

Центральная предельная теорема

октябрь–декабрь 2023 г.

Курс лекций «Теория вероятностей»

- Теорема Муавра – Лапласа
- Центральная предельная теорема
- Примеры применения ЦПТ
- Приближение биномиального распределения пуассоновским

Напоминание: сходимость по распределению

Определение (сходимость по распределению)

Последовательность $\{\xi_n, n = 1, 2, \dots\}$ случайных величин *сходится по распределению* к случайной величине ξ , если

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

в каждой точке непрерывности функции $F(x)$, где $F_n(x)$, $F(x)$ — функции распределения случайных величин ξ_n , ξ соответственно. Обозначение $\xi_n \xrightarrow{d} \xi$ или $\xi_n \xrightarrow{d} F(\cdot)$.

Сходимость по распределению к непрерывной случайной величине

Если предельное распределение $F(x)$ непрерывное, то для любого отрезка $[a, b] \subset \mathbb{R}$ предел

$$\lim_{n \rightarrow \infty} P(\xi_n \in [a, b]) = F(b) - F(a).$$

Теорема Муавра – Лапласа

Теорема (Муавр – Лаплас)

Пусть ξ_n подчиняются биномиальному распределению с параметрами p, n , где $p \in (0, 1), n = 1, 2, \dots$. Тогда

$$\frac{\xi_n - np}{\underbrace{\sqrt{np(1-p)}}_{\text{Var}(\xi_n)}} \xrightarrow{d} N(0, 1).$$

$\text{E}(\xi_n)$



Теорема Муавра – Лапласа: пример

Правильную монетку бросают 100 раз. Найти вероятность того, что число орлов не меньше 60.

Теорема Муавра – Лапласа: разбор примера

Точное значение этой вероятности равно

$$\sum_{i=60}^{100} C_{100}^i (1/2)^{100} = 0.02844.$$

$$P(B(p, n) = k) = \\ = C_n^k p^k (1-p)^{n-k} \\ k=0, 1, \dots, n$$

По теореме Муавра – Лапласа имеем

$$59.5 \rightarrow 0.0287$$

$$P(S_n \geq 60) = P\left(\frac{S_n - 100 \cdot 0.5}{\sqrt{100 \cdot 0.5 \cdot 0.5}} \geq \frac{60 - 100 \cdot 0.5}{\sqrt{100 \cdot 0.5 \cdot 0.5}}\right) \approx P(Z \geq 10/5) = 1 - \Phi(2) = 0.02275.$$

↑
здесь центральная теорема Муавра – Лапласа

↑
приближённо

Проверки на дискретность.

$$P(S_n \geq 60) = P(S_n > 59) \quad \text{т.к. дискретно}$$

$$S_n = 0, 1, \dots, 59, 60, 61, \dots$$

Проверка: В таких случаях берём

$$P(S_n > 59.5) \quad \left\{ \begin{array}{l} P(S_n \geq k) = P(S_n > k - \frac{1}{2}) \\ P(S_n > k) = P(S_n \geq k + \frac{1}{2}) \\ P(S_n \leq k) = P(S_n < k + \frac{1}{2}) \\ P(S_n < k) = P(S_n \leq k - \frac{1}{2}) \end{array} \right.$$

Теорема Муавра – Лапласа: разбор примера

А если монетку бросаем 10 раз и хотим найти вероятность того, что выпало не меньше 6 орлов? Точное значение этой вероятности равно

$$\sum_{i=6}^{10} C_{10}^i (1/2)^{10} = \underline{0.3770.}$$

По теореме Муавра – Лапласа имеем

$$P(S_n \geq 6) = P\left(\frac{S_n - 10 \cdot 0.5}{\sqrt{10 \cdot 0.5 \cdot 0.5}} \geq \frac{6 - 10 \cdot 0.5}{\sqrt{10 \cdot 0.5 \cdot 0.5}}\right) \approx P(Z \geq 1/1.5811) = 1 - \Phi(0.6325) = \underline{0.2635.}$$

→ 0.3759. (!!!)

Отличия существенные! При таких маленьких n нормальное приближение работает плохо.

Попробуй на симуляцию монеток.

Центральная предельная теорема

Пусть $\{\xi_n, n = 1, 2, \dots\}$ — последовательность независимых одинаково распределенных случайных величин, для которых

$$E(\xi_i) = m, \quad \text{Var}(\xi_i) = \sigma^2, \quad E|\xi_i^3| < +\infty.$$

Обозначим

$$S_n = \sum_{i=1}^n \xi_i, \quad T_n = \frac{S_n - nm}{\sqrt{n\sigma^2}} = \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}}.$$

Теорема (ЦПТ)

Последовательность случайных величин $\{T_n\}$ сходится по распределению к стандартной нормальной величине, т. е.

$$\lim_{n \rightarrow \infty} P(T_n \leq x) = \Phi(x)$$

для любого $x \in \mathbb{R}$.

ЦПТ: пример 1

В каждой партии некоторой игры игрок с вероятностью 0.7 проигрывает один рубль, с вероятностью 0.2 проигрывает два рубля и с вероятностью 0.1 выигрывает 10 рублей. Оцените вероятность того, что после 150 независимых партий

1. игрок суммарно окажется в проигрыше;
2. игрок выиграет суммарно не менее 40 рублей.

ЦПТ: разбор примера 1

Обозначим через X_i величину выигрыша или проигрыша игрока в i -й партии. Тогда $P(X_i = -1) = 0.7$, $P(X_i = -2) = 0.2$, $P(X_i = 10) = 0.1$.

Непосредственными вычислениями получаем:

$$m = E(X_i) = -0.1, \quad \text{Var}(X_i) = 11.49, \quad \sigma = \sqrt{\text{Var}(X_i)} = 3.39.$$

Обозначим $S = \sum_{i=1}^{150} X_i$ — величина суммарного выигрыша или проигрыша игрока за 150 партий. Воспользуемся центральной предельной теоремой:

$$1. P(S < 0) = P\left(\frac{S - nm}{\sigma\sqrt{n}} < \frac{0 + 0.1 \cdot 150}{3.39 \cdot 12.25}\right) \approx P(Z < 0.36) = 0.64.$$

$$2. P(S \geq 40) = P\left(\frac{S - nm}{\sigma\sqrt{n}} \geq \frac{40 + 0.1 \cdot 150}{3.39 \cdot 12.25}\right) \approx P(Z \geq 1.32) = 0.093.$$

$$\approx 1 - \Phi(1.32)$$

Астроном измеряет расстояние до некоторой звезды (в световых годах). Известно, что результаты измерений имеют среднее d (истинное значение расстояния) и стандартное отклонение 2 световых года. Сколько измерений нужно провести, чтобы с вероятностью 0.95 установить расстояние с точностью ± 0.5 световых года.

ЦПТ: разбор примера 2

Обозначим X_i — результат отдельного измерения, $S_n = \bar{X}$ — среднее значение.

Считаем измерения независимыми и воспользуемся ЦПТ:

$$\begin{aligned} 0.95 &= P(|S_n - d| \leq 0.5) = P\left(-0.5 \frac{\sqrt{n}}{2} \leq \frac{S_n - d}{2/\sqrt{n}} \leq 0.5 \frac{\sqrt{n}}{2}\right) \approx P\left(-0.5 \frac{\sqrt{n}}{2} \leq Z \leq 0.5 \frac{\sqrt{n}}{2}\right) \\ &= \Phi(\sqrt{n}/4) - \Phi(-\sqrt{n}/4) = 2\Phi(\sqrt{n}/4) - 1. \end{aligned}$$

Handwritten notes: "центрирована" (centered) above the fraction, "оценки" (estimation) with an arrow pointing to the absolute value, "N(0,1)" above the Z variable, and a red circle around the Z variable.

Приравняв $2\Phi(\sqrt{n}/4) - 1 = 0.95$, получаем $\Phi(\sqrt{n}/4) = 0.975$, $\sqrt{n}/4 = 1.96$ и $n = (7.84)^2 \approx 61.47$, то есть 62 наблюдений достаточно.

Оценка с помощью неравенства Чебышева дает

$$0.95 = P(|S_n - d| \leq 0.5) = 1 - P(|S_n - d| \geq 0.5) \leq \frac{4}{n \cdot (0.5)^2} = 1 - \frac{16}{n}.$$

Handwritten notes: "0.95" in red, a red arrow pointing to the inequality sign, and a red wavy line below the equation.

Откуда $n \geq 16/(1 - 0.95) = 320$. Это значение надежное, но обычно очень преувеличенное.

Игральный кубик бросают до тех пор, пока сумма выпавших очков не станет больше 300. Оцените вероятность того, что для этого понадобится не менее 80 бросков.

ЦПТ: разбор примера 3

1 2 3 4 5 6
1/6 1/6 1/6 1/6 1/6 1/6

$$E(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = 3.5$$

$$E(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \dots + 6^2 \cdot \frac{1}{6}; \quad \text{Var}(X) = E(X^2) - (E(X))^2$$

Переформулируем событие «для того, чтобы набрать больше 300 очков, потребовалось не менее 80 бросков» так: «за 79 бросков удалось набрать не больше 300 очков». Для игрального кубика $E(X_i) = 3.5$, $\text{Var}(X_i) = 2.917$.

Воспользуемся ЦПТ:

$$\begin{aligned} P\left(\sum_{i=1}^{79} X_i \leq 300\right) &= P\left(\frac{\sum_{i=1}^{79} X_i - 79 \cdot 3.5}{\sqrt{79 \cdot 2.917}} \leq \frac{300 - 79 \cdot 3.5}{\sqrt{79 \cdot 2.917}}\right) \approx P(Z \leq 1.548) \\ &= \Phi(1.548) = 0.9392. \end{aligned}$$

ЦПТ: пример 4

Фармацевтическая компания разрабатывает лекарства. Всего в работе находится $n = 1000$ препаратов. Вероятность того, что препарат окажется удачным, равна $p = 0.005$. Найдите вероятность того, что в конце концов удачными окажутся более 5 препаратов.

если $n \cdot p < 10$ и n большое (тысячи)
то норм. кривая, много
работает, а Макс Хоффман¹⁶

ЦПТ: разбор примера 4

Попробуем воспользоваться ЦПТ для биномиальной величины:

$$P\left(\sum_{i=1}^{1000} X_i > 5\right) = P\left(\frac{\sum_{i=1}^{1000} X_i - 5}{\sqrt{1000 \cdot 0.005 \cdot 0.095}} > \frac{5 - 5}{\sqrt{1000 \cdot 0.005 \cdot 0.095}}\right) \approx 1 - \Phi(0) = 0.5.$$

Handwritten notes: 0.411, 1 - \Phi(0.5/\sqrt{5}), and 5.5 with an arrow pointing to the numerator of the second fraction.

При этом точное значение этой вероятности равно 0.3840, то есть нормальное приближение не очень точное, несмотря на то, что число наблюдений большое.

Приближение биномиального распределения пуассоновским

Для таких перекошенных распределений вместо нормального приближения используют пуассоновское.

Теорема

Пусть дана последовательность биномиальных случайных величин ξ_n с параметрами p_n, n , где $\lim_{n \rightarrow \infty} np_n = \lambda > 0$. Тогда для любого $k = 0, 1, 2, \dots$



$$\lim_{n \rightarrow \infty} P(\xi_n = k) = e^{-\lambda} \frac{\lambda^k}{k!},$$

т. е. последовательность ξ_n по распределению сходится к пуассоновской случайной величине с параметром λ .

Доказательство.

$$P(\xi_n = k) = C_n^k p_n^k (1 - p_n)^{n-k} = \frac{n!}{k!(n-k)!} \cdot \frac{(np_n)^k}{n^k} \cdot \left(1 - \frac{np_n}{n}\right)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$



ЦПТ: продолжение разбора примера 4

Если воспользоваться пуассоновским приближением с $\lambda = 1000 \cdot 0.005 = 5$, то получим, что искомая вероятность

$$P\left(\sum_{i=1}^{1000} X_i > 5\right) \approx 1 - \sum_{k=0}^5 e^{-5} \frac{5^k}{k!} = \underline{\underline{0.3840}},$$

что совпало с истинным значением вплоть до 5-го знака после запятой.

- [1] Sh. Ross. *A First Course in Probability*. 9th ed. London: Pearson, 2014.
- [2] Б. В. Гнеденко. *Курс теории вероятностей*. М.: Издательство УРСС, 2005.
- [3] С. В. Головань, П. К. Катышев, А. А. Пересецкий. *Сборник задач по курсу теории вероятностей*. М.: Издательский дом «Дело» РАНХиГС, 2021.

Для заметок