HW3

Сметанин Александр

6 марта 2020 г.

1. С помощью оценки $X_{[1]}$ (то есть центральная статистика должна как-то быть основана на $X_{[1]}$) построить точный доверительный интервал для параметра θ по выборке объёма n из:

a. равномерного распределения на отрезке $[\theta, \theta+1]$;
Так как $X_i \sim U(\theta, \theta+1)$, то $Y_i = \theta + 1 - X_i \sim U[0,1]$.
Тогда

$$Y_{[n]} = \max(Y_1, \dots, Y_n) = \theta + 1 - \min(X_1, \dots, X_n) = \theta + 1 - X_{[1]} = \mathrm{G}(X, \theta)$$

будет иметь независящую от параметра θ функцию распределения

$$F_{Y_{[n]}}(x) = \begin{cases} 0, & x \le 0 \\ x^n, & 0 < x \le 1 \\ 1, & x > 1 \end{cases}$$

Тогда можно построить интервал:

$$P(a < \theta + 1 - X_{[1]} < b) = P(a - 1 + X_{[1]} < \theta < b - 1 + X_{[1]})$$

Ширина интервала $L = b-1+X_{[1]}-(a-1+X_{[1]}) = b-a$. Чтобы найти интервал минимальной ширины, надо подобрать a и b, расположенные максимально близко, при фиксированном уровне доверия $1-\varepsilon$. Так как плотность $Y_{[n]}$ монотонно возрастает на $[0,1]$, выберем $b=1$, значит $\int_a^1 nx^{n-1}dx = x^n|_a^1 = 1-a^n = 1-\varepsilon$, $a = \sqrt[n]{\varepsilon}$. Полученный интервал для уровня доверия $1-\varepsilon$:

$$(\sqrt[n]{\varepsilon} - 1 + X_{[1]}, X_{[1]})$$

b. равномерного распределения на отрезке $[\theta, 2\theta]$.
Так как $X_i \sim U(\theta, 2\theta)$, то $Y_i = \frac{X_i}{\theta} - 1 \sim U[0,1]$.
Тогда

$$Y_{[1]} = \min(Y_1, \dots, Y_n) = \frac{\min(X_1, \dots, X_n)}{\theta} - 1 = \frac{X_{[1]}}{\theta} - 1 = \mathrm{G}(X, \theta)$$

будет иметь независящую от параметра θ функцию распределения

$$F_{Y_{[1]}}(x) = \begin{cases} 0, & x \le 0 \\ 1-(1-x)^n, & 0 < x \le 1 \\ 1, & x > 1 \end{cases}$$

Тогда для любых положительных a и b можно построить интервал:

$$P(a < \frac{X_{[1]}}{\theta} - 1 < b) = P(\frac{X_{[1]}}{b+1} < \theta < \frac{X_{[1]}}{a+1})$$

Длина такого интервала $X_{[1]} \cdot \frac{b-a}{(a+1)(b+1)}$. Осталось выбрать такие a, b, чтобы длина этого интервала была наименьшей. Плотность распределения $Y_{[n]}$ на отрезке [0, 1] монотонно убывает.

Поэтому, фиксированная площадь $1-\varepsilon$ под графиком плотности, при минимальном интервале возможна при $a=0$, а b таком, что

$$\int_0^b n(1-x)^{n-1}dx = -(1-x)^n|_0^b = 1-(1-b)^n = 1-\varepsilon$$

Значит $b = 1-\sqrt[n]{\varepsilon}$. Полученный интервал для уровня доверия $1-\varepsilon$:

$$(\frac{X_{[1]}}{\sqrt[n]{\varepsilon}}, X_{[1]})$$

2. По выборке из нормального распределения со средним $\theta > 0$ и дисперсией θ^2 построить точный доверительный интервал для параметра θ уровня доверия $1-\varepsilon$. Так как $X_i \sim (\theta, \theta^2)$, то можно построить интервал:

$$P(-c < \sqrt{n}\frac{\bar{X}-\theta}{\theta} < c) = P(\frac{\bar{X}}{1+\frac{c}{\sqrt{n}}} < \theta < \frac{\bar{X}}{1-\frac{c}{\sqrt{n}}})$$

Подставляя для уровня доверия $1-\varepsilon$ нужный квантиль $a = \tau_{1-\frac{\varepsilon}{2}}$, получим интервал

$$\left(\frac{\bar{X}}{1+\frac{\tau_{1-\frac{\varepsilon}{2}}}{\sqrt{n}}}, \frac{\bar{X}}{1-\frac{\tau_{1-\frac{\varepsilon}{2}}}{\sqrt{n}}}\right)$$

3. В результате проверки $n = 400$ электрических лампочек $k = 40$ штук оказалось бракованными. Найти доверительный интервал уровня 0.99 для вероятности брака.
Имеем дело с выборкой размера n из испытания Бернулли с параметром p, для которого нужно построить доверительный интервал уровня $1-\varepsilon = 0.99$. Тогда количество брака K будет иметь биномиальное распределение $K \sim B(n,p)$ с мат. ожиданием $EK = np$ и дисперсией $\sigma^2 = np(1-p)$.
По теореме Муавра-Лапласа, при больших n:

$$K \sim N(np, \sigma)$$

$$P(-a < \frac{K-np}{\sigma} < a) = 1-\varepsilon$$

$$P(\frac{K-a\sigma}{n} < p < \frac{K+a\sigma}{n}) = 1-\varepsilon$$

Вместо стандартного отклонения возьмем его выборочную оценку:

$$\sigma = \sqrt{n\frac{k}{n}(1-\frac{k}{n})} = \sqrt{400\cdot\frac{40}{400}\cdot(1-\frac{40}{400})} = 6$$

Используя $a = 2.57$ для $\varepsilon = 0.01$, получим интервал:

$$(\frac{40-2.57\cdot 6}{400}, \frac{40-2.57\cdot 6}{400}) = (0.061, 0.139)$$

4. Пусть $X_1, \dots, X_n$ - выборка из распределения Коши с параметром сдвига a. Построить асимптотический доверительный интервал для a, используя выборочную медиану. Так как выборочная медиана является асимптотически нормальной оценкой теоретической медианы, а для распределения Коши медиана и параметр сдвига a совпадают, оценим сдвиг с помощью выборочной медианы, которую будем обозначать $\bar{med}$. Знаем, что

$$\sqrt{n}\left(\bar{med} - x_{1/2}\right) \xrightarrow{\mathrm{D}_{n\to\infty}} N\left(0, \frac{1}{4p^2\left(x_{1/2}\right)}\right).$$

Так как рассматриваем $C(a)$, получим

$$\sqrt{n}(\bar{med} - a) \xrightarrow{D_{n\to\infty}} N(0, \frac{\pi^2}{4})$$

$$P(-b < \sqrt{n}(\bar{med} - a) < b) = \Phi_{(0,\frac{\pi^2}{4})}(b) - \Phi_{(0,\frac{\pi^2}{4})}(-b) = \Phi\left(\frac{2b}{\pi}\right) - \Phi\left(\frac{-2b}{\pi}\right)$$

Переобозначим $c = \frac{2b}{\pi}$

$$P(-b < \sqrt{n}(\bar{med} - a) < b) = P(-\frac{c\pi}{2\sqrt{n}} < \bar{med} - a < \frac{c\pi}{2\sqrt{n}}) = P(\bar{med} - \frac{c\pi}{2\sqrt{n}} < a < \bar{med} + \frac{c\pi}{2\sqrt{n}})$$

Подставим для уровня доверия $1 - \varepsilon$ квантиль $c = \tau_{1-\frac{\varepsilon}{2}}$, получим интервал

$$(\bar{med} - \tau_{1-\frac{\varepsilon}{2}} \cdot \frac{\pi}{2\sqrt{n}}, \bar{med} + \tau_{1-\frac{\varepsilon}{2}} \cdot \frac{\pi}{2\sqrt{n}})$$

5. Пусть $X1, \ldots, X_n$ - выборка из нормального распределения со средним a и дисперсией σ^2, причем значение σ^2 известно.

a. Построить асимптотический доверительный интервал для a, используя выборочную медиану.

Так как выборочная медиана является асимптотически нормальной оценкой теоретической медианы, а для нормального распределения медиана и мат ожидание совпадают, значит можно матожидание оценивать выборочной медианой. Знаем, что

$$\sqrt{n}\left(\bar{med} - x_{1/2}\right) \xrightarrow{D_{n\to\infty}} N\left(0, \frac{1}{4p^2\left(x_{1/2}\right)}\right).$$

Так как рассматриваем $N(a, \sigma^2)$, получим

$$\sqrt{n}(\bar{med} - a) \xrightarrow{D_{n\to\infty}} N(0, \frac{\pi\sigma^2}{2})$$

$$P(-b < \sqrt{n}(\bar{med} - a) < b) = \Phi_{(0,\frac{\pi\sigma^2}{2})}(b) - \Phi_{(0,\frac{\pi\sigma^2}{2})}(-b) = \Phi\left(\frac{b}{\sigma}\sqrt{\frac{2}{\pi}}\right) - \Phi\left(\frac{-b}{\sigma}\sqrt{\frac{2}{\pi}}\right)$$

Переобозначим $c = \frac{b}{\sigma}\sqrt{\frac{2}{\pi}}$

$$P(-c\sigma\frac{\pi}{2} < \sqrt{n}(\bar{med} - a) < c\sigma\frac{\pi}{2}) = \Phi(c) - \Phi(-c) = 1 - \varepsilon$$

$$P(\bar{med} - c\sigma\sqrt{\frac{\pi}{2n}} < a < \bar{med} + c\sigma\sqrt{\frac{\pi}{2n}}) = \Phi(c) - \Phi(-c) = 1 - \varepsilon$$

Подставляя для уровня доверия $1 - \varepsilon$ нужный квантиль $c = \tau_{1-\frac{\varepsilon}{2}}$, получим интервал

$$(\bar{med} - \tau_{1-\frac{\varepsilon}{2}} \cdot \sigma\sqrt{\frac{\pi}{2n}}, \bar{med} + \tau_{1-\frac{\varepsilon}{2}} \cdot \sigma\sqrt{\frac{\pi}{2n}})$$

b. Сравнить полученный интервал с точным доверительным интервалом, построенным по выборочному среднему значению.

Построим доверительный интервал по выборочному среднему:

$$P(-a < \sqrt{n}\frac{\bar{X} - \mu}{\sigma} < a) = 1 - \varepsilon$$

$$(\bar{X} - a\frac{\sigma}{\sqrt{n}}, \bar{X} + a\frac{\sigma}{\sqrt{n}})$$

Подставляя для уровня доверия $1-\varepsilon$ нужный квантиль $a=\tau_{1-\frac{\varepsilon}{2}}$, получим интервал

$$(\bar{X}-\tau_{1-\frac{\varepsilon}{2}}\cdot\frac{\sigma}{\sqrt{n}},\bar{X}+\tau_{1-\frac{\varepsilon}{2}}\cdot\frac{\sigma}{\sqrt{n}})$$

Сравним длинну интервалов, построенных по выборочной медиане и выборочному среднему:

$$\frac{L_{\bar{med}}}{L_{\bar{X}}} \quad = \quad \frac{\bar{med}+\tau_{1-\frac{\varepsilon}{2}}\cdot\sigma\sqrt{\frac{\pi}{2n}}-(\bar{med}-\tau_{1-\frac{\varepsilon}{2}}\cdot\sigma\sqrt{\frac{\pi}{2n}})}{\bar{X}+\tau_{1-\frac{\varepsilon}{2}}\cdot\frac{\sigma}{\sqrt{n}}-(\bar{X}-\tau_{1-\frac{\varepsilon}{2}}\cdot\frac{\sigma}{\sqrt{n}})} \quad = \quad \frac{2\tau_{1-\frac{\varepsilon}{2}}\cdot\sigma\sqrt{\frac{\pi}{2n}}}{2\tau_{1-\frac{\varepsilon}{2}}\cdot\frac{\sigma}{\sqrt{n}}} \quad = \quad \sqrt{\frac{\pi}{2}}$$

Видим, что доверительный интервал, построенный по выборочной медиане в $\sqrt{\frac{\pi}{2}}$ раз шире, чем доверительный интервал, построенный по выборочному среднему.