Теория вероятностей и математическая статистика

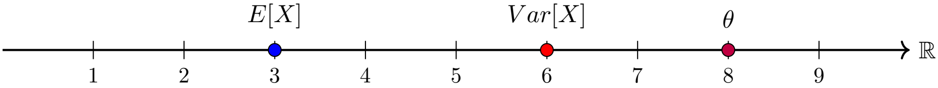
Выборочные распределения. Точечные оценки. Практическая сессия.

Глеб Карпов

ФКН ВШЭ

Напоминание: задача статистики

Идеальная ситуация: мы знаем параметры и/или характеристики



Реальная ситуация: параметры и/или характеристики неизвестны

Значения будто скрыты от нас туманом

- Параметры и характеристики случайной величины — числа, **точки** на числовой оси. Семейство методов для угадывания этих значений обычно называется **точечной оценкой**.

Напоминание: выьборочное среднее

- Пусть $\boxed{\mathcal{X} = (X_1, \dots, X_n)}$ — случайная выборка со средним $\mu = E[X_i]$ и дисперсией $\sigma^2 = Var[X_i] < \infty$.
- Одна из самых важных статистик — выборочное среднее:

$x = (x_1, \dots, x_n)$

$$\bar{X} = \frac{X_1 + \dots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Характеристики выборочного среднего $\bar{X}$:
  - $E[\bar{X}] = E\left[\frac{X_1}{n} + \dots + \frac{X_n}{n}\right] = \frac{1}{n}E\left[X_1 + \dots + X_n\right] = \frac{n\mu}{n} = \mu$
  - $\text{Var}\left[\bar{X}\right] = \text{Var}\left[\frac{X_1}{n} + \dots + \frac{X_n}{n}\right] = \frac{1}{n^2}\text{Var}\left[X_1 + \dots + X_n\right] = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$
- При выполнении условий ЦПТ $(n \geq 30)$ можем заявлять, что:

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Задачи

$E[X_i] = \mu = 0.5$

$X_1 \dots X_{36}$

$Var[X_i] = \sigma^2 = (0.02)^2$

Пример 1

Производственный процесс предназначен для изготовления деталей диаметром 0.5 дюйма. Каждый день выбирается случайная выборка из 36 деталей и записываются их диаметры. Если полученное выборочное среднее меньше 0.49 дюйма или больше 0.51 дюйма, процесс останавливается для настройки. Стандартное отклонение диаметра составляет 0.02 дюйма. Какова вероятность того, что производственная линия будет остановлена без необходимости?

$P(\{\overline{X} > 0.51\} \cup \{\overline{X} < 0.49\})$

$\overline{X} \underset{CLT}{\sim} N\left(0.5, \frac{(0.02)^2}{36}\right)$

Слайд для записей

$$\bar{X} \sim N\left(0.5, \frac{(0.02)^2}{36}\right)$$

CLT

$$P(\bar{X} > 0.51) = P\left(Z > \frac{(0.51-0.5)\cdot 6}{0.02}\right) = P(Z > 3) =$$

$$= 1 - F_Z(3) = 1 - 0.9987 =$$

$$= 0.0013$$

$$P(\bar{X} < 0.49) = P\left(Z < \frac{(0.49-0.5)\cdot 6}{0.02}\right) = P(Z < -3) = 0.0013$$

$$P = 2 \cdot 0.0013 = 0.0026$$

Слайд для записей

$X_1 \;\ldots\; X_n$

$P(X_i = 1) = p$

$P(X_i = 0) = 1 - p$

$S = \sum X_i$

$E[S] = np; \quad var[S] = npq$

$S \sim Bin(n, p)$

ЦПТ: $n \geq 30$, $S \sim N(np, npq)$

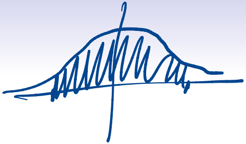
$E[\hat{p}] = E\left[\frac{S}{n}\right] = p$

$\hat{p} = \frac{S}{n}$ $\quad \hat{p} \sim N\left(p, \frac{pq}{n}\right)$

$Var[\hat{p}] = Var\left[\frac{S}{n}\right] = \frac{pq}{n}$

Задачи

$= 2\left(F_Z(2.18) - F_Z(0)\right) = 2\cdot(0.9854 - 0.5) = 0.9708$



Пример 2

Сообщается, что в крупном исследовании, проведённом в штате Нью-Йорк, примерно 30% участников исследования жили в пределах 1 мили от опасного места захоронения отходов. Пусть p обозначает долю всех жителей Нью-Йорка, которые живут в пределах 1 мили от такого места, и предположим, что $p = 0.3$.

1. Каковы математическое ожидание и стандартное отклонение выборочной доли $\hat{p}$ на основе случайной выборки размера $n = 400$?
2. При $n = 400$ чему равно $P(0.25 \leq \hat{p} \leq 0.35)$?

1. $\hat{p}$; $E[\hat{p}] = 0.3$; $var[\hat{p}] = \frac{0.3\cdot 0.7}{400}$

2. $P\left(\frac{0.25 - 0.3}{\sqrt{\frac{0.3\cdot 0.7}{400}}} < Z < \frac{0.35 - 0.3}{\sqrt{\frac{0.3\cdot 0.7}{400}}}\right) = P(-2.18 < Z < 2.18)$ =

Задачи

$P = 0.2$

$$\hat{p} \sim N\left(0.2, \frac{0.2 \cdot 0.8}{100}\right)$$

Пример 3

Кабельная компания решает, стоит ли заменить канал "магазин на диване" новой местной телепрограммой. Будет проведён опрос 100 абонентов. Кабельная компания решила оставить канал "магазин на диване", если выборочная доля окажется больше 0.25. Какова приближённая вероятность того, что кабельная компания все-таки его оставит,если истинная доля тех, кто его смотрит, составляет только 0.20?

$$P(\hat{p} > 0.25) = P\left(Z > \frac{0.25 - 0.2}{\sqrt{\frac{0.2 \cdot 0.8}{100}}}\right) = P(Z > 1.25) =$$

$$= 1 - F_Z(1.25) = 1 - 0.8944 = 0.1056$$