Статистическое моделирование и прогнозирование
Лекция 2. Выборочные характеристики и точечные оценки параметров распределения

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

Томский политехнический университет

16 марта 2020 г.

Содержание

Основные определения
- Точечная оценка
- Свойства точечной оценки
- Точечные оценки мат. ожидания, дисперсии, квантиля
- Оценки функций распределения и плотности вероятности

Лабораторная работа 2

Список использованных источников

Точечная оценка

Любые характеристики случайной величины X, полученные по её выборке $\{x_1, x_2, \ldots, x_n\}$, называются *выборочными* или *эмпирическими*. *Статической оценкой* называется выборочная характеристика, используемая в качестве приближённого значения неизвестной характеристики генеральной совокупности.

Статистическая оценка, представленная в виде числа (точки на числовой прямой), называется *точечной*. Тогда практическая применимость точечной оценки определяется такими её свойствами как несмещённость, состоятельность и эффективность.

Пусть $\{x_1, x_2, \ldots, x_n\}$ — случайная выборка, тогда $\theta_n(x_1, x_2, \ldots, x_n)$ – выборочная оценка некоторого параметра θ.

Свойства точечной оценки

Оценка θ_n называется *несмещённой*, если для любого фиксированного n верно, что $M(\theta_n) = \theta$.

Оценка θ_n называется *состоятельной*, если она сходится по вероятности к истинному значению параметра θ, то есть для любого $\epsilon > 0$ выполняется условие

$$\lim_{n\to\infty} \mathbb{P}\{|\theta_n - \theta| < \epsilon\} = 1. \tag{1}$$

Оценка θ_n называется *эффективной*, если она обладает наименьшей дисперсией, а значит и средним квадратическим отклонением от истинного значения параметра θ, по сравнению с любыми другими оценками данного класса.

Точечные оценки мат. ожидания, дисперсии, квантиля

Несмещёнными и состоятельными оценками для математического ожидания $M(X)$ и дисперсии $D(X)$ являются выборочное среднее $\bar{x}$ и исправленная выборочная дисперсия s_x^2:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{2}$$

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2. \tag{3}$$

Выборочный квантиль порядка p:

$$x_{p,n} = \begin{cases} x_{([np]+1)}, & \text{если } np \notin \mathbb{Z}; \\ \frac{1}{2}\left(x_{([np])} + x_{([np]+1)}\right), & \text{если } np \in \mathbb{Z}, \end{cases} \tag{4}$$

где $[a]$ – целая часть числа a.

Оценки функций распределения и плотности вероятности

Оценками функций распределения $F(x)$ и плотности вероятности $f(x)$ непрерывной случайной величины X будут построенные по её выборке эмпирическая функция распределения $F_n(x)$ и гистограмма $f_{n,h}(x)$:

$$F_n(x) \xrightarrow[n\to\infty]{\mathbb{P}} F(x), \quad f_{n,h}(x) \xrightarrow[n\to\infty]{\mathbb{P}} f(x) \tag{5}$$

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{(-\infty,x)}(x_i), \quad f_{n,h}(x) = \frac{1}{nh}\sum_{i=1}^{n} \mathbb{I}_{(kh,(k+1)h)}(x_i), \tag{6}$$

где $nh \to \infty$, $h \to 0$, $h = const$ — длина интервала группировки; $k = [x/h] \in \mathbb{Z}$ — номер интервала группировки.

Постановка заданий

- Вычислить выборочные характеристики (математическое ожидание, среднеквадратическое отклонение, квантили), построить графики эмпирической функции распределения $F_n(x)$ и гистограммы $f_{n,h}(x)$ для выборки $n = 100$ значений случайной величины X распределенной по нормальному закону. Сохранить скрипт в файле (*.r).
- Вычислить выборочные характеристики коэффициентов вариации, асимметрии и эксцесса. **Указание.** Установить пакет moments.
- Подготовить инструкцию использования функций: read.csv(), write.csv(), read.table(), write.table(), set.seed(), skewness(), kurtosis().
- Сгенерировать случайную величину $X \sim N(50, 25)$ в интервале $[0, 100]$. Вычислить выборочные характеристики. **Указание.** Установить пакет Runuran.
- **Написать функции импорта/экспорта выборочных данных в файл (*.csv).**

Результат

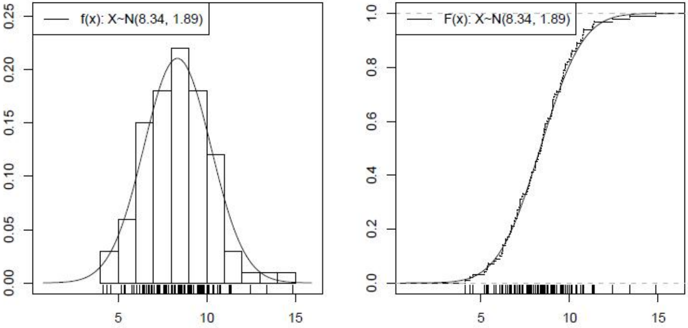

Рисунок 1 – Гистограмма $f_{n,h}(x)$ выборки значений и график функции плотности вероятности $f(x)$ (слева) и график функции распределения $F(x)$ (справа) с.в. $X \sim N(8.34, 1.89)$

Скрипт на R [Буховец et al., 2010]

Алгоритм 1 Графики эмпирической функции распределения и гистограмма

Вход: Выборка с.в..

```
source("samples.r")
n <- 100; x <- samples(n, seed=20100625)
a1 <- mean(x); s1 <- sd(x); quantile(x, c(0, .25, .5, .75, 1))
x2 <- seq(a1-4*s1, a1+4*s1, len=n); range(x2)
f1 <- dnorm(x2, a1, s1); F1 <- pnorm(x2, a1, s1)
ltext <- sprintf(”X~N(%.2f, %.2f)”,a1,s1); ltext
par(mfrow=c(1,2))
hist(x, breaks=”Scott”, xlim=range(x2), ylim=c(0, 1.2*max(f1)), freq=FALSE)
rug(x); lines(x2, f1); box()
legend(”topleft”, lty=1, legend=paste(”f(x):”, ltext))
plot(ecdf(x), pch=”.”, xlim=range(x2))
rug(x); lines(x2, F1)
legend(”topleft”, lty=1, legend=paste(”F(x):”, ltext))
```

Выход: Выборочные характеристики. Графики.

Скрипт на R [Буховец et al., 2010]

Алгоритм 2 Генерирование реализации случайной выборки с законом распределения, близким к нормальному

Вход: n – объём случайной выборки, seed – начальное состояние генератора п.с.ч.

```
samples <- function(n=100, seed=20100625) {
set.seed(seed)
a1 <- runif(1, min=-9, max=9)
s1 <- runif(1, min=0.1, max=3)
a2 <- runif(1, min=a1-0.7*s1, max=a1+0.7*s1)
s2 <- runif(1, min=0.1*s1, max=0.5*s1)
x <- rnorm(n, a1, s1) + rnorm(n, a2, s2)
}
```

Выход: Реализация случайной выборки

Скрипт на R

Алгоритм 3 Генерирование реализации случайной выборки с законом распределения, близким к нормальному с дополнительными ограничениями

Вход: n – объём случайной выборки, seed – начальное состояние генератора п.с.ч.

```
set.seed(20190625)
require("Runuran")
        # Normal distribution bounded between 0 and 100
d1 <- urnorm(n = 1000, mean = 50, sd = 25, lb = 0, ub = 100)
mean(d1)
sd(d1)
hist(d1)
```

Выход: Реализация случайной выборки $N(50, 25)$, выборочные характеристики, гистограмма

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