6 Состоятельность. Асимптотическая нормальность

На семинарах вы уже освоили такие свойства оценок, как состоятельность и асимптотическая нормальность, и научились их доказывать. Сформулируем еще некоторые полезные результаты в этой области (для более глубокого ознакомления с темой рекомендруем книгу М.Б. Лагутина "Наглядная математическая статистика").

Для начала напомним (или даже введем) некоторые понятия:

Определение 1. Выборочной медианой MED называют оценку

$$\mathrm{MED} = \begin{cases} X_{(k+1)}, & n = 2k+1, \\ \dfrac{X_{(k)} + X_{(k+1)}}{2}, & n = 2k, \end{cases},$$

усеченным среднем $\overline{X}_\alpha$

$$\overline{X}_\alpha = \frac{1}{n-2k}(X_{(k+1)} + \dots + X_{(n-k)}), \quad k = [\alpha n].$$

Выборочная медиана оценивает теоретическую медиану $x_{1/2} = F^{-1}(1/2)$. Усеченное среднее как правило используют для оценки центра симметрии у симметричных распределений.

Теорема 1. *Пусть распределение F таково, что $F(x_{1/2} + \varepsilon) > 1/2$ при всех $\varepsilon > 0$. Тогда* MED *будет состоятельной оценкой $x_{1/2}$.*

Теорема 2. *Пусть $X_1, \dots, X_n$ выборка из распределения с плотностью f, причем $f(x) > 0$ в некоторой окрестности $x_{1/2}$; здесь $x_{1/2}$ – медиана распеделения с.в. X_1. Тогда выборочная медиана* MED *является асимтотически нормальной оценкой $x_{1/2}$:*

$$\sqrt{n}(\mathrm{MED} - x_{1/2}) \xrightarrow{d} Z \sim \mathcal{N}\left(0, \frac{1}{4f^2(x_{1/2})}\right), \, n \to \infty.$$

Теорема 3. *Пусть $X_1, \dots, X_n \sim F(x - \theta)$, где F обладает следующими свойствами: найдется такое $0 < c \le +\infty$, что $F(-c) = 0$, $F(c) = 1$ и на $(-c, c)$ $F(x)$ имеет четную, непрерывную и положительную плотность $f(x)$.*

Тогда усеченное среднее $\overline{X}_\alpha$ при $0 < \alpha < 1/2$ является асимптотически нормальной оценкой θ:

$$\sqrt{n}(\overline{X}_\alpha - \theta) \xrightarrow{d} Z \sim \mathcal{N}\left(0, \sigma_\alpha^2\right), \, n \to \infty, \quad \sigma_\alpha^2 = \frac{2}{(1-2\alpha)^2}\left[\int_0^{x_{1-\alpha}} t^2 f(t)\, dt + \alpha x_{1-\alpha}^2\right],$$

где x_γ – решение уравнения $F(x_\gamma) = \gamma$.

Задачи

1. $X_1, ..., X_n \sim \mathrm{R}[0, \theta]$.
   - (a) Построить гистограммы для $X_{(n)}$ при разных n и сравнить с нормальной плотностью (с такими же математическим ожиданием и дисперсией, как у $X_{(n)}$).
   - (b) Построить гистограммы для $n(\theta - X_{(n)})$ при разных n и сравнить с плотностью распределения $\exp(1)$.
   - (c) * Изобразить гистограммы $\sqrt{n}(2\overline{X} - \theta)$ и $\sqrt{n}(2\overline{X}_\alpha - \theta)$ на одном графике, сравнить разбросы (обе ли оценки асимптотически нормальны, чья асимптотическая дисперсия меньше?).
   - (d) Сравнить, какая из оценок $((n+1)/n)X_{(n)}$ и $2\overline{X}$ чаще оказывается ближе к θ при разных n. Для этого смоделировать по 1000 реализаций (для каждого n) и найти, в какой доле из этих 1000 ближе оказалась $((n+1)/n)X_{(n)}$.

2. $X_1, \dots, X_n$ имеет распределение Коши $f_\theta(x) = (\pi(1 + (x - \theta)^2))^{-1}$.

   (a) Построить гистограммы для $\overline{X}$ при разных n. Является ли эта оценка состоятельной?
   *Сравнить гистограмму/оценку плотности с нормальной плотностью.

   (b) Построить гистограммы $\sqrt{n}(\mathrm{MED} - \theta)$, сравнить с соответсвующей нормальной плотностью (см. теорему 1).

3. * $X_1, ..., X_n \sim Bern(p)$, где а) $p = 1/3$ б) $p = 1/2$. Будет ли MED состоятельна? Асимптотически нормальна?
   Постройте гистограммы $\sqrt{n}(\mathrm{MED} - 1/2)$, похоже ли распределение на нормальное?

4. $X_1, ..., X_n \sim R([\theta - 2, \theta - 1] \cup [\theta + 1, \theta + 2])$. Будут ли выборочная медиана и усеченные средние а) состоятельны, б) асимптотически нормальны?

5. Смоделировать выборку из распределения Лапласа и численно сравнить асимптотическую дисперсию медианы, выборочного среднего, $\overline{X}_\alpha$ с $\alpha = 0.1$ и $\alpha = 0.3$. Для этого построить гистограммы или плотности каждого из распределений.