Статистическое моделирование и прогнозирование
Лекция 3. Интервальные оценки параметров распределения

Семёнов Михаил Евгеньевич
к. ф.-м. н., доцент ОЭФ ИЯТШ

Томский политехнический университет
16 марта 2020 г.

Содержание

Интервальная оценка

Интервальная оценка позволяет получить вероятностную характеристику точности оценивания неизвестного параметра θ.

Пусть имеется случайная выборка объёма n из непрерывного распределения случайной величины с неизвестным параметром θ для оценки которого строится интервал: (θ_n^-, θ_n^+), где θ_n^-, θ_n^+ — функции случайной выборки, такие, что верно равенство

$$\mathbb{P}\{\theta \in (\theta_n^-, \theta_n^+)\} = \gamma, \tag{1}$$

Тогда интервал (θ_n^-, θ_n^+) называют *доверительным* интервалом, накрывающим неизвестный параметр θ с заданной доверительной вероятностью γ или γ-доверительным интервалом. Обычно γ выбирают среди чисел: 0,9, 0,95, 0,975, 0.99.

Заметим, что при построении доверительных интервалов для дискретных случайных величин вместо равенства удаётся обеспечить лишь неравенство

$$\mathbb{P}\{\theta \in (\theta_n^-, \theta_n^+)\} \geq \gamma. \tag{2}$$

Метод построения доверительного интервала

Один из типичных методов построения доверительного интервала основан на использовании статистики $T(\theta)$, функция распределения которой $F(t)$ не зависит от оцениваемого параметра θ. При этом используются следующие предположения:

- Функция распределения статистики $F(t)$ является непрерывной и возрастающей;
- Для любой реализации выборки статистика $T(\theta)$ является непрерывной и монотонной функцией параметра θ.
- Задана доверительная вероятность γ.

Доверительный интервал для $M(X)$

При построении доверительного интервала для математического ожидания нормально распределённой случайной величины $X \sim \mathbb{N}(a, \sigma)$ по случайной выборке объёмом n используется статистика вида

$$T(a) = \frac{\bar{x} - a}{s_x}\sqrt{n}. \tag{3}$$

Действительно, если $\bar{x} \sim \mathbb{N}(a, \frac{\sigma}{\sqrt{n}})$, то $Z = \frac{\bar{x}-a}{s_x}\sqrt{n} \sim \mathbb{N}(0, 1)$.
В тоже время $V = s_x \frac{n-1}{\sigma}$ причём случайные величины Z и V независимы и статистика $T(a)$ может быть представлена в виде

$$T(a) = \frac{\bar{x} - a}{s_x}\sqrt{n} = Z\sqrt{\frac{n-1}{V}}. \tag{4}$$

Отсюда следует, что статистика $T(a)$ имеет распределение Стьюдента с $n-1$ числом степеней свободы.

Доверительный интервал для $M(X)$

Для убывающей по параметру a функции $T(a)$ определяющие доверительный интервал уравнения $T(\theta) = t_{\frac{1\pm\gamma}{2}}$ принимают вид

$$T(a_n^{\pm}) = \frac{\bar{x} - a_n^{\pm}}{s_x}\sqrt{n} = t_{\frac{1\pm\gamma}{2},n-1}. \tag{5}$$

Решая эти уравнения, находим нижнюю и верхнюю границы γ-доверительного интервала для математического ожидания $M(X) = a$ нормально распределённой случайной величины $X \sim \mathbb{N}(a, \sigma)$:

$$\left(\bar{x} - \frac{s_x}{\sqrt{n}} \cdot t_{\frac{1+\gamma}{2},n-1}, \quad \bar{x} - \frac{s_x}{\sqrt{n}} \cdot t_{\frac{1-\gamma}{2},n-1}\right), \tag{6}$$

где $t_{\frac{1\pm\gamma}{2},n-1}$ — квантили уровней $\frac{1\pm\gamma}{2}$ распределения Стьюдента с числом степеней свободы $n-1$.

Доверительный интервал для дисперсии $D(X)$

При построении доверительного интервала для дисперсии нормально распределённой случайной величины $X \sim \mathbb{N}(a, \sigma)$ по случайной выборке объёмом n используется статистика, имеющая χ^2 распределение с числом степеней свободы $n-1$

$$V(\sigma) = \frac{n-1}{\sigma^2} \cdot s_x^2 \sim \chi^2_{n-1}. \tag{7}$$

Для убывающей по параметру σ функции $V(\sigma)$ нижняя и верхняя границы γ-доверительного интервала определяются уравнениями

$$V(\sigma_n^{\pm}) = \frac{n-1}{\sigma_n^{2\pm}} \cdot s_x^2 = \chi^2_{\frac{1\pm\gamma}{2}, n-1}. \tag{8}$$

Доверительный интервал для дисперсии $D(X)$

Решая эти уравнения, находим γ-доверительный интервал для дисперсии $D(X) = \sigma^2$ нормально распределённой случайной величины $X \sim \mathbb{N}(a, \sigma)$:

$$\left(s_x^2 \cdot \frac{n-1}{\chi^2_{\frac{1+\gamma}{2}, n-1}}, \quad s_x^2 \cdot \frac{n-1}{\chi^2_{\frac{1-\gamma}{2}, n-1}}\right), \tag{9}$$

где $\chi^2_{\frac{1\pm\gamma}{2}, n-1}$ – квантили уровней $\frac{1\pm\gamma}{2}$ распределения Пирсона с числом степеней свободы $n-1$.

Постановка заданий

- Построить доверительные интервалы для математического ожидания $M(X)$ и дисперсии $D(X)$ стандартной нормально распределённой случайной величины $X \sim \mathbb{N}(0, 1)$ при различных значениях доверительной вероятности $\gamma \in [0.95, 0.99]$, шаг 0.01 и объёма выборок $n \in [100, 1000]$, шаг 50. Сохранить скрипт в файле (*.r).
- Подписать оси графиков, вывести два графика в одном графическом окне. Поменять оси графиков местами: объем выборки n – ось абсцисс, доверительный интервал γ – ось ординат.
- Установить начальное состояние генератора псевдослучайных чисел seed=ГГГГММДД, где ГГГГММДД – дата вашего рождения. Выполнить задание для случайной величины $X \sim \mathbb{N}(a, \sigma)$.
- Переписать скрипт построения интервальных оценок без операторов цикла for()/while().

Результат

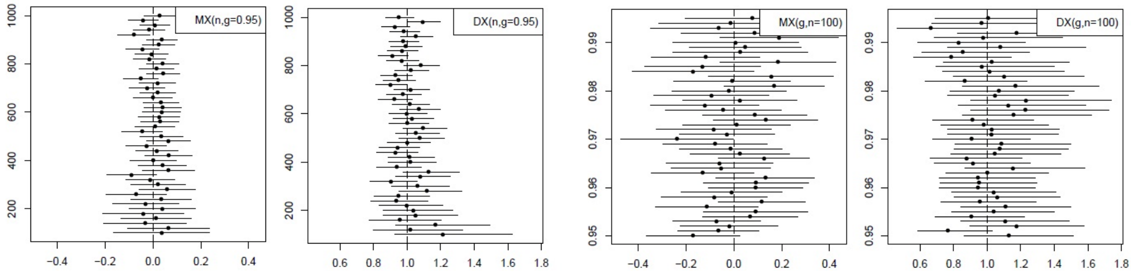

Рисунок 1 – Реализации доверительных интервалов для $M(X)$, $D(X)$

Скрипт на R [Буховец et al., 2010]

Алгоритм 1 Доверительные интервалы для $M(X)$, $D(X)$

Вход: Выборка с.в..

```
set.seed(20100625)
n <- seq(100, 1000, 20); g <- seq(0.95, 0.995, , length(n))
ciM <- function(x,n,g) mean(x)-sd(x)/sqrt(n)*qt((1+c(g,0,-g))/2, n-1)
ciD <- function(x,n,g) sd(x)*sd(x)*(n-1)/qchisq((1+c(g,0,-g))/2, n-1)
ciMn <- sapply(n, function(nn) ciM(rnorm(nn), nn, g[1]))
ciMg <- sapply(g, function(gg) ciM(rnorm(n[1]), n[1], gg))
txtMn <- sprintf("MX(n, g=%.3g)”, g[1]); txtMg <- sprintf("MX(g, n=%.0f)”, n[1])
cigraph <- function(x, y, point, text) { windows()
plot(range(x), range(y), type=”n”, xlab=” ”, ylab=” ”)
for(j in seq(length(y)))
lines(x[,j], rep(y[j], 3), lwd=2)
points(x[2,j], y[j], pch=16, lwd=2)
legend("topright legend=text, bg="white")
abline(v=point, lty=2, lwd=2) }
ci_graph(ciMn, n, 0, txtMn); ci_graph(ciDn, n, 1, txtDn)
```

Выход: Доверительные интервалы для $M(X)$, $D(X)$. Графики

Список использованных источников

Буховец, , Москалев, , Богатова, , and Бирючинская, (2010).
Статистический анализ данных в системе R. Учебное пособие.
ВГАУ, Воронеж.