Лекция 3. Различные асимптотические свойства

Кохович Дарья Игоревна

19 февраля 2024 г.

Обозначения $\mu_n(x)$

Определим для каждого действительного числа x случайную величину $\mu_n(x)$, равную числу элементов выборки $X = (X_1, \ldots, X_n)$, значения которых не превосходят x, то есть

$$\mu_n(x) = \sum_{i=1}^{n} \mathbb{I}(X_i \leq x).$$

Эмпирическая функция распределения

Мы ранее определяли э. ф. р.:

$$F_n(r) = \frac{\#\{X_i : X_i \leq r\}}{n}, \qquad -\infty < r < \infty.$$

Тогда верно:

$$\mathbb{P}\left\{F_n(x) = \frac{k}{n}\right\} = \mathbb{P}\{\mu_n(x) = k\}.$$

Или $F_n(x) = \frac{1}{n}\mu_n(x)$.

Эмпирическая функция распределения

С каждым X_i можно связать два события: $\{X_i \le x\}$ и $\{X_i > x\}$. Вероятности этих событий равны

$$p = \mathbb{P}\{X_i \le x\} = F(x), \qquad q = \mathbb{P}\{X_i > x\} = 1 - F(x)$$

для всех $i \in \{1, \dots, n\}$.
Если событие $\{X_i \le x\}$ назвать успехом, то $\mu_n(x)$ является числом успехов в n независимых испытаний Бернулли с вероятностью успеха p, поэтому $\mu_n(x)$ имеет биномиальное распределение $B(n, p)$ с $p = F(x)$. Следовательно,

$$\mathbb{P}\left\{F_n(x) = \frac{k}{n}\right\} = C_n^k F^k(x)(1 - F(x))^{n-k}, \quad k = 0, 1, \dots, n.$$

Скорость сближения для эмпирической функции распределения

Для э.ф.р. уже знаем, что это несмещенная и состоятельная оценка, а также доказали теорему Гливенко-Кантелли. Установим скорость сближения э.ф.р. $F_n(x)$ с $F(x)$.

$$\mathbb{E}F_n(x) = \frac{1}{n}\mathbb{E}\mu_n(x) = F(x),$$

$$\mathbb{D}F_n(x) = \frac{1}{n^2}\mathbb{D}\mu_n(x) = \frac{1}{n}F(x)(1 - F(x)).$$

Скорость сближения для эмпирической функции распределения

Вспомним неравенство Чебышева. Для любой случайной величины ξ верно $\mathbb{P}\{|\xi - \mathbb{E}\xi| > \epsilon\} \leq \mathbb{D}\xi/\epsilon^2$.
Тогда:

$$\mathbb{P}\{\sqrt{n}|F_n(x) - F(x)| > t\} \leq \frac{F(x)(1 - F(x))}{t^2} \leq \frac{1}{4t^2}.$$

Скорость сближения для эмпирической функции распределения

Полученная оценка универсальна, так как она не зависит ни от функции распределения $F(x)$, ни от точки x, ни от объема выборки n. В частности, верно:

$$\mathbb{P}\left\{|F_n(x) - F(x)| > \frac{5}{\sqrt{n}}\right\} \leq 0,01.$$

Теорема Муавра-Лапласа для э.ф.р.

Из курса ТВ мы знаем о сходимости биномиального распределения к нормальному при росте n (теорема Муавра-Лапласа). Если $0 < F(x) < 1$, то при $n \to \infty$ верно

$$\frac{\mu_n(x) - nF(x)}{\sqrt{nF(x)(1 - F(x))}} \to X, \quad X \sim N(0, 1).$$

Распределение порядковых статистик

Найдем функцию распределения $F^{(k)}(x) = \mathbb{P}\{X^{(k)} \le x\}$ и плотность $g_k(x) = (F^{(k)}(x))'$ произвольной порядковой статистики $X^{(k)}$ выборки $X = (X_1, \ldots, X_n)$.

Распределение порядковых статистик

Событие $\{X^{(k)} \le x\}$ означает, что не менее k элементов выборки X имеют значения, не превосходящие x, это есть событие $\{\mu_n(k) \ge x\}$. Следовательно

$$F^{(k)}(x) = \mathbb{P}\{\mu_n(k) \ge x\} = \sum_{i=k}^{n} C_n^i F^i(x)(1 - F(x))^{n-i}.$$

В частности:

$$F^{(1)}(x) = \mathbb{P}\{X^{(1)} \le x\} = 1 - (1 - F(x))^n,$$

$$F^{(n)}(x) = \mathbb{P}\{X^{(n)} \le x\} = F^n(x).$$

Распределение порядковых статистик

Неполная бета-функция определяется следующим образом:

$$B(z;a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^z t^{a-1}(1-t)^{b-1}dt$$

Она представляет собой функцию распределения закона $Be(a,b)$.

$$B(z;k,n-k+1) = \frac{n!}{(k-1)!(n-k)!} \int_0^z t^{k-1}(1-t)^{n-k}dt =$$

$$= \sum_{i=k}^{n} C_n^i z^i (1-z)^{n-i}.$$

Следовательно:

$$F^{(k)}(x) = B(F(x);k,n-k+1).$$

Распределение порядковых статистик

Соотношение $F^{(k)}(x) = B(F(x); k, n-k+1)$ можно продифференцировать. Тогда получим равенство для плотности $g_k(x)$

$$g_k(x) = \frac{n!}{(k-1)!(n-k)!} F^{k-1}(x)(1-F(x))^{n-k} f(x).$$

Асимптотическая нормальность

Def.
Оценка T_n называется асимптотически нормальной оценкой параметра θ с коэффициентом σ^2, если при $n \to \infty$ распределение случайной величины $\frac{\sqrt{n}(T_n - \theta)}{\sigma}$ сходится к стандартному нормальному распределению, то есть

$$\mathbb{P}\left(\frac{\sqrt{n}(T_n - \theta)}{\sigma} < x\right) \to \Phi(x), \quad n \to \infty.$$

Асимптотическая нормальность эмпирического среднего

Пусть дана выборка из распределения с конечной дисперсией $\mathbb{D}X_1 = \sigma^2$. Убедимся, что выборочное среднее $\bar{X}$ является асимптотически нормальной оценкой для истинного математического ожидания $m = EX_1$. При этом коэффициент асимптотической нормальности равен как раз $\sigma^2 = \mathbb{D}X_1$. Действительно, по центральной предельной теореме распределение членов последовательности

$$\frac{\sqrt{n}(\bar{X} - m)}{\sigma} = \frac{X_1 + \cdots + X_n - nm}{\sqrt{n}\sigma} = \frac{X_1 + \cdots + X_n - n\mathbb{E}X_1}{\sqrt{n\mathbb{D}X_1}}$$

сближается со стандартным нормальным распределением.

Пример не асимптотически нормальной оценки

Рассмотрим выборку объема n из равномерного на $[0, \theta]$ распределения. Рассмотрим оценку $T_n = X^{(n)}$ для θ.
Тогда величина $\sqrt{n}(X^{(n)} - \theta)$ при любом n может принимать только отрицательные значения, поэтому её распределение не может приближаться ни к какому нормальному закону.

Асимптотическая нормальность выборочных квантилей

Зафиксируем некоторое p, $0 < p < 1$, и рассмотрим порядковую статистику $X^{(k)}$ с $k = [np]$. Определим ζ_p как точку такую, что $F(\zeta_p) = p$. Верна следующая теорема:

Theorem

Если в некоторой окрестности точки ζ_p плотность $f(x)$ непрерывна вместе с производной и $f(\zeta_p) > 0$, то при $n \to \infty$

$$\mathcal{L}(X^{[np]}) \sim N\left(\zeta_p, \frac{pq}{nf^2(\zeta_p)}\right), \quad q = 1 - p.$$

Асимптотическая нормальность выборочной медианы

Найдем асимптотическое распределение выборочной медианы $X^{[n/2]}$ для выборки X из нормального распределения $N(\mu, \sigma^2)$.
Из теоремы:

$$\mathcal{L}(X^{[n/2]}) \sim N\left(\zeta_{1/2}, \frac{1}{4nf^2(\zeta_{1/2})}\right).$$

Вспомним, что $f(x)$ симметрична относительно точки μ, поэтому $\zeta_{1/2} = \mu$. Вычислим $f(\mu)$:

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\mu-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}\sigma} \cdot 1 = \frac{1}{\sqrt{2\pi}\sigma}$$

Асимптотическая нормальность выборочной медианы

Следовательно:

$$\mathcal{L}(X^{[n/2]}) \sim N\Big(\mu, \frac{\pi\sigma^2}{2n}\Big).$$

Асимптотическое поведение крайних порядковых статистик

Theorem

Для выборки из абсолютно непрерывного распределения статистики $X^{(r)}$ *и* $X^{(n-s+1)}$ *при любых фиксированных* $r, s \geq 1$ *и* $n \to \infty$ *асимптотически независимы и при этом*

$$\mathcal{L}(nF(X^{(r)})) \to \Gamma(1, r),$$

$$\mathcal{L}(n(1 - F(X^{(n-s+1)}))) \to \Gamma(1, s).$$

Определим преобразованные статистики:

$$\kappa_n = nF(X^{(r)}), \qquad \nu_n = n(1 - F(X^{(n-s+1)})).$$

Асимптотическое поведение крайних порядковых статистик

Заметим, что теорема определяет вид предельных распределений не самих порядковых статистик, а некоторых функций от них. То есть

$$X^{(r)} = F^{-1}\left(\frac{\kappa_n}{n}\right), \quad X^{(n-s+1)} = F^{-1}\left(1 - \frac{\nu_n}{n}\right).$$

Асимптотическое поведение крайних порядковых статистик

Пусть $X_i \sim \Gamma(1,1)$ – стандартное показательное распределение.
То есть $F(x) = 1 - e^{-x}$.
Следовательно $F^{-1}(t) = -\ln(1-t)$, $0 < t < 1$.
Тогда $X^{(n-s+1)} = \ln n - \ln \nu_n$. Тогда по теореме:

$$\mathbb{P}\{X^{(n-s+1)} - \ln n \leq x\} = \mathbb{P}\{\nu_n \geq e^{-x}\} \to \frac{1}{(s-1)!} \int_{e^{-x}}^{\infty} e^t t^{s-1} dt.$$

Асимптотическое поведение крайних порядковых статистик

Обозначим $\pi_k(t) = e^{-t} t^k / k!$. Не сложно проверить, что

$$\frac{1}{(s-1)!} \int_{e^{-x}}^{\infty} e^t t^{s-1} dt = \int_{e^{-x}}^{\infty} \pi_{s-1}(y) dy = \sum_{k=0}^{s-1} \pi_k(e^{-x}).$$

В частности:

$$\mathbb{P}\{X^{(n)} - \ln n \leq x\} \to \sum_{i=0}^{\infty} \pi_k(e^{-x}) = e^{-e^{-x}}, \quad -\infty < x < \infty.$$