$$

Условие $\sum_{t=1}^n x_{n,t}^2 \xrightarrow{p} 1$ сводится к тому, чтобы процесс $\sum_{t=1}^n (x_t - \mathbb{E}[x_t])^2$ удовлетворял ЗБЧ. Достаточными условиями этого являются: $\{(x_t - \mathbb{E}[x_t])^2 - \sigma_t^2, \mathfrak{F}_t\}$ есть мартингал-разность, где $\mathbb{E}((x_t - \mathbb{E}[x_t])^2 - \sigma_t^2)^2$ постоянно равномерно по t .

Условие $\max_{1 \leq t \leq n} |x_{n,t}| \xrightarrow{p} 0$ является немного избыточным, хотя само по себе достаточно слабо. Следующий результат представлен в работе White (1984, следствие 5.25). Идея в том, что даже если $\{x_t, \mathfrak{F}_t\}$ неоднороден (например, не обладает постоянными дисперсиями), пока неоднородность не носит взрывного характера (то есть ни одно наблюдение не доминирует над всеми остальными), центральная предельная теорема остается верной.

Теорема 6.2 Пусть $\{x_t, \mathfrak{F}_t\}$ есть мартингал-разность с конечными дисперсиями $\{\sigma_t^2\}$, и $1/n \sum_{t=1}^n \sigma_t^2 \rightarrow \sigma^2$, $\sup_{t \in \mathbb{Z}} \mathbb{E}|x_t|^r < \infty$ для некоторого $r > 2$ и $1/n \sum_{t=1}^n x_t^2 \xrightarrow{p} \sigma^2$, тогда $1/\sqrt{n} \sum_{t=1}^n x_t \xrightarrow{d} N(0, \sigma^2)$.

6.2 Почти эпохально зависимые массивы

Следующий результат представляет собой несколько усеченную версию теоремы 2 из работы de Jong (1997).

Теорема 6.3 Пусть $\{x_{n,t}\}$ есть треугольный массив с нулевым средним, и предположим следующее:

- Существует последовательность положительных чисел $\{c_{n,t}\}$, такая что $\{x_{n,t}/c_{n,t}\}$ является L_r -ограниченным для некоторого $r > 2$, равномерно по t и n .
- $\{x_{n,t}\}$ обладает L_2 -ПЭЗ-свойством размера $1/2$ с константами $\{d_{n,t}\}$ и коэффициентами $\{\varphi_m\}$. База ϵ_t равномерно перемешивающая размера $r/[2(r-1)]$ или сильно

перемешивающая размера $r/(r - 2)$.

с. Числа $\{c_{n,t}\}$ удовлетворяют $\sum_{t=1}^n c_{n,t}^2 = O(1)$.

Тогда $\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, 1)$.

Замечание. Так как в теореме 6.3 требуется лишь, чтобы $\{x_{n,t}\}$ обладал ПЭЗ-свойством на равномерно или сильно перемешивающем процессе, перемешивающие процессы соответствующего размера, очевидно, удовлетворяют сформулированным условиям.

Рассмотрим упрощенное следствие.

Следствие 6.1 Пусть $\{x_t\}$ слабо стационарен и L_r -ограничен для некоторого $r > 2$. Предположим также, что $\{x_t\}$ обладает L_2 -ПЭЗ-свойством размера $1/2$ с константами $d_t = d$ и коэффициентами φ_m и ПЭЗ-базой $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$. Наконец, пусть

$$\liminf_{n \geq 1} \mathbb{E} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\} \right)^2 > 0.$$

Запишем

$$x_{n,t} = \frac{1}{\sqrt{n}} \left(\frac{x_t - \mathbb{E}[x_t]}{\sqrt{\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}^2)}} \right).$$

Тогда $\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, 1)$ и $\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}^2) = O(1)$.

Замечание. Заметим, что мы предположили невырожденность стандартного отклонения $\liminf_{n \geq 1} \mathbb{E}(1/\sqrt{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}^2) > 0$, а сформулированный результат гарантирует, что оно также ограничено: $\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}^2) \leq K$. Следовательно, $\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n \{x_t - \mathbb{E}[x_t]\}^2) \rightarrow v > 0$, где $v < \infty$.

6.3 Метод Крамера–Волда

Обобщение вышеприведенных ЦПТ на многомерный случай $x_t \in \mathbb{R}^k$ как правило не сложнее, чем доказательство ЦПТ для линейной комбинации $\sum_{i=1}^k r_i x_{i,t}$.

Теорема 6.4 (Метод Крамера–Волда) Если $\{x_{n,t}\}$ есть \mathbb{R}^k -значный стохастический массив, и $\forall r \in \mathbb{R}^k$

$$\sum_{t=1}^n r' x_{n,t} \xrightarrow{d} N(0, r'r),$$

то $\sum_{t=1}^n x_{n,t} \xrightarrow{d} N(0, I_k)$.

Следствие 6.2 Если $\{x_t\}$ суть \mathbb{R}^k -значный случайный процесс $\forall r \in \mathbb{R}^k$, и

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n r' x_t \xrightarrow{d} N(0, r'\Sigma r),$$

то $1/\sqrt{n} \sum_{t=1}^n x_t \xrightarrow{d} N(0, \Sigma)$, где $\Sigma = \lim_{n \rightarrow \infty} \mathbb{E}(1/\sqrt{n} \sum_{t=1}^n x_t)^2$.

Взятие линейных комбинаций сохраняет и свойство перемешиваемости, и ПЭЗ-свойство. В свете этого теорема 6.4 и следствие 6.2 крайне полезны для получения предельных распределений многомерных оценок (например квази-ММП, МНК, ОММ).

7 Теория слабых пределов и броуновское движение

В настоящем разделе мы представим концепции функциональных пределов, полезные для характеристики предельных распределений оценок, связанных с процессами с единичными корнями. Мы демонстрируем главные результаты на примере вывода предельного распределения МНК-оценки коэффициента наклона в модели AR(1) с единичным корнем.

7.1 Случайное блуждание без сноса: пример

Рассмотрим случайное блуждание без сноса:

$$x_t = x_{t-1} + \epsilon_t, \quad \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2).$$

Рассмотрим также МНК-оценку $\hat{\rho}$ коэффициента наклона ρ в модели AR(1)

$$x_t = \rho x_{t-1} + \epsilon_t, \quad \text{где } \rho = 1.$$

Тогда

$$\hat{\rho} = \frac{\sum_{t=2}^n x_t x_{t-1}}{\sum_{t=2}^n x_{t-1}^2} = 1 + \frac{\sum_{t=2}^n \epsilon_t x_{t-1}}{\sum_{t=2}^n x_{t-1}^2}.$$

Таким образом, $\hat{\rho} \xrightarrow{p} 1$, если правильным образом масштабированные суммы $\sum_{t=2}^n \epsilon_t x_{t-1} \xrightarrow{p} 0$. Но учитывая что

$$x_t = \sum_{i=0}^{\infty} \epsilon_{t-i}$$

обладает бесконечной дисперсией

$$\mathbb{E}[x_t^2] = \sum_{i=0}^{\infty} \sigma^2 = \infty, \quad \text{поскольку } \epsilon_t \stackrel{iid}{\sim} (0, \sigma^2),$$

каковы пределы по вероятности и по распределению процесса $\sum_{t=2}^n \epsilon_t x_{t-1}$? Принимая во внимание то, что $\mathbb{E}[x_t^2] = \infty$, можем ли мы использовать множитель $1/n$ в обеих компонентах отношения (например $1/n \sum_{t=2}^n x_{t-1}^2$)? Быстрые ответы таковы: *вероятно негауссовские; нет.*

Мы получим предельное распределение оценки $\hat{\rho}$ последовательно, сначала рассмотрев функциональную центральную предельную теорию, а потом решая эту задачу для $\hat{\rho}$ непосредственно.

7.2 Функциональная предельная теория

Как и все термины в настоящем эссе, этот термин имеет свою аббревиатуру. Изложение будет касаться только тех результатов, которые нужны для определения предела по распределению оценки $\hat{\rho}$ раздела 7.1.

Пусть $\{x_t\}$ — произвольный случайный процесс с нулевым средним. Если $x_t \stackrel{iid}{\sim} (0, \sigma^2)$, то (например, по ЦПТ Линдберга–Леви)

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n x_t \xrightarrow{d} N(0, \sigma^2).$$

Здесь n совершенно произвольно, пока предел берется при $n \rightarrow \infty$. Результат остается верным для любой доли $[\lambda n]$ от n , где $\lambda \in [0, 1]$, а $[z]$ обозначает целую часть числа z . Простое масштабирование и тот факт, что $[\lambda n]/\lambda n \rightarrow 1$, позволяют написать

$$Z_n(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[\lambda n]} x_t = \frac{\sqrt{[\lambda n]}}{\sqrt{n}} \times \frac{1}{\sqrt{[\lambda n]}} \sum_{t=1}^{[\lambda n]} x_t \xrightarrow{d} N(0, \lambda) = Z(\lambda),$$

скажем, где предельное распределение $Z(\lambda)$ является нормальным с нулевым средним и дисперсией λ . Таким образом, «функция» $Z_n(\lambda)$, вычисленная в точке $\lambda \in [0, 1]$ сходится к нормальному закону с нулевым средним $Z(\lambda)$, который сам является функцией λ . Следовательно,

$$Z_n(\lambda) \xrightarrow{d} Z(\lambda) \text{ поточечно на } [0, 1].$$

Так как гауссовский закон полностью характеризуется средним и дисперсией (Doob 1953), λ полностью характеризует это гауссовское распределение с нулевым средним.

Теперь для каждого n мы считаем

$$\{Z_n(\lambda)\} = \{Z_n(\lambda) : 0 \leq \lambda \leq 1\}$$

случайным процессом, индексированным параметром λ . Каковы свойства этого процесса? Каковы его траектории (то есть как процесс $Z_n(\lambda)$ выглядит для каждого λ для конкретной реализации $\{x_t\}_{t=1}^n$)? К чему сходится процесс $\{Z_n(\lambda)\}$, и сходится ли вообще?

Если процесс $\{Z_n(\lambda)\}$ имеет предел, скажем, $\{Z(\lambda)\}$, то это *слабый предел*¹, или *функциональный предел*. Мы обозначаем тот факт, что $\{Z(\lambda)\}$ является слабым пределом $\{Z_n(\lambda)\}$, символами

$$Z_n(\lambda) \implies Z(\lambda).$$

Заметим, что здесь есть существенные отличия от обычной поточечной сходимости: $Z_n(\lambda) \xrightarrow{d} Z(\lambda)$ для каждого $\lambda \in [0, 1]$. К сожалению, можно привести примеры, когда поточечная сходимость выполняется, но процесс $\{Z_n(\lambda)\}$, являющийся случайным выбором из распределения траекторий, не сходится.

Если $Z(\lambda)$ имеет определенные свойства, то говорят, что $Z_n(\lambda)$ сходится в пространстве с этими свойствами. Например, если $Z(\lambda)$ — непрерывная функция λ , то $Z_n(\lambda)$ сходится на $C[0, 1]$, пространстве непрерывных вещественнозначных функций $[0, 1]$.

Оказывается, для iid, не-iid и вырожденных функций не-iid случайных процессов $\{x_t\}$ процесс частичных сумм $\{Z_n(\lambda)\}$ сходится к предельному процессу $\{Z(\lambda)\}$, который идентичен *броуновскому движению*. Это верно для мартингал-разностей, перемешивающих и ПЭЗ-процессов $\{x_t\}$ (Davidson 1994, de Jong & Davidson 2000, Hill 2009).

Мы рассмотрим только iid-случай, но читатель может посмотреть формулировки аналогичных результатов для мартингал-разностей в соответствующей литературе. Дадим сначала определения броуновского движения и меры Винера (Wiener 1923).

Броуновское движение (винеровский процесс) *Броуновским движением или винеровским процессом* $\{Z(\lambda)\} = \{Z(\lambda) : 0 \leq \lambda \leq 1\}$ называется непрерывный процесс на $[0, 1]$, такой что

$$(i) Z(0) = 0 \text{ почти наверное;}$$

¹Важно не путать это слово «слабый» с состоятельностью, которую часто называют «слабым пределом по вероятности», или слабым пределом.

(ii) $[Z(\lambda_2) - Z(\lambda_1), \dots, Z(\lambda_k) - Z(\lambda_{k-1})]$ для любого $k \in \mathbb{N}$ имеет многомерное нормальное распределение с диагональной ковариационной матрицей для любых $0 \leq \lambda_1 < \lambda_2 < \dots < \lambda_k \leq 1$ и гауссовскими приращениями $Z(\lambda) - Z(\lambda') \sim N(0, \lambda - \lambda')$ для любых $\lambda > \lambda'$;

(iii) $Z(\lambda)$ непрерывен по λ почти наверное.

Замечание 1. Так как $Z(\lambda)$ почти наверное непрерывная функция на отрезке $[0, 1]$, ее траектории принадлежат пространству $C[0, 1]$ с вероятностью единица. Траекторией называется реализация $\{Z(\lambda, \omega) : 0 \leq \lambda \leq 1\}$, где ω — случайный исход из Ω . Реализация броуновского движения изображена на Рис. 1.

Замечание 2. Многомерное нормальное распределение имеет диагональную ковариационную матрицу тогда и только тогда, когда его скалярные компоненты независимы: приращение $Z(\lambda_i) - Z(\lambda_{i-1})$ независимо от приращения $Z(\lambda_j) - Z(\lambda_{j-1}) \forall \lambda_{i-1} < \lambda_i < \lambda_{j-1} < \lambda_j$.

Замечание 3. Ограничение $\lambda \in [0, 1]$ — условность, которой мы пользуемся ниже. Вообще говоря, λ может принадлежать любому компакту $A \subset \mathbb{R}$, или $[0, \infty)$, и соответствующий процесс существует на $C[A]$ или $C[0, \infty)$ соответственно. См. Billingsley (1999).

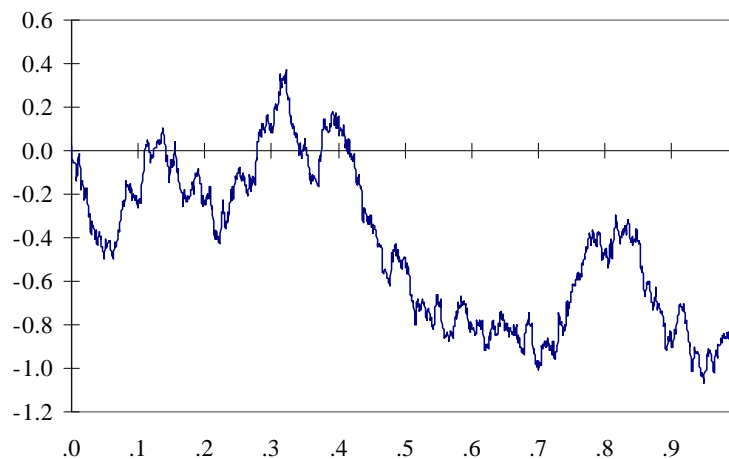


Рис. 1: Реализация броуновского движения

Замечание 4. На самом деле, на Рис. 1 изображено лишь кусочное приближение к траектории броуновского движения, так как мы изобразили не совсем непрерывную функцию (мы соединили лишь конечное множество точек!). А именно, на рисунке изображен

$$x_t = \frac{1}{\sqrt{n}} \sum_{i=0}^{t-1} \epsilon_{t-i} \text{ для } t = 1, \dots, 1000, \text{ где } \epsilon_t \stackrel{iid}{\sim} N(0, 1),$$

что можно переписать как

$$Z_n(\lambda) := \frac{1}{\sqrt{n}} \sum_{i=0}^{[n\lambda]} \epsilon_{t-i} \text{ для } \lambda \in \left\{ \frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n} \right\} = \{0,001; 0,002; \dots, 1\}.$$

Заметим, что $Z_n(\lambda)$ обладает всеми свойствами броуновского движения, кроме непрерывности: мы можем утверждать лишь, что этот процесс непрерывен всюду за исключением счетного числа точек $[0, 1]$. При гладком увеличении λ процесс $Z_n(\lambda)$ претерпевает небольшие скачки, когда λ пересекает точки $1/n, 2/n$ и т. д. Скачки асимптотически исчезают с вероятностью 1. Более того, несмотря на то, что (нулевые с вероятностью 1) скачки всегда асимптотически присутствуют, их всего счетное число. Это может казаться большим

количеством, однако счетное множество имеет нулевую меру. Асимптотически, вероятность скачка нулевая. Иными словами, траектории почти наверное непрерывны, и значит принадлежат $C[0, 1]$ с вероятностью 1. Этого достаточно для доказательства широкого спектра асимптотических результатов, которые мы приводим и используем ниже.

Винеровской мерой называется вероятностная мера (распределение вероятностей) броуновского движения $\{Z(\lambda)\}$.

Мера Винера Мерой Винера W называется вероятностная мера на $C[0, 1]$ со свойствами

$$(i) W(Z(0) = 0) = 1;$$

$$(ii) W(Z(\lambda) \leq a) = \int_{-\infty}^a (\sqrt{2\pi\lambda})^{-1} \exp\{0,5 Z(\lambda)^2/\lambda\};$$

(iii) если $\{Z(\lambda)\} \sim W$, то $Z(\lambda_2) - Z(\lambda_1)$ независим от $Z(\lambda_4) - Z(\lambda_3)$ для всех $0 \leq \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4 \leq 1$.

Замечание 1. Таким образом, мера Винера W есть вероятностная мера, соответствующая броуновскому движению, или винеровскому процессу.

Замечание 2. Легко показать, что определение броуновского движения, данное нами, избыточно: любой процесс $\{Z(\lambda)\} \sim W$ имеет точно те же свойства, которым характеризуется броуновское движение. Однако очень трудно доказать, что винеровская мера существует (то есть процесс $\{Z(\lambda)\}$ действительно обладает соответствующей вероятностной мерой).

Функциональная центральная предельная теорема (ФЦПТ), или теорема о слабом пределе, представленная ниже, позволяет асимптотически характеризовать траектории $\{Z_n(\lambda)\}$, такие как на Рис. 1, а также получить предельное распределение $\hat{\rho}$ из раздела 7.1.

Теорема 7.1 Пусть $\{x_t\}$ есть iid-процесс с нулевым средним и $\mathbb{E}[x_t^2] = \sigma^2 < \infty$, и положим $Z_n(\lambda) = 1/\sqrt{n} \sum_{t=1}^{\lfloor \lambda n \rfloor} x_t$, $\lambda \in [0, 1]$. Тогда

$$Z_n(\lambda) \Longrightarrow Z(\lambda),$$

где $Z(\lambda)$ обладает почти наверное непрерывными траекториями. В частности, $Z(\lambda)$ суть броуновское движение.

Вспомним теорему о непрерывном отображении: если некоторая последовательность $Z_n \xrightarrow{d} Z$, то для непрерывных функций $g(Z_n) \xrightarrow{d} g(Z)$. Фактически, данная теорема позволяет менять местами символы непрерывной функции и слабого предела (Billingsley 1999):

$$g(Z_n(\lambda)) \Longrightarrow g(Z(\lambda)).$$

Следующий полезный результат доказывается с помощью теоремы 7.1 и является обобщением теоремы о непрерывном отображении.

Теорема 7.2 Рассмотрим любую функцию $g: \mathbb{R} \rightarrow \mathbb{R}$, которая непрерывна всюду, за исключением, быть может, счетного числа точек своего носителя. В условиях теоремы 7.1

$$\int_0^1 g(Z_n(\lambda)) d\lambda \xrightarrow{d} \int_0^1 g(Z(\lambda)) d\lambda.$$

7.3 Случайные блуждания со сносом: асимптотическая теория МНК

Вернемся к проблеме МНК-инференции для процесса с единичным корнем. Предположим, что время отсчитывается с момента $t = 1$: $\epsilon_t = 0$ почти наверное $\forall t \leq 0$. Простой подстановкой убеждаемся в том, что

$$x_t = \sum_{i=0}^{t-1} \epsilon_{t-i} = \sum_{\tau=1}^t \epsilon_{\tau}.$$

Мы получили известный результат (Phillips 1987).

Теорема 7.3 *Рассмотрим процесс случайного блуждания из раздела 7.1. МНК-оценка коэффициента наклона удовлетворяет условию*

$$n(\hat{\rho} - 1) \xrightarrow{d} \frac{1}{2} \frac{Z(\lambda)^2 - 1}{\int_0^1 Z(\lambda)^2 d\lambda}.$$

Замечание. Справедливы следующие два важных утверждения: во-первых, предел очевидно негауссов; во-вторых, сходимость теперь имеет скорость n . Таким образом, $\hat{\rho} \xrightarrow{p} 1$ с необычно высокой скоростью².

Список литературы

- An, H.Z. & F.C. Huang (1996). The geometrical ergodicity of nonlinear autoregressive models. *Statistica Sinica* 6, 943–956.
- Andrews, D.W.K. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930–934.
- Bierens, H.J. (1981). *Robust Methods and Asymptotic Theory in Nonlinear Econometrics*. Springer-Verlag.
- Billingsley, P.J. (1999). *Convergence of Probability Measure*. Wiley: New York.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press: Oxford.
- Davidson, J. & R.M. de Jong (2000). The functional central limit theorem and weak convergence to stochastic integrals. *Econometric Theory* 16, 621–642.
- de Jong, R.M. (1997). Central limit theorems for dependent heterogeneous random variables. *Econometric Theory* 13, 353–367.
- Dedecker, J., P. Doukhan, G. Lang, J.R. Leon, S. Louhichi & C. Prieur (2007). *Weak Dependence: with Examples and Applications*. Springer.
- Doob, J. L. (1953). *Stochastic Processes*. New York: Wiley.
- Gallant, A.R. & H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell: Oxford.
- Guegan D. & S. Ladoucette (2001). Non-mixing properties of long memory processes. *Comptes Rendus de l'Academie des Sciences Series I Mathematics* 333, 373–376.
- Hill, J.B. (2009). On functional central limit theorems for dependent, heterogenous arrays with applications to tail index and tail dependence estimation. *Journal of Statistical Planning and Inference* 139, 2091–2110.
- Ibragimov, I.A. (1962). Some limit theorems for stationary processes *Theory of Probability and its Applications* 7, 349–382.
- Leibschner, E. (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *Journal of Time Series Analysis* 26, 669–689.

²На самом деле это неудивительно: МНК-оценки суперсостоятельны (то есть сходятся со скоростью большей, чем \sqrt{n}) для большого количества процессов с бесконечной дисперсией. Процесс с единичным корнем представляет собой простой частный случай процесса с бесконечной дисперсией.

Lévy, P. (1925). *Calcul de Probabilités*. Gauthier-Villars: Paris.

Lévy, P. (1954). *Théorie de l'Addition des Variables Eléatoires*. Gauthier-Villars: Paris.

McLeish, D.L. (1974). Dependent central limit theorems. *Annals of Probability* 2, 620–628.

McLeish, D.L. (1975). A maximal inequality and dependent strong law. *Annals of Probability* 3, 329–339.

Phillips, P.C.B. (1987). Time series regression with a unit root. *Econometrica* 55, 277–301.

Serfling, R.J. (1968). Contributions to central limit theory for dependent variables. *Annals of Mathematical Statistics* 39, 1158–1175.

Wiener, N. (1923). Differential space. *Journal of Mathematics and Physics* 2, 131–174.

White, H. (1984). *Asymptotic Theory for Econometricians*. Academic Press: Orlando.

Приложение: избранные доказательства

Доказательство леммы 2.1. См. Davidson (1994). ■

Доказательство теоремы 2.1. Ясно, что

$$\begin{aligned}\mathbb{E}|x_t|^{1+\delta} &= \mathbb{E}|x_t|^{1+\delta} I(|x_t| \geq M) + \mathbb{E}|x_t|^{1+\delta} I(|x_t| < M) \\ &\geq \mathbb{E}|x_t|^{1+\delta} I(|x_t| \geq M) \geq M^\delta \times \mathbb{E}|x_t| I(|x_t| \geq M).\end{aligned}$$

Следовательно,

$$\mathbb{E}|x_t| I(|x_t| \geq M) \leq \mathbb{E}|x_t|^{1+\delta} / M^\delta \rightarrow 0$$

при $M \rightarrow \infty$ в силу $\mathbb{E}|x_t|^{1+\delta} < \infty$ из-за $L_{1+\delta}$ -ограниченности. ■

Доказательство теоремы 3.1. См. теорему 13.13 в работе Davidson (1994). ■

Доказательство теоремы 3.3. См. следствие 13.14 в работе Davidson (1994). ■

Доказательство теоремы 3.4. Используя слабую стационарность и неравенство Чебышева, можно записать, что для любого $\varepsilon > 0$

$$\begin{aligned}\mathbb{P}\left\{\left|\frac{1}{n}\sum_{t=1}^n\{x_t - \mathbb{E}[x_t]\}\right| > \varepsilon\right\} &\leq \mathbb{E}\left(\frac{1}{n}\sum_{t=1}^n\{x_t - \mathbb{E}[x_t]\}\right)^2 \varepsilon^{-2} \\ &= \varepsilon^{-2} \times \frac{1}{n^2} \sum_{s,t=1}^n \mathbb{C}(x_s, x_t) \\ &= \varepsilon^{-2} \times \left(\frac{\sigma^2}{n} + \frac{1}{n} \sum_{i=1}^{n-1} (1 - i/n) \times \mathbb{C}(x_{i+1}, x_1)\right) \\ &\leq \varepsilon^{-2} \times \left(\frac{\sigma^2}{n} + \frac{1}{n} \sum_{i=1}^n |\mathbb{C}(x_{i+1}, x_1)|\right),\end{aligned}$$

так как $1 - i/n \in [0, 1]$. Из слабой стационарности следует, что $\sigma^2 < \infty$, и значит

$$\frac{1}{n} \sum_{i=1}^n |\mathbb{C}(x_{i+1}, x_1)| \rightarrow 0,$$

чего достаточно для

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{t=1}^n\{x_t - \mathbb{E}[x_t]\}\right| > \varepsilon\right\} \rightarrow 0.$$

Абсолютная суммируемость $\sum_{i=0}^{\infty} |\mathbb{C}(x_{i+1}, x_1)| < \infty$ влечет $\mathbb{C}(x_{N+1}, x_1) \rightarrow 0$ при $N \rightarrow \infty$. Но если $\{\mathbb{C}(x_{i+1}, x_1)\}$ сходится к нулю, то для суммирования Чезаро верно

$$1/n \sum_{i=1}^n |\mathbb{C}(x_{i+1}, x_1)| \rightarrow 0$$

(Davidson 1994, теорема 2.26). ■

Доказательство леммы 3.1. Так как процесс $x_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$ имеет нулевое среднее, и $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$, автоковариации $\gamma(t-s) := \mathbb{E}[x_s x_t]$, $t \geq s$, имеют вид

$$\sum_{i,j=0}^{\infty} \phi^i \phi^j \mathbb{E}[\epsilon_{s-i} \epsilon_{t-j}] = \sigma^2 \sum_{j=i+t-s}^{\infty} \phi^i \phi^j = \sigma^2 \sum_{i=0}^{\infty} \phi^i \phi^{i+t-s} = \sigma^2 \phi^{t-s} / (1 - \phi^2),$$

то есть зависят только от смещения $h := t - s$, но не конкретных значений s и t . Значит, $\gamma(h) = \sigma^2 \phi^h / (1 - \phi^2)$, и

$$\sum_{i=1}^{\infty} |\gamma(h)| = \frac{\sigma^2}{1 - \phi^2} \sum_{i=1}^{\infty} |\phi|^h = \frac{\sigma^2}{(1 - \phi^2)(1 - |\phi|)} < \infty.$$

Абсолютная суммируемость ковариаций влечет эргодичность для среднего, ср. с теоремой 3.4. ■

Доказательство леммы 4.2. См. Ibragimov (1962), а также теорему 14.2 и следствие 14.3 в работе Davidson (1994). ■

Доказательство леммы 4.3. См. Serfling (1968), а также теорему 14.4 и следствие 14.5 в работе Davidson (1994). ■

Доказательство теоремы 4.1. Геометрическое сильное перемешивание и $L_{2+\delta}$ -ограниченность означают, что для некоторого $K > 0$ и малого $\iota > 0$

$$|\mathbb{C}(x_t, x_{t-m})| \leq K \alpha_m^\iota \leq K \rho^{\iota m}.$$

Следовательно,

$$\sum_{m=0}^{\infty} |\mathbb{C}(x_t, x_{t-m})| \leq K (1 - \rho^\iota)^{-1} < \infty.$$

Теперь можно применить теорему 3.4. ■

Доказательство леммы 4.4. См. теорему 14.1 в работе Davidson (1994). ■

Доказательство леммы 4.5. Необходимо применить теорему 3.4. ■

Доказательство леммы 4.6. Положим $F_t := \sigma(\epsilon_\tau : \tau \leq t)$ и заметим, что

$$\mathbb{E}[x_t | F_{t-m}^{t+m}] = \mathbb{E}\left[\sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} | F_{t-m}^t\right] = \sum_{i=0}^m \psi_i \epsilon_{t-i} + \sum_{i=m+1}^{\infty} \psi_i \mathbb{E}[\epsilon_{t-i} | F_{t-m}^{t+m}].$$

Применяя неравенство Минковского и условное неравенство Йенсена, получаем

$$\begin{aligned} \|x_t - \mathbb{E}[x_t | F_{t-m}^{t+m}]\|_2 &= \left\| \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i} - \sum_{i=0}^m \psi_i \epsilon_{t-i} - \sum_{i=m+1}^{\infty} \psi_i \mathbb{E}[\epsilon_{t-i} | F_{t-m}^{t+m}] \right\|_p \\ &= \left\| \sum_{i=m+1}^{\infty} \psi_i (\epsilon_{t-i} - \mathbb{E}[\epsilon_{t-i} | F_{t-m}^{t+m}]) \right\|_p \\ &\leq \sum_{i=m+1}^{\infty} |\psi_i| \times \|\epsilon_{t-i}\|_p \leq \sup_{s \in \mathbb{Z}} \|\epsilon_s\|_p \times \sum_{i=m+1}^{\infty} |\psi_i| = d_t \times \varphi_m, \end{aligned}$$

скажем, где $d_t = \sup_{s \in \mathbb{Z}} \|\epsilon_s\|_p < \infty \forall t$. Ясно, что из $\sum_{i=0}^{\infty} |\psi_i| < \infty$ следует $\sum_{i=m+1}^{\infty} |\psi_i| \rightarrow 0$ при $m \rightarrow \infty$. Если $\psi_i = O(i^{-1-\lambda-\iota})$, то

$$m^\lambda \varphi_m \leq K \sum_{i=m+1}^{\infty} i^{-1-\iota} < \infty,$$

и значит $m^\lambda \varphi_m \rightarrow 0$, то есть ПЭЗ-размер $\lambda > 0$. ■

Доказательство леммы 4.8. См. теорему 17.9 в работе Davidson (1994). ■

Доказательство леммы 4.7. Используя неравенство Минковского и ПЭЗ-свойство, получаем

$$\begin{aligned} \left\| \sum_{i=1}^k \psi_i x_{i,t} - \mathbb{E} \left[\sum_{i=1}^k \psi_i x_{i,t} | F_{t-m}^{t+m} \right] \right\|_p &= \left\| \sum_{i=1}^k \psi_i (x_{i,t} - \mathbb{E}[x_{i,t} | F_{t-m}^{t+m}]) \right\|_p \\ &\leq \sum_{i=1}^k |\psi_i| \|(x_{i,t} - \mathbb{E}[x_{i,t} | F_{t-m}^{t+m}])\|_p \\ &\leq \sum_{i=1}^k |\psi_i| d_{i,t} \varphi_{i,m} \leq \left(\sum_{i=1}^k |\psi_i| d_{i,t} \right) \left(\sum_{i=1}^k \varphi_{i,m} \right) = d_t \times \varphi_m. \end{aligned}$$

■

Доказательство теоремы 5.1. Используя телескопическое свойство математического ожидания и свойство мартигал-разности, получаем, что для всех $h \geq 1$

$$\begin{aligned} \gamma(h) &= \mathbb{E}[x_t x_{t-h}] = \mathbb{E}(x_{t-h} \mathbb{E}[x_t | \mathfrak{S}_{t-h}]) \\ &= \mathbb{E}(x_{t-h} \mathbb{E}(\mathbb{E}[x_t | \mathfrak{S}_{t-1}] | \mathfrak{S}_{t-h})) = \mathbb{E}(x_{t-h} \mathbb{E}(0 | \mathfrak{S}_{t-h})) = 0. \end{aligned}$$

Тогда по предположению

$$\frac{1}{n^2} \sum_{t=1}^n \sigma_t^2 + \frac{2}{n} \sum_{i=1}^{n-1} (1 - i/n) \times \gamma(i) = \frac{1}{n^2} \sum_{t=1}^{n-1} \sigma_t^2 = o(1).$$

■

Доказательство леммы 5.1. См. теорему 15.14 в работе Davidson (1994). ■

Доказательство леммы 5.2. См. следствие 20.2 в работе Davidson (1994). ■

Доказательство теоремы 5.2. Положим $T_m := \sum_{t=1}^m x_t/a_t$ и заметим, что $\{T_{n,m} := T_n - T_m, \mathfrak{S}_n\}$ является мартигалом, так как

$$\begin{aligned} \mathbb{E}[T_{n,m} | \mathfrak{S}_{n-1}] &= \mathbb{E} \left(\sum_{t=m+1}^n x_t/a_t | \mathfrak{S}_{n-1} \right) \\ &= \sum_{t=m+1}^n \mathbb{E}[x_t | \mathfrak{S}_{n-1}] / a_t = \sum_{t=m+1}^{n-1} x_t/a_t = T_{n-1,m}. \end{aligned}$$

Более того, используя телескопическое свойство математического ожидания и свойства мартигал-разности, получаем

$$\mathbb{E}(T_n - T_m)^2 = \sum_{t=m+1}^n \sigma_t^2 / a_t^2.$$

Из леммы 5.1 следует, что

$$\mathbb{P} \left\{ \max_{1 \leq k \leq n} |T_k - T_m| > \varepsilon \right\} \leq \frac{\mathbb{E}(T_n - T_m)^2}{\varepsilon^2} = \varepsilon^{-2} \sum_{t=m+1}^n \sigma_t^2 / a_t^2.$$

Теперь необходимо применить лемму 5.2 с $c_t^2 = \sigma_t^2/a_t^2$, чтобы получить

$$T_n = \sum_{t=1}^m x_t/a_t \xrightarrow{a.s.} T,$$

для некоторого T . Использование леммы Кронекера³ завершает доказательство. Действительно, если $\sum_{t=1}^m x_t/a_t \xrightarrow{a.s.} T$, то $\sum_{t=1}^m x_t/a_n \xrightarrow{a.s.} 0$. ■

Доказательство теоремы 5.3. См. теорему 20.11 в работе Davidson (1994). ■

Доказательство леммы 5.3. Достаточно показать, что $1/n \sum_{i=1}^n B_i \rightarrow 0$. Но это следует из леммы Кронекера: $\sum_{i=1}^{\infty} i^{-1} B_i < \infty \implies 1/n \sum_{i=1}^n B_i \rightarrow 0$. ■

Доказательство теоремы 5.4. Следует из лемм 4.2 и 5.3. ■

Доказательство теоремы 5.5. При условиях (a) и (b) теорема 17.5 в работе Davidson (1994) утверждает, что $\{x_{n,t}, F_t\}$ образует набор L_2 -миксингалов размера $1/2$ с миксингал-константами $c_{n,t} \leq \max\{\|x_{n,t}\|_r, d_{n,t}\}$. Следовательно (см. McLeish 1975),

$$\mathbb{E} \left(\sum_{t=1}^n x_{n,t} \right)^2 = O \left(\sum_{t=1}^n c_{n,t}^2 \right) = O \left(\sum_{t=1}^n (\max\{\|x_{n,t}\|_r, d_{n,t}\})^2 \right).$$

При условии (c)

$$\mathbb{E} \left(\sum_{t=1}^n x_{n,t} \right)^2 = O(1).$$

Но это означает, что для любой последовательности чисел $\{a_n\}$, $a_n \nearrow \infty$,

$$\mathbb{P} \left\{ \left| \frac{1}{a_n} \sum_{t=1}^n x_{n,t} \right| > \varepsilon \right\} \leq \varepsilon^{-2} \mathbb{E} \left(\frac{1}{a_n} \sum_{t=1}^n x_{n,t} \right)^2 = O(1/a_n^2) = o(1).$$

■

Доказательство следствия 5.1. Достаточно проверить выполнение условий теоремы 5.5 для

$$x_{n,t} := \frac{1}{\sqrt{n}} (x_t - \mathbb{E}[x_t]).$$

Ясно, что $\mathbb{E}[x_{n,t}] = 0$, и $x_{n,t}$ является L_r -ограниченным в силу неравенств Минковского и Ляпунова:

$$\|x_{n,t}\|_r \leq \frac{1}{\sqrt{n}} (\|x_t\|_r + \|x_t\|_1) \leq \frac{2}{\sqrt{n}} \|x_t\|_r < \infty.$$

Значит, условие (a) теоремы 5.5 выполнено.

Заметим, что $\{x_{n,t}\}$ является L_2 -ПЭЗ размера $1/2$ на $\{\epsilon_t\}$ с константами $d_{n,t} = d/\sqrt{n}$ и коэффициентами φ_m , поскольку

$$\|x_{n,t} - \mathbb{E}[x_{n,t}|F_{t-m}^{t+m}]\|_2 = (1/\sqrt{n}) \|x_t - \mathbb{E}[x_t|F_{t-m}^{t+m}]\|_2 \leq (d/\sqrt{n}) \varphi_m.$$

Значит, условие (b) теоремы 5.5 выполнено.

³См. Davidson (1994, лемма 2.35).

Из сильной стационарности и L_r -ограниченности заключаем, что

$$\begin{aligned} \sum_{t=1}^n (\max \{ \|x_{n,t}\|_r, d_{n,t} \})^2 &\leq \sum_{t=1}^n (\max \{ \| (1/\sqrt{n}) (x_t - \mathbb{E}[x_t]) \|_r, (d/\sqrt{n}) \})^2 \\ &\leq K \max \{ \|x_t\|_r, d \} \sum_{t=1}^n 1/n = K \max \{ \|x_t\|_r, d \} < \infty, \end{aligned}$$

где равенство выполнено равномерно по t . Значит, условие (c) теоремы 5.5 выполнено. ■

Доказательство следствия 6.2. Используя аргументацию из доказательств теоремы 5.5 и следствия 5.1, заключаем, что $\{1/\sqrt{n}(x_t - \mathbb{E}[x_t])\}$ является L_2 -ПЭЗ размера $1/2$ на $\{F_t\}$ с константами $\tilde{d}_{n,t} = d/\sqrt{n}$ и коэффициентами φ_m . Так как база состоит из независимых и одинаково распределенных случайных величин, она является перемешивающей, и значит вследствие теоремы 17.5 работы Davidson (1994) $\{1/\sqrt{n}(x_t - \mathbb{E}[x_t]), F_t\}$ является L_2 -миксингал-массивом с константами, для которых

$$\tilde{c}_{n,t} \leq \max \{ \|x_{n,t}\|_r, \tilde{d}_{n,t} \} \leq K/\sqrt{n}$$

выполняется равномерно по t . Следовательно, (см. McLeish 1975)

$$\mathbb{E} \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n (x_t - \mathbb{E}[x_t]) \right)^2 = O \left(\sum_{t=1}^n c_{n,t}^2 \right) = O \left(\sum_{t=1}^n 1/n \right) = O(1).$$

Учитывая $\liminf_{n \geq 1} \mathbb{E}(1/\sqrt{n} \sum_{t=1}^n (x_t - \mathbb{E}[x_t]))^2 > 0$, заключаем, что

$$x_{n,t} = \frac{1}{\sqrt{n}} \left(\frac{x_t - \mathbb{E}[x_t]}{\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n x_t)^2} \right)$$

является корректно определенным треугольным массивом.

Все условия теоремы выполняются, если положить $\tilde{c}_{n,t} = K/\sqrt{n}$ для любого конечного $K > 0$. Например, очевидно, что $\{x_{n,t}/c_{n,t}\}$ L_r -ограничен, так как

$$\left\| \frac{x_{n,t}}{c_{n,t}} \right\|_r = K \left\| \left(\frac{x_t - \mathbb{E}[x_t]}{\mathbb{E}(1/\sqrt{n} \sum_{t=1}^n x_t)^2} \right) \right\|_r \leq K \|x_t - \mathbb{E}[x_t]\|_r \leq K \|x_t\|_r,$$

где первое неравенство следует из $\liminf_{n \geq 1} \mathbb{E}(1/\sqrt{n} \sum_{t=1}^n (x_t - \mathbb{E}[x_t]))^2 > 0$, а второе — из неравенств Минковского и Ляпунова. ■

Доказательство теоремы 6.4. См. теоремы 25.5 и 25.6 в работе Davidson (1994). ■

Доказательство теоремы 7.3. Запишем

$$n(\hat{\rho} - 1) = \frac{1/n \sum_{t=2}^n \epsilon_t x_{t-1}}{1/n^2 \sum_{t=2}^n x_{t-1}^2}.$$

Достаточно доказать, что

$$\frac{1}{n} \sum_{t=2}^n \epsilon_t x_{t-1} \xrightarrow{d} \sigma^2 \frac{1}{2} (Z(\lambda)^2 - 1) \quad \text{и} \quad \frac{1}{n^2} \sum_{t=2}^n x_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 Z(\lambda)^2 d\lambda.$$

Шаг 1. Так как $x_t^2 = (x_{t-1} + \epsilon_t)^2 = x_{t-1}^2 + 2x_{t-1}\epsilon_t + \epsilon_t^2$, можно записать

$$x_{t-1}\epsilon_t = \frac{1}{2} (x_t^2 - x_{t-1}^2) - \frac{1}{2} \epsilon_t^2,$$

и значит

$$\sum_{t=2}^n \epsilon_t x_{t-1} = \frac{1}{2} \sum_{t=2}^n (x_t^2 - x_{t-1}^2) - \frac{1}{2} \sum_{t=1}^n \epsilon_t^2 = \frac{1}{2} \left(x_n^2 - \sum_{t=1}^n \epsilon_t^2 \right).$$

Так как $\epsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$, легко показать, что $1/n \sum_{t=1}^n \epsilon_t^2 \xrightarrow{P} \sigma^2$. Рассмотрим член

$$bx_n^2 = \left(\sum_{\tau=1}^n \epsilon_\tau \right)^2.$$

В силу ЦПТ для независимых одинаково распределенных случайных величин и теоремы о непрерывном отображении

$$\frac{1}{n} x_n^2 = \sigma^2 \left(\frac{1}{\sqrt{n}} \sum_{\tau=1}^n \epsilon_\tau / \sigma \right)^2 \xrightarrow{d} \sigma^2 \times Z(1)^2,$$

где $Z(1) \sim N(0, 1)$. Следовательно,

$$\frac{1}{n} \sum_{t=2}^n \epsilon_t x_{t-1} \xrightarrow{d} \frac{1}{2} \left(\sigma^2 Z(1)^2 - \sigma^2 \right) = \sigma^2 \frac{1}{2} \left(Z(1)^2 - 1 \right).$$

Шаг 2. Чтобы доказать, что $1/n^2 \sum_{t=2}^n x_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 Z(\lambda)^2 d\lambda$, где $\{Z(\lambda) : 0 \leq \lambda \leq 1\}$ есть броуновское движение, заметим, что $1/n \sum_{t=2}^n (x_{t-1}^2/n)$ приближает интеграл от x_{t-1}^2/n , и для любого t подстановкой обратно по времени получаем

$$x_t / \sqrt{n} = \frac{1}{\sqrt{n}} \sum_{\tau=1}^{[n \times t/n]} \epsilon_\tau.$$

Предел $1/n \sum_{t=2}^n (x_{t-1}^2/n)$ легко посчитать, если определить кусочный процесс

$$\begin{aligned} S_n(\lambda) &= 0 \text{ если } 0 \leq \lambda < 1/n, \\ &= x_1^2/n \text{ если } 1/n \leq \lambda < 2/n, \\ &\dots \\ &= x_{n-1}^2/n \text{ если } (n-1)/n \leq \lambda < n/n = 1, \\ &= x_n^2/n \text{ если } \lambda = 1. \end{aligned}$$

Следовательно, по определению,

$$\int_0^1 S_n(\lambda) d\lambda = \frac{1}{n} \sum_{t=2}^n x_{t-1}^2/n = \frac{1}{n^2} \sum_{t=2}^n x_{t-1}^2.$$

Далее, согласно теоремам 7.1 и 7.2 имеем слабую сходимость на $C[0, 1]$:

$$S_n(\lambda) = \sigma^2 \times \left(\frac{1}{\sqrt{n}} \sum_{t=1}^{[n\lambda]} \epsilon_t / \sigma \right)^2 \implies \sigma^2 \times Z(\lambda)^2.$$

Используя теорему о непрерывном отображении, заключаем, что

$$\int_0^1 S_n(\lambda) d\lambda \xrightarrow{d} \sigma^2 \int_0^1 Z(\lambda)^2 d\lambda.$$

■

Dependence and stochastic limit theory

Jonathan Hill

University of North Carolina, Chapel Hill, USA

In this essay we provide the basic asymptotic theory that serves as background theory for estimators in time series. We outline concepts of dependence used for stochastic limit theory, covering mixing, mixingale and near epoch dependence properties. We then detail some of the most general probability and distribution limit theorems for these processes popularly employed for time series theory and applications.

