Точечные оценки

Выборочная и описательная статистика

Статистика в целом подразделяется на два основных раздела: **описательную статистику** и **выборочную (инференциальную) статистику**. Первый раздел был рассмотрен в Лекции 1. Описательная статистика предназначена для представления, обобщения и наглядного описания имеющихся данных. Второй раздел является центральным и наиболее значимым в теории статистики. Его ***цель — формулировать выводы, выходящие за рамки имеющихся данных, позволяя делать более широкие обобщения.***

Вывод (inference) — это заключение, сделанное на основе логических рассуждений. В статистике под выводом обычно понимается формулирование утверждений о генеральной совокупности на основании анализа выборочных данных. В большинстве случаев исследовать всю генеральную совокупность невозможно, и единственный способ изучить её свойства — анализ доступной её части, то есть **выборки**. Инференциальные методы

Инференциальные методы (методы статистического вывода) – это совокупность статистических методов, позволяющих делать обобщения о генеральной совокупности на основе выборочных данных.

позволяют распространять результаты выборочного исследования на всю генеральную совокупность с использованием методов **индукции**. Именно поэтому особое внимание уделяется процессам формирования выборки, сбора данных и оценки **выборочной ошибки**. От качества выполнения этих этапов зависит корректность обобщения выборочных результатов на всю совокупность.

Статистический вывод включает в себя два основных направления: **оценивание параметров** и **проверку статистических гипотез**. Настоящая глава посвящена **оцениванию параметров**.

Оценивание

Оценивание — это процесс нахождения численного приближения для интересующего параметра **генеральной совокупности**. Примером такого параметра может быть **истинное среднее значение заработной платы** IT-специалистов в России. Оценивание решает проблему **неполных данных**: мы не можем вычислить истинное среднее значение напрямую, поскольку данные обо всей совокупности недоступны. Поэтому значение параметра совокупности оценивается на основе **выборочных данных**.

Оценка (Estimator) — это правило, которое позволяет вычислить численное значение, приближённое к параметру совокупности. Обычно оценка представляет собой

функцию от выборочных данных. Например, **выборочное среднее** является оценкой для **генерального среднего** и вычисляется как функция наблюдений в выборке:

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

Оценка представляет собой **"стратегию угадывания"**, используемую для поиска истинного значения параметра.

Точечная оценка (Point Estimate) — это конкретное числовое значение, полученное с помощью оценки. Например, если $\bar{X}$ — это оценка генерального среднего, а вычисленное по выборке значение равно **85 000**, то **85 000 — это точечная оценка среднего уровня зарплат**.

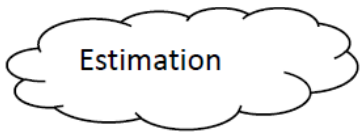
Параметры vs. Оценки

Важно различать **параметр генеральной совокупности** и его **оценку**, которая стремится его приблизить.

- **Параметр совокупности** – это **константа**, значение которой фиксировано, но, как правило, неизвестно.
- **Оценка (Estimator)** – это **случайная величина**, вычисляемая на основе случайной выборки для приближения неизвестного параметра.

Как и любая случайная величина, **оценка** имеет **распределение вероятностей**, а также такие характеристики, как **математическое ожидание** и **дисперсия**.

Примеры параметров и соответствующих им оценок приведены в таблице ниже:



Examples of estimators $\hat{\theta}$	Examples of population parameters θ
sample mean $\mathbf{\bar{X}}$	population mean $\boldsymbol{\mu}$
sample proportion $\hat{p}$	population proportion **p**
Sample standard deviation **s**	Population standard deviation $\boldsymbol{\sigma}$

Как видно, в левой колонке таблицы содержатся $\bar{X}$, $\hat{p}$ и S^2, которые являются **выборочными статистиками**. Однако **не каждая оценка** может называться выборочной статистикой. Разница между этими терминами будет объяснена в конце главы.

Можно представить **параметр генеральной совокупности** и его **оценку** как **мишень и оружие**:

- **Параметр генеральной совокупности** — это **мишень** исследования, неизвестное **константное значение**.
- **Оценка** — это **инструмент**, с помощью которого мы пытаемся «попасть» в истинное значение параметра.

Обозначения:

- **Параметр совокупности (мишень)** традиционно обозначается символами, такими как μ (среднее), σ^2 (дисперсия), p (доля).
- **Оценка параметра (инструмент)** удобно обозначается символом с «шляпкой» (circumflex), например, $\hat{\mu}$, $\hat{\sigma}^2$ $\hat{p}$.

Свойства оценок

Как выбрать оценку? Строго говоря, для любого параметра генеральной совокупности можно построить **бесконечное множество различных оценок**. Однако существуют **критерии**, которые позволяют выделить **хорошие** оценки.

Три основных свойства **качественной оценки**:

1. **Несмещённость (Unbiasedness)**
   - Оценка $\hat{\theta}$ называется **несмещённой**, если её математическое ожидание совпадает с истинным значением оцениваемого параметра:

$$E[\hat{\theta}] = \theta$$

   - Это означает, что в среднем оценка не завышает и не занижает истинное значение параметра.

2. **Эффективность (Efficiency)**

- Среди всех **несмещённых оценок** предпочтение отдаётся **той, у которой наименьшая дисперсия**, так как она меньше «разбросана» относительно истинного значения.
- Если две оценки $\widehat{\theta_1}$ и $\widehat{\theta_2}$ несмещённые, но $Var(\widehat{\theta_1}) < Var(\widehat{\theta_2})$, то $\widehat{\theta_1}$ **эффективнее**.

3. **Состоятельность (Consistency)**
   - Оценка $\hat{\theta}$ называется **состоятельной**, если при увеличении объёма выборки она **с высокой вероятностью стремится** к истинному значению

$$\hat{\theta}_n \xrightarrow{P} \theta, \quad \text{при } n \to \infty$$

- Это означает, что чем больше выборка, тем **точнее** оценка приближает параметр.

Хорошая оценка должна быть **как минимум состоятельной**, желательно **несмещённой**, а среди нескольких несмещённых оценок предпочтительна **наиболее эффективная**.

Как найти несмещённые оценки?

В статистике один из ключевых вопросов — **как найти хорошую оценку параметра**. Мы уже знаем, что **несмещённость** — это одно из важных свойств оценок. Давайте разберёмся, как такие оценки можно находить и подведём к **методу максимального правдоподобия (ММП)**.

1. Способы нахождения несмещённых оценок

Существует несколько методов нахождения несмещённых оценок, но основные два:

1. **Метод моментов**
2. **Метод максимального правдоподобия (MLE, ММП)**

Метод моментов — более интуитивный, но менее точный. **Метод максимального правдоподобия — более строгий и часто даёт хорошие оценки параметров.**

Метод моментов в математической статистике

Метод моментов (ММ) — это один из основных способов нахождения **оценок параметров распределения**. Его идея проста: **приравниваем теоретические моменты (известные из распределения) к их выборочным аналогам** и решаем полученную систему уравнений.

Метод моментов менее точен, чем **метод максимального правдоподобия (ММП)**, но его проще применять, особенно когда функция правдоподобия сложна.

1. Основная идея метода моментов

Пусть у нас есть случайная величина X, чьё распределение зависит от **неизвестного параметра** θ. Мы знаем, что у этого распределения есть **теоретические моменты**:

$$E[X^k] = \mu_k(\theta), \quad k = 1, 2, 3, \ldots$$

Но у нас нет полной информации о распределении, а есть только **выборка** $X_1, X_2, \ldots, X_n$. Из неё мы можем вычислить **выборочные моменты**:

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k.$$

➡️ В методе моментов мы **приравниваем** выборочные моменты к теоретическим:

$$\hat{\mu}_k = \mu_k(\theta).$$

Решая это уравнение относительно θ, мы находим **оценку параметра**.

2. Простые примеры метода моментов

Пример 1: Оценка математического ожидания

Допустим, мы знаем, что случайная величина X имеет **некоторое распределение** с **неизвестным** математическим ожиданием μ.

Для генеральной совокупности:

$$E[X] = \mu.$$

Выборочное среднее (выборочный первый момент):

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Применяя метод моментов:

$$\bar{X} = \mu \quad \Rightarrow \quad \hat{\mu} = \bar{X}.$$

То есть **выборочное среднее — это оценка математического ожидания по методу моментов**.

Пример 2: Оценка параметров равномерного распределения $U[a,b]$

Пусть $X \sim U[a,b]$ — равномерное распределение на отрезке $[a,b]$, но **параметры a и b неизвестны**.

Мы знаем теоретические моменты:

$$E[X] = \frac{a+b}{2}, \quad E[X^2] = \frac{a^2+ab+b^2}{3}.$$

А выборочные моменты:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Приравниваем:

$$\bar{X} = \frac{a+b}{2}, \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \frac{a^2+ab+b^2}{3}.$$

Решая систему, находим оценки $\hat{a}$ и $\hat{b}$.

Пример 3: Оценка параметра экспоненциального распределения

Пусть $X \sim \mathrm{Exp}(\lambda)$, т.е. плотность:

$$f(x|\lambda) = \lambda e^{-\lambda x}, \quad x > 0.$$

Мы знаем, что:

$$E[X] = \frac{1}{\lambda}.$$

Выборочный момент:

$$\bar{X} = \frac{1}{n}\sum X_i.$$

Приравниваем:

$$\bar{X} = \frac{1}{\lambda} \quad \Rightarrow \quad \hat{\lambda} = \frac{1}{\bar{X}}.$$

➡️ Таким образом, оценка λ методом моментов:

$$\hat{\lambda} = \frac{1}{\bar{X}}.$$

3. Достоинства и недостатки метода моментов

✅ **Плюсы:**

- **Простота: не требует сложных вычислений, подходит для интуитивного подхода.**
- **Работает даже при сложных распределениях, где ММП может быть трудно вычислить.**
- **Часто даёт несмещённые оценки.**

❌ **Минусы:**

- **Может давать неэффективные оценки (например, с высокой дисперсией).**
- **Не всегда даёт статистически состоятельные оценки (может расходиться при больших n).**
- **Не учитывает всю информацию о распределении, в отличие от метода максимального правдоподобия.**

Метод максимального правдоподобия (MLE, ММП)

Идея ММП: найти такие параметры распределения, при которых вероятность наблюдаемых данных максимальна.

Допустим, у нас есть выборка $X_1, X_2, ..., X_n$ из распределения с неизвестным параметром θ, и мы знаем **функцию плотности вероятности (или функцию вероятности)** $f(X|\theta)$.

Тогда правдоподобие выборки:

$$L(\theta) = P(X_1, X_2, ..., X_n|\theta) = \prod_{i=1}^{n} f(X_i|\theta).$$

Мы ищем **такое значение $\hat{\theta}$, которое максимизирует эту функцию**:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta).$$

Часто удобнее работать с **логарифмом правдоподобия**:

$$\ln L(\theta) = \sum_{i=1}^{n} \ln f(X_i|\theta).$$

Пример нахождения несмещённой оценки методом ММП

Пример: Оценка среднего для нормального распределения

Пусть $X_1, X_2, ..., X_n \sim N(\mu, \sigma^2)$. Тогда плотность распределения:

$$f(X|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X-\mu)^2}{2\sigma^2}\right).$$

Запишем функцию правдоподобия:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X_i-\mu)^2}{2\sigma^2}\right).$$

Возьмём логарифм:

$$\ln L(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2.$$

Найдём производную по μ и приравняем к нулю:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0.$$

Отсюда:

$$\sum_{i=1}^{n}(X_i-\mu) = 0 \quad \Rightarrow \quad \hat{\mu} = \frac{1}{n}\sum X_i = \bar{X}.$$

Эта оценка **совпадает с выборочным средним** и является **несмещённой**.

Оценка дисперсии

Аналогично, найдём оценку для σ^2:

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i - \mu)^2 = 0.$$

Решая уравнение, получаем:

$$\hat{\sigma}^2_{\mathrm{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Мы уже **доказывали, что эта оценка является смещённой**! Чтобы получить **несмещённую** оценку, надо домножить на $\frac{n}{n-1}$, и тогда:

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

Итог

1. Метод моментов может дать несмещённую оценку, но не всегда эффективную.
2. Метод максимального правдоподобия (MLE) даёт самую правдоподобную оценку, но она может быть смещённой (как оценка дисперсии).
3. Если оценка получена методом ММП и оказалась смещённой, её можно корректировать.

Метод максимального правдоподобия — один из самых мощных способов нахождения оценок параметров в статистике и машинном обучении.

Эффективность точечных оценок

Эффективность статистической оценки — это свойство, характеризующее **качество оценки параметра** с точки зрения **минимальной дисперсии**. Оценка называется **эффективной**, если среди всех несмещённых оценок она имеет **наименьшую возможную дисперсию**.

Формально:
Пусть у нас есть две **несмещённые оценки** параметра θ: $\widehat{\theta_1}$ и $\widehat{\theta_2}$. Если для любого θ выполняется:

$$D(\hat{\theta}_1) \leq D(\hat{\theta}_2),$$

то $\widehat{\theta_1}$ **эффективнее**, чем $\widehat{\theta_2}$.

Максимально возможную эффективность среди несмещённых оценок задаёт **граница Рао-Крамера**.

Как определить эффективность оценки?

1. **Проверить несмещённость**:

$$E[\hat{\theta}] = \theta$$

Если оценка смещённая, её эффективность нельзя сравнивать по критерию дисперсии.

2. **Вычислить дисперсию оценки**:

$$D(\hat{\theta}) = E[(\hat{\theta} - \theta)^2].$$

3. Сравнить дисперсию с границей Рао-Крамера (если доступна информация Фишера). Оценка эффективна, если её дисперсия достигает границы Рао-Крамера.

Информация Фишера

Информация Фишера — это мера **"чувствительности" функции правдоподобия** к изменениям параметра θ. Чем больше информации Фишера, тем точнее можно оценить θ.

Определяется как **математическое ожидание второй производной логарифма функции правдоподобия**:

$$I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\ln L(X|\theta)\right].$$

Где:

- $L(X|\theta)$ — функция правдоподобия,
- Чем больше $I(\theta)$, тем **больше информации о параметре содержится в выборке**.

Альтернативно, можно записать:

$$I(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\ln L(X|\theta)\right)^2\right].$$

Интерпретация:

- Если **информация Фишера высока**, то **оценка будет более точной** (т.к. меньше её дисперсия).
- Если **информация Фишера мала**, то даже при большой выборке оценка будет "разбросанной".

Пример:

- Пусть $X_1, \ldots, X_n \sim N(\theta, 1)$ (нормальное распределение со средним θ и известной дисперсией 1).
- Тогда информация Фишера:

$$I(\theta) = n.$$

➡️ Чем больше n, тем больше информации, а значит, **точность оценки θ растёт**.

Неравенство Крамера — Рао

Неравенство Краме́ра — Ра́о — неравенство, которое при некоторых условиях на статистическую модель даёт нижнюю границу для дисперсии оценки неизвестного параметра, выражая её через информацию Фишера.

Названо по именам шведского математика Харальда Крамера и индийского математика Кальямпуди Рао, но независимо от них устанавливалось также Фреше, Дармуа (фр. Georges Darmois), Айткеном (англ. Alexander Aitken) и Сильверстоуном (Harold Silverstone). Известно обобщение в квантовой теории оценивания — квантовое неравенство Крамера — Рао.

Это фундаментальное утверждение о **нижней границе дисперсии несмещённых оценок**. Оно говорит, что **никакая несмещённая оценка не может иметь дисперсию меньше, чем величина, обратная информации Фишера**.

$$Var(\hat{\theta}) \geq \frac{1}{I_n(\theta)}$$

Если оценка достигает этой границы, то она **эффективна**.

Пример: оценка среднего нормального распределения

Пусть $X_1, \ldots, X_n \sim N(\theta, \sigma^2)$, где σ^2 **известно**.

Мы знаем, что **несмещённая оценка** для θ — это выборочное среднее:

$$\hat{\theta} = \bar{X} = \frac{1}{n} \sum X_i.$$

Теперь найдем **границу Рао-Крамера**.

1. Функция правдоподобия:

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(X_i - \theta)^2}{2\sigma^2}\right).$$

2. Логарифмическая функция правдоподобия:

$$\ln L(\theta) = -\frac{n}{2} \ln(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(X_i - \theta)^2}{2\sigma^2}.$$

3. Первая производная:

$$\frac{\partial}{\partial\theta}\ln L(\theta)=\sum_{i=1}^{n}\frac{X_i-\theta}{\sigma^2}.$$

4. Вторая производная:

$$\frac{\partial^2}{\partial\theta^2}\ln L(\theta)=-\frac{n}{\sigma^2}.$$

5. Информация Фишера:

$$I(\theta)=-E\left[\frac{\partial^2}{\partial\theta^2}\ln L(\theta)\right]=\frac{n}{\sigma^2}.$$

6. Граница Рао-Крамера:

$$D(\hat{\theta})\geq\frac{1}{nI(\theta)}=\frac{\sigma^2}{n}.$$

Но мы знаем, что **дисперсия выборочного среднего**:

$$D(\bar{X})=\frac{\sigma^2}{n}.$$

Так как $D(\bar{X})$ **достигает границы Рао-Крамера**, значит, $\bar{X}$ **— эффективная оценка**.

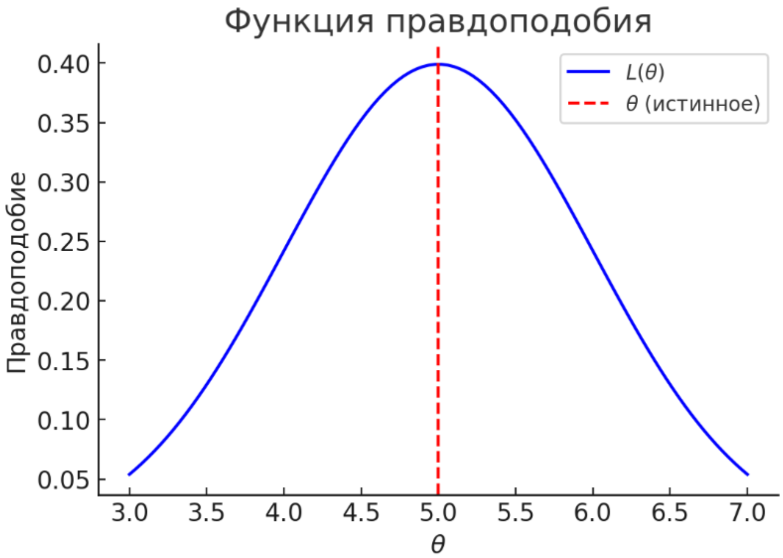
Итог

1. **Эффективная оценка** — это такая, у которой **минимальная дисперсия среди несмещённых оценок**.
2. **Информация Фишера** показывает, **сколько информации о параметре содержится в выборке**.
3. **Неравенство Рао-Крамера** даёт **нижнюю границу дисперсии несмещённых оценок**.
4. **Если оценка достигает этой границы, она эффективна** (пример: выборочное среднее при нормальном распределении).

Графическая иллюстрация примера:

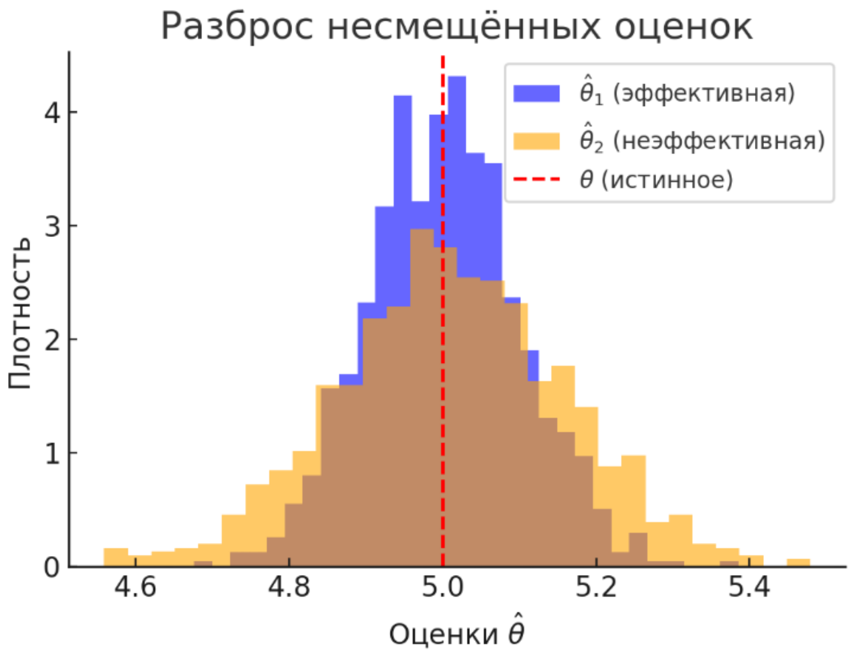
Графическая интерпретация неравенства Рао-Крамера поможет лучше понять его смысл. Давайте построим несколько иллюстраций:

1. **Функция правдоподобия и информация Фишера**
2. **Разброс несмещённых оценок**
3. **Граница Рао-Крамера**



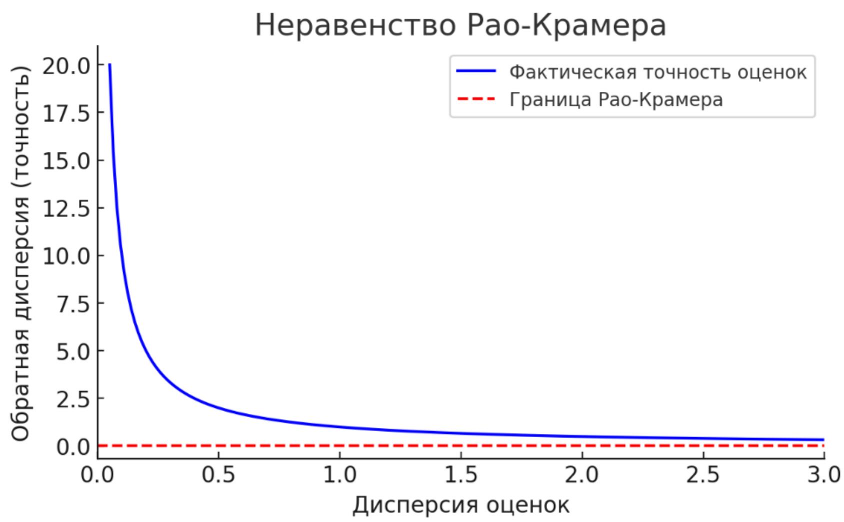
На этом графике показана функция правдоподобия $L(\theta)$, достигающая максимума в истинном значении параметра θ.

Чем острее пик (чем больше информация Фишера), тем точнее оценка — её дисперсия меньше. Теперь рассмотрим неравенство Рао-Крамера.



На этом графике две оценки параметра θ:

- **Синяя гистограмма** (θ_1) соответствует эффективной оценке с минимальной дисперсией.
- **Оранжевая гистограмма** (θ_2) имеет большую дисперсию, что делает её менее эффективной.



На этом графике показано, что фактическая точность оценок (синяя кривая) не может превысить границу, заданную неравенством Рао-Крамера (красная пунктирная линия).

Если оценка достигает этой границы, она называется **эффективной**. В противном случае она имеет избыточную дисперсию и может быть улучшена.

Такое графическое представление помогает интуитивно понять, почему мы стремимся к оценкам с минимальной дисперсией, но не ниже предела, заданного неравенством Рао-Крамера.

Состоятельность статистической оценки

Состоятельность — это свойство оценки, означающее, что при увеличении размера выборки она сходится по вероятности к истинному значению параметра. Это фундаментальное требование к хорошим оценкам: они должны становиться точнее при увеличении объёма данных.

Определение

Оценка $\hat{\theta}_n$ параметра θ называется **состоятельной**, если:

$$\hat{\theta}_n \xrightarrow{P} \theta, \quad \text{при } n \to \infty.$$

Это означает, что с ростом n вероятность отклонения оценки $\hat{\theta}_n$ от θ сколь угодно мало стремится к нулю.

Виды состоятельности

1. **Состоятельность по вероятности (слабая состоятельность)**

$$\forall \varepsilon > 0: \quad P(|\hat{\theta}_n - \theta| > \varepsilon) \to 0, \quad \text{при } n \to \infty.$$

   Это стандартное определение.

2. **Состоятельность почти наверное (сильная состоятельность)**

$$P\left(\lim_{n\to\infty} \hat{\theta}_n = \theta\right) = 1.$$

   Это более строгий вариант: оценка с вероятностью 1 сходится к θ.

3. **Состоятельность в среднем квадрате (по среднему квадратичному)**

$$E[(\hat{\theta}_n - \theta)^2] \to 0, \quad \text{при } n \to \infty.$$

   Этот критерий включает не только сходимость по вероятности, но и требует, чтобы дисперсия оценки стремилась к нулю.

Теорема Чебышёва

Теорема Чебышёва (неравенство Чебышёва) — один из ключевых результатов теории вероятностей, который позволяет оценить вероятность отклонения случайной величины от её математического ожидания. Она полезна при доказательстве состоятельности статистических оценок.

Формулировка теоремы Чебышёва

Пусть X — случайная величина с конечным математическим ожиданием $E[X]$ и конечной дисперсией $\mathrm{Var}(X)$. Тогда для любого $\varepsilon > 0$ выполняется неравенство:

$$P(|X - E[X]| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

где:

- $P(|X - E[X]| \geq \varepsilon)$ — вероятность того, что отклонение X от математического ожидания $E[X]$ больше или равно ε,
- $\mathrm{Var}(X)$ — дисперсия случайной величины X,
- ε — произвольное положительное число.

Интуитивный смысл

Это неравенство говорит, что чем больше дисперсия случайной величины, тем выше вероятность её больших отклонений от математического ожидания.

Следствие:
Если дисперсия уменьшается при увеличении выборки, то вероятность больших отклонений становится малой, что используется при доказательстве состоятельности оценок.

Доказательство теоремы Чебышёва

Используем **неравенство Маркова**, которое утверждает, что для любой неотрицательной случайной величины Y и $a > 0$:

$$P(Y \geq a) \leq \frac{E[Y]}{a}$$

Теперь применим его к $Y = (X - E[X])^2$ и возьмём $a = \varepsilon^2$:

$$P((X - E[X])^2 \geq \varepsilon^2) \leq \frac{E[(X - E[X])^2]}{\varepsilon^2}$$

Но $E[(X - E[X])^2] = \mathrm{Var}(X)$, поэтому:

$$P(|X - E[X]| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

Что и требовалось доказать.

Применение в статистике

1. Доказательство слабой состоятельности

Пусть $\hat{\theta}_n$ — оценка параметра θ, и пусть $E[\hat{\theta}_n] = \theta$, а дисперсия $\mathrm{Var}(\hat{\theta}_n) \to 0$ при $n \to \infty$. Тогда, применяя теорему Чебышёва:

$$P(|\hat{\theta}_n - \theta| \geq \varepsilon) \leq \frac{\mathrm{Var}(\hat{\theta}_n)}{\varepsilon^2} \to 0, \quad \text{при } n \to \infty.$$

Это означает, что $\hat{\theta}_n$ сходится по вероятности к θ, то есть является состоятельной.

2. Пример: оценка среднего

Рассмотрим выборочное среднее:

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

из выборки $X_1, X_2, \ldots, X_n$, где $E[X_i] = \mu$ и $\mathrm{Var}(X_i) = \sigma^2$. Тогда:

$$E[\bar{X}_n] = \mu, \quad \mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}.$$

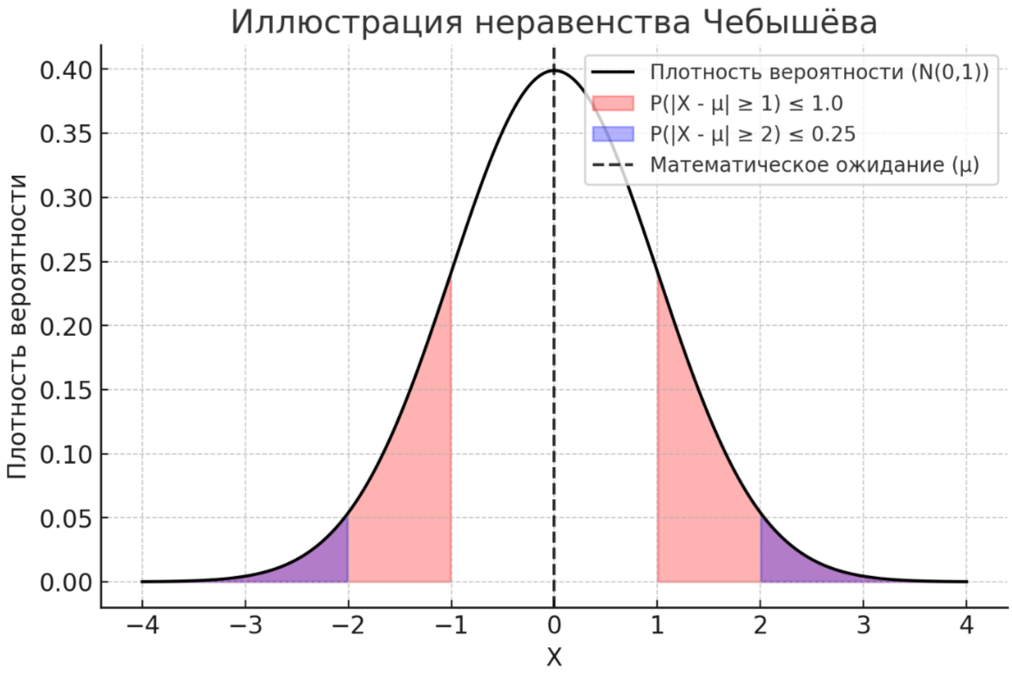
Применяя теорему Чебышёва:

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0, \quad \text{при } n \to \infty.$$

Таким образом, выборочное среднее является состоятельной оценкой математического ожидания μ.

Графическая интерпретация

Построим график, иллюстрирующий теорему Чебышёва: покажем, как вероятность больших отклонений уменьшается при увеличении выборки.



На графике показано, как вероятность больших отклонений уменьшается при увеличении ε. Цветные области иллюстрируют вероятность отклонения от математического ожидания, ограниченную сверху неравенством Чебышёва.

Чем больше ε, тем меньше вероятность отклонений за его пределы. Это подтверждает, что дисперсия контролирует разброс случайной величины.

Теоремы о совместной плотности нескольких случайных величин

Совместная плотность вероятности описывает, как две или более случайные величины распределены вместе. Рассмотрим основные теоремы и свойства.

1. Определение совместной плотности вероятности

Пусть X и Y — две непрерывные случайные величины. Если существует функция $f_{X,Y}(x,y)$, такая что для любого множества A:

$$P((X,Y) \in A) = \int \int_A f_{X,Y}(x,y)\,dx\,dy,$$

то $f_{X,Y}(x,y)$ называется **совместной функцией плотности вероятности (joint probability density function, joint PDF)** случайных величин X и Y.

Свойства совместной плотности:

1. $f_{X,Y}(x,y) \geq 0$ для всех x, y.
2. **Нормировка:**

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1.$$

3. **Вероятность попадания в область:**

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x,y)\,dx\,dy.$$

2. Маргинальные (предельные) плотности

Предельные плотности случайных величин X и Y получаются интегрированием совместной плотности по одной из переменных:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx.$$

Эти плотности описывают распределение каждой случайной величины отдельно.

3. Условная плотность вероятности

Если $f_Y(y) > 0$, то **условная плотность** случайной величины X при заданном $Y = y$ определяется как:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

Аналогично:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

Эти функции показывают, как распределяется X, если известны значения Y, и наоборот.

4. Независимость случайных величин

Случайные величины X и Y независимы, если их совместная плотность разлагается в произведение предельных плотностей:

$$f_{X,Y}(x,y) = f_X(x) f_Y(y), \quad \forall x, y.$$

То есть, знание одной случайной величины не даёт информации о другой.

5. Теорема о связи функции распределения и плотности

Пусть $F_{X,Y}(x,y)$ — совместная функция распределения:

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y).$$

Если X и Y непрерывны, то совместная плотность связана с функцией распределения:

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y).$$

6. Теорема о свертке (вычисление плотности суммы двух случайных величин)

Если X и Y — независимые случайные величины с плотностями $f_X(x)$ и $f_Y(y)$, то плотность их суммы $Z = X + Y$ вычисляется по формуле свёртки:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\, dx.$$

Эта формула используется для нахождения распределений сумм случайных величин.

7. Теорема о линейном преобразовании случайных величин

Если $X_1, X_2, \ldots, X_n$ — случайные величины с известной совместной плотностью, а $Y_1, Y_2, \ldots, Y_n$ получены линейным преобразованием:

$$Y_i = \sum_{j=1}^{n} a_{ij} X_j + b_i,$$

то совместная плотность новых величин определяется формулой:

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \frac{1}{|\det(A)|} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n),$$

где A — матрица коэффициентов a_{ij}.

Выводы

- Совместная плотность вероятности описывает, как случайные величины распределены вместе.
- По ней можно получить частные и условные плотности.
- Независимость случайных величин упрощает выражение совместной плотности.
- Формула факторизации помогает выражать плотности через условные распределения.
- Формула свёртки позволяет находить распределение суммы случайных величин.