Центральная предельная теорема

❗ Исследуйте центральную предельную теорему

Это краткий обзор центральной предельной теоремы (ЦПТ), целью которого является придать конкретный смысл природе сходимости (по распределению, *а не* по вероятности).
Эта страница содержит (в конце) интерактивный контент, который вы можете свободно изучать и изменять. Для более сложных изменений рекомендуется запустить и отредактировать интерактивный Jupyter Notebook. Вы можете:

- Запустить его на компьютере с установленным Python/Jupyter-lab (требуются `ipywidgets`, `numpy`, `matplotlib`).
- Запустить его с помощью онлайн-сервиса. Официальный сайт Jupyter предоставляет бесплатный и открытый сервис. Если вам нужна большая вычислительная мощность, вы можете импортировать блокнот в Google Colab (требуется учётная запись Google).

Результаты о сходимости

Рассмотрим некоторые основные результаты о сходимости для сумм центрированных случайных величин с конечной дисперсией.

Классическая формулировка

Теорема 0.1 (Центральная предельная теорема). *Пусть $(X_n)_{n\geq 1}$ — последовательность независимых одинаково распределённых (i.i.d.) вещественнозначных случайных величин с $\mathbb{E}[|X_n|^2] < \infty$. Обозначим $m := \mathbb{E}[X_n]$, $\sigma := \sqrt{\mathbb{D}[X_n]}$ и*

$$S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - m}{\sigma} \tag{1}$$

Тогда S_n сходится ***по распределению*** *к стандартной нормальной случайной величине, скажем $Z \sim \mathcal{N}(0,1)$. Другими словами, для каждой ограниченной измеримой функции $f : \mathbb{R} \to \mathbb{R}$, непрерывной почти всюду, выполняется*

$$\lim_{n\to\infty} \mathbb{E}[f(S_n)] = \frac{1}{\sqrt{2\pi}} \int f(x) e^{-x^2/2} dx$$

В частности, из предыдущей теоремы можно вывести равномерную сходимость функции распределения

$$\lim_{n\to\infty} \sup_{a<b} |\mathbb{P}(a < S_n \leq b) - \mathbb{P}(a < Z \leq b)| = 0$$

Количественная версия

Предыдущая Теорема 0.1 не затрагивает скорость сходимости.

Теорема 0.2 (Количественная центральная предельная теорема)**.** *Пусть $(Y_n)_{n\geq 1}$ — последовательность* ***независимых*** *случайных величин с $\mathbb{E}[Y_n] = 0$ и $\mathbb{E}[Y_n^2] = 1$, $\mathbb{E}[|Y_n|^3] < \infty$. Пусть*

$$S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$$

Для $g \in C_b^3(\mathbb{R})$ с $C := \sup_x |g'''(x)|$ выполняется

$$|\mathbb{E}[g(S_n)] - \mathbb{E}[g(Z)]| \leq \frac{C}{6\sqrt{n}} \left(\frac{2^{3/2}}{\sqrt{\pi}} + \frac{1}{n} \sum_{k=1}^{n} \mathbb{E}[|Y_k|^3] \right)$$

где $Z \sim \mathcal{N}(0,1)$; а именно

$$\mathbb{E}[g(Z)] = \tfrac{1}{\sqrt{2\pi}} \int g(x) e^{-x^2/2} dx$$

Лемма 0.1. *Пусть V, Y и Z — три случайные величины, такие что*

- *V и Y независимы; V и Z независимы.*
- *Y и Z имеют конечный третий момент.*
- *$\mathbb{E}[Y] = \mathbb{E}[Z]$ и $\mathbb{E}[Y^2] = \mathbb{E}[Z^2]$.*

Тогда для любой $g \in C_b^3$, полагая $C := \sup_{x\in\mathbb{R}} |g'''(x)|$, выполняется следующее неравенство:

$$|\mathbb{E}[g(V+Y)] - \mathbb{E}[g(V+Z)]| \leq \frac{C}{6} \left(\mathbb{E}[|Y|^3] + \mathbb{E}[|Z|^3] \right)$$

Лемма 0.1. По формуле Тейлора, для трёх точек $v, y, z \in \mathbb{R}$ выполняется

$$g(v+y) - g(v+z) = g'(v)(y-z) + \tfrac{1}{2} g''(v)(y^2 - z^2) + R(v,y) - R(v,z)$$

где остаточные члены $R(v,\cdot)$ ограничены как $|R(v,x)| \leq C|x|^3/6$. Теперь вычислим предыдущую формулу для каждого $\omega \in \Omega$ в точках $v = V(\omega)$, $y = Y(\omega)$ и $z = Z(\omega)$, и возьмём математическое ожидание. Тогда $\mathbb{E}[g'(V)(Y-Z)] = \mathbb{E}[g'(V)]\mathbb{E}[(Y-Z)] = 0$ (используя гипотезы о независимости и равенстве математических ожиданий). Аналогично для члена $g''(V)(Y^2 - Z^2)$. Таким образом

$$|\mathbb{E}[g(V+Y) - g(V+Z)]| = |\mathbb{E}[R(V,Y) - R(V,Z)]| \leq \frac{C}{6} \left(\mathbb{E}[|Y|^3] + \mathbb{E}[|Z|^3] \right)$$

□

Лемма 0.2. *Пусть $g \in C_b^3(\mathbb{R})$, пусть $Y_1, \ldots, Y_n$ — независимые случайные величины, и $Z_1, \ldots, Z_n$ — другой набор независимых случайных величин. Предположим, что $\mathbb{E}[Y_i] = \mathbb{E}[Z_i]$ и $\mathbb{E}[Y_i^2] = \mathbb{E}[Z_i^2] < \infty$. Пусть C определено как в Лемма 0.1. Тогда*

$$\left| \mathbb{E}\left[g\left(\tfrac{Y_1 + \ldots + Y_n}{\sqrt{n}} \right) - g\left(\tfrac{Z_1 + \ldots + Z_n}{\sqrt{n}} \right) \right] \right| \leq \frac{C}{6n^{3/2}} \sum_{k=1}^{n} \left(\mathbb{E}[|Y_k|^3] + \mathbb{E}[|Z_k|^3] \right)$$

В частности, если Y_i являются i.i.d. и Z_i являются i.i.d. (в общем случае с другим распределением)

$$\left|\mathbb{E}\left[g\left(\tfrac{Y_1+\ldots+Y_n}{\sqrt{n}}\right)-g\left(\tfrac{Z_1+\ldots+Z_n}{\sqrt{n}}\right)\right]\right| \le \frac{C\left(\mathbb{E}[|Y_1|^3]+\mathbb{E}[|Z_1|^3]\right)}{6\sqrt{n}}$$

Лемма 0.2. Без ограничения общности, можно предположить, что все $Y_1, \ldots, Y_n, Z_1, \ldots, Z_n$ являются независимыми случайными величинами. Тогда запишем $V_k = (Y_1 + \ldots + Y_{k-1} + Z_{k+1} + \ldots + Z_n)/\sqrt{n}$, чтобы получить

$$\begin{aligned}\left|\mathbb{E}\left[g\left(\tfrac{Y_1+\ldots+Y_n}{\sqrt{n}}\right)-g\left(\tfrac{Z_1+\ldots+Z_n}{\sqrt{n}}\right)\right]\right| &= \left|\sum_{k=1}^{n}\mathbb{E}\left[g\left(V_k+\tfrac{Y_k}{\sqrt{n}}\right)-g\left(V_k+\tfrac{Z_k}{\sqrt{n}}\right)\right]\right| \\ &\le \sum_{k=1}^{n}\frac{C}{6}\left(\mathbb{E}[|Y_k/\sqrt{n}|^3]+\mathbb{E}[|Z_k/\sqrt{n}|^3]\right)\end{aligned}$$

где в последнем неравенстве мы использовали Лемма 0.1 n раз. □

Теорема 0.2. Если Z_i являются i.i.d. стандартными нормальными, то $(Z_1+\ldots+Z_n)/\sqrt{n}$ также является стандартной нормальной величиной, и, следовательно, её распределение не зависит от n. Теорема 0.2, таким образом, является следствием Лемма 0.2 и тождества $\mathbb{E}[|Z_i|^3] = 2^{3/2}/\sqrt{\pi}$. □

Упражнение 0.1. Пусть $Z \sim \mathcal{N}(0,1)$ — стандартная нормальная случайная величина. Пусть $(Y_n)_{n\ge1}$ — i.i.d. последовательность с $\mathbb{E}[Y_i^k] = \mathbb{E}[Z^k]$ для $k = 1, \ldots, \ell$. Пусть $g \in C_b^\ell(\mathbb{R})$. Докажите, что существует константа C (зависящая от g и распределения Y_i), такая что

$$|\mathbb{E}[g(S_n)] - \mathbb{E}[g(Z)]| \le Cn^{-(\ell-1)/2}$$

Мартингальная версия

Стоит упомянуть, что Центральная предельная теорема выходит далеко за рамки независимых случайных величин. В конечном счёте, для такого типа результата даже не требуется, чтобы величины были определены на линейном пространстве (например, совершение малых случайных шагов на многообразии, по мере уменьшения шагов, будет сходиться к распределению на пространстве непрерывных кривых на многообразии, называемому броуновским движением). Таким образом, существуют сильные локальные версии ЦПТ, версии для метрических пространств, эргодические версии и так далее. Интересный пример, требующий лишь элементарных гипотез, охватывает случай **мартингалов**.

Теорема 0.3 (Центральная предельная теорема для мартингалов). *Пусть $(X_n)_{n\ge1}$ — последовательность вещественнозначных случайных величин и пусть*

$$M_n := X_1 + \ldots + X_n$$

Предположим, что

- $\mathbb{E}[X_n|M_{n-1}] = 0$.
- *Для $Q_n := \mathbb{E}[X_n^2|M_{n-1}]$, выполняется $\sum_{n=1}^\infty Q_n = \infty$ п.н. (почти наверное).*
- $\mathbb{E}\left[\sup_n \mathbb{E}[|X_n|^3|M_{n-1}]\right] < \infty$.

Пусть $\tau_\ell := \inf\{N \in \mathbb{N} : \sum_{n=1}^N Q_n \ge \ell\}$. Тогда $M_{\tau_\ell}/\sqrt{\ell}$ сходится к стандартной нормальной величине по распределению при $\ell \to \infty$.

Результаты о несходимости

Пока всё хорошо. Если Y_n — центрированные i.i.d. случайные величины с конечной дисперсией, то S_n сходится к нормальному пределу. **По распределению**. Будет ли эта сходимость иметь место по вероятности или даже п.н.?

Утверждение 0.1. *Независимо от вероятностного пространства и распределения Y_n, Центральная предельная теорема **не** выполняется по вероятности, даже для подпоследовательностей.*

Суть в том, что если две последовательности $(S_n), (S'_n)$ сходятся по вероятности, то $S_n + S'_n$ также сходится по вероятности (по неравенству треугольника). То же утверждение неверно для сходимости по распределению, так как сходимость по распределению касается не самих случайных величин, а только их распределений. Таким образом, сходимость S_n или S'_n ничего не говорит об их совместном распределении.

Утверждение 0.1. Любая предельная точка (вдоль некоторой подпоследовательности) по вероятности S для S_n будет иметь стандартное нормальное распределение. В частности, $\sqrt{2}S_{2n} - S_n$ сходилась бы к $(\sqrt{2} - 1)S \sim \mathcal{N}(0, 3 - 2\sqrt{2})$ по вероятности (вдоль той же подпоследовательности). Но

$$S'_n := \sqrt{2}S_{2n} - S_n = \frac{1}{\sqrt{n}} \sum_{i=n+1}^{2n} Y_i$$

является суммой n i.i.d. величин, делённой на $\sqrt{n}$, следовательно, Теорема 0.1 применима к S'_n. А именно, для любой предельной точки S, величина $(\sqrt{2} - 1)S$ также должна иметь закон $\mathcal{N}(0, 1)$. Следовательно, предельных точек не существует. □

Визуализация сходимости

Что означает, что последовательность сходится по распределению? Зафиксируем μ, центрированную вероятностную меру на $\mathbb{R}$, и некоторое значение n, 'достаточно большое' (как мы видели, насколько большое, зависит от μ, например, от её третьего момента), и рассмотрим i.i.d. величины $X_1, \dots, X_n$ с законом μ и соответствующую им S_n, как в Уравнение 1. Мы можем многократно, скажем N раз, независимо сгенерировать выборку $(X_1, \dots, X_n)$ и, следовательно, S_n. Центральная предельная теорема говорит нам, что с большой вероятностью доля выборок, для которых S_n попадает в заданный интервал $[a, b]$, *приблизительно* равна интегралу Гаусса по $[a, b]$. Здесь *приблизительно* означает, что вероятность этого события сходится к 1 по мере роста N и n.

При большом n вероятность нахождения выборки в заданном интервале сходится к интегралу Гаусса по этому интервалу. В этом и заключается содержание центральной предельной теоремы. Здесь мы берём N выборок, строим гистограмму их распределения по интервалам и сравниваем результат с теоретической плотностью Гаусса.

С другой стороны, тот факт, что S_n *не* сходится п.н., означает, что если мы зафиксируем одну реализацию (так сказать, одно ω) и будем отслеживать значение S_n в зависимости от n, оно не сойдётся ни к какому значению.

Каждая отдельная реализация не сходится как функция от n. Мы генерируем X_i как i.i.d. и строим график S_n как функции от n. Даже при больших n график колеблется, и сходимость не наступает.

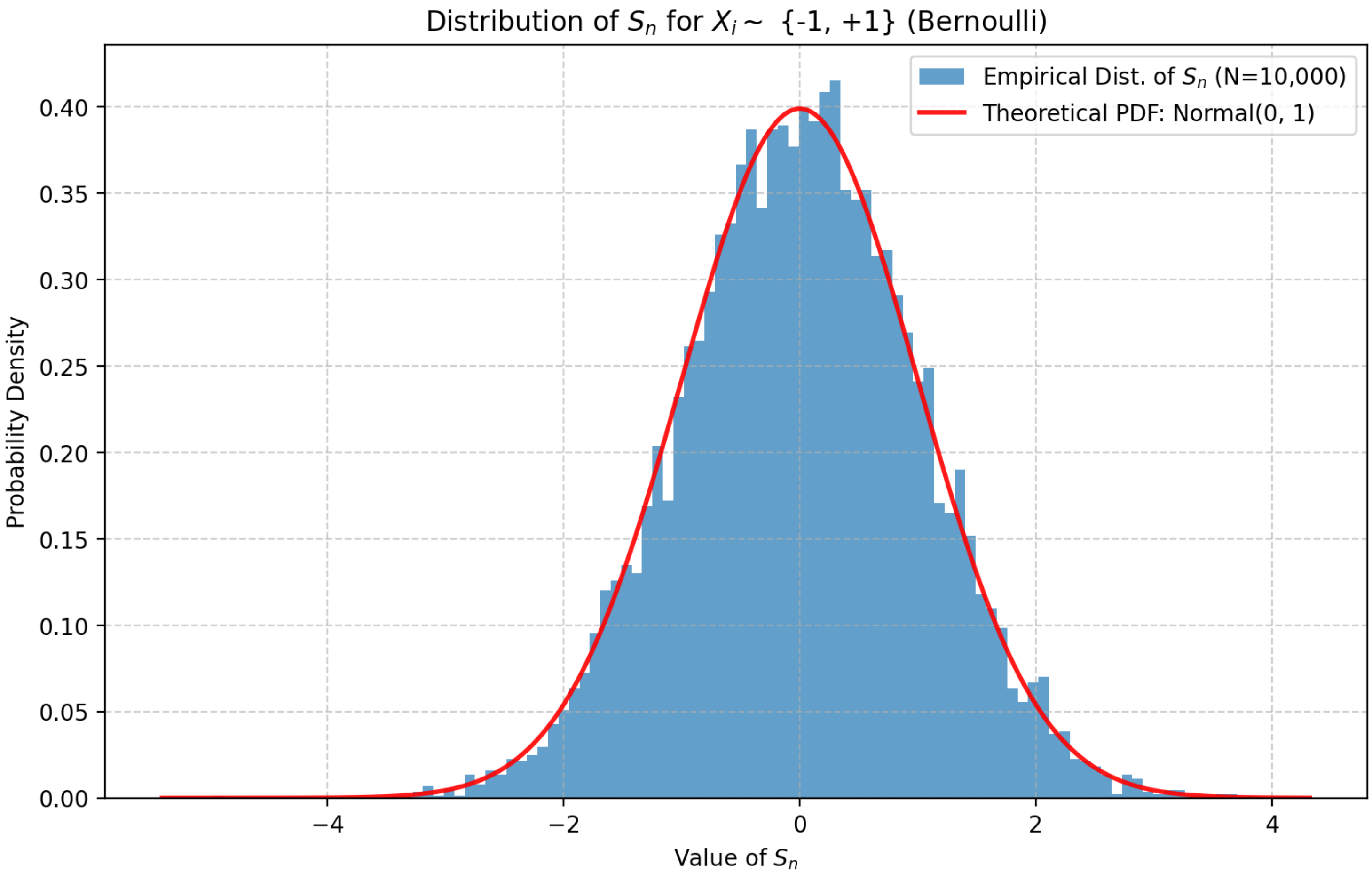

Рисунок 1: При большом n вероятность нахождения выборки в заданном интервале сходится к интегралу Гаусса по этому интервалу. В этом и заключается содержание центральной предельной теоремы. Здесь мы берём N выборок, строим гистограмму их распределения по интервалам и сравниваем результат с теоретической плотностью Гаусса.

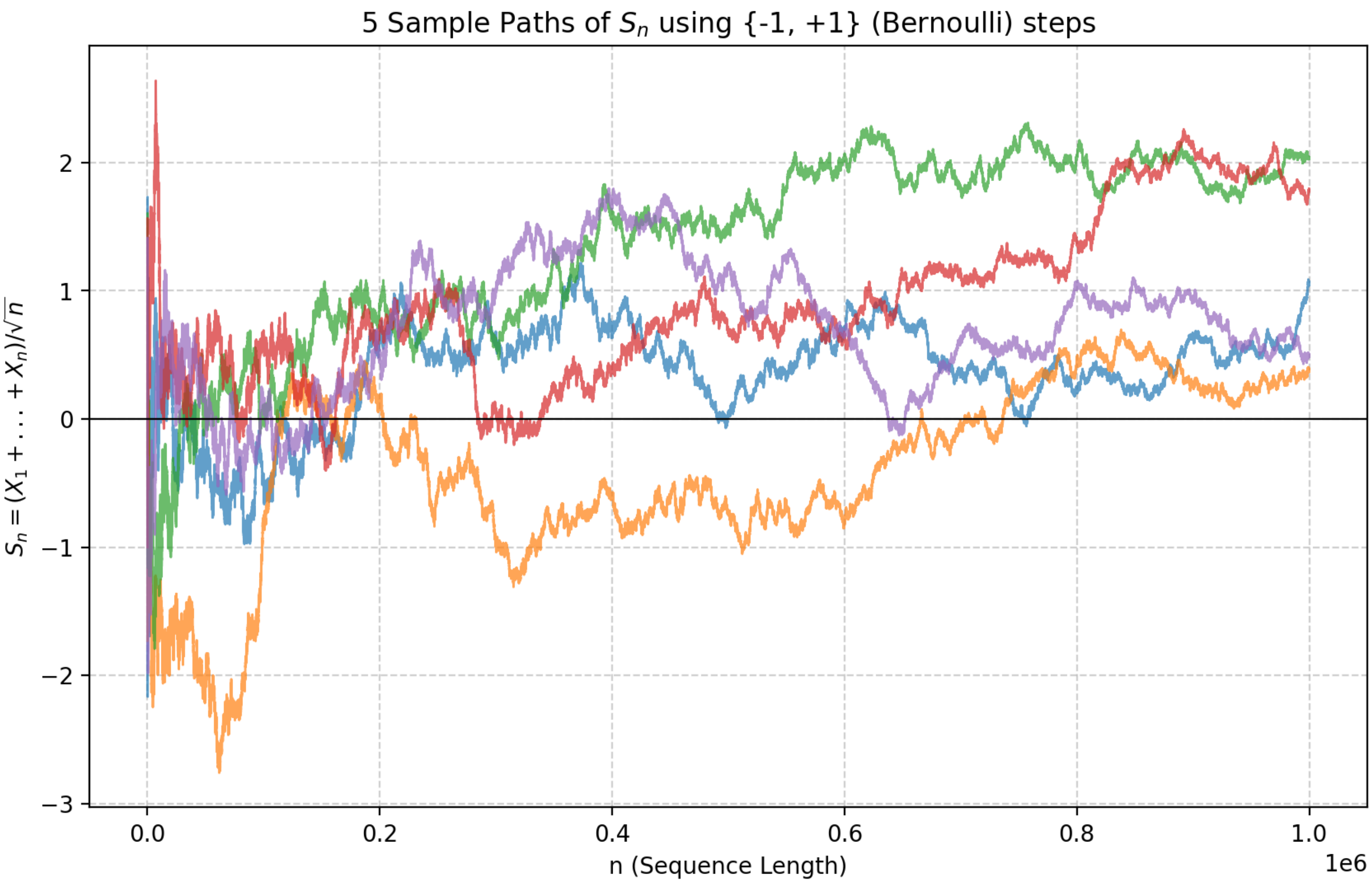

Рисунок 2: Каждая отдельная реализация не сходится как функция от n. Мы генерируем X_i как i.i.d. и строим график S_n как функции от n. Даже при больших n график колеблется, и сходимость не наступает. Здесь мы строим $N = 5$ различных реализаций до $n = 10^6$.