Теория Вероятностей и Статистика

Статистические оценки и их свойства

Потанин Богдан Станиславович

старший преподаватель, кандидат экономических наук

2021

- Информация о распределении многих случайных величин, таких как зарплата случайно взятого индивида, дневная посещаемость сайта или цена акции, может быть крайне полезна на практике.

- Информация о распределении многих случайных величин, таких как зарплата случайно взятого индивида, дневная посещаемость сайта или цена акции, может быть крайне полезна на практике.
- В теории вероятностей мы обычно предполагаем, что распределение случайных величин нам известно.

- Информация о распределении многих случайных величин, таких как зарплата случайно взятого индивида, дневная посещаемость сайта или цена акции, может быть крайне полезна на практике.
- В теории вероятностей мы обычно предполагаем, что распределение случайных величин нам известно.
- К сожалению, в реальности, распределение соответствующих случайных величин нам, как правило, неизвестно. Поэтому необходимо его аппроксимировать (оценить) при помощи имеющихся у нас данных: о зарплатах опрошенных индивидов, о посещаемости сайта и ценах акций в предшествующие дни и т.д.

- Информация о распределении многих случайных величин, таких как зарплата случайно взятого индивида, дневная посещаемость сайта или цена акции, может быть крайне полезна на практике.
- В теории вероятностей мы обычно предполагаем, что распределение случайных величин нам известно.
- К сожалению, в реальности, распределение соответствующих случайных величин нам, как правило, неизвестно. Поэтому необходимо его аппроксимировать (оценить) при помощи имеющихся у нас данных: о зарплатах опрошенных индивидов, о посещаемости сайта и ценах акций в предшествующие дни и т.д.
- Если аппроксимация неизвестного распределения или, по крайней мере, его отдельных характеристик (математическое ожидание, дисперсия и т.д.) окажется достаточно точна, то на практике вместо неизвестного истинного распределения и его характеристик можно использовать соответствующие аппроксимации.

- Информация о распределении многих случайных величин, таких как зарплата случайно взятого индивида, дневная посещаемость сайта или цена акции, может быть крайне полезна на практике.
- В теории вероятностей мы обычно предполагаем, что распределение случайных величин нам известно.
- К сожалению, в реальности, распределение соответствующих случайных величин нам, как правило, неизвестно. Поэтому необходимо его аппроксимировать (оценить) при помощи имеющихся у нас данных: о зарплатах опрошенных индивидов, о посещаемости сайта и ценах акций в предшествующие дни и т.д.
- Если аппроксимация неизвестного распределения или, по крайней мере, его отдельных характеристик (математическое ожидание, дисперсия и т.д.) окажется достаточно точна, то на практике вместо неизвестного истинного распределения и его характеристик можно использовать соответствующие аппроксимации.
- Математическая статистика, в частности, позволяет находить подобного рода аппроксимации (оценки) и определять их качество (насколько они точны).

Выборка

Определение

- Последовательность независимых, одинаково распределенных случайных величин $X_1, X_2, ..., X_n$ будем именовать **выборкой** объема $n \in N$.

Выборка

Определение

- Последовательность независимых, одинаково распределенных случайных величин $X_1, X_2, ..., X_n$ будем именовать **выборкой** объема $n \in N$.
- Элементы выборки X_i, где $i \in \{1, ..., n\}$, именуются **наблюдениями**.

Выборка

Определение

- Последовательность независимых, одинаково распределенных случайных величин $X_1, X_2, ..., X_n$ будем именовать **выборкой** объема $n \in N$.
- Элементы выборки X_i, где $i \in \{1, ..., n\}$, именуются **наблюдениями**.
- Если $X_i \sim \Theta(\theta)$, то выборка была получена из распределения Θ с вектором параметров θ.

Выборка
Определение

- Последовательность независимых, одинаково распределенных случайных величин $X_1, X_2, ..., X_n$ будем именовать **выборкой** объема $n \in N$.
- Элементы выборки X_i, где $i \in \{1, ..., n\}$, именуются **наблюдениями**.
- Если $X_i \sim \Theta(\theta)$, то выборка была получена из распределения Θ с вектором параметров θ.
- Конкретные значения $x_1, ..., x_n$, которые приняли наблюдения в выборке, именуются **реализациями**.

Выборка
Определение

- Последовательность независимых, одинаково распределенных случайных величин $X_1, X_2, ..., X_n$ будем именовать **выборкой** объема $n \in N$.
- Элементы выборки X_i, где $i \in \{1, ..., n\}$, именуются **наблюдениями**.
- Если $X_i \sim \Theta(\theta)$, то выборка была получена из распределения Θ с вектором параметров θ.
- Конкретные значения $x_1, ..., x_n$, которые приняли наблюдения в выборке, именуются **реализациями**.
- Через $X = (X_1, ..., X_n)$ и $x = (x_1, ..., x_n)$ обозначим векторы, состоящие из наблюдений и их реализаций соответственно.

Выборка

Определение

- Последовательность независимых, одинаково распределенных случайных величин $X_1, X_2, ..., X_n$ будем именовать **выборкой** объема $n \in N$.
- Элементы выборки X_i, где $i \in \{1, ..., n\}$, именуются **наблюдениями**.
- Если $X_i \sim \Theta(\theta)$, то выборка была получена из распределения Θ с вектором параметров θ.
- Конкретные значения $x_1, ..., x_n$, которые приняли наблюдения в выборке, именуются **реализациями**.
- Через $X = (X_1, ..., X_n)$ и $x = (x_1, ..., x_n)$ обозначим векторы, состоящие из наблюдений и их реализаций соответственно.

Пример:

- Лаврентий 3 раза подбрасывает обычную правильную монетку.

Выборка
Определение

- Последовательность независимых, одинаково распределенных случайных величин $X_1, X_2, ..., X_n$ будем именовать **выборкой** объема $n \in N$.
- Элементы выборки X_i, где $i \in \{1, ..., n\}$, именуются **наблюдениями**.
- Если $X_i \sim \Theta(\theta)$, то выборка была получена из распределения Θ с вектором параметров θ.
- Конкретные значения $x_1, ..., x_n$, которые приняли наблюдения в выборке, именуются **реализациями**.
- Через $X = (X_1, ..., X_n)$ и $x = (x_1, ..., x_n)$ обозначим векторы, состоящие из наблюдений и их реализаций соответственно.

  Пример:
- Лаврентий 3 раза подбрасывает обычную правильную монетку.
- Число орлов, которое выпадет при i-м броске (ноль или один), где $i \in \{1, 2, 3\}$, является случайной величиной $X_i \sim Ber(0.6)$ с параметром $p = 0.5$.

- Последовательность независимых, одинаково распределенных случайных величин $X_1, X_2, ..., X_n$ будем именовать **выборкой** объема $n \in N$.
- Элементы выборки X_i, где $i \in \{1, ..., n\}$, именуются **наблюдениями**.
- Если $X_i \sim \Theta(\theta)$, то выборка была получена из распределения Θ с вектором параметров θ.
- Конкретные значения $x_1, ..., x_n$, которые приняли наблюдения в выборке, именуются **реализациями**.
- Через $X = (X_1, ..., X_n)$ и $x = (x_1, ..., x_n)$ обозначим векторы, состоящие из наблюдений и их реализаций соответственно.
  Пример:
- Лаврентий 3 раза подбрасывает обычную правильную монетку.
- Число орлов, которое выпадет при i-м броске (ноль или один), где $i \in \{1, 2, 3\}$, является случайной величиной $X_i \sim Ber(0.6)$ с параметром $p = 0.5$.
- Поскольку X_1, X_2 и X_3 независимы и одинаково распределены, то они формируют выборку объема $n = 3$ из распределения $Ber(0.5)$.

Выборка
Определение

- Последовательность независимых, одинаково распределенных случайных величин $X_1, X_2, ..., X_n$ будем именовать **выборкой** объема $n \in N$.
- Элементы выборки X_i, где $i \in \{1, ..., n\}$, именуются **наблюдениями**.
- Если $X_i \sim \Theta(\theta)$, то выборка была получена из распределения Θ с вектором параметров θ.
- Конкретные значения $x_1, ..., x_n$, которые приняли наблюдения в выборке, именуются **реализациями**.
- Через $X = (X_1, ..., X_n)$ и $x = (x_1, ..., x_n)$ обозначим векторы, состоящие из наблюдений и их реализаций соответственно.

Пример:

- Лаврентий 3 раза подбрасывает обычную правильную монетку.
- Число орлов, которое выпадет при i-м броске (ноль или один), где $i \in \{1, 2, 3\}$, является случайной величиной $X_i \sim Ber(0.6)$ с параметром $p = 0.5$.
- Поскольку X_1, X_2 и X_3 независимы и одинаково распределены, то они формируют выборку объема $n = 3$ из распределения $Ber(0.5)$.
- Допустим, что первые два раза выпал орел, а последний раз – решка. В таком случае реализации выборки будут иметь вид $x_1 = 1, x_2 = 1$ и $x_3 = 0$, что можно кратко записать как $x = (1, 1, 0)$.

- **Выборочное среднее** является средним значением по выборке $X_1, ..., X_n$:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Выборочное среднее

Определение и некоторые свойства

- **Выборочное среднее** является средним значением по выборке $X_1, ..., X_n$:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Реализация выборочного среднего рассчитывается по реализациям наблюдений:

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Выборочное среднее

Определение и некоторые свойства

- **Выборочное среднее** является средним значением по выборке $X_1, ..., X_n$:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Реализация выборочного среднего рассчитывается по реализациям наблюдений:

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Пользуясь независимостью и одинаковой распределенностью элементов выборки нетрудно показать:

$$E(\overline{X}_n) = E(X_1) \qquad Var(\overline{X}_n) = \frac{Var(X_1)}{n}$$

Выборочное среднее

Определение и некоторые свойства

- **Выборочное среднее** является средним значением по выборке $X_1, ..., X_n$:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Реализация выборочного среднего рассчитывается по реализациям наблюдений:

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Пользуясь независимостью и одинаковой распределенностью элементов выборки нетрудно показать:

$$E(\overline{X}_n) = E(X_1) \qquad Var(\overline{X}_n) = \frac{Var(X_1)}{n}$$

Пример:
Имеется выборка объема $n = 3$ из экспоненциального распределения с параметром $\lambda = 0.2$. Найдем математическое ожидание, дисперсию и реализацию выборочного среднего, если известно, что $x = (1, 6, 3)$.

Выборочное среднее

Определение и некоторые свойства

- **Выборочное среднее** является средним значением по выборке $X_1, ..., X_n$:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Реализация выборочного среднего рассчитывается по реализациям наблюдений:

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Пользуясь независимостью и одинаковой распределенностью элементов выборки нетрудно показать:

$$E(\overline{X}_n) = E(X_1) \qquad Var(\overline{X}_n) = \frac{Var(X_1)}{n}$$

Пример:
Имеется выборка объема $n = 3$ из экспоненциального распределения с параметром $\lambda = 0.2$. Найдем математическое ожидание, дисперсию и реализацию выборочного среднего, если известно, что $x = (1, 6, 3)$.

$$E(\overline{X}_3) = E(X_1) = 1/0.2 = 5$$

Выборочное среднее

Определение и некоторые свойства

- **Выборочное среднее** является средним значением по выборке $X_1, ..., X_n$:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Реализация выборочного среднего рассчитывается по реализациям наблюдений:

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Пользуясь независимостью и одинаковой распределенностью элементов выборки нетрудно показать:

$$E(\overline{X}_n) = E(X_1) \qquad Var(\overline{X}_n) = \frac{Var(X_1)}{n}$$

Пример:
Имеется выборка объема $n = 3$ из экспоненциального распределения с параметром $\lambda = 0.2$. Найдем математическое ожидание, дисперсию и реализацию выборочного среднего, если известно, что $x = (1, 6, 3)$.

$$E(\overline{X}_3) = E(X_1) = 1/0.2 = 5 \qquad Var(\overline{X}_3) = (1/0.2^2)/3 = 25/3$$

Выборочное среднее

Определение и некоторые свойства

- **Выборочное среднее** является средним значением по выборке $X_1, ..., X_n$:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Реализация выборочного среднего рассчитывается по реализациям наблюдений:

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

- Пользуясь независимостью и одинаковой распределенностью элементов выборки нетрудно показать:

$$E(\overline{X}_n) = E(X_1) \qquad Var(\overline{X}_n) = \frac{Var(X_1)}{n}$$

Пример:
Имеется выборка объема $n = 3$ из экспоненциального распределения с параметром $\lambda = 0.2$. Найдем математическое ожидание, дисперсию и реализацию выборочного среднего, если известно, что $x = (1, 6, 3)$.

$$E(\overline{X}_3) = E(X_1) = 1/0.2 = 5 \qquad Var(\overline{X}_3) = (1/0.2^2)/3 = 25/3 \qquad \overline{x}_3 = (1+6+3)/3 = 10/3$$

Статистическая оценка

Определение

- Любая функция от выборки $T(X_1, ..., X_n)$ именуется **статистикой**.

Статистическая оценка
Определение

- Любая функция от выборки $T(X_1, ..., X_n)$ именуется **статистикой**.
- Пусть $X \sim \Theta(\theta)$, где $\theta \in R$. Статистика $\hat{\theta}(X_1, ..., X_n)$ может рассматриваться в качестве **оценки** параметра θ.

Статистическая оценка

Определение

- Любая функция от выборки $T(X_1,...,X_n)$ именуется **статистикой**.
- Пусть $X \sim \Theta(\theta)$, где $\theta \in R$. Статистика $\hat{\theta}\left(X_1,...,X_n\right)$ может рассматриваться в качестве **оценки** параметра θ.
- Для краткости обозначим $\hat{\theta} = \hat{\theta}\left(X_1,...,X_n\right)$ и $\hat{\theta}(x) = \hat{\theta}\left(x_1,...,x_n\right)$.

Статистическая оценка
Определение

- Любая функция от выборки $T(X_1, ..., X_n)$ именуется **статистикой**.
- Пусть $X \sim \Theta(\theta)$, где $\theta \in R$. Статистика $\hat{\theta}(X_1, ..., X_n)$ может рассматриваться в качестве **оценки** параметра θ.
- Для краткости обозначим $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$ и $\hat{\theta}(x) = \hat{\theta}(x_1, ..., x_n)$.
- **Практический смысл**: реализация оценки $\hat{\theta}(x)$ может рассматриваться в качестве приблизительного значения параметра θ.

Статистическая оценка

Определение

- Любая функция от выборки $T(X_1, ..., X_n)$ именуется **статистикой**.
- Пусть $X \sim \Theta(\theta)$, где $\theta \in R$. Статистика $\hat{\theta}(X_1, ..., X_n)$ может рассматриваться в качестве **оценки** параметра θ.
- Для краткости обозначим $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$ и $\hat{\theta}(x) = \hat{\theta}(x_1, ..., x_n)$.
- **Практический смысл**: реализация оценки $\hat{\theta}(x)$ может рассматриваться в качестве приблизительного значения параметра θ.
  Пример:
- Лаврентий 5 раза подбрасывает монетку, выпадающую орлом с вероятностью p.

Статистическая оценка

Определение

- Любая функция от выборки $T(X_1, ..., X_n)$ именуется **статистикой**.
- Пусть $X \sim \Theta(\theta)$, где $\theta \in R$. Статистика $\hat{\theta}(X_1, ..., X_n)$ может рассматриваться в качестве **оценки** параметра θ.
- Для краткости обозначим $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$ и $\hat{\theta}(x) = \hat{\theta}(x_1, ..., x_n)$.
- **Практический смысл**: реализация оценки $\hat{\theta}(x)$ может рассматриваться в качестве приблизительного значения параметра θ.
  Пример:
- Лаврентий 5 раза подбрасывает монетку, выпадающую орлом с вероятностью p.
- Число орлов, которое выпадет при i-м броске (ноль или один), где $i \in \{1, ..., 5\}$, является случайной величиной $X_i \sim Ber(p)$ с параметром $p \in (0, 1)$.

Статистическая оценка
Определение

- Любая функция от выборки $T(X_1, ..., X_n)$ именуется **статистикой**.
- Пусть $X \sim \Theta(\theta)$, где $\theta \in R$. Статистика $\hat{\theta}(X_1, ..., X_n)$ может рассматриваться в качестве **оценки** параметра θ.
- Для краткости обозначим $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$ и $\hat{\theta}(x) = \hat{\theta}(x_1, ..., x_n)$.
- **Практический смысл**: реализация оценки $\hat{\theta}(x)$ может рассматриваться в качестве приблизительного значения параметра θ.
  Пример:
- Лаврентий 5 раза подбрасывает монетку, выпадающую орлом с вероятностью p.
- Число орлов, которое выпадет при i-м броске (ноль или один), где $i \in \{1, ..., 5\}$, является случайной величиной $X_i \sim Ber(p)$ с параметром $p \in (0, 1)$.
- Броски Лаврентия формируют выборку $X_1, ..., X_5$ объема $n = 5$ из распределения Бернулли $Ber(p)$ с параметром $p \in (0, 1)$.

Статистическая оценка

Определение

- Любая функция от выборки $T(X_1, ..., X_n)$ именуется **статистикой**.
- Пусть $X \sim \Theta(\theta)$, где $\theta \in R$. Статистика $\hat{\theta}(X_1, ..., X_n)$ может рассматриваться в качестве **оценки** параметра θ.
- Для краткости обозначим $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$ и $\hat{\theta}(x) = \hat{\theta}(x_1, ..., x_n)$.
- **Практический смысл**: реализация оценки $\hat{\theta}(x)$ может рассматриваться в качестве приблизительного значения параметра θ.
  Пример:
- Лаврентий 5 раза подбрасывает монетку, выпадающую орлом с вероятностью p.
- Число орлов, которое выпадет при i-м броске (ноль или один), где $i \in \{1, ..., 5\}$, является случайной величиной $X_i \sim Ber(p)$ с параметром $p \in (0, 1)$.
- Броски Лаврентия формируют выборку $X_1, ..., X_5$ объема $n = 5$ из распределения Бернулли $Ber(p)$ с параметром $p \in (0, 1)$.
- Рассмотрим оценку $\hat{p} = \overline{X}_5$ и ее реализацию $\hat{p} = \overline{x}_5$.

- Любая функция от выборки $T(X_1, ..., X_n)$ именуется **статистикой**.
- Пусть $X \sim \Theta(\theta)$, где $\theta \in R$. Статистика $\hat{\theta}(X_1, ..., X_n)$ может рассматриваться в качестве **оценки** параметра θ.
- Для краткости обозначим $\hat{\theta} = \hat{\theta}(X_1, ..., X_n)$ и $\hat{\theta}(x) = \hat{\theta}(x_1, ..., x_n)$.
- **Практический смысл**: реализация оценки $\hat{\theta}(x)$ может рассматриваться в качестве приблизительного значения параметра θ.
  Пример:
- Лаврентий 5 раза подбрасывает монетку, выпадающую орлом с вероятностью p.
- Число орлов, которое выпадет при i-м броске (ноль или один), где $i \in \{1, ..., 5\}$, является случайной величиной $X_i \sim Ber(p)$ с параметром $p \in (0, 1)$.
- Броски Лаврентия формируют выборку $X_1, ..., X_5$ объема $n = 5$ из распределения Бернулли $Ber(p)$ с параметром $p \in (0, 1)$.
- Рассмотрим оценку $\hat{p} = \overline{X}_5$ и ее реализацию $\hat{p} = \overline{x}_5$.
- Пусть монетка 3 раза выпала орлом и 2 раза – решкой, например, $x = (1, 1, 0, 1, 0)$. Тогда $\hat{p}(x) = (1 + 1 + 0 + 1 + 0)/5 = 0.6$.

Статистическая оценка

Определение

- Любая функция от выборки $T(X_1, ..., X_n)$ именуется **статистикой**.
- Пусть $X \sim \Theta(\theta)$, где $\theta \in R$. Статистика $\hat{\theta}\,(X_1, ..., X_n)$ может рассматриваться в качестве **оценки** параметра θ.
- Для краткости обозначим $\hat{\theta} = \hat{\theta}\,(X_1, ..., X_n)$ и $\hat{\theta}(x) = \hat{\theta}\,(x_1, ..., x_n)$.
- **Практический смысл**: реализация оценки $\hat{\theta}\,(x)$ может рассматриваться в качестве приблизительного значения параметра θ.
  Пример:
- Лаврентий 5 раза подбрасывает монетку, выпадающую орлом с вероятностью p.
- Число орлов, которое выпадет при i-м броске (ноль или один), где $i \in \{1, ..., 5\}$, является случайной величиной $X_i \sim Ber(p)$ с параметром $p \in (0, 1)$.
- Броски Лаврентия формируют выборку $X_1, ..., X_5$ объема $n = 5$ из распределения Бернулли $Ber(p)$ с параметром $p \in (0, 1)$.
- Рассмотрим оценку $\hat{p} = \overline{X}_5$ и ее реализацию $\hat{p} = \overline{x}_5$.
- Пусть монетка 3 раза выпала орлом и 2 раза – решкой, например, $x = (1, 1, 0, 1, 0)$. Тогда $\hat{p}(x) = (1 + 1 + 0 + 1 + 0)/5 = 0.6$.
- Если Лаврентий не знает истинной вероятности p, с которой монетка выпадает орлом, то он может предположить, что она приблизительно совпадает с реализацией ее оценки $\hat{p}$, то есть $p \approx \hat{p} = 0.6$.

Свойства статистических оценок

Несмещенность

- Оценка $\hat{\theta}$ параметра θ является **несмещенной**, если при любом допустимом для распределения Θ (из которого была получена выборка) значении параметра θ:

$$E(\hat{\theta}) = \theta$$

Свойства статистических оценок
Несмещенность

- Оценка $\hat{\theta}$ параметра θ является **несмещенной**, если при любом допустимом для распределения Θ (из которого была получена выборка) значении параметра θ:

$$E(\hat{\theta}) = \theta$$

- Разницу $E(\hat{\theta}) - \theta$ часто именуют **величиной смещения** оценки $\hat{\theta}$.

Свойства статистических оценок

Несмещенность

- Оценка $\hat{\theta}$ параметра θ является **несмещенной**, если при любом допустимом для распределения Θ (из которого была получена выборка) значении параметра θ:

$$E(\hat{\theta}) = \theta$$

- Разницу $E(\hat{\theta}) - \theta$ часто именуют **величиной смещения** оценки $\hat{\theta}$.
  Примеры:
- В примере с Лаврентием оценка $\hat{p}$ является несмещенной оценкой параметра p:

$$E(\hat{p}) = E(\overline{X}_5) = p$$

Свойства статистических оценок

Несмещенность

- Оценка $\hat{\theta}$ параметра θ является **несмещенной**, если при любом допустимом для распределения Θ (из которого была получена выборка) значении параметра θ:

$$E(\hat{\theta}) = \theta$$

- Разницу $E(\hat{\theta}) - \theta$ часто именуют **величиной смещения** оценки $\hat{\theta}$.

Примеры:

- В примере с Лаврентием оценка $\hat{p}$ является несмещенной оценкой параметра p:

$$E(\hat{p}) = E(\overline{X}_5) = p$$

Покажем, что другая оценка $\hat{p}^* = X_1 \times ... \times X_5$ окажется смещенной:

$$E(\hat{p}^*) = E(X_1) \times ... \times E(X_5) = p \times p \times p \times p \times p = p^5 \neq p$$

Свойства статистических оценок

Несмещенность

- Оценка $\hat{\theta}$ параметра θ является **несмещенной**, если при любом допустимом для распределения Θ (из которого была получена выборка) значении параметра θ:

$$E(\hat{\theta}) = \theta$$

- Разницу $E(\hat{\theta}) - \theta$ часто именуют **величиной смещения** оценки $\hat{\theta}$.
  Примеры:
- В примере с Лаврентием оценка $\hat{p}$ является несмещенной оценкой параметра p:

$$E(\hat{p}) = E(\overline{X}_5) = p$$

  Покажем, что другая оценка $\hat{p}^* = X_1 \times ... \times X_5$ окажется смещенной:

$$E(\hat{p}^*) = E(X_1) \times ... \times E(X_5) = p \times p \times p \times p \times p = p^5 \neq p$$

- Имеется выборка объема $n = 3$ из распределения Пуассона с параметром λ. Проверьте, являются ли несмещенными оценки $\hat{\lambda}_1 = (X_1 + X_2 + X_3)$, $\hat{\lambda}_2 = (X_1 + X_2 - X_3)$ и $\hat{\lambda}_3 = (X_1X_2 + X_3)$.

- Оценка $\hat{\theta}$ параметра θ является **несмещенной**, если при любом допустимом для распределения Θ (из которого была получена выборка) значении параметра θ:

$$E(\hat{\theta}) = \theta$$

- Разницу $E(\hat{\theta}) - \theta$ часто именуют **величиной смещения** оценки $\hat{\theta}$.

Примеры:

- В примере с Лаврентием оценка $\hat{p}$ является несмещенной оценкой параметра p:

$$E(\hat{p}) = E(\overline{X}_5) = p$$

Покажем, что другая оценка $\hat{p}^* = X_1 \times ... \times X_5$ окажется смещенной:

$$E(\hat{p}^*) = E(X_1) \times ... \times E(X_5) = p \times p \times p \times p \times p = p^5 \neq p$$

- Имеется выборка объема $n = 3$ из распределения Пуассона с параметром λ. Проверьте, являются ли несмещенными оценки $\hat{\lambda}_1 = (X_1 + X_2 + X_3)$, $\hat{\lambda}_2 = (X_1 + X_2 - X_3)$ и $\hat{\lambda}_3 = (X_1 X_2 + X_3)$.

Решение:

$$E(\hat{\lambda}_1) = E(X_1) + E(X_2) + E(X_3) = \lambda + \lambda + \lambda = 3\lambda \neq \lambda \implies \text{ смещенная}$$

Свойства статистических оценок

Несмещенность

- Оценка $\hat{\theta}$ параметра θ является **несмещенной**, если при любом допустимом для распределения Θ (из которого была получена выборка) значении параметра θ:

$$E(\hat{\theta}) = \theta$$

- Разницу $E(\hat{\theta}) - \theta$ часто именуют **величиной смещения** оценки $\hat{\theta}$.
  Примеры:
- В примере с Лаврентием оценка $\hat{p}$ является несмещенной оценкой параметра p:

$$E(\hat{p}) = E(\overline{X}_5) = p$$

  Покажем, что другая оценка $\hat{p}^* = X_1 \times ... \times X_5$ окажется смещенной:

$$E(\hat{p}^*) = E(X_1) \times ... \times E(X_5) = p \times p \times p \times p \times p = p^5 \neq p$$

- Имеется выборка объема $n = 3$ из распределения Пуассона с параметром λ. Проверьте, являются ли несмещенными оценки $\hat{\lambda}_1 = (X_1 + X_2 + X_3)$, $\hat{\lambda}_2 = (X_1 + X_2 - X_3)$ и $\hat{\lambda}_3 = (X_1 X_2 + X_3)$.
  Решение:

$$E(\hat{\lambda}_1) = E(X_1) + E(X_2) + E(X_3) = \lambda + \lambda + \lambda = 3\lambda \neq \lambda \implies \text{ смещенная}$$

$$E(\hat{\lambda}_2) = E(X_1) + E(X_2) - E(X_3) = \lambda + \lambda - \lambda = \lambda \implies \text{ несмещенная}$$

Свойства статистических оценок

Несмещенность

- Оценка $\hat{\theta}$ параметра θ является **несмещенной**, если при любом допустимом для распределения Θ (из которого была получена выборка) значении параметра θ:
$$E(\hat{\theta}) = \theta$$
- Разницу $E(\hat{\theta}) - \theta$ часто именуют **величиной смещения** оценки $\hat{\theta}$.
Примеры:
- В примере с Лаврентием оценка $\hat{p}$ является несмещенной оценкой параметра p:
$$E(\hat{p}) = E(\overline{X}_5) = p$$
Покажем, что другая оценка $\hat{p}^* = X_1 \times ... \times X_5$ окажется смещенной:
$$E(\hat{p}^*) = E(X_1) \times ... \times E(X_5) = p \times p \times p \times p \times p = p^5 \neq p$$
- Имеется выборка объема $n = 3$ из распределения Пуассона с параметром λ. Проверьте, являются ли несмещенными оценки $\hat{\lambda}_1 = (X_1 + X_2 + X_3)$, $\hat{\lambda}_2 = (X_1 + X_2 - X_3)$ и $\hat{\lambda}_3 = (X_1X_2 + X_3)$.
Решение:
$$E(\hat{\lambda}_1) = E(X_1) + E(X_2) + E(X_3) = \lambda + \lambda + \lambda = 3\lambda \neq \lambda \implies \text{смещенная}$$
$$E(\hat{\lambda}_2) = E(X_1) + E(X_2) - E(X_3) = \lambda + \lambda - \lambda = \lambda \implies \text{несмещенная}$$
$$E(\hat{\lambda}_3) = E(X_1)E(X_2) + E(X_3) = \lambda^2 + \lambda \neq \lambda \implies \text{смещенная}$$

- Рассмотрим бесконечную последовательность оценок $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), ...$, где n-я оценка $\hat{\theta}_n(X_1, ..., X_n)$ получена по выборке $X_1, ..., X_n$.

Асимптотическая несмещенность

- Рассмотрим бесконечную последовательность оценок $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), ...$, где n-я оценка $\hat{\theta}_n(X_1, ..., X_n)$ получена по выборке $X_1, ..., X_n$.
- Данная последовательность оценок именуется **асимптотически несмещенной**, если:

$$\lim_{n \to \infty} E(\hat{\theta}_n) = \theta$$

- Рассмотрим бесконечную последовательность оценок $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), ...$, где n-я оценка $\hat{\theta}_n(X_1, ..., X_n)$ получена по выборке $X_1, ..., X_n$.
- Данная последовательность оценок именуется **асимптотически несмещенной**, если:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

- Последовательность несмещенных оценок всегда будет асимптотически несмещенной, а обратное – не всегда верно.

- Рассмотрим бесконечную последовательность оценок $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), ...,$ где n-я оценка $\hat{\theta}_n(X_1, ..., X_n)$ получена по выборке $X_1, ..., X_n$.
- Данная последовательность оценок именуется **асимптотически несмещенной**, если:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

- Последовательность несмещенных оценок всегда будет асимптотически несмещенной, а обратное – не всегда верно.
  Пример:
  Вы формируете выборку, записывая число посетителей магазина в выходные дни, которые подчиняется распределению Пуассона с параметром λ. Определите, будет ли ваша последовательность оценок $\hat{\lambda}_n = \frac{n}{n+1}\overline{X}_n$ асимптотически несмещенной, а также будут ли оценки этой последовательности несмещенными.

Свойства статистических оценок

Асимптотическая несмещенность

- Рассмотрим бесконечную последовательность оценок $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), \ldots$, где n-я оценка $\hat{\theta}_n(X_1, \ldots, X_n)$ получена по выборке $X_1, \ldots, X_n$.
- Данная последовательность оценок именуется **асимптотически несмещенной**, если:

$$\lim_{n \to \infty} E(\hat{\theta}_n) = \theta$$

- Последовательность несмещенных оценок всегда будет асимптотически несмещенной, а обратное – не всегда верно.
  Пример:
  Вы формируете выборку, записывая число посетителей магазина в выходные дни, которые подчиняется распределению Пуассона с параметром λ. Определите, будет ли ваша последовательность оценок $\hat{\lambda}_n = \frac{n}{n+1}\overline{X}_n$ асимптотически несмещенной, а также будут ли оценки этой последовательности несмещенными.
  Решение: асимптотическая несмещенность соблюдается, поскольку:

$$\lim_{n \to \infty} E(\hat{\lambda}_n) = \lim_{n \to \infty} E\left(\frac{n}{n+1}\overline{X}_n\right) = \lim_{n \to \infty} \frac{n}{n+1} E\left(\overline{X}_n\right) = 1 \times \lambda = \lambda$$

- Рассмотрим бесконечную последовательность оценок $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), ...$, где n-я оценка $\hat{\theta}_n(X_1, ..., X_n)$ получена по выборке $X_1, ..., X_n$.
- Данная последовательность оценок именуется **асимптотически несмещенной**, если:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

- Последовательность несмещенных оценок всегда будет асимптотически несмещенной, а обратное – не всегда верно.
  Пример:
  Вы формируете выборку, записывая число посетителей магазина в выходные дни, которые подчиняется распределению Пуассона с параметром λ. Определите, будет ли ваша последовательность оценок $\hat{\lambda}_n = \frac{n}{n+1}\overline{X}_n$ асимптотически несмещенной, а также будут ли оценки этой последовательности несмещенными.
  Решение: асимптотическая несмещенность соблюдается, поскольку:

$$\lim_{n\to\infty} E(\hat{\lambda}_n) = \lim_{n\to\infty} E\left(\frac{n}{n+1}\overline{X}_n\right) = \lim_{n\to\infty} \frac{n}{n+1} E\left(\overline{X}_n\right) = 1 \times \lambda = \lambda$$

  При этом оценки данной последовательности являются смещенными:

$$E(\hat{\lambda}_n) = \frac{n}{n+1}\lambda \neq \lambda$$

- Рассмотрим бесконечную последовательность оценок $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), ...$, где n-я оценка $\hat{\theta}_n(X_1, ..., X_n)$ получена по выборке $X_1, ..., X_n$.

- Рассмотрим бесконечную последовательность оценок $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), ...$, где n-я оценка $\hat{\theta}_n(X_1, ..., X_n)$ получена по выборке $X_1, ..., X_n$.
- Данная последовательность оценок именуется **состоятельной**, если $\hat{\theta}_n \xrightarrow{p} \theta$. Напомним, что для этого достаточно выполнения следующих условий:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$
$$\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$$

- Рассмотрим бесконечную последовательность оценок $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), ...$, где n-я оценка $\hat{\theta}_n(X_1, ..., X_n)$ получена по выборке $X_1, ..., X_n$.
- Данная последовательность оценок именуется **состоятельной**, если $\hat{\theta}_n \xrightarrow{p} \theta$. Напомним, что для этого достаточно выполнения следующих условий:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$
$$\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$$

Пример:

- Лаврентий бесконечное число раз подбрасывает монетку. Проверьте, будут ли состоятельными последовательности оценок $\hat{p}_n = \overline{X}_n$ и $\hat{p}_n^* = X_1 \times ... \times X_n$.

Свойства статистических оценок

Состоятельность

- Рассмотрим бесконечную последовательность оценок $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), ...$, где n-я оценка $\hat{\theta}_n(X_1, ..., X_n)$ получена по выборке $X_1, ..., X_n$.
- Данная последовательность оценок именуется **состоятельной**, если $\hat{\theta}_n \xrightarrow{p} \theta$. Напомним, что для этого достаточно выполнения следующих условий:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$
$$\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$$

Пример:

- Лаврентий бесконечное число раз подбрасывает монетку. Проверьте, будут ли состоятельными последовательности оценок $\hat{p}_n = \overline{X}_n$ и $\hat{p}_n^* = X_1 \times ... \times X_n$.

Решение: первая последовательность оценок состоятельная, поскольку соблюдены оба условия:

$$\lim_{n\to\infty} E(\hat{p}_n) = \lim_{n\to\infty} E(X_n) = \lim_{n\to\infty} p = p$$

- Рассмотрим бесконечную последовательность оценок $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), ...,$ где n-я оценка $\hat{\theta}_n(X_1, ..., X_n)$ получена по выборке $X_1, ..., X_n$.
- Данная последовательность оценок именуется **состоятельной**, если $\hat{\theta}_n \xrightarrow{p} \theta$. Напомним, что для этого достаточно выполнения следующих условий:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$

$$\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$$

Пример:

- Лаврентий бесконечное число раз подбрасывает монетку. Проверьте, будут ли состоятельными последовательности оценок $\hat{p}_n = \overline{X}_n$ и $\hat{p}_n^* = X_1 \times ... \times X_n$.

Решение: первая последовательность оценок состоятельная, поскольку соблюдены оба условия:

$$\lim_{n\to\infty} E(\hat{p}_n) = \lim_{n\to\infty} E(X_n) = \lim_{n\to\infty} p = p$$

$$\lim_{n\to\infty} Var(\hat{p}_n) = \lim_{n\to\infty} Var(\overline{X}_n) = \lim_{n\to\infty} Var(X_1)/n = 0$$

- Рассмотрим бесконечную последовательность оценок $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), ...$, где n-я оценка $\hat{\theta}_n(X_1, ..., X_n)$ получена по выборке $X_1, ..., X_n$.
- Данная последовательность оценок именуется **состоятельной**, если $\hat{\theta}_n \xrightarrow{p} \theta$. Напомним, что для этого достаточно выполнения следующих условий:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta$$
$$\lim_{n\to\infty} Var(\hat{\theta}_n) = 0$$

Пример:

- Лаврентий бесконечное число раз подбрасывает монетку. Проверьте, будут ли состоятельными последовательности оценок $\hat{p}_n = \overline{X}_n$ и $\hat{p}_n^* = X_1 \times ... \times X_n$.

Решение: первая последовательность оценок состоятельная, поскольку соблюдены оба условия:

$$\lim_{n\to\infty} E(\hat{p}_n) = \lim_{n\to\infty} E(X_n) = \lim_{n\to\infty} p = p$$
$$\lim_{n\to\infty} Var(\hat{p}_n) = \lim_{n\to\infty} Var(\overline{X}_n) = \lim_{n\to\infty} Var(X_1)/n = 0$$

Вторая последовательность оценок несостоятельная, так как не соблюдено одно из условий:

$$\lim_{n\to\infty} E(\hat{p}_n^*) = \lim_{n\to\infty} E(X_1) \times ... \times E(X_n) = \lim_{n\to\infty} p^n = 0 \neq p$$

- **Эффективность** оценок отражает их точность в соответствии с критерием ожидаемого среднеквадратического отклонения (MSE) от истинного значения параметра:

$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = Var(\hat{\theta}) + E(\hat{\theta} - \theta)^2$$

- **Эффективность** оценок отражает их точность в соответствии с критерием ожидаемого среднеквадратического отклонения (MSE) от истинного значения параметра:

$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = Var(\hat{\theta}) + E(\hat{\theta} - \theta)^2$$

Чем меньше $MSE(\hat{\theta})$, тем выше эффективность оценки $\hat{\theta}$.

- **Эффективность** оценок отражает их точность в соответствии с критерием ожидаемого среднеквадратического отклонения (MSE) от истинного значения параметра:
$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = Var(\hat{\theta}) + E(\hat{\theta} - \theta)^2$$
Чем меньше $MSE(\hat{\theta})$, тем выше эффективность оценки $\hat{\theta}$.
- Если оценка $\hat{\theta}$ несмещенная, то $MSE(\hat{\theta}) = Var(\hat{\theta})$.

- **Эффективность** оценок отражает их точность в соответствии с критерием ожидаемого среднеквадратического отклонения (MSE) от истинного значения параметра:

$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = Var(\hat{\theta}) + E(\hat{\theta} - \theta)^2$$

  Чем меньше $MSE(\hat{\theta})$, тем выше эффективность оценки $\hat{\theta}$.
- Если оценка $\hat{\theta}$ несмещенная, то $MSE(\hat{\theta}) = Var(\hat{\theta})$.
- Рассмотрим множество оценок $\mathcal{K}$, часто именуемых **классом**. Оценка $\hat{\theta}$ именуется **эффективной**, если она обладает наибольшей эффективностью (наименьшим MSE) среди всех оценок при любом допустимом θ:

$$MSE(\hat{\theta}) \leq MSE(\hat{\theta}^*), \forall \hat{\theta}^* \in \mathcal{K}$$

- **Эффективность** оценок отражает их точность в соответствии с критерием ожидаемого среднеквадратического отклонения (MSE) от истинного значения параметра:

$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = Var(\hat{\theta}) + E(\hat{\theta} - \theta)^2$$

  Чем меньше $MSE(\hat{\theta})$, тем выше эффективность оценки $\hat{\theta}$.
- Если оценка $\hat{\theta}$ несмещенная, то $MSE(\hat{\theta}) = Var(\hat{\theta})$.
- Рассмотрим множество оценок $\mathcal{K}$, часто именуемых **классом**. Оценка $\hat{\theta}$ именуется **эффективной**, если она обладает наибольшей эффективностью (наименьшим MSE) среди всех оценок при любом допустимом θ:

$$MSE(\hat{\theta}) \leq MSE(\hat{\theta}^*), \forall \hat{\theta}^* \in \mathcal{K}$$

  Пример:
- В примере с Лаврентием рассмотрим оценки $\hat{p}_1 = (X_1 + X_2)/2$, $\hat{p}_2 = (2X_1 - X_2)$ и $\hat{p}_3 = 3X_1$, формирующие класс $\mathcal{K} = \{\hat{p}_1, \hat{p}_2, \hat{p}_3\}$. Найдите эффективную оценку.

- **Эффективность** оценок отражает их точность в соответствии с критерием ожидаемого среднеквадратического отклонения (MSE) от истинного значения параметра:
$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = Var(\hat{\theta}) + E(\hat{\theta} - \theta)^2$$
Чем меньше $MSE(\hat{\theta})$, тем выше эффективность оценки $\hat{\theta}$.
- Если оценка $\hat{\theta}$ несмещенная, то $MSE(\hat{\theta}) = Var(\hat{\theta})$.
- Рассмотрим множество оценок $\mathcal{K}$, часто именуемых **классом**. Оценка $\hat{\theta}$ именуется **эффективной**, если она обладает наибольшей эффективностью (наименьшим MSE) среди всех оценок при любом допустимом θ:
$$MSE(\hat{\theta}) \leq MSE(\hat{\theta}^*), \forall \hat{\theta}^* \in \mathcal{K}$$
Пример:
- В примере с Лаврентием рассмотрим оценки $\hat{p}_1 = (X_1 + X_2)/2$, $\hat{p}_2 = (2X_1 - X_2)$ и $\hat{p}_3 = 3X_1$, формирующие класс $\mathcal{K} = \{\hat{p}_1, \hat{p}_2, \hat{p}_3\}$. Найдите эффективную оценку.
Решение: для краткости воспользуемся несмещенностью первых двух оценок:
$$MSE(\hat{p}_1) = Var(\hat{p}) = (Var(X_1) + Var(X_2))/4 = 0.5p(1 - p)$$

- **Эффективность** оценок отражает их точность в соответствии с критерием ожидаемого среднеквадратического отклонения (MSE) от истинного значения параметра:
$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = Var(\hat{\theta}) + E(\hat{\theta} - \theta)^2$$
Чем меньше $MSE(\hat{\theta})$, тем выше эффективность оценки $\hat{\theta}$.
- Если оценка $\hat{\theta}$ несмещенная, то $MSE(\hat{\theta}) = Var(\hat{\theta})$.
- Рассмотрим множество оценок $\mathcal{K}$, часто именуемых **классом**. Оценка $\hat{\theta}$ именуется **эффективной**, если она обладает наибольшей эффективностью (наименьшим MSE) среди всех оценок при любом допустимом θ:
$$MSE(\hat{\theta}) \leq MSE(\hat{\theta}^*), \forall \hat{\theta}^* \in \mathcal{K}$$
Пример:
- В примере с Лаврентием рассмотрим оценки $\hat{p}_1 = (X_1 + X_2)/2$, $\hat{p}_2 = (2X_1 - X_2)$ и $\hat{p}_3 = 3X_1$, формирующие класс $\mathcal{K} = \{\hat{p}_1, \hat{p}_2, \hat{p}_3\}$. Найдите эффективную оценку.
Решение: для краткости воспользуемся несмещенностью первых двух оценок:
$$MSE(\hat{p}_1) = Var(\hat{p}) = (Var(X_1) + Var(X_2))/4 = 0.5p(1-p)$$
$$MSE(\hat{p}_2) = Var(\hat{p}_2) = 4Var(X_1) + Var(X_2) = 5p(1-p)$$

- **Эффективность** оценок отражает их точность в соответствии с критерием ожидаемого среднеквадратического отклонения (MSE) от истинного значения параметра:

$$MSE(\hat{\theta}) = E((\hat{\theta}-\theta)^2) = Var(\hat{\theta}) + E(\hat{\theta}-\theta)^2$$

Чем меньше $MSE(\hat{\theta})$, тем выше эффективность оценки $\hat{\theta}$.

- Если оценка $\hat{\theta}$ несмещенная, то $MSE(\hat{\theta}) = Var(\hat{\theta})$.
- Рассмотрим множество оценок $\mathcal{K}$, часто именуемых **классом**. Оценка $\hat{\theta}$ именуется **эффективной**, если она обладает наибольшей эффективностью (наименьшим MSE) среди всех оценок при любом допустимом θ:

$$MSE(\hat{\theta}) \leq MSE(\hat{\theta}^*), \forall \hat{\theta}^* \in \mathcal{K}$$

Пример:

- В примере с Лаврентием рассмотрим оценки $\hat{p}_1 = (X_1 + X_2)/2$, $\hat{p}_2 = (2X_1 - X_2)$ и $\hat{p}_3 = 3X_1$, формирующие класс $\mathcal{K} = \{\hat{p}_1, \hat{p}_2, \hat{p}_3\}$. Найдите эффективную оценку.

Решение: для краткости воспользуемся несмещенностью первых двух оценок:

$$MSE(\hat{p}_1) = Var(\hat{p}) = (Var(X_1) + Var(X_2))/4 = 0.5p(1-p)$$

$$MSE(\hat{p}_2) = Var(\hat{p}_2) = 4Var(X_1) + Var(X_2) = 5p(1-p)$$

$$MSE(\hat{p}_3) = Var(3X_1) + E(3X_1 - p)^2 = 9p(1-p) + (3p-p)^2 = 5p(1-p) + 4p$$

- **Эффективность** оценок отражает их точность в соответствии с критерием ожидаемого среднеквадратического отклонения (MSE) от истинного значения параметра:

$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = Var(\hat{\theta}) + E(\hat{\theta} - \theta)^2$$

  Чем меньше $MSE(\hat{\theta})$, тем выше эффективность оценки $\hat{\theta}$.
- Если оценка $\hat{\theta}$ несмещенная, то $MSE(\hat{\theta}) = Var(\hat{\theta})$.
- Рассмотрим множество оценок $\mathcal{K}$, часто именуемых **классом**. Оценка $\hat{\theta}$ именуется **эффективной**, если она обладает наибольшей эффективностью (наименьшим MSE) среди всех оценок при любом допустимом θ:

$$MSE(\hat{\theta}) \leq MSE(\hat{\theta}^*), \forall \hat{\theta}^* \in \mathcal{K}$$

Пример:

- В примере с Лаврентием рассмотрим оценки $\hat{p}_1 = (X_1 + X_2)/2$, $\hat{p}_2 = (2X_1 - X_2)$ и $\hat{p}_3 = 3X_1$, формирующие класс $\mathcal{K} = \{\hat{p}_1, \hat{p}_2, \hat{p}_3\}$. Найдите эффективную оценку.
  Решение: для краткости воспользуемся несмещенностью первых двух оценок:

$$MSE(\hat{p}_1) = Var(\hat{p}) = (Var(X_1) + Var(X_2))/4 = 0.5p(1-p)$$
$$MSE(\hat{p}_2) = Var(\hat{p}_2) = 4Var(X_1) + Var(X_2) = 5p(1-p)$$
$$MSE(\hat{p}_3) = Var(3X_1) + E(3X_1 - p)^2 = 9p(1-p) + (3p-p)^2 = 5p(1-p) + 4p$$

  Поскольку $0.5p(1-p) \leq 5p(1-p) \leq 5p(1-p) + 4p$ при любом допустимом значении параметра p, то есть при $p \in (0,1)$, то оценка $\hat{p}_1$ является эффективной в классе $\mathcal{K}$.

- Рассмотрим оценку $\hat{g}(\theta)$ функции $g(\theta)$ от параметра θ. Ее свойства определяются по аналогии с рассмотренными ранее свойствами оценки параметра.

Статистическая оценка функции от параметра

Определение и свойства

- Рассмотрим оценку $\hat{g}(\theta)$ функции $g(\theta)$ от параметра θ. Ее свойства определяются по аналогии с рассмотренными ранее свойствами оценки параметра.
- Обычно в качестве функций от параметров рассматривают различные характеристики распределений, такие как математическое ожидание, дисперсия, мода, медиана, квантили, вероятности и т.д.

Статистическая оценка функции от параметра

Определение и свойства

- Рассмотрим оценку $\hat{g}(\theta)$ функции $g(\theta)$ от параметра θ. Ее свойства определяются по аналогии с рассмотренными ранее свойствами оценки параметра.
- Обычно в качестве функций от параметров рассматривают различные характеристики распределений, такие как математическое ожидание, дисперсия, мода, медиана, квантили, вероятности и т.д.
- Если бесконечная последовательность оценок $\hat{\theta}_1, \hat{\theta}_2, \ldots$ является состоятельной для параметра $g(\theta)$, то, по теореме Манна-Вальда, последовательность непрерывных функций от этих оценок $g(\hat{\theta})_1, g(\hat{\theta})_2, \ldots$ будет состоятельной для функции от параметра $g(\theta)$. То есть для непрерывной функции $g(\theta)$ из $\hat{\theta}_n \xrightarrow{p} \theta$ следует $g(\hat{\theta})_n \xrightarrow{p} g(\theta)$.

Статистическая оценка функции от параметра

Определение и свойства

- Рассмотрим оценку $\hat{g}(\theta)$ функции $g(\theta)$ от параметра θ. Ее свойства определяются по аналогии с рассмотренными ранее свойствами оценки параметра.
- Обычно в качестве функций от параметров рассматривают различные характеристики распределений, такие как математическое ожидание, дисперсия, мода, медиана, квантили, вероятности и т.д.
- Если бесконечная последовательность оценок $\hat{\theta}_1, \hat{\theta}_2, ...$ является состоятельной для параметра $g(\theta)$, то, по теореме Манна-Вальда, последовательность непрерывных функций от этих оценок $g(\hat{\theta})_1, g(\hat{\theta})_2, ...$ будет состоятельной для функции от параметра $g(\theta)$. То есть для непрерывной функции $g(\theta)$ из $\hat{\theta}_n \xrightarrow{p} \theta$ следует $g(\hat{\theta})_n \xrightarrow{p} g(\theta)$.

  Пример:
- Вернемся к примеру с Лаврентием и найдем состоятельную оценку дисперсии числа орлов, выпадающих при одном броске, то есть оценку $\widehat{Var}(X_i)$ дисперсии $Var(X_i)$.

Статистическая оценка функции от параметра

Определение и свойства

- Рассмотрим оценку $\hat{g}(\theta)$ функции $g(\theta)$ от параметра θ. Ее свойства определяются по аналогии с рассмотренными ранее свойствами оценки параметра.
- Обычно в качестве функций от параметров рассматривают различные характеристики распределений, такие как математическое ожидание, дисперсия, мода, медиана, квантили, вероятности и т.д.
- Если бесконечная последовательность оценок $\hat{\theta}_1, \hat{\theta}_2, \ldots$ является состоятельной для параметра $g(\theta)$, то, по теореме Манна-Вальда, последовательность непрерывных функций от этих оценок $g(\hat{\theta})_1, g(\hat{\theta})_2, \ldots$ будет состоятельной для функции от параметра $g(\theta)$. То есть для непрерывной функции $g(\theta)$ из $\hat{\theta}_n \xrightarrow{p} \theta$ следует $g(\hat{\theta})_n \xrightarrow{p} g(\theta)$.

  Пример:
- Вернемся к примеру с Лаврентием и найдем состоятельную оценку дисперсии числа орлов, выпадающих при одном броске, то есть оценку $\widehat{Var}(X_i)$ дисперсии $Var(X_i)$.
- Поскольку $X_i \sim Ber(p)$, то $g(p) = Var(X_i) = p(1-p)$.

- Рассмотрим оценку $\hat{g}(\theta)$ функции $g(\theta)$ от параметра θ. Ее свойства определяются по аналогии с рассмотренными ранее свойствами оценки параметра.
- Обычно в качестве функций от параметров рассматривают различные характеристики распределений, такие как математическое ожидание, дисперсия, мода, медиана, квантили, вероятности и т.д.
- Если бесконечная последовательность оценок $\hat{\theta}_1, \hat{\theta}_2, ...$ является состоятельной для параметра $g(\theta)$, то, по теореме Манна-Вальда, последовательность непрерывных функций от этих оценок $g(\hat{\theta})_1, g(\hat{\theta})_2, ...$ будет состоятельной для функции от параметра $g(\theta)$. То есть для непрерывной функции $g(\theta)$ из $\hat{\theta}_n \xrightarrow{p} \theta$ следует $g(\hat{\theta})_n \xrightarrow{p} g(\theta)$.

  Пример:
- Вернемся к примеру с Лаврентием и найдем состоятельную оценку дисперсии числа орлов, выпадающих при одном броске, то есть оценку $\widehat{Var}(X_i)$ дисперсии $Var(X_i)$.
- Поскольку $X_i \sim Ber(p)$, то $g(p) = Var(X_i) = p(1-p)$.
- Так как $\hat{p} = \overline{X}_n$ является состоятельной оценкой для параметра p и функция $g(p) = p(1-p)$ непрерывна, то по теореме Манна-Вальда состоятельная оценка этой функции, то есть дисперсии $Var(X_i)$, будет иметь вид:

$$\widehat{Var}(X_i) = g(\hat{p}) = \hat{p}(1-\hat{p}) = \overline{X}_n(1-\overline{X}_n)$$

- Рассмотрим оценку $\hat{g}(\theta)$ функции $g(\theta)$ от параметра θ. Ее свойства определяются по аналогии с рассмотренными ранее свойствами оценки параметра.
- Обычно в качестве функций от параметров рассматривают различные характеристики распределений, такие как математическое ожидание, дисперсия, мода, медиана, квантили, вероятности и т.д.
- Если бесконечная последовательность оценок $\hat{\theta}_1, \hat{\theta}_2, ...$ является состоятельной для параметра $g(\theta)$, то, по теореме Манна-Вальда, последовательность непрерывных функций от этих оценок $g(\hat{\theta})_1, g(\hat{\theta})_2, ...$ будет состоятельной для функции от параметра $g(\theta)$. То есть для непрерывной функции $g(\theta)$ из $\hat{\theta}_n \xrightarrow{p} \theta$ следует $g(\hat{\theta})_n \xrightarrow{p} g(\theta)$.
  Пример:
- Вернемся к примеру с Лаврентием и найдем состоятельную оценку дисперсии числа орлов, выпадающих при одном броске, то есть оценку $\widehat{Var}(X_i)$ дисперсии $Var(X_i)$.
- Поскольку $X_i \sim Ber(p)$, то $g(p) = Var(X_i) = p(1-p)$.
- Так как $\hat{p} = \overline{X}_n$ является состоятельной оценкой для параметра p и функция $g(p) = p(1-p)$ непрерывна, то по теореме Манна-Вальда состоятельная оценка этой функции, то есть дисперсии $Var(X_i)$, будет иметь вид:

$$\widehat{Var}(X_i) = g(\hat{p}) = \hat{p}(1-\hat{p}) = \overline{X}_n(1-\overline{X}_n)$$

- Полученная оценка дисперсии является смещенной, поскольку:

$$E(\widehat{Var}(X_i)) = E(\overline{X}_n(1-\overline{X}_n)) = E(\overline{X}_n) - E(\overline{X}_n^2) = E(\overline{X}_n) - Var(\overline{X}_n) - E(\overline{X}_n)^2 = p - \frac{p}{n} - p^2 \neq p(1-p)$$

Дополнительный пример

Доходы населения

- Доход случайно взятого индивида является случайной величиной с функцией плотности:

$$f_{X_i}(x) = \begin{cases} 2x/\theta^2, \text{ при } x \in [0, \theta] \\ 0, \text{ в противном случае} \end{cases}, \text{ где } \theta > 0$$

Из доходов случайно взятых индивидов была сформирована выборка $X_1, ..., X_n$.

Дополнительный пример

Доходы населения

- Доход случайно взятого индивида является случайной величиной с функцией плотности:

$$f_{X_i}(x) = \begin{cases} 2x/\theta^2, \text{ при } x \in [0, \theta] \\ 0, \text{ в противном случае} \end{cases}, \text{ где } \theta > 0$$

Из доходов случайно взятых индивидов была сформирована выборка $X_1, ..., X_n$.

- Покажем, что $\hat{\theta}_n = 1.5\overline{X}_n$ является несмещенной оценкой параметра θ:

$$E(\hat{\theta}_n) = E(1.5\overline{X}_n) = 1.5E(\overline{X}_n) = 1.5E(X_i) = 1.5\int_0^{\theta}(2x/\theta^2)x dx = \theta$$

Дополнительный пример

Доходы населения

- Доход случайно взятого индивида является случайной величиной с функцией плотности:

$$f_{X_i}(x) = \begin{cases} 2x/\theta^2, \text{ при } x \in [0, \theta] \\ 0, \text{ в противном случае} \end{cases}, \text{ где } \theta > 0$$

Из доходов случайно взятых индивидов была сформирована выборка $X_1, ..., X_n$.

- Покажем, что $\hat{\theta}_n = 1.5\overline{X}_n$ является несмещенной оценкой параметра θ:

$$E(\hat{\theta}_n) = E(1.5\overline{X}_n) = 1.5E(\overline{X}_n) = 1.5E(X_i) = 1.5\int_0^{\theta}(2x/\theta^2)xdx = \theta$$

- Бесконечная последовательность оценок $\hat{\theta}_1, \hat{\theta}_2, ...$ является состоятельной, так как:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \lim_{n\to\infty} \theta = \theta$$

Дополнительный пример

Доходы населения

- Доход случайно взятого индивида является случайной величиной с функцией плотности:

$$f_{X_i}(x) = \begin{cases} 2x/\theta^2, \text{ при } x \in [0, \theta] \\ 0, \text{ в противном случае} \end{cases}, \text{ где } \theta > 0$$

Из доходов случайно взятых индивидов была сформирована выборка $X_1, ..., X_n$.

- Покажем, что $\hat{\theta}_n = 1.5\overline{X}_n$ является несмещенной оценкой параметра θ:

$$E(\hat{\theta}_n) = E(1.5\overline{X}_n) = 1.5E(\overline{X}_n) = 1.5E(X_i) = 1.5\int_0^{\theta}(2x/\theta^2)xdx = \theta$$

- Бесконечная последовательность оценок $\hat{\theta}_1, \hat{\theta}_2, ...$ является состоятельной, так как:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \lim_{n\to\infty}\theta = \theta$$

$$\lim_{n\to\infty} Var(\hat{\theta}_n) = \lim_{n\to\infty} Var(1.5\overline{X}) = \lim_{n\to\infty}\frac{1.5^2}{n}Var(X_i) = 0$$

Дополнительный пример

Доходы населения

- Доход случайно взятого индивида является случайной величиной с функцией плотности:

$$f_{X_i}(x) = \begin{cases} 2x/\theta^2, \text{ при } x \in [0, \theta] \\ 0, \text{ в противном случае} \end{cases}, \text{ где } \theta > 0$$

Из доходов случайно взятых индивидов была сформирована выборка $X_1, ..., X_n$.

- Покажем, что $\hat{\theta}_n = 1.5\overline{X}_n$ является несмещенной оценкой параметра θ:

$$E(\hat{\theta}_n) = E(1.5\overline{X}_n) = 1.5E(\overline{X}_n) = 1.5E(X_i) = 1.5\int_0^{\theta}(2x/\theta^2)xdx = \theta$$

- Бесконечная последовательность оценок $\hat{\theta}_1, \hat{\theta}_2, ...$ является состоятельной, так как:

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \lim_{n\to\infty} \theta = \theta$$

$$\lim_{n\to\infty} Var(\hat{\theta}_n) = \lim_{n\to\infty} Var(1.5\overline{X}) = \lim_{n\to\infty} \frac{1.5^2}{n} Var(X_i) = 0$$

- Нетрудно показать, что медиана X_i равна $m = \theta/\sqrt{2}$. Поскольку речь идет о непрерывной функции от параметра, то по теореме Манна-Вальда состоятельная последовательность оценок медианы заработков случайно взятого индивида будет иметь вид:

$$\hat{m}_n = \hat{\theta}_n/\sqrt{2} = 1.5\overline{X}_n/\sqrt{2}$$

.