

Теория Вероятностей и Статистика

Оценивание характеристик распределения

Потанин Богдан Станиславович

старший преподаватель, кандидат экономических наук

2021

- Часто, сделать заранее достаточно точное предположение о семействе распределений, из которого была получена выборка, весьма затруднительно.

- Часто, сделать заранее достаточно точное предположение о семействе распределений, из которого была получена выборка, весьма затруднительно.
- Тем не менее, некоторые отдельные характеристики (вероятности, медиана, математическое ожидание и т.д.) можно оценить, не накладывая существенных допущений о распределении и даже не учитывая его параметры.

- Часто, сделать заранее достаточно точное предположение о семействе распределений, из которого была получена выборка, весьма затруднительно.
- Тем не менее, некоторые отдельные характеристики (вероятности, медиана, математическое ожидание и т.д.) можно оценить, не накладывая существенных допущений о распределении и даже не учитывая его параметры.
- Рассмотрим методы, позволяющие оценивать различные характеристики распределения без предварительного оценивания параметров и без предположений о конкретном семействе распределений.

Оценивание вероятностей

Безусловные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.

Оценивание вероятностей

Безусловные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.
- Необходимо оценить вероятность $P(X_1 \in A)$, где $A \subset R$.

Оценивание вероятностей

Безусловные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.
- Необходимо оценить вероятность $P(X_1 \in A)$, где $A \subset R$.
- Несмещенная и состоятельная оценка данной вероятности может быть получена как доля наблюдений в выборке, принадлежащих множеству A :

$$\hat{P}(X_1 \in A) = \frac{1}{n} \sum_{i=1}^n I(X_i \in A) \sim \frac{1}{n} B(n, P(X_1 \in A))$$

$$\text{где: } I(X_i \in A) = \begin{cases} 1, & \text{если } X_i \in A \\ 0, & \text{в противном случае} \end{cases} \sim \text{Ber}(P(X_1 \in A))$$

Оценивание вероятностей

Безусловные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.
- Необходимо оценить вероятность $P(X_1 \in A)$, где $A \subset R$.
- Несмещенная и состоятельная оценка данной вероятности может быть получена как доля наблюдений в выборке, принадлежащих множеству A :

$$\hat{P}(X_1 \in A) = \frac{1}{n} \sum_{i=1}^n I(X_i \in A) \sim \frac{1}{n} B(n, P(X_1 \in A))$$

$$\text{где: } I(X_i \in A) = \begin{cases} 1, & \text{если } X_i \in A \\ 0, & \text{в противном случае} \end{cases} \sim \text{Ber}(P(X_1 \in A))$$

- Подставляя вместо наблюдений X_i их реализации x_i можно получить реализацию оценки соответствующей вероятности.

Оценивание вероятностей

Безусловные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.
- Необходимо оценить вероятность $P(X_1 \in A)$, где $A \subset R$.
- Несмещенная и состоятельная оценка данной вероятности может быть получена как доля наблюдений в выборке, принадлежащих множеству A :

$$\hat{P}(X_1 \in A) = \frac{1}{n} \sum_{i=1}^n I(X_i \in A) \sim \frac{1}{n} B(n, P(X_1 \in A))$$

$$\text{где: } I(X_i \in A) = \begin{cases} 1, & \text{если } X_i \in A \\ 0, & \text{в противном случае} \end{cases} \sim \text{Ber}(P(X_1 \in A))$$

- Подставляя вместо наблюдений X_i их реализации x_i можно получить реализацию оценки соответствующей вероятности.

Доказательство: докажем несмещенность и состоятельность рассматриваемой оценки вероятности:

$$E\left(\hat{P}(X_1 \in A)\right) = \frac{1}{n} \sum_{i=1}^n E(I(X_i \in A)) = \frac{1}{n} \sum_{i=1}^n P(X_i \in A) = \frac{1}{n} \times nP(X_1 \in A) = P(X_1 \in A)$$

Оценивание вероятностей

Безусловные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.
- Необходимо оценить вероятность $P(X_1 \in A)$, где $A \subset R$.
- Несмещенная и состоятельная оценка данной вероятности может быть получена как доля наблюдений в выборке, принадлежащих множеству A :

$$\hat{P}(X_1 \in A) = \frac{1}{n} \sum_{i=1}^n I(X_i \in A) \sim \frac{1}{n} B(n, P(X_1 \in A))$$

$$\text{где: } I(X_i \in A) = \begin{cases} 1, & \text{если } X_i \in A \\ 0, & \text{в противном случае} \end{cases} \sim \text{Ber}(P(X_1 \in A))$$

- Подставляя вместо наблюдений X_i их реализации x_i можно получить реализацию оценки соответствующей вероятности.

Доказательство: докажем несмещенность и состоятельность рассматриваемой оценки вероятности:

$$E(\hat{P}(X_1 \in A)) = \frac{1}{n} \sum_{i=1}^n E(I(X_i \in A)) = \frac{1}{n} \sum_{i=1}^n P(X_i \in A) = \frac{1}{n} \times nP(X_1 \in A) = P(X_1 \in A)$$

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{P}(X_1 \in A)) = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i=1}^n \text{Var}(I(X_i \in A)) = \lim_{n \rightarrow \infty} \frac{n \times \text{Var}(I(X_1 \in A))}{n^2} = 0$$

Оценивание вероятностей

Пример оценивания обычных вероятностей

- Имеется выборка с реализацией $x = (5, 3, 1, -2, 3)$. Найдите реализации оценок вероятностей $P(X_1 = 3)$, $P(X_1 \geq 2.5)$, $P(X_1 = 2)$, $P(-10 \leq X_1 < 5)$, $P(X_1 \in \{2, -2, 1\})$ и $P(X_1 \leq 100)$.

Оценивание вероятностей

Пример оценивания обычных вероятностей

- Имеется выборка с реализацией $x = (5, 3, 1, -2, 3)$. Найдите реализации оценок вероятностей $P(X_1 = 3)$, $P(X_1 \geq 2.5)$, $P(X_1 = 2)$, $P(-10 \leq X_1 < 5)$, $P(X_1 \in \{2, -2, 1\})$ и $P(X_1 \leq 100)$.

Решение:

$$\hat{P}(X_1 = 3)(x) = \frac{1}{5} (I(x_1 = 3) + I(x_2 = 3) + I(x_3 = 3) + I(x_4 = 3) + I(x_5 = 3)) =$$

Оценивание вероятностей

Пример оценивания обычных вероятностей

- Имеется выборка с реализацией $x = (5, 3, 1, -2, 3)$. Найдите реализации оценок вероятностей $P(X_1 = 3)$, $P(X_1 \geq 2.5)$, $P(X_1 = 2)$, $P(-10 \leq X_1 < 5)$, $P(X_1 \in \{2, -2, 1\})$ и $P(X_1 \leq 100)$.

Решение:

$$\begin{aligned}\hat{P}(X_1 = 3)(x) &= \frac{1}{5} (I(x_1 = 3) + I(x_2 = 3) + I(x_3 = 3) + I(x_4 = 3) + I(x_5 = 3)) = \\ &= \frac{1}{5} (0 + 1 + 0 + 0 + 1) = \frac{2}{5} = 0.4\end{aligned}$$

Оценивание вероятностей

Пример оценивания обычных вероятностей

- Имеется выборка с реализацией $x = (5, 3, 1, -2, 3)$. Найдите реализации оценок вероятностей $P(X_1 = 3)$, $P(X_1 \geq 2.5)$, $P(X_1 = 2)$, $P(-10 \leq X_1 < 5)$, $P(X_1 \in \{2, -2, 1\})$ и $P(X_1 \leq 100)$.

Решение:

$$\begin{aligned}\hat{P}(X_1 = 3)(x) &= \frac{1}{5} (I(x_1 = 3)) + I(x_2 = 3)) + I(x_3 = 3)) + I(x_4 = 3)) + I(x_5 = 3))) = \\ &= \frac{1}{5} (0 + 1 + 0 + 0 + 1) = \frac{2}{5} = 0.4 \\ \hat{P}(X_1 \geq 2.5)(x) &= \frac{1}{5} (1 + 1 + 0 + 0 + 1) = \frac{3}{5} = 0.6\end{aligned}$$

Оценивание вероятностей

Пример оценивания обычных вероятностей

- Имеется выборка с реализацией $x = (5, 3, 1, -2, 3)$. Найдите реализации оценок вероятностей $P(X_1 = 3)$, $P(X_1 \geq 2.5)$, $P(X_1 = 2)$, $P(-10 \leq X_1 < 5)$, $P(X_1 \in \{2, -2, 1\})$ и $P(X_1 \leq 100)$.

Решение:

$$\begin{aligned}\hat{P}(X_1 = 3)(x) &= \frac{1}{5} (I(x_1 = 3) + I(x_2 = 3) + I(x_3 = 3) + I(x_4 = 3) + I(x_5 = 3)) = \\ &= \frac{1}{5} (0 + 1 + 0 + 0 + 1) = \frac{2}{5} = 0.4\end{aligned}$$

$$\hat{P}(X_1 \geq 2.5)(x) = \frac{1}{5} (1 + 1 + 0 + 0 + 1) = \frac{3}{5} = 0.6$$

$$\hat{P}(X_1 = 2)(x) = \frac{1}{5} (0 + 0 + 0 + 0 + 0) = 0$$

Оценивание вероятностей

Пример оценивания обычных вероятностей

- Имеется выборка с реализацией $x = (5, 3, 1, -2, 3)$. Найдите реализации оценок вероятностей $P(X_1 = 3)$, $P(X_1 \geq 2.5)$, $P(X_1 = 2)$, $P(-10 \leq X_1 < 5)$, $P(X_1 \in \{2, -2, 1\})$ и $P(X_1 \leq 100)$.

Решение:

$$\begin{aligned}\hat{P}(X_1 = 3)(x) &= \frac{1}{5} (I(x_1 = 3) + I(x_2 = 3) + I(x_3 = 3) + I(x_4 = 3) + I(x_5 = 3)) = \\ &= \frac{1}{5} (0 + 1 + 0 + 0 + 1) = \frac{2}{5} = 0.4\end{aligned}$$

$$\hat{P}(X_1 \geq 2.5)(x) = \frac{1}{5} (1 + 1 + 0 + 0 + 1) = \frac{3}{5} = 0.6$$

$$\hat{P}(X_1 = 2)(x) = \frac{1}{5} (0 + 0 + 0 + 0 + 0) = 0$$

$$\hat{P}(-10 \leq X_1 < 5)(x) = \frac{1}{5} (0 + 1 + 1 + 1 + 1) = \frac{4}{5} = 0.8$$

Оценивание вероятностей

Пример оценивания обычных вероятностей

- Имеется выборка с реализацией $x = (5, 3, 1, -2, 3)$. Найдите реализации оценок вероятностей $P(X_1 = 3)$, $P(X_1 \geq 2.5)$, $P(X_1 = 2)$, $P(-10 \leq X_1 < 5)$, $P(X_1 \in \{2, -2, 1\})$ и $P(X_1 \leq 100)$.

Решение:

$$\begin{aligned}\hat{P}(X_1 = 3)(x) &= \frac{1}{5} (I(x_1 = 3) + I(x_2 = 3) + I(x_3 = 3) + I(x_4 = 3) + I(x_5 = 3)) = \\ &= \frac{1}{5} (0 + 1 + 0 + 0 + 1) = \frac{2}{5} = 0.4\end{aligned}$$

$$\hat{P}(X_1 \geq 2.5)(x) = \frac{1}{5} (1 + 1 + 0 + 0 + 1) = \frac{3}{5} = 0.6$$

$$\hat{P}(X_1 = 2)(x) = \frac{1}{5} (0 + 0 + 0 + 0 + 0) = 0$$

$$\hat{P}(-10 \leq X_1 < 5)(x) = \frac{1}{5} (0 + 1 + 1 + 1 + 1) = \frac{4}{5} = 0.8$$

$$\hat{P}(X_1 \in \{2, -2, 1\})(x) = \frac{1}{5} (0 + 0 + 1 + 1 + 0) = \frac{2}{5} = 0.4$$

Оценивание вероятностей

Пример оценивания обычных вероятностей

- Имеется выборка с реализацией $x = (5, 3, 1, -2, 3)$. Найдите реализации оценок вероятностей $P(X_1 = 3)$, $P(X_1 \geq 2.5)$, $P(X_1 = 2)$, $P(-10 \leq X_1 < 5)$, $P(X_1 \in \{2, -2, 1\})$ и $P(X_1 \leq 100)$.

Решение:

$$\begin{aligned}\hat{P}(X_1 = 3)(x) &= \frac{1}{5} (I(x_1 = 3)) + I(x_2 = 3)) + I(x_3 = 3)) + I(x_4 = 3)) + I(x_5 = 3))) = \\ &= \frac{1}{5} (0 + 1 + 0 + 0 + 1) = \frac{2}{5} = 0.4\end{aligned}$$

$$\hat{P}(X_1 \geq 2.5)(x) = \frac{1}{5} (1 + 1 + 0 + 0 + 1) = \frac{3}{5} = 0.6$$

$$\hat{P}(X_1 = 2)(x) = \frac{1}{5} (0 + 0 + 0 + 0 + 0) = 0$$

$$\hat{P}(-10 \leq X_1 < 5)(x) = \frac{1}{5} (0 + 1 + 1 + 1 + 1) = \frac{4}{5} = 0.8$$

$$\hat{P}(X_1 \in \{2, -2, 1\})(x) = \frac{1}{5} (0 + 0 + 1 + 1 + 0) = \frac{2}{5} = 0.4$$

$$\hat{P}(X_1 \leq 100)(x) = \frac{1}{5} (1 + 1 + 1 + 1 + 1) = \frac{5}{5} = 1$$

Оценивание вероятностей

Условные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.

Оценивание вероятностей

Условные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.
- Необходимо оценить условную вероятность $P(X_1 \in A | X_1 \in B)$, где $A, B \subset R$ и $P(X_1 \in B) > 0$.

Оценивание вероятностей

Условные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.
- Необходимо оценить условную вероятность $P(X_1 \in A | X_1 \in B)$, где $A, B \subset R$ и $P(X_1 \in B) > 0$.
- Несмещенная и состоятельная оценка данной вероятности может быть получена как доля наблюдений в выборке, принадлежащих множеству $A \cap B$, среди наблюдений, принадлежащих B :

$$\hat{P}(X_1 \in A | X_1 \in B) = \frac{1}{m} \sum_{i=1}^n I(X_i \in A \cap B) = \frac{\hat{P}(X_1 \in A \cap B)}{\hat{P}(X_1 \in B)}, \text{ где } m = \sum_{i=1}^n I(X_i \in B)$$

Оценивание вероятностей

Условные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.
- Необходимо оценить условную вероятность $P(X_1 \in A | X_1 \in B)$, где $A, B \subset R$ и $P(X_1 \in B) > 0$.
- Несмещенная и состоятельная оценка данной вероятности может быть получена как доля наблюдений в выборке, принадлежащих множеству $A \cap B$, среди наблюдений, принадлежащих B :

$$\hat{P}(X_1 \in A | X_1 \in B) = \frac{1}{m} \sum_{i=1}^n I(X_i \in A \cap B) = \frac{\hat{P}(X_1 \in A \cap B)}{\hat{P}(X_1 \in B)}, \text{ где } m = \sum_{i=1}^n I(X_i \in B)$$

- Подставляя вместо наблюдений X_i их реализации x_i можно получить реализацию оценки соответствующей условной вероятности.

Оценивание вероятностей

Условные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.
- Необходимо оценить условную вероятность $P(X_1 \in A | X_1 \in B)$, где $A, B \subset R$ и $P(X_1 \in B) > 0$.
- Несмещенная и состоятельная оценка данной вероятности может быть получена как доля наблюдений в выборке, принадлежащих множеству $A \cap B$, среди наблюдений, принадлежащих B :

$$\hat{P}(X_1 \in A | X_1 \in B) = \frac{1}{m} \sum_{i=1}^n I(X_i \in A \cap B) = \frac{\hat{P}(X_1 \in A \cap B)}{\hat{P}(X_1 \in B)}, \text{ где } m = \sum_{i=1}^n I(X_i \in B)$$

- Подставляя вместо наблюдений X_i их реализации x_i можно получить реализацию оценки соответствующей условной вероятности.

Доказательство: докажем состоятельность, используя теорему Слущкого:

$$\begin{cases} \hat{P}(X_1 \in A \cap B) \xrightarrow{P} P(X_1 \in A \cap B) \\ \hat{P}(X_1 \in B) \xrightarrow{P} P(X_1 \in B) \end{cases} \implies \hat{P}(X_1 \in A | X_1 \in B) \xrightarrow{P} \frac{P(X_1 \in A \cap B)}{P(X_1 \in B)} = P(X_1 \in A | X_1 \in B)$$

Оценивание вероятностей

Условные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.
- Необходимо оценить условную вероятность $P(X_1 \in A | X_1 \in B)$, где $A, B \subset R$ и $P(X_1 \in B) > 0$.
- Несмещенная и состоятельная оценка данной вероятности может быть получена как доля наблюдений в выборке, принадлежащих множеству $A \cap B$, среди наблюдений, принадлежащих B :

$$\hat{P}(X_1 \in A | X_1 \in B) = \frac{1}{m} \sum_{i=1}^n I(X_i \in A \cap B) = \frac{\hat{P}(X_1 \in A \cap B)}{\hat{P}(X_1 \in B)}, \text{ где } m = \sum_{i=1}^n I(X_i \in B)$$

- Подставляя вместо наблюдений X_i их реализации x_i можно получить реализацию оценки соответствующей условной вероятности.

Доказательство: докажем состоятельность, используя теорему Слущкого:

$$\begin{cases} \hat{P}(X_1 \in A \cap B) \xrightarrow{P} P(X_1 \in A \cap B) \\ \hat{P}(X_1 \in B) \xrightarrow{P} P(X_1 \in B) \end{cases} \implies \hat{P}(X_1 \in A | X_1 \in B) \xrightarrow{P} \frac{P(X_1 \in A \cap B)}{P(X_1 \in B)} = P(X_1 \in A | X_1 \in B)$$

Пример: имеется реализация выборки $x = (5, 3, 1, -2, 3)$, найдем реализацию оценки $P(X_1 \leq 3 | X_1 > 0)$.

$$m = I(X_1 > 0) + \dots + I(X_5 > 0) = 1 + 1 + 1 + 0 + 1 = 4$$

Оценивание вероятностей

Условные вероятности

- Имеется выборка X_1, \dots, X_n из распределения $\Theta(\theta)$.
- Необходимо оценить условную вероятность $P(X_1 \in A | X_1 \in B)$, где $A, B \subset R$ и $P(X_1 \in B) > 0$.
- Несмещенная и состоятельная оценка данной вероятности может быть получена как доля наблюдений в выборке, принадлежащих множеству $A \cap B$, среди наблюдений, принадлежащих B :

$$\hat{P}(X_1 \in A | X_1 \in B) = \frac{1}{m} \sum_{i=1}^n I(X_i \in A \cap B) = \frac{\hat{P}(X_1 \in A \cap B)}{\hat{P}(X_1 \in B)}, \text{ где } m = \sum_{i=1}^n I(X_i \in B)$$

- Подставляя вместо наблюдений X_i их реализации x_i можно получить реализацию оценки соответствующей условной вероятности.

Доказательство: докажем состоятельность, используя теорему Слущкого:

$$\begin{cases} \hat{P}(X_1 \in A \cap B) \xrightarrow{p} P(X_1 \in A \cap B) \\ \hat{P}(X_1 \in B) \xrightarrow{p} P(X_1 \in B) \end{cases} \implies \hat{P}(X_1 \in A | X_1 \in B) \xrightarrow{p} \frac{P(X_1 \in A \cap B)}{P(X_1 \in B)} = P(X_1 \in A | X_1 \in B)$$

Пример: имеется реализация выборки $x = (5, 3, 1, -2, 3)$, найдем реализацию оценки $P(X_1 \leq 3 | X_1 > 0)$.

$$m = I(X_1 > 0) + \dots + I(X_5 > 0) = 1 + 1 + 1 + 0 + 1 = 4$$

$$\hat{P}(X_1 \leq 3 | X_1 > 0)(x) = \frac{1}{4} (I(0 < x_1 \leq 3) + \dots + I(0 < x_5 \leq 3)) = \frac{1}{4} (0 + 1 + 1 + 0 + 1) = \frac{3}{4}$$

- **Выборочная (эмпирическая) функция распределения** определяется как:

$$\hat{F}_n(t) = \hat{P}(X_1 \leq t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t) \sim \frac{1}{n} B(n, P(X_1 \leq t))$$

- **Выборочная (эмпирическая) функция распределения** определяется как:

$$\hat{F}_n(t) = \hat{P}(X_1 \leq t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t) \sim \frac{1}{n} B(n, P(X_1 \leq t))$$

- Выборочная функция распределения в точке t является несмещенной и состоятельной оценкой функции распределения в этой же точке, то есть $\hat{F}_n(t) \xrightarrow{P} F_{X_1}(t), \forall t \in R$.

- **Выборочная (эмпирическая) функция распределения** определяется как:

$$\hat{F}_n(t) = \hat{P}(X_1 \leq t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t) \sim \frac{1}{n} B(n, P(X_1 \leq t))$$

- Выборочная функция распределения в точке t является несмещенной и состоятельной оценкой функции распределения в этой же точке, то есть $\hat{F}_n(t) \xrightarrow{P} F_{X_1}(t), \forall t \in R$.

Пример: Найдём реализацию выборочной функции распределения для выборки с реализацией $x = (5, 3, 1, -2, 3)$.

Оценивание вероятностей

Выборочная (эмпирическая) функция распределения

- **Выборочная (эмпирическая) функция распределения** определяется как:

$$\hat{F}_n(t) = \hat{P}(X_1 \leq t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t) \sim \frac{1}{n} B(n, P(X_1 \leq t))$$

- Выборочная функция распределения в точке t является несмещенной и состоятельной оценкой функции распределения в этой же точке, то есть $\hat{F}_n(t) \xrightarrow{P} F_{X_1}(t), \forall t \in R$.

Пример: Найдём реализацию выборочной функции распределения для выборки с реализацией $x = (5, 3, 1, -2, 3)$.

$$F_n(t)|(X = x) = \left\{ \right.$$

Оценивание вероятностей

Выборочная (эмпирическая) функция распределения

- **Выборочная (эмпирическая) функция распределения** определяется как:

$$\hat{F}_n(t) = \hat{P}(X_1 \leq t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t) \sim \frac{1}{n} B(n, P(X_1 \leq t))$$

- Выборочная функция распределения в точке t является несмещенной и состоятельной оценкой функции распределения в этой же точке, то есть $\hat{F}_n(t) \xrightarrow{P} F_{X_1}(t), \forall t \in R$.

Пример: Найдём реализацию выборочной функции распределения для выборки с реализацией $x = (5, 3, 1, -2, 3)$.

$$F_n(t)|(X = x) = \begin{cases} (0 + 0 + 0 + 0 + 0)/5 = 0, & \text{если } t < -2 \\ \end{cases}$$

- Выборочная (эмпирическая) функция распределения определяется как:

$$\hat{F}_n(t) = \hat{P}(X_1 \leq t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t) \sim \frac{1}{n} B(n, P(X_1 \leq t))$$

- Выборочная функция распределения в точке t является несмещенной и состоятельной оценкой функции распределения в этой же точке, то есть $\hat{F}_n(t) \xrightarrow{P} F_{X_1}(t), \forall t \in R$.

Пример: Найдём реализацию выборочной функции распределения для выборки с реализацией $x = (5, 3, 1, -2, 3)$.

$$F_n(t)|(X = x) = \begin{cases} (0 + 0 + 0 + 0 + 0)/5 = 0, & \text{если } t < -2 \\ (0 + 0 + 0 + 1 + 0)/5 = 0.2, & \text{если } -2 \leq t < 1 \end{cases}$$

Оценивание вероятностей

Выборочная (эмпирическая) функция распределения

- Выборочная (эмпирическая) функция распределения определяется как:

$$\hat{F}_n(t) = \hat{P}(X_1 \leq t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t) \sim \frac{1}{n} B(n, P(X_1 \leq t))$$

- Выборочная функция распределения в точке t является несмещенной и состоятельной оценкой функции распределения в этой же точке, то есть $\hat{F}_n(t) \xrightarrow{P} F_{X_1}(t), \forall t \in R$.

Пример: Найдём реализацию выборочной функции распределения для выборки с реализацией $x = (5, 3, 1, -2, 3)$.

$$F_n(t)|(X = x) = \begin{cases} (0 + 0 + 0 + 0 + 0)/5 = 0, & \text{если } t < -2 \\ (0 + 0 + 0 + 1 + 0)/5 = 0.2, & \text{если } -2 \leq t < 1 \\ (0 + 0 + 1 + 1 + 0)/5 = 0.4, & \text{если } 1 \leq t < 3 \end{cases}$$

Оценивание вероятностей

Выборочная (эмпирическая) функция распределения

- Выборочная (эмпирическая) функция распределения определяется как:

$$\hat{F}_n(t) = \hat{P}(X_1 \leq t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t) \sim \frac{1}{n} B(n, P(X_1 \leq t))$$

- Выборочная функция распределения в точке t является несмещенной и состоятельной оценкой функции распределения в этой же точке, то есть $\hat{F}_n(t) \xrightarrow{P} F_{X_1}(t), \forall t \in R$.

Пример: Найдём реализацию выборочной функции распределения для выборки с реализацией $x = (5, 3, 1, -2, 3)$.

$$F_n(t)|(X = x) = \begin{cases} (0 + 0 + 0 + 0 + 0)/5 = 0, & \text{если } t < -2 \\ (0 + 0 + 0 + 1 + 0)/5 = 0.2, & \text{если } -2 \leq t < 1 \\ (0 + 0 + 1 + 1 + 0)/5 = 0.4, & \text{если } 1 \leq t < 3 \\ (0 + 1 + 1 + 1 + 1)/5 = 0.8, & \text{если } 3 \leq t < 5 \end{cases}$$

Оценивание вероятностей

Выборочная (эмпирическая) функция распределения

- Выборочная (эмпирическая) функция распределения определяется как:

$$\hat{F}_n(t) = \hat{P}(X_1 \leq t) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq t) \sim \frac{1}{n} B(n, P(X_1 \leq t))$$

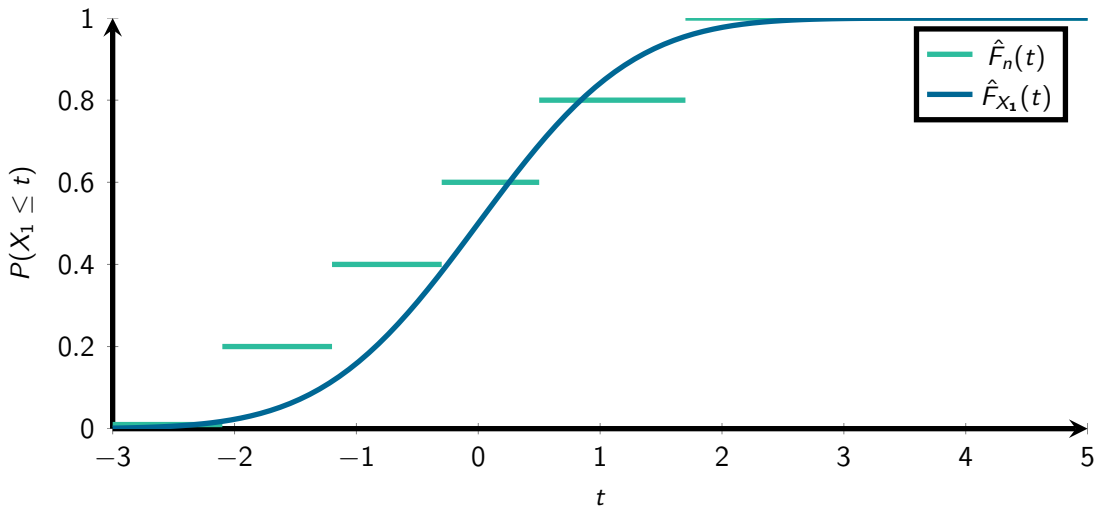
- Выборочная функция распределения в точке t является несмещенной и состоятельной оценкой функции распределения в этой же точке, то есть $\hat{F}_n(t) \xrightarrow{P} F_{X_1}(t), \forall t \in R$.

Пример: Найдём реализацию выборочной функции распределения для выборки с реализацией $x = (5, 3, 1, -2, 3)$.

$$F_n(t)|(X = x) = \begin{cases} (0 + 0 + 0 + 0 + 0)/5 = 0, & \text{если } t < -2 \\ (0 + 0 + 0 + 1 + 0)/5 = 0.2, & \text{если } -2 \leq t < 1 \\ (0 + 0 + 1 + 1 + 0)/5 = 0.4, & \text{если } 1 \leq t < 3 \\ (0 + 1 + 1 + 1 + 1)/5 = 0.8, & \text{если } 3 \leq t < 5 \\ (1 + 1 + 1 + 1 + 1)/5 = 1, & \text{если } t \geq 5 \end{cases}$$

Оценивание вероятностей

График эмпирической функции распределения на фоне теоретической функции распределения



Оценивание вероятностей

Верхняя граница погрешности

- Погрешность при использовании выборочной функции распределения вместо истинной можно записать как $|\hat{F}_n(t) - F_{X_1}(t)|$.

Оценивание вероятностей

Верхняя граница погрешности

- Погрешность при использовании выборочной функции распределения вместо истинной можно записать как $|\hat{F}_n(t) - F_{X_1}(t)|$.
- При помощи неравенства Чебышева можно найти верхнюю границу для вероятности того, что соответствующая погрешность превысит определенное значение:

$$P(|\hat{F}_n(t) - F_{X_1}(t)| > \varepsilon) \leq \frac{F_{X_1}(t)(1 - F_{X_1}(t))}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

Оценивание вероятностей

Верхняя граница погрешности

- Погрешность при использовании выборочной функции распределения вместо истинной можно записать как $|\hat{F}_n(t) - F_{X_1}(t)|$.
- При помощи неравенства Чебышева можно найти верхнюю границу для вероятности того, что соответствующая погрешность превысит определенное значение:

$$P(|\hat{F}_n(t) - F_{X_1}(t)| > \varepsilon) \leq \frac{F_{X_1}(t)(1 - F_{X_1}(t))}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

- Чем больше объем выборки n , тем меньше максимально возможная вероятность получить погрешность, превышающую ε .

Оценивание вероятностей

Верхняя граница погрешности

- Погрешность при использовании выборочной функции распределения вместо истинной можно записать как $|\hat{F}_n(t) - F_{X_1}(t)|$.
- При помощи неравенства Чебышева можно найти верхнюю границу для вероятности того, что соответствующая погрешность превысит определенное значение:

$$P(|\hat{F}_n(t) - F_{X_1}(t)| > \varepsilon) \leq \frac{F_{X_1}(t)(1 - F_{X_1}(t))}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

- Чем больше объем выборки n , тем меньше максимально возможная вероятность получить погрешность, превышающую ε .
- Например, если объем выборки составляет $n = 1000000$ миллион наблюдений, то вероятность того, что погрешность превысит $\varepsilon = 0.01$ окажется меньше, чем:

$$\frac{1}{4 \times 1000000 \times 0.01^2} = 0.0025$$

Выборочные характеристики

Выборочные моменты

- Выборочное среднее \bar{X}_n является несмещенной и состоятельной оценкой математического ожидания $E(X_1)$.

Выборочные характеристики

Выборочные моменты

- Выборочное среднее \bar{X}_n является несмещенной и состоятельной оценкой математического ожидания $E(X_1)$.

Доказательство:

$$E(\bar{X}_n) = E(X_1)$$

Выборочные характеристики

Выборочные моменты

- Выборочное среднее \bar{X}_n является несмещенной и состоятельной оценкой математического ожидания $E(X_1)$.

Доказательство:

$$E(\bar{X}_n) = E(X_1)$$

$$\lim_{n \rightarrow \infty} \text{Var}(\bar{X}_n) = \lim_{n \rightarrow \infty} \text{Var}(X_1)/n = 0$$

Выборочные характеристики

Выборочные моменты

- Выборочное среднее \bar{X}_n является несмещенной и состоятельной оценкой математического ожидания $E(X_1)$.

Доказательство:

$$E(\bar{X}_n) = E(X_1)$$

$$\lim_{n \rightarrow \infty} \text{Var}(\bar{X}_n) = \lim_{n \rightarrow \infty} \text{Var}(X_1)/n = 0$$

- По аналогии несмещенные и состоятельные оценки начальных моментов $E(X_1^k)$, где $k \in N$, могут быть получены с использованием выборочных начальных моментов соответствующего порядка:

$$\hat{m}_k = \bar{X}_n^k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

Выборочные характеристики

Выборочные моменты

- Выборочное среднее \bar{X}_n является несмещенной и состоятельной оценкой математического ожидания $E(X_1)$.

Доказательство:

$$E(\bar{X}_n) = E(X_1)$$

$$\lim_{n \rightarrow \infty} \text{Var}(\bar{X}_n) = \lim_{n \rightarrow \infty} \text{Var}(X_1)/n = 0$$

- По аналогии несмещенные и состоятельные оценки начальных моментов $E(X_1^k)$, где $k \in N$, могут быть получены с использованием выборочных начальных моментов соответствующего порядка:

$$\hat{m}_k = \overline{X^k}_n = \frac{1}{n} \sum_{i=1}^n X_i^k$$

Пример: рассчитаем реализацию третьего начального выборочного момента по выборке с реализациями $x = (5, 2, 5)$:

$$\hat{m}_k(x) = \frac{1}{3} (5^3 + 2^3 + 5^3) = 86$$

Выборочные характеристики

Выборочная дисперсия

- **Выборочная дисперсия** рассчитывается как:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Выборочные характеристики

Выборочная дисперсия

- **Выборочная дисперсия** рассчитывается как:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

- Она является смещенной, но состоятельной оценкой дисперсии $\text{Var}(X_1)$.

Выборочные характеристики

Выборочная дисперсия

- Выборочная дисперсия рассчитывается как:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

- Она является смещенной, но состоятельной оценкой дисперсии $\text{Var}(X_1)$.

Доказательство: покажем, что оценка является смещенной:

Выборочные характеристики

Выборочная дисперсия

- Выборочная дисперсия рассчитывается как:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

- Она является смещенной, но состоятельной оценкой дисперсии $\text{Var}(X_1)$.

Доказательство: покажем, что оценка является смещенной:

$$E(S_n^2) = \frac{1}{n} \sum_{i=1}^n E((X_i - \bar{X}_n)^2) = \frac{1}{n} \times n E((X_1 - \bar{X}_n)^2) = E((X_1 - \bar{X}_n)^2) =$$

Выборочные характеристики

Выборочная дисперсия

- Выборочная дисперсия рассчитывается как:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

- Она является смещенной, но состоятельной оценкой дисперсии $Var(X_1)$.

Доказательство: покажем, что оценка является смещенной:

$$\begin{aligned} E(S_n^2) &= \frac{1}{n} \sum_{i=1}^n E((X_i - \bar{X}_n)^2) = \frac{1}{n} \times n E((X_1 - \bar{X}_n)^2) = E((X_1 - \bar{X}_n)^2) = \\ &= Var(X_1 - \bar{X}_n) + E(X_1 - \bar{X}_n)^2 = Var(X_1) + Var(\bar{X}_n) - 2Cov(X_1, \bar{X}_n) + 0^2 = \end{aligned}$$

Выборочные характеристики

Выборочная дисперсия

- Выборочная дисперсия рассчитывается как:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

- Она является смещенной, но состоятельной оценкой дисперсии $\text{Var}(X_1)$.

Доказательство: покажем, что оценка является смещенной:

$$\begin{aligned} E(S_n^2) &= \frac{1}{n} \sum_{i=1}^n E((X_i - \bar{X}_n)^2) = \frac{1}{n} \times n E((X_1 - \bar{X}_n)^2) = E((X_1 - \bar{X}_n)^2) = \\ &= \text{Var}(X_1 - \bar{X}_n) + E(X_1 - \bar{X}_n)^2 = \text{Var}(X_1) + \text{Var}(\bar{X}_n) - 2\text{Cov}(X_1, \bar{X}_n) + 0^2 = \\ &= \text{Var}(X_1) + \frac{\text{Var}(X_1)}{n} - 2\text{Cov}\left(X_1, \frac{1}{n} X_1\right) = \text{Var}(X_1) \times \left(1 + \frac{1}{n} - 2\frac{1}{n}\right) = \frac{n-1}{n} \text{Var}(X_1) \end{aligned}$$

Выборочные характеристики

Выборочная дисперсия

- Выборочная дисперсия рассчитывается как:

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

- Она является смещенной, но состоятельной оценкой дисперсии $Var(X_1)$.

Доказательство: покажем, что оценка является смещенной:

$$\begin{aligned} E(S_n^2) &= \frac{1}{n} \sum_{i=1}^n E((X_i - \bar{X}_n)^2) = \frac{1}{n} \times n E((X_1 - \bar{X}_n)^2) = E((X_1 - \bar{X}_n)^2) = \\ &= Var(X_1 - \bar{X}_n) + E(X_1 - \bar{X}_n)^2 = Var(X_1) + Var(\bar{X}_n) - 2Cov(X_1, \bar{X}_n) + 0^2 = \\ &= Var(X_1) + \frac{Var(X_1)}{n} - 2Cov\left(X_1, \frac{1}{n}X_1\right) = Var(X_1) \times \left(1 + \frac{1}{n} - 2\frac{1}{n}\right) = \frac{n-1}{n} Var(X_1) \end{aligned}$$

- Состоятельная и несмещенная оценка дисперсии может быть получена с использованием исправленной (скорректированной) выборочной дисперсии:

$$\hat{\sigma}^2 = \frac{n}{n-1} S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Выборочные характеристики

Вариационный ряд

- Обозначим через $X_{(i)}$ наблюдение, которое является i -м по величине в выборке. Оно называется i -й **порядковой статистикой**. Статистики $X_{(1)} = \min(X_1, \dots, X_n)$ и $X_{(n)} = \max(X_1, \dots, X_n)$ именуются **минимальной и максимальной порядковыми статистиками** соответственно.

Выборочные характеристики

Вариационный ряд

- Обозначим через $X_{(i)}$ наблюдение, которое является i -м по величине в выборке. Оно называется i -й **порядковой статистикой**. Статистики $X_{(1)} = \min(X_1, \dots, X_n)$ и $X_{(n)} = \max(X_1, \dots, X_n)$ именуются **минимальной и максимальной порядковыми статистиками** соответственно.
- Последовательность $X_{(1)}, \dots, X_{(n)}$ **именуется вариационным рядом**.

Выборочные характеристики

Вариационный ряд

- Обозначим через $X_{(i)}$ наблюдение, которое является i -м по величине в выборке. Оно называется i -й **порядковой статистикой**. Статистики $X_{(1)} = \min(X_1, \dots, X_n)$ и $X_{(n)} = \max(X_1, \dots, X_n)$ именуется **минимальной и максимальной порядковыми статистиками** соответственно.
- Последовательность $X_{(1)}, \dots, X_{(n)}$ **именуется вариационным рядом**.
- Распределение максимальной экстремальной статистики определяется как:

$$F_{X_{(n)}}(x) = F_{\max(X_1, \dots, X_n)}(x) = P(\max(X_1, \dots, X_n) \leq x) = P(X_1 \leq x, \dots, X_n \leq x) =$$

Выборочные характеристики

Вариационный ряд

- Обозначим через $X_{(i)}$ наблюдение, которое является i -м по величине в выборке. Оно называется i -й **порядковой статистикой**. Статистики $X_{(1)} = \min(X_1, \dots, X_n)$ и $X_{(n)} = \max(X_1, \dots, X_n)$ именуется **минимальной и максимальной порядковыми статистиками** соответственно.
- Последовательность $X_{(1)}, \dots, X_{(n)}$ **именуется вариационным рядом**.
- Распределение максимальной экстремальной статистики определяется как:

$$\begin{aligned} F_{X_{(n)}}(x) &= F_{\max(X_1, \dots, X_n)}(x) = P(\max(X_1, \dots, X_n) \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \\ &= P(X_1 \leq x) \times \dots \times P(X_n \leq x) = F_{X_1}(x) \times \dots \times F_{X_n}(x) = (F_{X_1}(x))^n \end{aligned}$$

Выборочные характеристики

Вариационный ряд

- Обозначим через $X_{(i)}$ наблюдение, которое является i -м по величине в выборке. Оно называется i -й **порядковой статистикой**. Статистики $X_{(1)} = \min(X_1, \dots, X_n)$ и $X_{(n)} = \max(X_1, \dots, X_n)$ именуется **минимальной и максимальной порядковыми статистиками** соответственно.
- Последовательность $X_{(1)}, \dots, X_{(n)}$ **именуется вариационным рядом**.
- Распределение максимальной экстремальной статистики определяется как:

$$\begin{aligned} F_{X_{(n)}}(x) &= F_{\max(X_1, \dots, X_n)}(x) = P(\max(X_1, \dots, X_n) \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \\ &= P(X_1 \leq x) \times \dots \times P(X_n \leq x) = F_{X_1}(x) \times \dots \times F_{X_n}(x) = (F_{X_1}(x))^n \end{aligned}$$

- Распределение минимальной экстремальной статистики определяется как:

$$F_{X_{(1)}}(x) = F_{\min(X_1, \dots, X_n)}(x) = P(\min(X_1, \dots, X_n) \leq x) = 1 - P(\min(X_1, \dots, X_n) > x) =$$

Выборочные характеристики

Вариационный ряд

- Обозначим через $X_{(i)}$ наблюдение, которое является i -м по величине в выборке. Оно называется **i -й порядковой статистикой**. Статистики $X_{(1)} = \min(X_1, \dots, X_n)$ и $X_{(n)} = \max(X_1, \dots, X_n)$ именуются **минимальной и максимальной порядковыми статистиками** соответственно.
- Последовательность $X_{(1)}, \dots, X_{(n)}$ **именуется вариационным рядом**.
- Распределение максимальной экстремальной статистики определяется как:

$$\begin{aligned} F_{X_{(n)}}(x) &= F_{\max(X_1, \dots, X_n)}(x) = P(\max(X_1, \dots, X_n) \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \\ &= P(X_1 \leq x) \times \dots \times P(X_n \leq x) = F_{X_1}(x) \times \dots \times F_{X_n}(x) = (F_{X_1}(x))^n \end{aligned}$$

- Распределение минимальной экстремальной статистики определяется как:

$$\begin{aligned} F_{X_{(1)}}(x) &= F_{\min(X_1, \dots, X_n)}(x) = P(\min(X_1, \dots, X_n) \leq x) = 1 - P(\min(X_1, \dots, X_n) > x) = \\ &= 1 - P(X_1 > x) \times \dots \times P(X_n > x) = 1 - (1 - F_{X_1}(x)) \times \dots \times (1 - F_{X_n}(x)) = 1 - (1 - F_{X_1}(x))^n \end{aligned}$$

Выборочные характеристики

Вариационный ряд

- Обозначим через $X_{(i)}$ наблюдение, которое является i -м по величине в выборке. Оно называется i -й **порядковой статистикой**. Статистики $X_{(1)} = \min(X_1, \dots, X_n)$ и $X_{(n)} = \max(X_1, \dots, X_n)$ именуется **минимальной и максимальной порядковыми статистиками** соответственно.
- Последовательность $X_{(1)}, \dots, X_{(n)}$ **именуется вариационным рядом**.
- Распределение максимальной экстремальной статистики определяется как:

$$\begin{aligned} F_{X_{(n)}}(x) &= F_{\max(X_1, \dots, X_n)}(x) = P(\max(X_1, \dots, X_n) \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \\ &= P(X_1 \leq x) \times \dots \times P(X_n \leq x) = F_{X_1}(x) \times \dots \times F_{X_n}(x) = (F_{X_1}(x))^n \end{aligned}$$

- Распределение минимальной экстремальной статистики определяется как:

$$\begin{aligned} F_{X_{(1)}}(x) &= F_{\min(X_1, \dots, X_n)}(x) = P(\min(X_1, \dots, X_n) \leq x) = 1 - P(\min(X_1, \dots, X_n) > x) = \\ &= 1 - P(X_1 > x) \times \dots \times P(X_n > x) = 1 - (1 - F_{X_1}(x)) \times \dots \times (1 - F_{X_n}(x)) = 1 - (1 - F_{X_1}(x))^n \end{aligned}$$

- Распределение i -й порядковой статистики определяется как:

$$F_{X_{(i)}}(x) = \sum_{k=i}^n C_n^k (1 - F_{X_1}(x))^{n-k} (F_{X_1}(x))^k$$

Выборочные характеристики

Вариационный ряд

- Обозначим через $X_{(i)}$ наблюдение, которое является i -м по величине в выборке. Оно называется **i -й порядковой статистикой**. Статистики $X_{(1)} = \min(X_1, \dots, X_n)$ и $X_{(n)} = \max(X_1, \dots, X_n)$ именуется **минимальной и максимальной порядковыми статистиками** соответственно.
- Последовательность $X_{(1)}, \dots, X_{(n)}$ **именуется вариационным рядом**.
- Распределение максимальной экстремальной статистики определяется как:

$$\begin{aligned} F_{X_{(n)}}(x) &= F_{\max(X_1, \dots, X_n)}(x) = P(\max(X_1, \dots, X_n) \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = \\ &= P(X_1 \leq x) \times \dots \times P(X_n \leq x) = F_{X_1}(x) \times \dots \times F_{X_n}(x) = (F_{X_1}(x))^n \end{aligned}$$

- Распределение минимальной экстремальной статистики определяется как:

$$\begin{aligned} F_{X_{(1)}}(x) &= F_{\min(X_1, \dots, X_n)}(x) = P(\min(X_1, \dots, X_n) \leq x) = 1 - P(\min(X_1, \dots, X_n) > x) = \\ &= 1 - P(X_1 > x) \times \dots \times P(X_n > x) = 1 - (1 - F_{X_1}(x)) \times \dots \times (1 - F_{X_n}(x)) = 1 - (1 - F_{X_1}(x))^n \end{aligned}$$

- Распределение i -й порядковой статистики определяется как:

$$F_{X_{(i)}}(x) = \sum_{k=i}^n C_n^k (1 - F_{X_1}(x))^{n-k} (F_{X_1}(x))^k$$

Пример: найдем вероятность того, что в выборке из $n = 5$ экспоненциальных случайных величин с параметром $\lambda = 0.2$ наибольшее значение не превысит 10:

$$F_{X_{(5)}}(10) = (1 - e^{-0.2 \times 10})^5 \approx 0.483$$

Оценивание функции плотности

Гистограмма

- Гистограмма является оценкой функции плотности.

Оценивание функции плотности

Гистограмма

- Гистограмма является оценкой функции плотности.
- Разобьем выборку на m интервалов равной длины h , где интервал обозначим как b_k .

Оценивание функции плотности

Гистограмма

- Гистограмма является оценкой функции плотности.
- Разобьем выборку на m интервалов равной длины h , где интервал обозначим как b_k .
- Гистограмма предполагает следующую (как правило смещенную) оценку функции плотности в точке $t \in b_k$:

$$\hat{f}_{X_i}(t) = \frac{1}{nh} \sum_{i=1}^n I(X_i \in b_k)$$

Оценивание функции плотности

Гистограмма

- Гистограмма является оценкой функции плотности.
- Разобьем выборку на m интервалов равной длины h , где интервал обозначим как b_k .
- Гистограмма предполагает следующую (как правило смещенную) оценку функции плотности в точке $t \in b_k$:

$$\hat{f}_{X_i}(t) = \frac{1}{nh} \sum_{i=1}^n I(X_i \in b_k)$$

- Эффективность оценки зависит от параметра h . Существуют различные подходы к подбору оптимального значения данного параметра, например, по правилу Райса (Rice rule) $h = 2n^{-1.5}$.

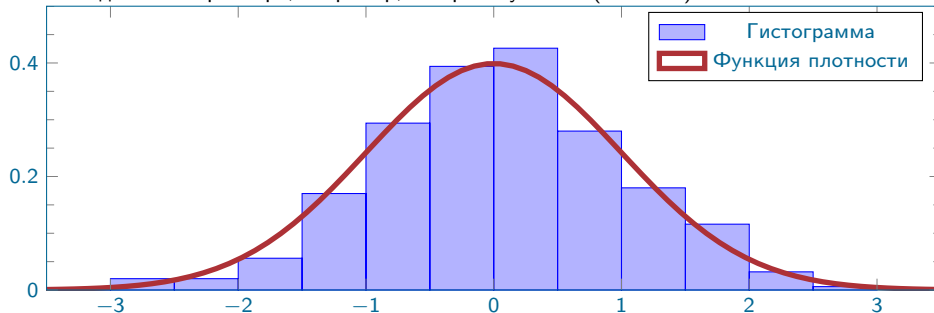
Оценивание функции плотности

Гистограмма

- Гистограмма является оценкой функции плотности.
- Разобьем выборку на m интервалов равной длины h , где интервал обозначим как b_k .
- Гистограмма предполагает следующую (как правило смещенную) оценку функции плотности в точке $t \in b_k$:

$$\hat{f}_{X_i}(t) = \frac{1}{nh} \sum_{i=1}^n I(X_i \in b_k)$$

- Эффективность оценки зависит от параметра h . Существуют различные подходы к подбору оптимального значения данного параметра, например, по правилу Райса (Rice rule) $h = 2n^{-1.5}$.



Оценивание функции плотности

Ядерное оценивание

- Функция плотности при помощи ядер оценивается следующим образом:

$$\hat{f}_{X_1}(t) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{t - x_i}{h}\right)$$

Оценивание функции плотности

Ядерное оценивание

- Функция плотности при помощи ядер оценивается следующим образом:

$$\hat{f}_{X_1}(t) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{t - x_i}{h}\right)$$

- Функция K является неотрицательной и именуется **ядром**. В качестве нее часто используют функцию плотности стандартного нормального распределения $K = \phi(t)$ (гауссовское ядро).

Оценивание функции плотности

Ядерное оценивание

- Функция плотности при помощи ядер оценивается следующим образом:

$$\hat{f}_{X_1}(t) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{t - x_i}{h}\right)$$

- Функция K является неотрицательной и именуется **ядром**. В качестве нее часто используют функцию плотности стандартного нормального распределения $K = \phi(t)$ (гауссовское ядро).
- Параметр h именуется **шириной окна** и его подбирают из соображения максимизации эффективности.

Оценивание функции плотности

Ядерное оценивание

- Функция плотности при помощи ядер оценивается следующим образом:

$$\hat{f}_{X_1}(t) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{t - x_i}{h}\right)$$

- Функция K является неотрицательной и именуется **ядром**. В качестве нее часто используют функцию плотности стандартного нормального распределения $K = \phi(t)$ (гауссовское ядро).
- Параметр h именуется **шириной окна** и его подбирают из соображения максимизации эффективности.

