Центральная предельная теорема (ЦПТ) обещает нам, что сумма независимых одинаково распределенных слагаемых примерно нормально распределена. Эти заметки посвещены доказательству ЦПТ без использования характеристических функций.

1

Для аккуратной формулировки и доказательства вспомним сначала определение сходимости по распределению.

Сходимость по распределению

Последовательность случайных величин (R_n) сходится к R по распределению, если

$$\lim_{n\to\infty} \mathbb{P}(R_n \leq x) = \mathbb{P}(R \leq x) = F(x)$$

в любой точке x, где функция распределения F величины R непрерывна.

Перед доказательством ЦПТ нам потребуется лемма. Эта лемма позволяет от пределов вероятностей перейти к изучению пределов ожиданий гладких функций. Казалось бы, вероятности проще, чем ожидания, да ещё каких-то ненаписанных явно гладких функций! Однако для гладких функций применима мощнейшая идея разложения в ряд Тейлора.

Лемма

Для того, чтобы последовательность случайных величин (R_n) сходилась к R по распределению достаточно того, что для любой бесконечно дифференцируемой функций h с ограниченными производными выполнено условие

$$\lim_{n\to\infty} \mathbb{E}(h(R_n)) = \mathbb{E}(h(R)).$$

Доказательство леммы

Нам надо доказать, что при большом n вероятность $\mathbb{P}(R_n \leq x)$ не может слишком сильно отличаться от вероятности $\mathbb{P}(R \leq x)$ ни в большую, ни в меньшую сторону.

Докажем половину утверждения, вторая половина доказывается по аналогии. Основная идея доказательства такова: вероятность $\mathbb{P}(R_n \leq x)$ можно заменить на ожидание $\mathbb{E}(I(R_n \leq x))$, а «ступенчатый» индикатор I можно сколь угодно точно приблизить гладкой много раз дифференцируемой функцией.

Поехали. Выбираем произвольное положительное ε. Наша цель — доказать, что начиная с некоторого n вероятность $\mathbb{P}(R_n \leq x) > \mathbb{P}(R \leq x) - \varepsilon$.

С помощью ожидания индикатора и функции распределения F величины R наша цель записывается так:

$$\mathbb{E}(I(R_n \leq x)) > F(x) - \varepsilon.$$

Отступим от точки x чуть-чуть влево, в точку $x - \delta$. В силу непрерывности F в точке x размер оступа δ можно выбрать так, что $F(x-\delta) > F(x) - \varepsilon/2$.

Теперь придумаем гладкую функцию h, которая чуть-чуть занижает индикатор $I(R_n \leq x)$. А именно, левее $x-\delta$ функция h равна 1, правее x функция h равна нулю, а на отрезке $[x-\delta, x]$ функция h плавно спускается от 1 к 0. По построению,

$$I(R_n \leq x - \delta) \leq h(R_n) \leq I(R_n \leq x).$$

Делаем первый шаг по замене индикатора на не превосходящую его гладкую функцию h:

$$\mathbb{P}(R_n \leq x) = \mathbb{E}(I(R_n \leq x)) \geq \mathbb{E}(h(R_n)).$$

Теперь выберем n достаточно большим, так, чтобы

$$\mathbb{E}(h(R_n)) \geq \mathbb{E}(h(R)) - \varepsilon/2.$$

Теперь заменяем гладкую функцию h на не превосходящий её индикатор,

$$\mathbb{E}(h(R)) - \varepsilon/2 \geq \mathbb{E}(I(R \leq x - \delta)) - \varepsilon/2 = F(x-\delta) - \varepsilon/2.$$

Вспоминаем, что точку $x-\delta$ мы выбрали недалеко от x и получаем в итоге, что начиная с некоторого n

$$\mathbb{P}(X_n \leq x) \geq F(x) - \varepsilon.$$

Аналогично доказывается и вторая половина. На этот раз надо отступать от x вправо в точку $x+\delta$, и заменять индикатор $I(R_n \leq x)$ мажорирующей его гладкой функцией h.

По доказательству видно, что лемма остается верна, если расширить класс функций до просто непрерывных или до трижды дифференцируемых с конечными производными.

При желании можно сконструировать используемую в доказательстве функцию h явно, например, на базе бесконечно плавно стартующей из нуля функции

$$g(t) = \begin{cases} \exp(1/t) \text{ при } t > 0, \\ 0, \text{ при } t \leq 0. \end{cases}$$

Упражнение к лемме

Докажите, что для любого ε начиная с некоторого n выполнено неравенство

$$\mathbb{P}(X_n \leq x) \leq F(x) + \varepsilon.$$

2

Вспомним одну из формулировок ЦПТ.

Центральная предельная теорема

Если величины Q_1, Q_2, ..., независимы и одинаково распределены с конечным ожиданием μ и дисперсией σ^2, то отмасштабированная сумма

$$Z_n = \frac{\sum_{i=1}^n Q_i - \mathbb{E}(\sum_{i=1}^n Q_i)}{\sqrt{\operatorname{Var}(\sum_{i=1}^n Q_i)}}$$

стремится по распределению к $\mathcal{N}(0; 1)$.

3

Для начала представим S_n в виде отмасштабированных слагаемых.

$$Z_n = \frac{Q_1 - \mu}{\sigma\sqrt{n}} + \ldots + \frac{Q_{n-1} - \mu}{\sigma\sqrt{n}} + \frac{Q_n - \mu}{\sigma\sqrt{n}} = X_1 + \ldots + X_{n-1} + X_n$$

Замечаем, что $\mathbb{E}(X_i) = 0$, $\mathrm{Var}(X_i) = 1/n$.

3.1

Теперь потихоньку начнем менять слагаемые в правом хвосте на независимые слагаемые Y_i с таким же нулевым ожиданием, такой же дисперсией $1/n$, но нормально распределенные:

Удалим X_n, добавим Y_n, удалим X_{n-1}, добавим Y_{n-1}, и так далее…

Промежуточную сумму до удаления очередного X_i обозначим с помощью $Z_{n,i}$, а после удаления очередного X_i — с помощью $S_{n,i}$.

Для трёх величин схема выглядит так:

$$X_1 + X_2 + X_3 = Z_{3,3} \xrightarrow{-X_3} S_{3,3} \xrightarrow{+Y_3} Z_{3,2} \xrightarrow{-X_2} S_{3,2} \xrightarrow{+Y_2} Z_{3,1} \xrightarrow{-X_1} S_{3,1} \xrightarrow{+Y_1} Z_{3,0} = Y_1 + Y_2 + Y_3$$

Величина $Z_{n,i}$ будет своими первыми i слагаемыми содержать иксы, а оставшимися слагаемыми — игреки. В сумме $S_{n,i}$ полностью отсутствует i-е слагаемое, слагаемые с меньшими номерами — это $X_1, \ldots, X_{i-1}$, слагаемые с большими номерами — это $Y_{i+1}, \ldots, Y_n$.

Для наглядного примера,

$$Z_{5,3} = X_1 + X_2 + X_3 + Y_4 + Y_5,$$
$$S_{5,3} = X_1 + X_2 + 0 + Y_4 + Y_5,$$

В общем виде схема выглядит так:

$$\sum_{i=1}^{n} X_i = Z_{n,n} \xrightarrow{-X_n} S_{n,n} \xrightarrow{+Y_n} Z_{n,n-1} \xrightarrow{-X_{n-1}} \ldots \xrightarrow{-X_2} S_{n,2} \xrightarrow{+Y_2} Z_{n,1} \xrightarrow{-X_1} S_{n,1} \xrightarrow{+Y_1} Z_{n,0} = \sum_{i=1}^{n} Y_i$$

В схеме n шагов, каждый из которых состоит из двух полушагов, удаления X_i и добавления Y_i.

Заметим, что $S_{n,i}$ не зависит ни от X_i, ни от Y_i. Это пригодится.

Замечаем также, что $Z_{n,0} = Y_1 + \ldots + Y_n \sim \mathcal{N}(0;1)$.

В силу леммы нам достаточно доказать, что для любой бесконечно дифференцируемой h с ограниченными производными $\mathbb{E}(h(Z_{n,n})) \to \mathbb{E}(h(Z_{n,0}))$.

3.2

Поехали. Выбираем произвольное положительное ε. Наша цель — доказать, что начиная с некоторого n отличие этих двух ожиданий невелико,

$$\mathbb{E}(h(Z_{n,n})) - \mathbb{E}(h(Z_{n,0})) \in [-\varepsilon; +\varepsilon].$$

Посмотрим на нашу схему подмен

$$h\left(\sum_{i=1}^{n} X_i\right) = h(Z_{n,n}) \xrightarrow{-X_n} h(S_{n,n}) \xrightarrow{+Y_n} h(Z_{n,n-1}) \xrightarrow{-X_{n-1}} \ldots \xrightarrow{+Y_2} h(Z_{n,1}) \xrightarrow{-X_1} h(S_{n,1}) \xrightarrow{+Y_1} h(Z_{n,0}) = h\left(\sum_{i=1}^{n} Y_i\right)$$

С ростом n цепочка растет, а каждый шаг по идее должен становится всё меньше. Если мы докажем, что начиная с некоторого n разница

$$\mathbb{E}(h(Z_{n,i})) - \mathbb{E}(h(Z_{n,i-1}))$$

от **каждого** шага становится по модулю меньше ε/n, то дело будет в шляпе!

3.3

Остановимся на первой паре полушагов,

$$h(Z_{n,n}) \xrightarrow{-X_n} h(S_{n,n}) \xrightarrow{+Y_n} h(Z_{n,n-1})$$

Доказательство для других пар полушагов полностью аналогично.

Наша разница $h(Z_{n,n}) - h(Z_{n,n-1})$ разбивается в два полушага,

$$\mathbb{E}\, h(Z_{n,n}) - \mathbb{E}\, h(Z_{n,n-1}) = \left(\mathbb{E}\, h(Z_{n,n}) - \mathbb{E}\, h(S_{n,n})\right) - \left(\mathbb{E}\, h(Z_{n,n-1}) - \mathbb{E}\, h(S_{n,n})\right).$$

Заглянем в будущее, чтобы осознать план действий. Оказывается, что обе полушаговых разницы, $\left(\mathbb{E}\, h(Z_{n,n}) - \mathbb{E}\, h(S_{n,n})\right)$ и $\left(\mathbb{E}\, h(Z_{n,n-1}) - \mathbb{E}\, h(S_{n,n})\right)$, очень похожи на некоторую общую величину. Эта величина окажется равной $\mathbb{E}\left(\frac{h''(S_{n,n})}{2n}\right)$, но это не важно. Важно, что начиная с некоторого n отличие каждой полушаговой разницы от этой общей величины будет меньше $\varepsilon/2n$. При вычитании двух разниц общая величина уничтожится, и разница для целого шага окажется по модулю меньше ε/n.

Проведем доказательство для разницы первого полушага, $\left(\mathbb{E}\, h(Z_{n,n}) - \mathbb{E}\, h(S_{n,n})\right)$. Доказательство для разницы второго полушага, $\left(\mathbb{E}\, h(Z_{n,n-1}) - \mathbb{E}\, h(S_{n,n})\right)$, аналогично.

В этот момент можно уже не писать индекс (n, n) у Z и S :)

3.4

Выполним линеаризацию функции $h(Z_{n,n})$ в окрестности точки $S_{n,n}$. Заметим предварительно, что эти точки отличаются ровно на X_n, $Z_{n,n} = S_{n,n} + X_n$.

$$h(Z_{n,n}) \approx h(S_{n,n}) + h'(S_{n,n})(Z_{n,n} - S_{n,n}) = h(S_{n,n}) + h'(S_{n,n})X_n.$$

Для доказательства потребуется вспомнить точный смысл примерного равенства, а именно, остаток в форме Лагранжа. Найдётся такая точка C между $S_{n,n}$ и $Z_{n,n}$, что

$$h(Z_{n,n}) = h(S_{n,n}) + h'(S_{n,n})X_n + \frac{h''(C)}{2!}X_n^2.$$

Выделяем нужную нам разницу,

$$h(Z_{n,n}) - h(S_{n,n}) = h'(S_{n,n})X_n + \frac{h''(C)}{2!}X_n^2.$$

Прибавим и вычтем справа в числителе $h''(S_{n,n})$,

$$h(Z_{n,n}) - h(S_{n,n}) = h'(S_{n,n})X_n + \frac{h''(S_{n,n})}{2!}X_n^2 + \frac{h''(C) - h''(S_{n,n})}{2!}X_n^2.$$

Берём математическое ожидание, вспомнив, что $\mathbb{E}(X_n) = 0$, $\mathrm{Var}(X_n) = 1/n$, а X_n не зависит от S_n,

$$\mathbb{E}\, h(Z_{n,n}) - \mathbb{E}\, h(S_{n,n}) = 0 + \mathbb{E}\left(\frac{h''(S_{n,n})}{2n}\right) + \mathbb{E}\left(\frac{h''(C) - h''(S_{n,n})}{2}X_n^2\right).$$

3.5

Сосредоточимся на последнем слагаемом и рассмотрим два случая, в зависимости от того, больше ли $|X_n|$ чем δ.

$$\frac{h''(C) - h''(S_{n,n})}{2!}X_n^2 = \left(\frac{h''(C) - h''(S_{n,n})}{2!}X_n^2 I(|X_n| \leq \delta)\right) + \left(\frac{h''(C) - h''(S_{n,n})}{2!}X_n^2 I(|X_n| > \delta)\right).$$

Изучаем первое слагаемое. Вспомним, что точка C находится между $S_{n,n}$ и $Z_{n,n}$, а $S_{n,n} + X_n = Z_{n,n}$. Поэтому $\left|C - S_{n,n}\right| \leq \delta$, если $|X_n| \leq \delta$.

У функции h ограничена третья производная, выберем δ настолько маленьким, чтобы зажать разницу $h''(C) - h''(S_{n,n})$ до величины меньшей $\varepsilon/2$.

Получаем ограничение для первого слагаемого,

$$\mathbb{E}\left|\frac{h''(C) - h''(S_{n,n})}{2!}X_n^2 I(|X_n| \leq \delta)\right| \leq E\left(\frac{\varepsilon}{4}X_n^2\right) = \frac{\varepsilon}{4n}$$

Изучаем второе слагаемое. У функции h ограничена вторая производная константой M.

$$\mathbb{E}\left|\frac{h''(C)-h''(S_{n,n})}{2!}X_n^2 I(|X_n|>\delta)\right| \leq E\left(\frac{M+M}{4}X_n^2 I(|X_n|>\delta)\right) = \frac{M}{2n}\,\mathbb{E}(X_n^2 I(|X_n|>\delta))$$

Подберем n настолько большим, что $\mathbb{E}(X_n^2 I(|X_n|>\delta)) < \varepsilon/2M$. При этом второе слагаемое будет также ограничено величиной $\varepsilon/4n$. Тем самым мы доказали, что начиная с некоторого n

$$\mathbb{E}\left|\frac{h''(C)-h''(S_{n,n})}{2!}X_n^2\right| \leq \frac{\varepsilon}{4n}+\frac{\varepsilon}{4n}=\frac{\varepsilon}{2n}$$

То есть,

$$\mathbb{E}\,h(Z_{n,n}) - \mathbb{E}\,h(S_{n,n}) \in \left[\mathbb{E}\left(\frac{h''(S_{n,n})}{2n}\right)-\frac{\varepsilon}{2n};\mathbb{E}\left(\frac{h''(S_{n,n})}{2n}\right)+\frac{\varepsilon}{2n}\right].$$

В этот же диапазон попадает и величина $\mathbb{E}\,h(Z_{n,n-1}) - \mathbb{E}\,h(S_{n,n})$, поэтому

$$\mathbb{E}\,h(Z_{n,n}) - \mathbb{E}\,h(Z_{n,n-1}) \in \left[-\frac{\varepsilon}{n};+\frac{\varepsilon}{n}\right].$$

3.6

Вспомним наш долгий путь. Сначала мы разбили разницу $\mathbb{E}\,h(Z_{n,n})-\mathbb{E}\,h(Z_{n,0})$ на $2n$ полушагов. Каждый шаг состоит из полушага удаления X_i и полушага добавления Y_i. Изменение $\mathbb{E}\,h$, вызванное каждым шагом, состоит из разницы изменений вызванных полушагами. А изменение от каждого полушага при больших n не отличается от общей константы более чем на $\varepsilon/2n$. Поэтому каждый шаг даёт изменение не больше ε/n, и вся разница $\mathbb{E}\,h(Z_{n,n})-\mathbb{E}\,h(Z_{n,0})$ начиная с некоторого момента меньше ε.

Упражнение к теореме

Докажите, что для любого ε начиная с некоторого n выполнено условие

$$\mathbb{E}\,h(Z_{n,n-1}) - \mathbb{E}\,h(S_{n,n}) \in \left[\mathbb{E}\left(\frac{h''(S_{n,n})}{2!}\right)-\frac{\varepsilon}{2n};\mathbb{E}\left(\frac{h''(S_{n,n})}{2!}\right)+\frac{\varepsilon}{2n}\right].$$

Решение упражнения к теореме

Линеаризовать также надо в окрестности точки $S_{n,n}$, а разница $Z_{n,n-1}-S_{n,n}$ окажется равной Y_n. По ожиданию и дисперсии Y_n ничем не отличается от X_n. Поэтому остаётся лишь полностью скопировать доказательство с заменой X_n на Y_n.

3.7

В основном изложение следует статье (Chin 2022). Постарался сделать изложение более «мотивированным», чтобы перед шагами яснее была видна цель. Также излагаю один случай из повторяющихся. С одной стороны, это облегчает понимание, с другой стороны аналогичный случай можно решать в виде упражнения.

Chin, Calvin Wooyoung. 2022. «A Short and Elementary Proof of the Central Limit Theorem by Individual Swapping». *The American Mathematical Monthly* 129: 374–80. https://arxiv.org/abs/2106.00871.