Оценивание параметров распределения

Схема математической статистики

Выборка: $x_1, \ldots, x_n$ Параметр: θ

$\hat{\theta} \longrightarrow f_{\hat{\theta}}(t)$

Точность оценки, прогнозов

доверительные интервалы

Ответы на вопросы

проверка гипотез

Как оценить

- Метод моментов
- Метод максимального правдоподобия

Хорошие свойства

- Несмещенная
- Состоятельная
- Эффективная

Союзники

Асимптотические (при большом n)

- ЦПТ
- Дельта-метод

Точные

- Теорема Фишера
- $\chi^2_n, t_n, F_{n,k}$
- Ещё союзники!

План

- Что такое методов моментов и метод максимального правдоподобия
- Примеры для дискретных и непрерывных распределений
- Несмещенность, состоятельность и эффективность
- Информация Фишера, почему это информация и откуда она берётся

Метод моментов

Метод моментов

$X_1, \dots, X_n$ одинаково независимо распределены $(\ iid\ )$

Момент $\mathbb{E}\left(X_i^k\right)$ зависит от неизвестного параметра θ:

$$\mathbb{E}\left(X_i^k\right) = f(\theta)$$

Оценкой метода моментов называется случайная величина:

$$\hat{\theta}_{MM} = f^{-1}(\overline{X^k})$$

То есть оценка получается решением уравнения

$$\mathbb{E}\left(X_i^k\right) \approx \frac{\sum x_i^k}{n}$$

Метод моментов

Чаще всего хватает первого момента и берут $k = 1$, то есть решают уравнение:

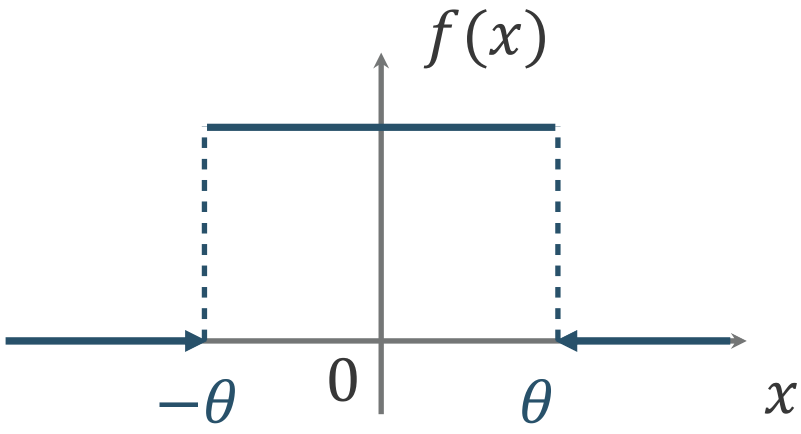
$$\mathbb{E}(X_i) \approx \frac{\sum x_i}{n}$$

Метод моментов

Если оказывается, что $\mathbb{E}(X_i) = 0$, тогда используют моменты более высоких порядков:

$$X_1, \dots, X_n \sim iid\ U[-\theta; \theta]$$

$$\mathbb{E}(X_i) = 0 \Rightarrow \bar{x} = 0$$



$$\mathbb{E}(X_i) = 0$$

❗ Используя первый момент, нельзя получить оценку

$$\mathbb{E}\left(X_i^2\right) = \frac{\theta^2}{3} \Rightarrow \overline{x^2} = \frac{\theta^2}{3}$$

$$\Rightarrow \hat{\theta}_{MM} = \left(3\overline{x^2}\right)^{0.5}$$

Метод моментов

Если у распределения несколько параметров, используют несколько моментов:

$$X_1, \dots, X_n \sim iid\ N(\mu, \sigma^2)$$

Нужно оценить два параметра: дисперсию и математическое ожидание, используем два момента:

$$\begin{cases} \mathbb{E}(X_i) \approx \bar{x} \\ \mathbb{E}\left(X_i^2\right) \approx \overline{x^2} \end{cases} \Leftrightarrow \begin{cases} \mu = \bar{x} \\ \sigma^2 + \mu^2 = \overline{x^2} \end{cases} \Leftrightarrow \begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \overline{x^2} - \bar{x}^2 \end{cases}$$

Резюме

- Метод моментов позволяет оценить параметр неизвестного распределения по выборке
- Перед тем, как использовать метод моментов, мы должны предположить, из какого распределения выборка была получена
- Обычно для оценки достаточно первого момента
- Если у распределения есть несколько параметров, используют несколько моментов

Чего хочет статистик?

Схема математической статистики

Выборка: $x_1, \ldots, x_n$ Параметр: θ

$$\hat{\theta} \longrightarrow f_{\hat{\theta}}(t)$$

Точность оценки, прогнозов

доверительные интервалы

Ответы на вопросы

проверка гипотез

Как оценить

- Метод моментов
- Метод максимального правдоподобия

Хорошие свойства

- Несмещенная
- Состоятельная
- Эффективная

Союзники

Асимптотические (при большом n)

- ЦПТ
- Дельта-метод

Точные

- Теорема Фишера
- $\chi^2_n, t_n, F_{n,k}$
- Ещё союзники!

Схема математической статистики

Выборка: $x_1, \dots, x_n$ Параметр: θ

$\hat{\theta} \longrightarrow f_{\hat{\theta}}(t)$

Точность оценки, прогнозов

доверительные интервалы

Ответы на вопросы

проверка гипотез

Как оценить

- Метод моментов
- Метод максимального правдоподобия

Союзники

Асимптотические (при большом n)

- ЦПТ
- Дельта-метод

Точные

- Теорема Фишера
- $\chi^2_n, t_n, F_{n,k}$
- Ещё союзники!

Хорошие свойства

- Несмещенная
- Состоятельная
- Эффективная

Несмещённость

Оценка называется несмещённой, если её математическое ожидание равно оцениваемому параметру:

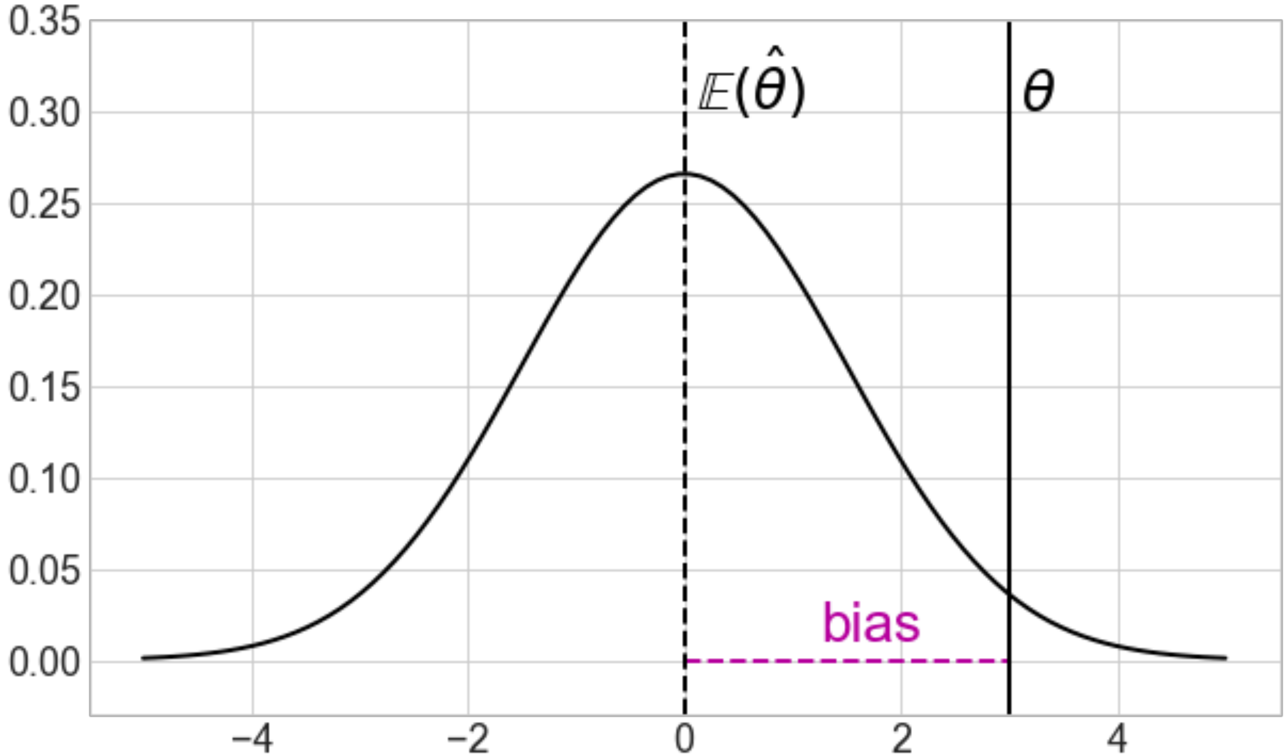
$$\mathbb{E}(\hat{\theta}) = \theta$$

Смещение оценки это разница между её математическим ожиданием и её реальным значением:

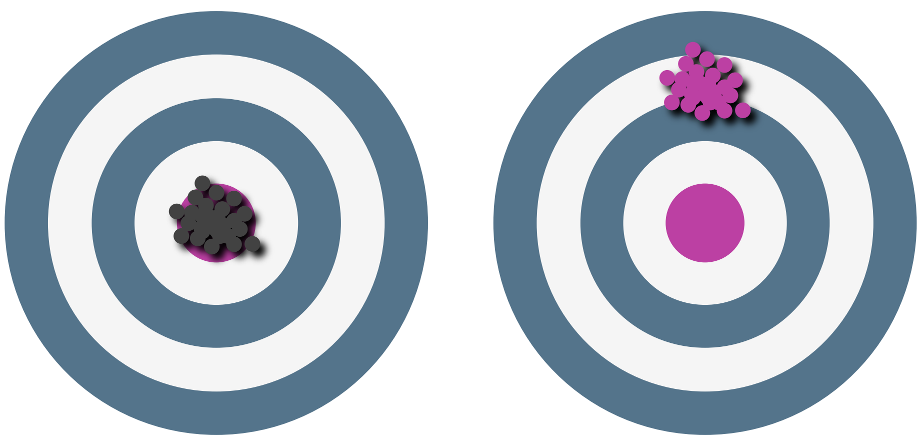
$$bias(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$$

Простым языком: если при фиксированном n мы постоянно используем нашу оценку, в среднем мы не ошибаемся

Несмещённость



Оценка 1



Оценка 2

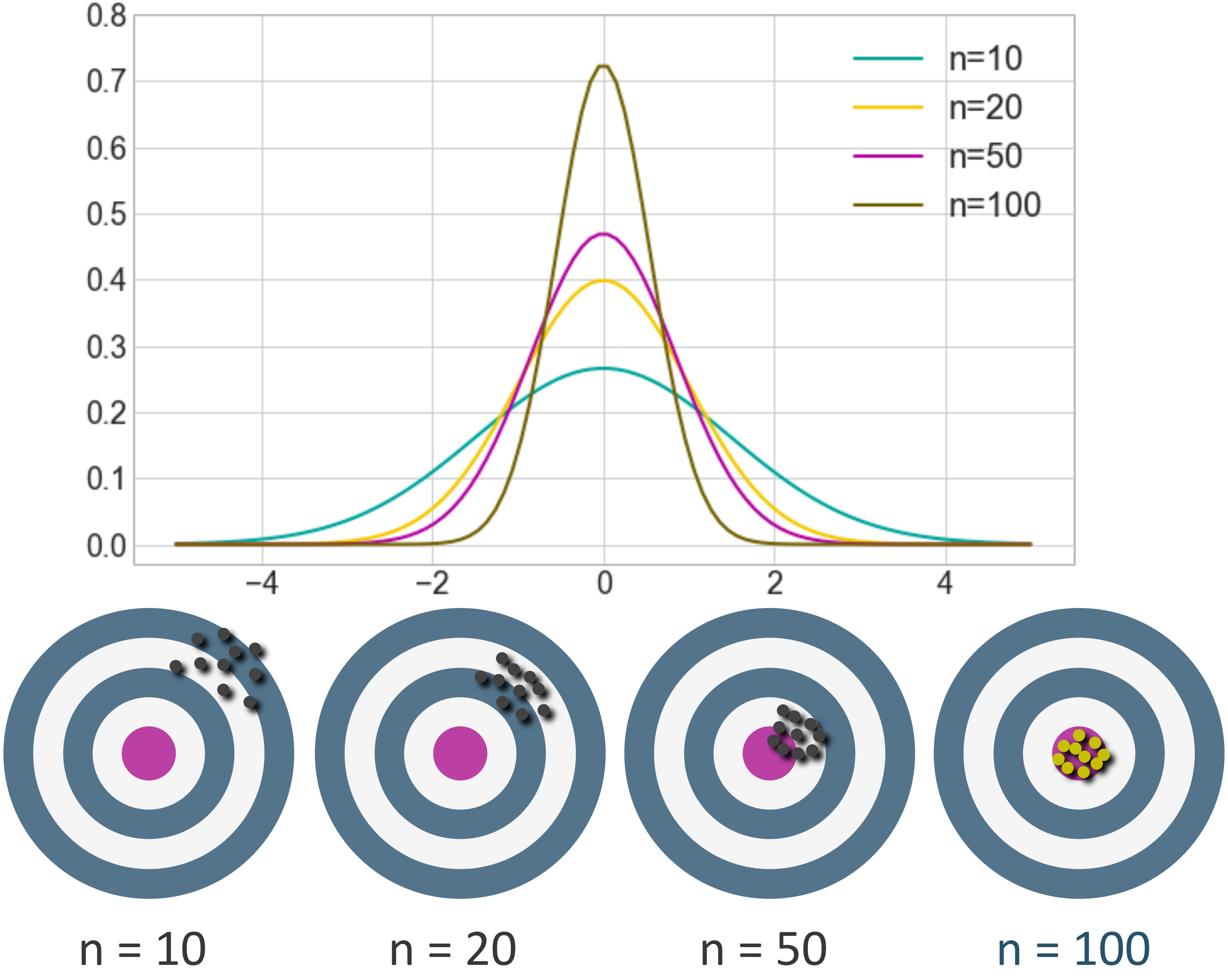
Состоятельность

Оценка называется состоятельной, если она сходится по вероятности к истинному значению параметра при $n \to \infty$

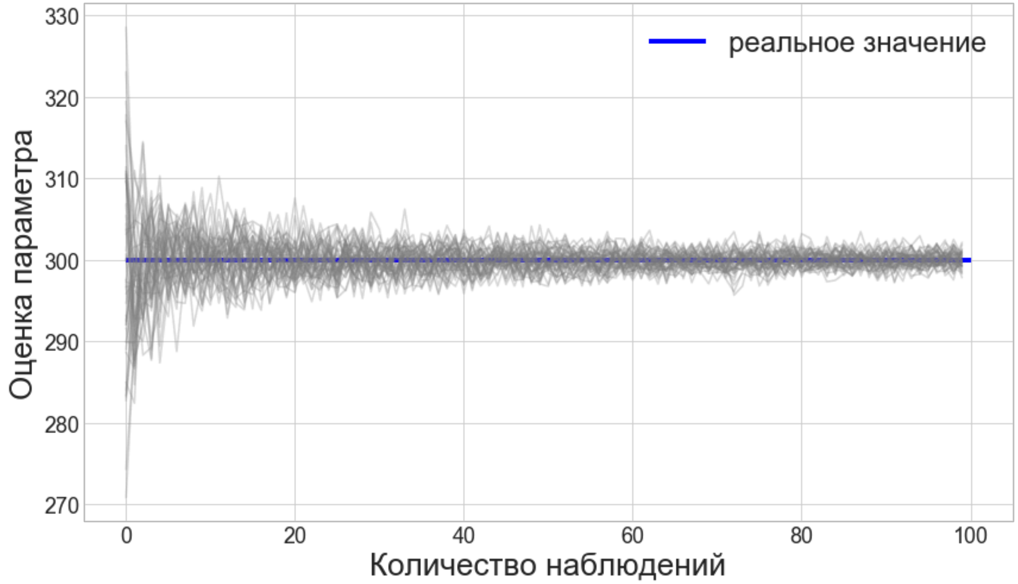
$$\hat{\theta} \xrightarrow{p} \theta$$

Простым языком: чем больше наблюдений, тем мы ближе к истине

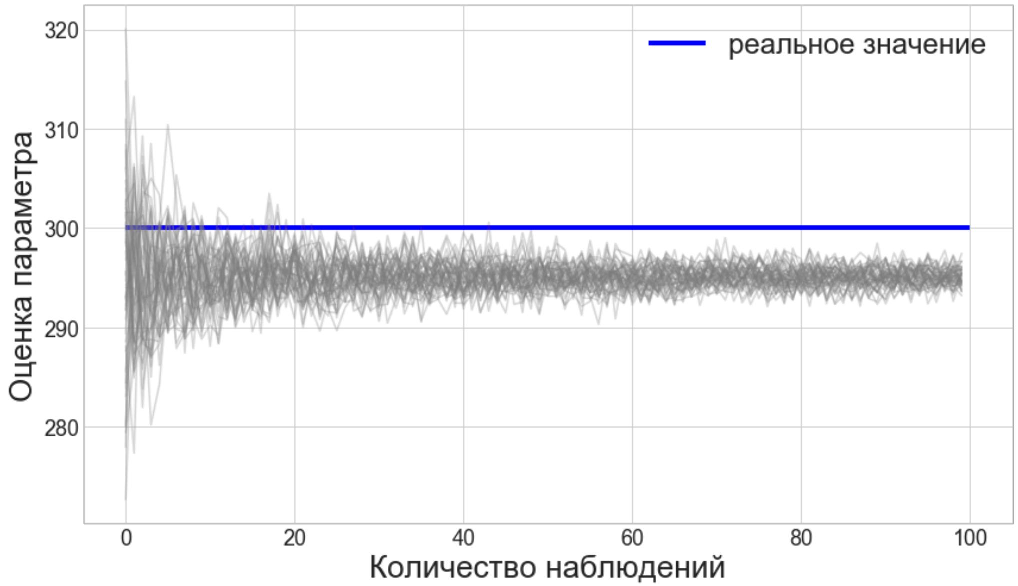
Состоятельность



Состоятельность



Состоятельная оценка



Несостоятельная оценка

Асимптотическая несмещённость

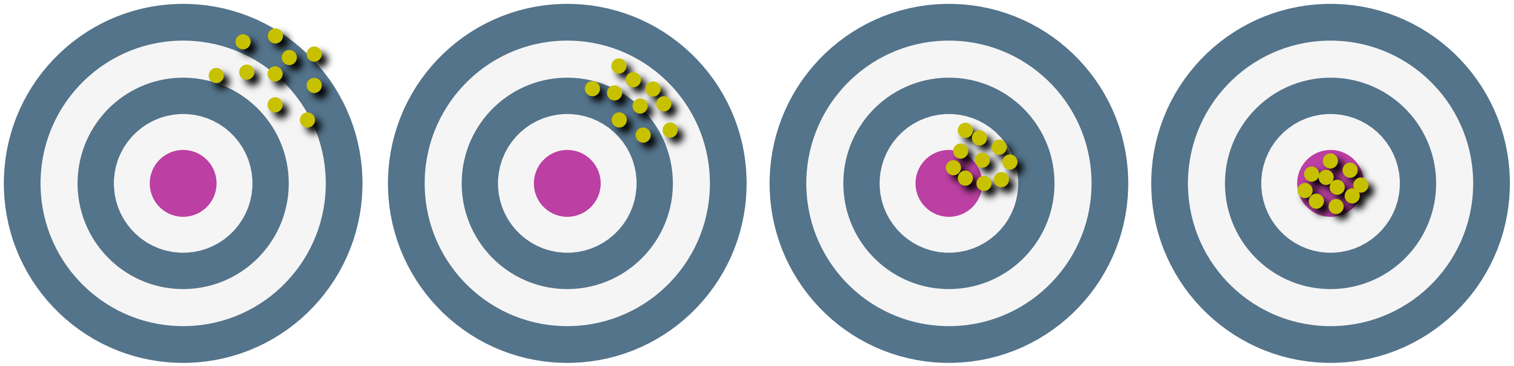
Оценка называется асимптотически несмещённой, если её математическое ожидание сходится к оцениваемому параметру при $n \to \infty$:

$$\mathbb{E}(\hat{\theta}) \to \theta$$

Простым языком: если мы постоянно используем нашу оценку, в среднем, при очень больших n, мы не ошибаемся

Состоятельность VS асимптотическая несмёщенность

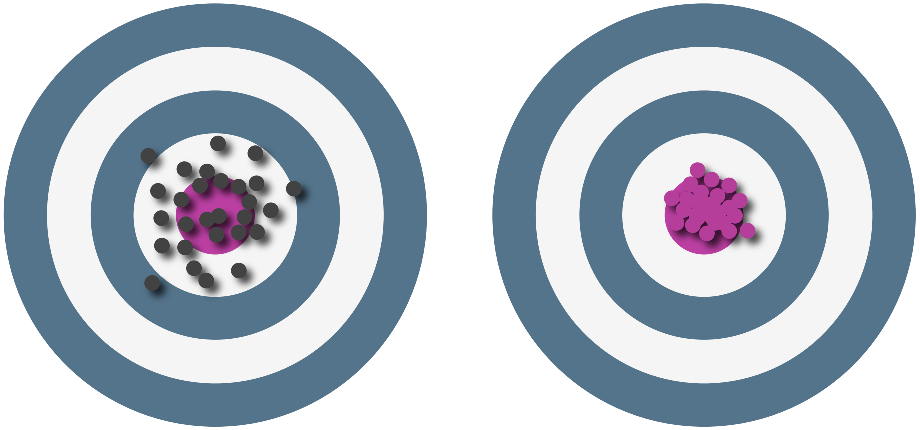
Асимптотически несмещённая и состоятельная оценка



Сравнение оценок

Несмещённых и состоятельных оценок может оказаться несколько ⇒ нужно научиться их сравнивать

Оценка 1



Оценка 2

Сравнение оценок

Несмещённых и состоятельных оценок может оказаться несколько $\Rightarrow$ нужно научиться их сравнивать

Обычно оценки между собой сравнивают с помощью квадратичной ошибки:

$$MSE = \mathbb{E}\left(\hat{\theta} - \theta\right)^2$$

Для несмещённых оценок MSE совпадает с дисперсией оценки

Простым языком: чем более предсказуема оценка, тем точнее прогноз (уже доверительный интервал)

Резюме

Статистик хочет получить:

- несмещённую оценку – хочет в среднем не ошибаться при фиксированном размере выборки
- состоятельную оценку – хочет при большом числе наблюдений быть близко к реальности
- оценку с маленькой средней квадратичной ошибкой

Великая дилемма: смещение против разброса

Сравнение оценок

Несмещённых и состоятельных оценок может оказаться несколько $\Rightarrow$ нужно научиться их сравнивать

Обычно оценки между собой сравнивают с помощью квадратичной ошибки:

$$MSE = \mathbb{E}\left(\hat{\theta} - \theta\right)^2$$

Для несмещённых оценок MSE совпадает с дисперсией оценки

Простым языком: чем более предсказуема оценка, тем точнее прогноз (уже доверительный интервал)

Разложение на смещение и разброс

Для того, чтобы сравнить оценки можно выбрать любую другую функцию потерь, но у MSE есть несколько хороших свойств

MSE можно представить в виде суммы смещения и разброса:

Разложение на смещение и разброс

Для того, чтобы сравнить оценки можно выбрать любую другую функцию потерь, но у MSE есть несколько хороших свойств

MSE можно представить в виде суммы смещения и разброса:

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

Разложение на смещение и разброс

Для того, чтобы сравнить оценки можно выбрать любую другую функцию потерь, но у MSE есть несколько хороших свойств

MSE можно представить в виде суммы смещения и разброса:

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

$$= \mathbb{E}\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right)^2 + 2 \cdot \mathbb{E}\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right) \cdot \mathbb{E}\big(\mathbb{E}(\hat{\theta}) - \theta\big) + \mathbb{E}\big(\mathbb{E}(\hat{\theta}) - \theta\big)^2 =$$

Разложение на смещение и разброс

Для того, чтобы сравнить оценки можно выбрать любую другую функцию потерь, но у MSE есть несколько хороших свойств

MSE можно представить в виде суммы смещения и разброса:

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

$$= \mathbb{E}\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right)^2 + 2 \cdot \mathbb{E}\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right) \cdot \mathbb{E}(\mathbb{E}(\hat{\theta}) - \theta) + \mathbb{E}(\mathbb{E}(\hat{\theta}) - \theta)^2 =$$

$$= Var(\hat{\theta}) + 2 \cdot \left(\mathbb{E}(\hat{\theta}) - \mathbb{E}(\hat{\theta})\right) \cdot (\mathbb{E}(\hat{\theta}) - \theta) + (\mathbb{E}(\hat{\theta}) - \theta)^2 =$$

Разложение на смещение и разброс

Для того, чтобы сравнить оценки можно выбрать любую другую функцию потерь, но у MSE есть несколько хороших свойств

MSE можно представить в виде суммы смещения и разброса:

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

$$= \mathbb{E}\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right)^2 + 2 \cdot \mathbb{E}\left(\hat{\theta} - \mathbb{E}(\hat{\theta})\right) \cdot \mathbb{E}(\mathbb{E}(\hat{\theta}) - \theta) + \mathbb{E}(\mathbb{E}(\hat{\theta}) - \theta)^2 =$$

$$= Var(\hat{\theta}) + 2 \cdot \left(\mathbb{E}(\hat{\theta}) - \mathbb{E}(\hat{\theta})\right) \cdot (\mathbb{E}(\hat{\theta}) - \theta) + (\mathbb{E}(\hat{\theta}) - \theta)^2 =$$

$$= Var(\hat{\theta}) + \text{bias}^2(\hat{\theta})$$

Bias-variance decomposition

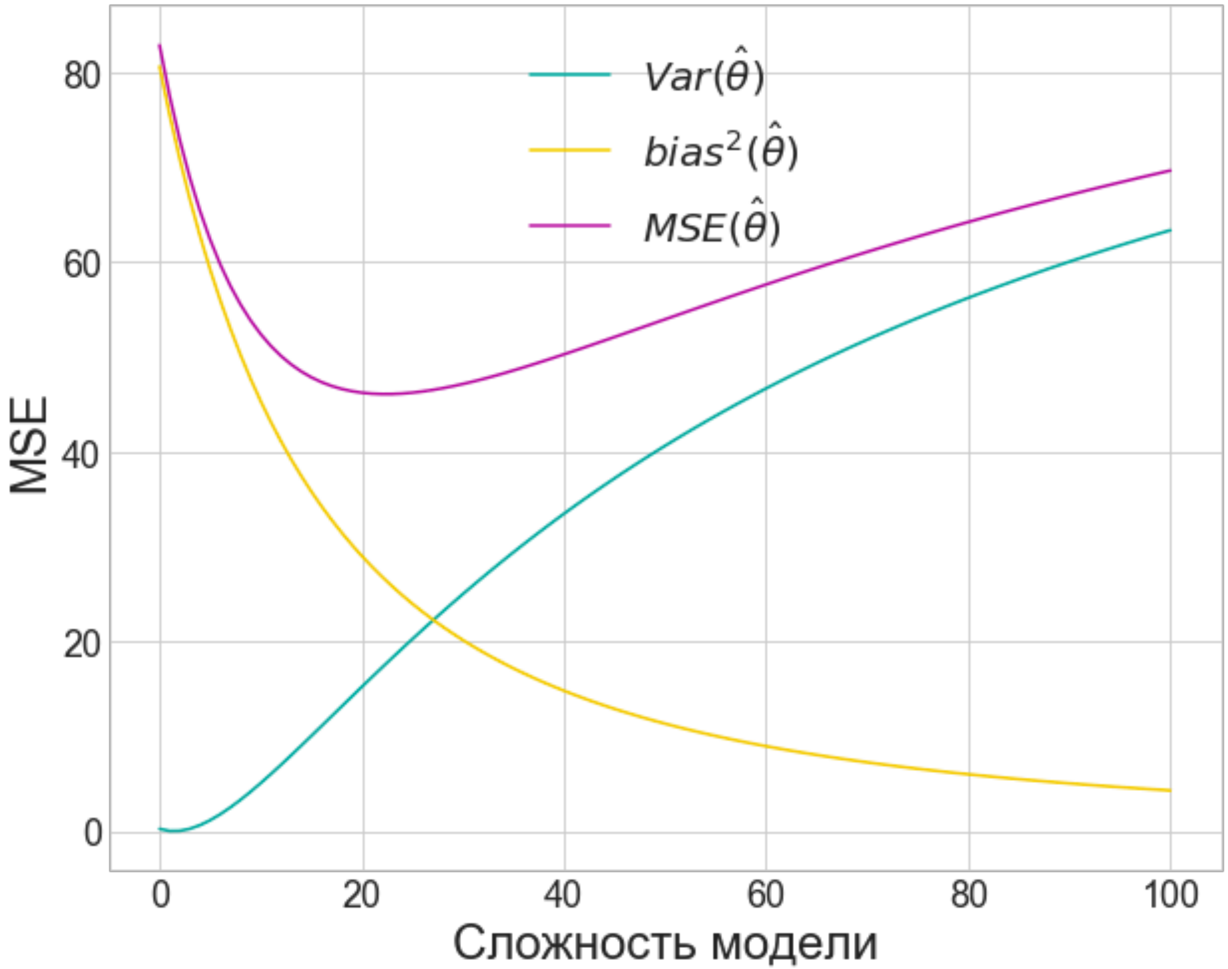
Между смещением и разбросом можно искать компромисс, это позволяет уменьшить среднеквадратичную ошибку

$$MSE = \mathbb{E}\left(\hat{\theta} - \theta\right)^2 = Var\left(\hat{\theta}\right) + \mathrm{bias}^2(\hat{\theta})$$

Bias-variance decomposition

$$MSE = \mathbb{E}\left(\hat{\theta} - \theta\right)^2 = Var\left(\hat{\theta}\right) + \text{bias}^2(\hat{\theta})$$

Bias-variance decomposition



$$MSE = \mathbb{E}\left(\hat{\theta} - \theta\right)^2 = Var\left(\hat{\theta}\right) + \mathrm{bias}^2(\hat{\theta})$$

Эффективность оценок

Эффективность

Между смещением и разбросом можно искать компромисс, это позволяет уменьшить среднеквадратичную ошибку

$$MSE = \mathbb{E}\left(\hat{\theta} - \theta\right)^2 = Var\left(\hat{\theta}\right) + \mathrm{bias}^2(\hat{\theta})$$

В классе всех возможных оценок наилучшей в смысле среднеквадратического подхода не существует

Можно попробовать зафиксировать смещение и найти оценку с наименьшей дисперсией

Эффективность

Можно попробовать зафиксировать смещение и найти оценку с наименьшей дисперсией

Такая оценка называется эффективной в классе со смещением $\text{bias}(\hat{\theta})$

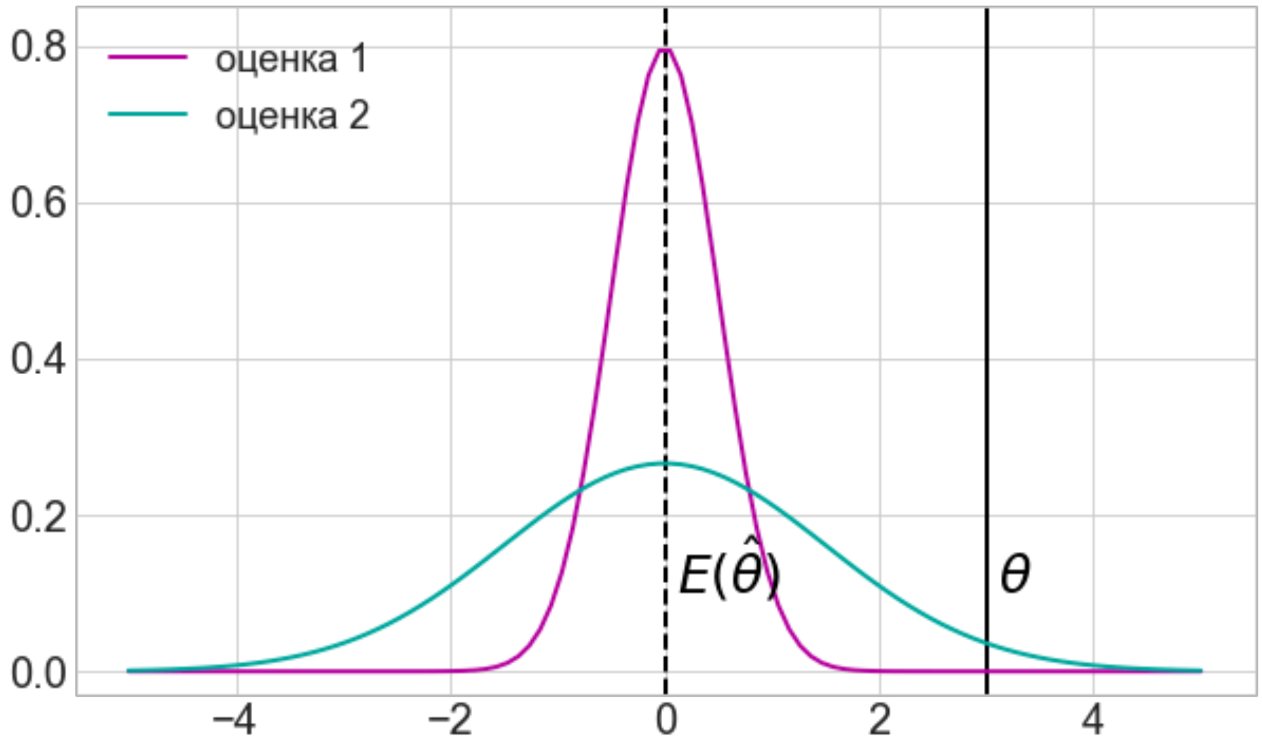
Нас будут интересовать несмещённые эффективные оценки

Простым языком: эффективная оценка обладает самым узким доверительным интервалов в своём классе

Эффективность

У оценок одинаковое смещение (класс), но при этом у оценки 1 дисперсия меньше

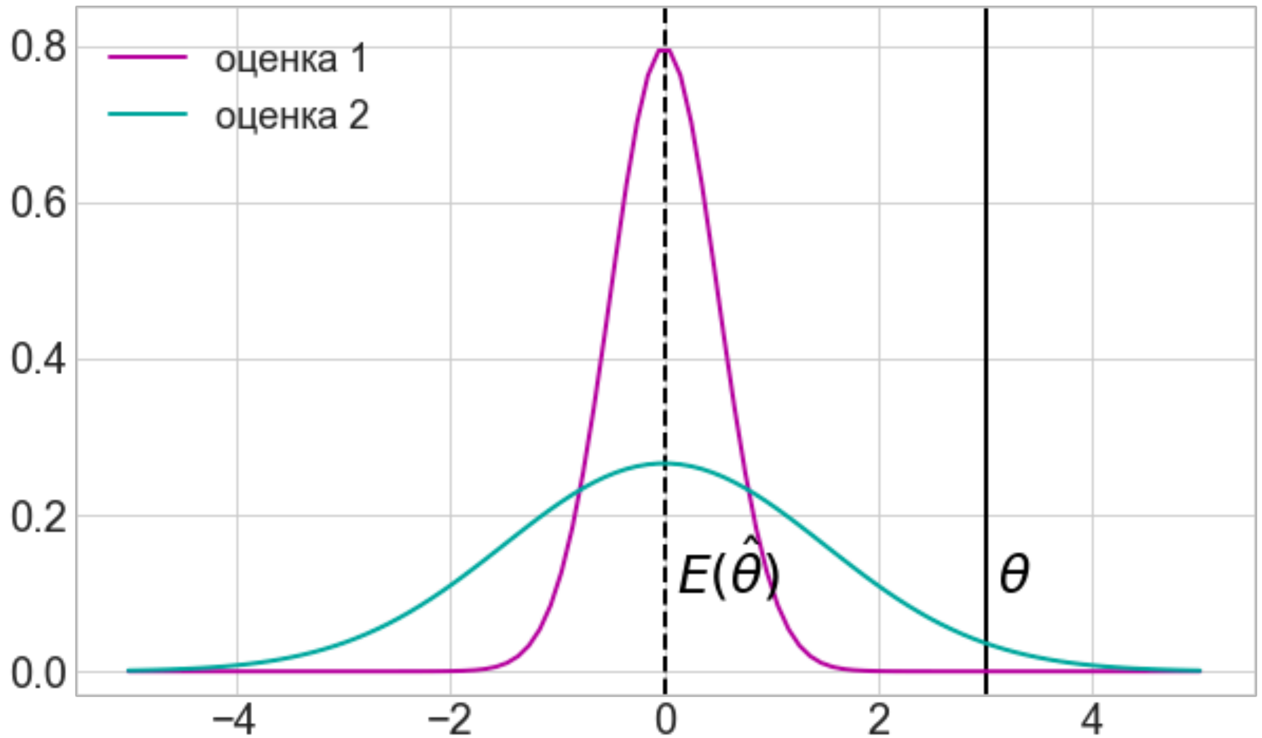
Если у оценки 1 самая маленькая дисперсия из всех существующих ⇒ она для нас самая предпочтительная



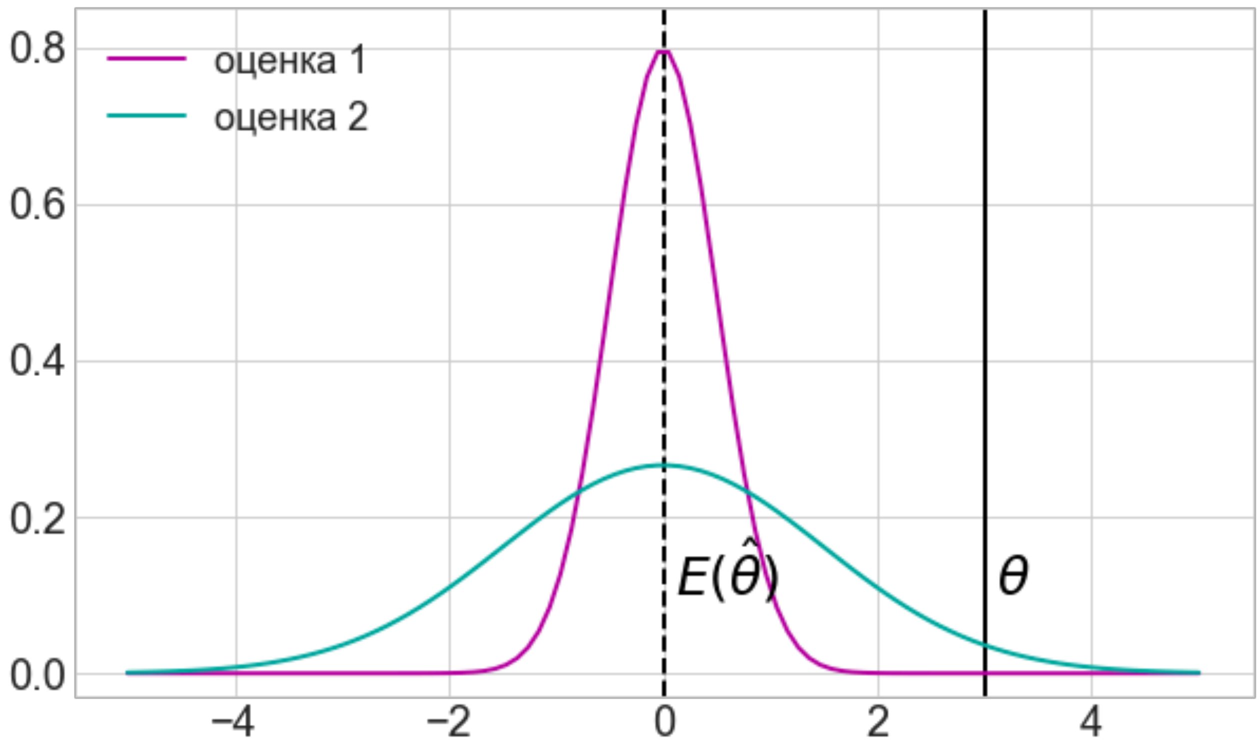
Эффективность

То есть оценка 1 эффективная в классе с таким смещением

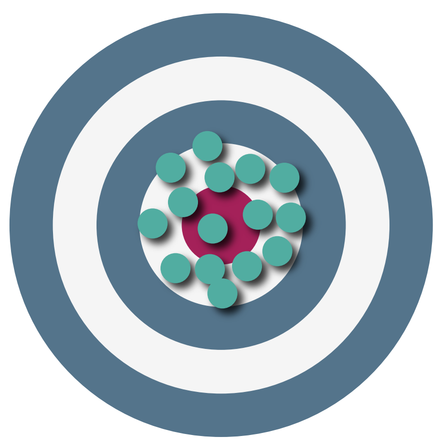
Нас будут интересовать несмещённые эффективные оценки



Эффективность



Оценка 1



Оценка 2

Неравенство Рао-Фреше-Крамера

Для функции потерь MSE существует теоретическая нижняя граница, её называют неравенством Рао Фреше Крамера

Неравенство Рао-Фреше-Крамера

Если оценка параметра несмещена и выполнены условия регулярности:

1. Область определения случайной величины не зависит от параметра θ
2. Сложное техническое условие, разрешающее брать производные (обычно формулируется по-разному)
3. Существует конечная положительная информация Фишера

$$J(\theta) = \mathbb{E}\left(\frac{\partial \ln f(x,\theta)}{\partial \theta}\right)^2$$

$f(x,\theta)$ – плотность распределения для непрерывных случайных величин и вероятность для дискретных

Неравенство Рао-Фреше-Крамера

Тогда для дисперсии оценки выполняется неравенство Рао Фреше Крамера:

$$Var(\hat{\theta}) \geq \frac{1}{n \cdot J(\theta)}$$

Если оказалось, что $Var(\hat{\theta}) = \frac{1}{n \cdot J(\theta)}$, тогда оценка эффективна

Точно такое же неравенство можно выписать для смещённых оценок:

$$Var(\hat{\theta}) \geq \frac{(1 + bias'_\theta)^2}{n \cdot J(\theta)}$$

Резюме

Резюме

Несмещённость – много раз используя оценку, при фиксированном размере выборки, в среднем, мы не ошибаемся

Как проверить:

По-честному найти $\mathbb{E}(\hat{\theta})$ и сравнить его с θ

Резюме

Состоятельность – последовательность оценок при увеличении числа наблюдений сходится к истинному значению параметра

Как проверить:

- используя ЗБЧ найти к чему сходится оценка
- использовать условие Чебышёва:

Если оценка несмещённая и её дисперсия при росте n стремится к нулю ⇒ она состоятельная:

$$\mathbb{E}(\bar{x}) = \mu$$

$$Var(\bar{x}) = \frac{\sigma^2}{n} \to 0 \text{ при } n \to \infty$$

Резюме

Оценки можно сравнивать между собой с помощью различных функций потерь, обычно используют MSE:

$$MSE = \mathbb{E}\left(\hat{\theta} - \theta\right)^2 = Var\left(\hat{\theta}\right) + \text{bias}^2(\hat{\theta})$$

Поиск компромисса между смещением и разбросом позволяет уменьшить MSE, этим часто пользуются в машинном обучении **(регуляризация)**

В классе всех возможных оценок наилучшей в смысле MSE оценки не существует

Обычно нас интересуют несмещённые оценки

Резюме

Эффективность – хотим самый узкий доверительный интервал ⇒ ищем оценку с самой маленькой дисперсией в каком-то классе

Как проверить:

- иногда помогает неравенство Рао-Фреше-Крамера

$$Var(\hat{\theta}) \geq \frac{1}{n \cdot J(\theta)}$$

Если мы получили равенство, оценка эффективна.
Если нет, мы не можем сказать про неё ничего конкретного,
и нужна более мощная процедура
для проверки

Метод максимального правдоподобия

Схема математической статистики

Выборка: $x_1, \ldots, x_n$ Параметр: θ

$\hat{\theta}$ → $f_{\hat{\theta}}(t)$

Точность оценки, прогнозов

доверительные интервалы

Ответы на вопросы

проверка гипотез

Как оценить

- Метод моментов
- Метод максимального правдоподобия

Хорошие свойства

- Несмещенная
- Состоятельная
- Эффективная

Союзники

Асимптотические (при большом n)

- ЦПТ
- Дельта-метод

Точные

- Теорема Фишера
- $\chi^2_n, t_n, F_{n,k}$
- Ещё союзники!

Схема математической статистики

Выборка: $x_1, \ldots, x_n$ Параметр: θ

$\hat{\theta} \longrightarrow f_{\hat{\theta}}(t)$

Точность оценки, прогнозов

доверительные интервалы

Ответы на вопросы

проверка гипотез

Как оценить

- Метод моментов
- Метод максимального правдоподобия

Хорошие свойства

- Несмещенная
- Состоятельная
- Эффективная

Союзники

Асимптотические (при большом n)

- ЦПТ
- Дельта-метод

Точные

- Теорема Фишера
- $\chi^2_n, t_n, F_{n,k}$
- Ещё союзники!

Задача о фонтане

- Юра приехал в южный город и увидел, что там есть фонтан и он работает

$$1$$

- А как часто он работает?

Гипотезы:	Вероятности:
Фонтан работает раз в год	$\frac{1}{365}$
Фонтан работает каждые выходные	$\sim\frac{1}{3}$
Фонтан работает всегда	1

Максимальная вероятность увидеть фонтан при верности гипотезы

vk.com

Задача о фонтане

- Параметр θ – работоспособность фонтана
- Наблюдение x_1 – сегодня фонтан работал

Задача: Максимизировать вероятность появления выборки по значению θ:

$$\mathbb{P}(\, x_1 = \text{работает} \mid \theta \,) \to \max_{\theta}$$

$$\mathbb{P}(\, x_1 = \text{работает} \mid \theta = \text{раз в году} \,) = \frac{1}{365}$$

$$\mathbb{P}(\, x_1 = \text{работает} \mid \theta = \text{по выходным} \,) = \frac{1}{3}$$

$$\boxed{\mathbb{P}(\, x_1 = \text{работает} \mid \theta = \text{каждый день} \,) = 1}$$

Правдоподобие

Правдоподобие (likelihood function) – вероятность получить наблюдаемую выборку при конкретном значении параметра

Оценка максимального правдоподобия – значение параметра, которое максимизирует правдоподобие

Правдоподобие

$$\mathrm{X}$$

Выборка: $x_1, \ldots, x_n$

Предположение: выборка пришла из распределения с плотностью $f(\, x \mid \theta \,)$.

Параметр θ (константа) мы не знаем и хотим оценить по выборке.

Правдоподобие

Правдоподобие выборки:

$$\mathrm{L}(\theta \mid x_1, \dots, x_n) = \mathbb{P}(x_1, \dots, x_n \mid \theta) = f(x_1, \dots, x_n \mid \theta)$$
$$= f(x_1 \mid \theta) \cdot f(x_2 \mid \theta) \cdot \ldots \cdot f(x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

При разных значениях θ мы получаем большую или меньшую вероятность получить наблюдаемые данные

Если выполнено неравенство

$$\mathrm{L}(\theta_1 \mid x_1, \dots, x_n) > \mathrm{L}(\theta_2 \mid x_1, \dots, x_n)$$

значение параметра θ_1 называют “более правдоподобным”

Метод максимального правдоподобия

Метод максимального правдоподобия состоит в выборе в качестве оценки $\hat{\theta}$ значения, при котором правдоподобие достигает максимума:

$$\mathrm{L}(\theta \mid x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i \mid \theta) \to \max_{\theta}$$

Оценка максимального правдоподобия (maximum likelihood estimation):

$$\hat{\theta}^{MLE} = \operatorname*{argmax}_{\theta} \mathrm{L}(\theta \mid x_1, \dots, x_n)$$

Метод максимального правдоподобия

С логарифмической функцией работать удобнее, поэтому правдоподобие обычно логарифмируют и ищут максимум:

$$\ln \mathrm{L}(\theta \mid x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f(x_i \mid \theta) \to \max_{\theta}$$

Возьмём производную и приравняем её к нулю:

$$\frac{\partial \ln L}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \ln f(x_i \mid \theta)}{\partial \theta} = 0$$

Решив это уравнение, получим оценку максимального правдоподобия

Резюме

- Метод максимального правдоподобия заключается в максимизации вероятности получить наблюдаемые данные по неизвестным параметрам
- Возможны ситуации, в которых функция правдоподобия не ограничена и MLE не существует
- Возможны ситуации, в которых функция правдоподобия достигает глобального максимума для нескольких θ
- Метод нельзя использовать, если не выполнены условия регулярности (область зависит от параметра или функция недифференцируема)

Дискретный пример

Пример: распределение Бернулли

$$X_i = \begin{cases} 1, \text{если любит кофе} \\ 0, \text{если не любит кофе} \end{cases}$$

X_i	0	1
$\mathbb{P}(X_i = k)$	$1 - p$	p

$$x_1 = 1, x_2 = 0, x_3 = 1, \ldots, x_n = 0 \sim iid\ Bern(p)$$

Задача: найти ML-оценку для p

$$L(\, p \mid x_1, \ldots, x_n \,) = \mathbb{P}(\, x_1, \ldots, x_n \mid p \,) =$$
$$= \mathbb{P}(x_1 \mid p) \cdot \mathbb{P}(x_2 \mid p) \cdot \mathbb{P}(x_3 \mid p) \cdot \ldots \cdot \mathbb{P}(x_n \mid p) =$$
$$= p \cdot (1 - p) \cdot p \cdot \ldots \cdot (1 - p) =$$
$$= p^{\sum x_i} \cdot (1 - p)^{n - \sum x_i} \to \max_p$$

Прологарифмируем:

$$\ln \mathrm{L} = \sum x_i \cdot \ln p + (n - \sum x_i) \cdot \ln(1 - p) \to \max_p$$

Пример: распределение Бернулли

Прологарифмируем:

$$\ln \mathrm{L} = \sum x_i \cdot \ln p + (n - \sum x_i) \cdot \ln(1 - p) \to \max_p$$

Возьмём производную:

$$\frac{\partial \ln L}{\partial p} = \frac{\sum x_i}{p} - \frac{(n - \sum x_i)}{1 - p}$$

Колпачки появляются после приравнивания к нулю

$$\frac{\sum x_i}{\hat{p}} - \frac{(n - \sum x_i)}{1 - \hat{p}} = 0$$

$$\sum x_i - \cancel{\hat{p} \cdot \sum x_i} = n \cdot \hat{p} - \cancel{\hat{p} \cdot \sum x_i}$$

$$\hat{p}^{ML} = \frac{1}{n} \sum x_i = \bar{x}$$

Непрерывный пример

Пример: нормальное распределение

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad x_1, x_2, \dots, x_n \sim iid\ N(\mu, \sigma^2)$$

Задача: найти ML-оценку для μ и σ^2

$$\mathrm{L}(\mu, \sigma^2 \mid x_1, \dots, x_n) = f(x_1, \dots, x_n \mid \mu, \sigma^2) =$$

$$= f(x_1 \mid \mu, \sigma^2) \cdot f(x_2 \mid \mu, \sigma^2) \cdot \dots \cdot f(x_n \mid \mu, \sigma^2) =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdot \dots \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x_n-\mu)^2}{2\sigma^2}} =$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \cdot e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}} \to \max_{\mu,\sigma^2}$$

Прологарифмируем:

$$\ln \mathrm{L} = -\frac{n}{2} \cdot \ln 2\pi - \frac{n}{2} \cdot \ln \sigma^2 - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2 \to \max_{\mu,\sigma^2}$$

Пример: нормальное распределение

Прологарифмируем:

$$\ln \mathrm{L} = -\frac{n}{2} \cdot \ln 2\,\pi - \frac{n}{2} \cdot \ln \sigma^2 - \frac{1}{2\,\sigma^2} \sum (x_i - \mu)^2 \to \max_{\mu, \sigma^2}$$

Возьмём производную:

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{2\,\sigma^2} \cdot (-2) \cdot \sum (x_i - \mu)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\,\sigma^4} \sum (x_i - \mu)^2$$

$$\begin{cases} -\dfrac{1}{2\,\hat{\sigma}^2} \cdot (-2) \cdot \sum (x_i - \hat{\mu}) = 0 \\ -\dfrac{n}{2} \cdot \dfrac{1}{\hat{\sigma}^2} + \dfrac{1}{2\,\hat{\sigma}^4} \sum (x_i - \hat{\mu})^2 = 0 \end{cases}$$

Пример: нормальное распределение

Прологарифмируем:

$$\ln \mathrm{L} = -\frac{n}{2} \cdot \ln 2\,\pi - \frac{n}{2} \cdot \ln \sigma^2 - \frac{1}{2\,\sigma^2} \sum (x_i - \mu)^2 \to \max_{\mu, \sigma^2}$$

Возьмём производную:

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{2\,\sigma^2} \cdot (-2) \cdot \sum (x_i - \mu)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\,\sigma^4} \sum (x_i - \mu)^2$$

$$\begin{cases} \sum (x_i - \hat{\mu}) = 0 \\ -n + \frac{1}{\hat{\sigma}^2} \sum (x_i - \hat{\mu})^2 = 0 \end{cases}$$

Пример: нормальное распределение

Прологарифмируем:

$$\ln \mathrm{L} = -\frac{n}{2} \cdot \ln 2\,\pi - \frac{n}{2} \cdot \ln \sigma^2 - \frac{1}{2\,\sigma^2} \sum (x_i - \mu)^2 \to \max_{\mu,\sigma^2}$$

Возьмём производную:

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{2\,\sigma^2} \cdot (-2) \cdot \sum (x_i - \mu)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\,\sigma^4} \sum (x_i - \mu)^2$$

$$\begin{cases} \sum x_i = n\,\hat{\mu} \\ -n + \dfrac{1}{\hat{\sigma}^2} \sum (x_i - \hat{\mu})^2 = 0 \end{cases}$$

Пример: нормальное распределение

Прологарифмируем:

$$\ln \mathrm{L} = -\frac{n}{2} \cdot \ln 2\,\pi - \frac{n}{2} \cdot \ln \sigma^2 - \frac{1}{2\,\sigma^2} \sum (x_i - \mu)^2 \to \max_{\mu, \sigma^2}$$

Возьмём производную:

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{2\,\sigma^2} \cdot (-2) \cdot \sum (x_i - \mu)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\,\sigma^4} \sum (x_i - \mu)^2$$

$$\begin{cases} \sum x_i = n\,\hat{\mu} \\ -n + \frac{1}{\hat{\sigma}^2} \sum (x_i - \hat{\mu})^2 = 0 \end{cases} \qquad \hat{\mu}_{ML} = \bar{x}$$

Пример: нормальное распределение

Прологарифмируем:

$$\ln \mathrm{L} = -\frac{n}{2} \cdot \ln 2\,\pi - \frac{n}{2} \cdot \ln \sigma^2 - \frac{1}{2\,\sigma^2} \sum (x_i - \mu)^2 \to \max_{\mu,\sigma^2}$$

Возьмём производную:

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{2\,\sigma^2} \cdot (-2) \cdot \sum (x_i - \mu)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\,\sigma^4} \sum (x_i - \mu)^2$$

$$\begin{cases} \sum x_i = n\,\hat{\mu} \\ -n + \frac{1}{\hat{\sigma}^2} \sum (x_i - \bar{x})^2 = 0 \end{cases} \qquad \hat{\mu}_{ML} = \bar{x}$$

Пример: нормальное распределение

Прологарифмируем:

$$\ln \mathrm{L} = -\frac{n}{2} \cdot \ln 2\,\pi - \frac{n}{2} \cdot \ln \sigma^2 - \frac{1}{2\,\sigma^2} \sum (x_i - \mu)^2 \to \max_{\mu,\sigma^2}$$

Возьмём производную:

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{2\,\sigma^2} \cdot (-2) \cdot \sum (x_i - \mu)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\,\sigma^4} \sum (x_i - \mu)^2$$

$$\begin{cases} \sum x_i = n\,\hat{\mu} \\ \frac{1}{\hat{\sigma}^2} \sum (x_i - \bar{x})^2 = n \end{cases} \qquad \hat{\mu}_{ML} = \bar{x}$$

Пример: нормальное распределение

Прологарифмируем:

$$\ln \mathrm{L} = -\frac{n}{2} \cdot \ln 2\,\pi - \frac{n}{2} \cdot \ln \sigma^2 - \frac{1}{2\,\sigma^2} \sum (x_i - \mu)^2 \to \max_{\mu,\sigma^2}$$

Возьмём производную:

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{2\,\sigma^2} \cdot (-2) \cdot \sum (x_i - \mu)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\,\sigma^4} \sum (x_i - \mu)^2$$

$$\begin{cases} \sum x_i = n\,\hat{\mu} \\ \dfrac{1}{\hat{\sigma}^2} \sum (x_i - \bar{x})^2 = n \end{cases} \qquad \begin{array}{c} \hat{\mu}_{ML} = \bar{x} \\ \hat{\sigma}^2_{ML} = \dfrac{1}{n} \sum (x_i - \bar{x})^2 \end{array}$$

Пример: нормальное распределение

Прологарифмируем:

$$\ln \mathrm{L} = -\frac{n}{2} \cdot \ln 2\,\pi - \frac{n}{2} \cdot \ln \sigma^2 - \frac{1}{2\,\sigma^2} \sum (x_i - \mu)^2 \to \max_{\mu,\sigma^2}$$

Возьмём производную:

$$\frac{\partial \ln L}{\partial \mu} = -\frac{1}{2\,\sigma^2} \cdot (-2) \cdot \sum (x_i - \mu)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2\,\sigma^4} \sum (x_i - \mu)^2$$

$$\begin{cases} \sum x_i = n\,\hat{\mu} \\ \frac{1}{\hat{\sigma}^2} \sum (x_i - \bar{x})^2 = n \end{cases}$$

$$\boxed{\begin{gathered} \hat{\mu}_{ML} = \bar{x} \\ \hat{\sigma}^2_{ML} = \frac{1}{n} \sum (x_i - \bar{x})^2 \end{gathered}}$$

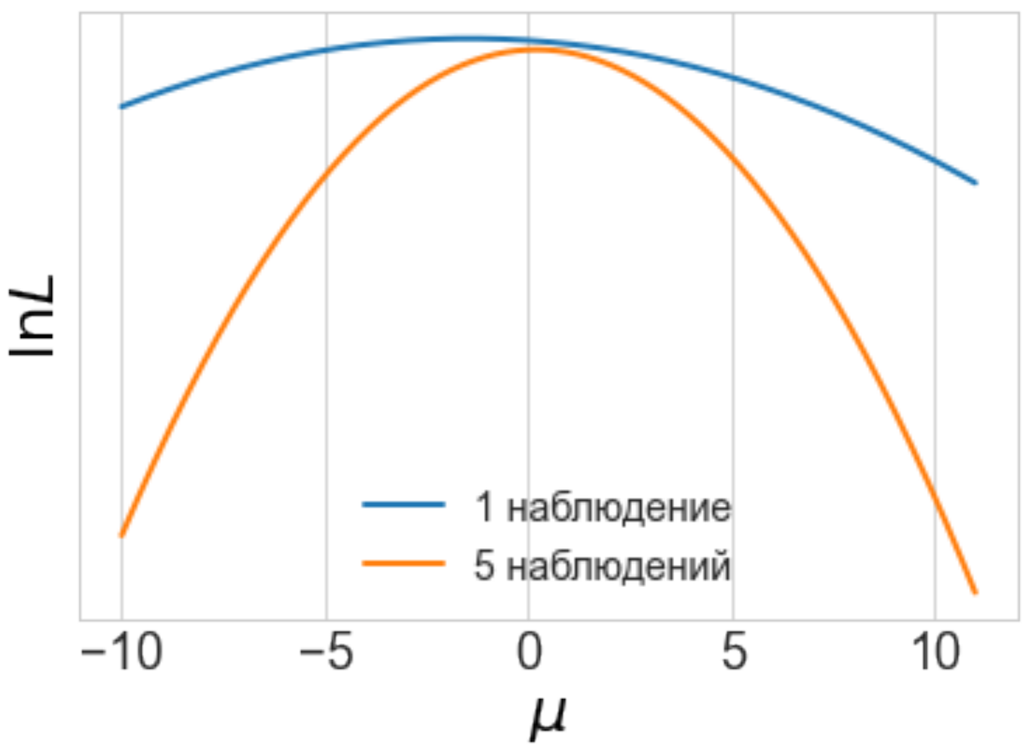
Информация Фишера

Точка максимума

- Одним из важнейших аспектов функции правдоподобия является её **поведение вблизи точки максимума**
- Если вблизи максимума функция достаточно плоская, то имеющиеся наблюдения **мало говорят о значениях параметров**
- Те же самые данные можно наблюдать с близкими вероятностями при разных значениях параметров
- Если функция имеет ярко выраженный пик, **данные имеют больше информации о параметрах**

Пример: $N(\mu, 1)$

- Для синей ситуации у нас мало информации, функция плоская
- Для красной ситуации у нас более яркий пик и более чёткая оценка



Накопление информации

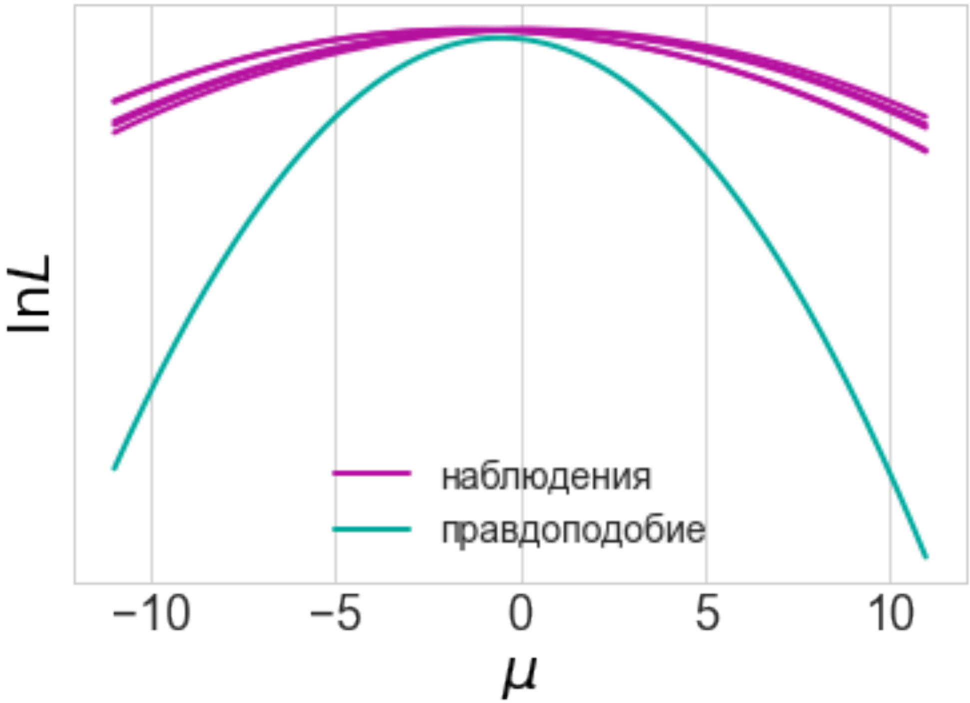
Логарифм правдоподобия:

$$\ln \mathrm{L}(\theta \mid x_1, \dots, x_n) = \sum_{i=1}^{n} \ln f(x_i \mid \theta)$$

- Одно слагаемое можно проинтерпретировать, как логарифм правдоподобия, вычисленный на основе одного наблюдения

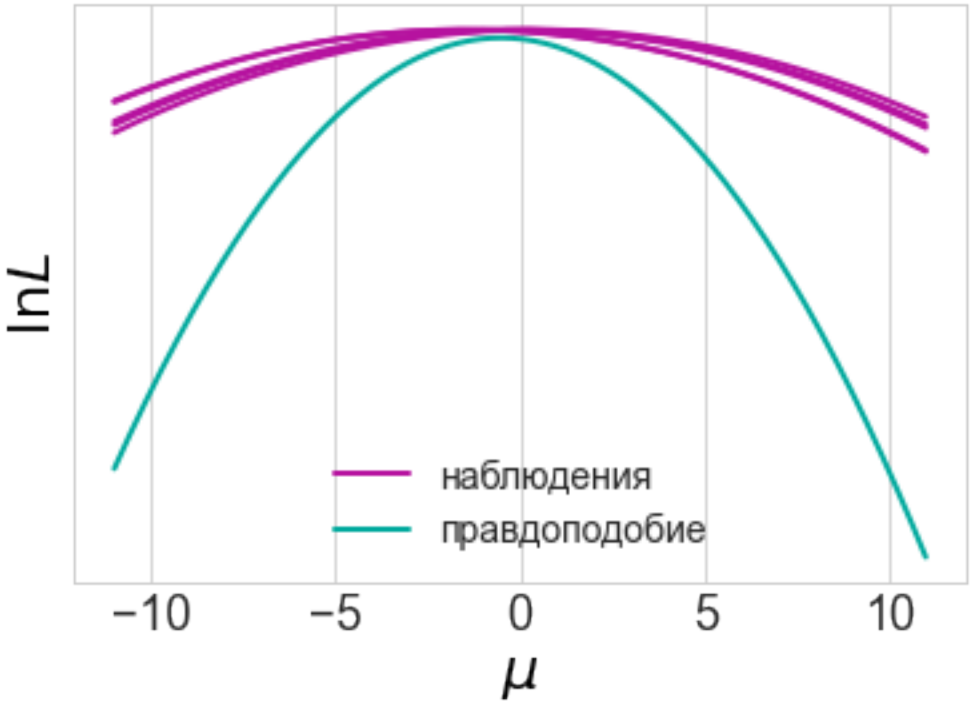
- Дополнительные слагаемые дают информацию о том, как ведёт себя правдоподобие для новых наблюдений

Пример: $N(\mu, 1)$



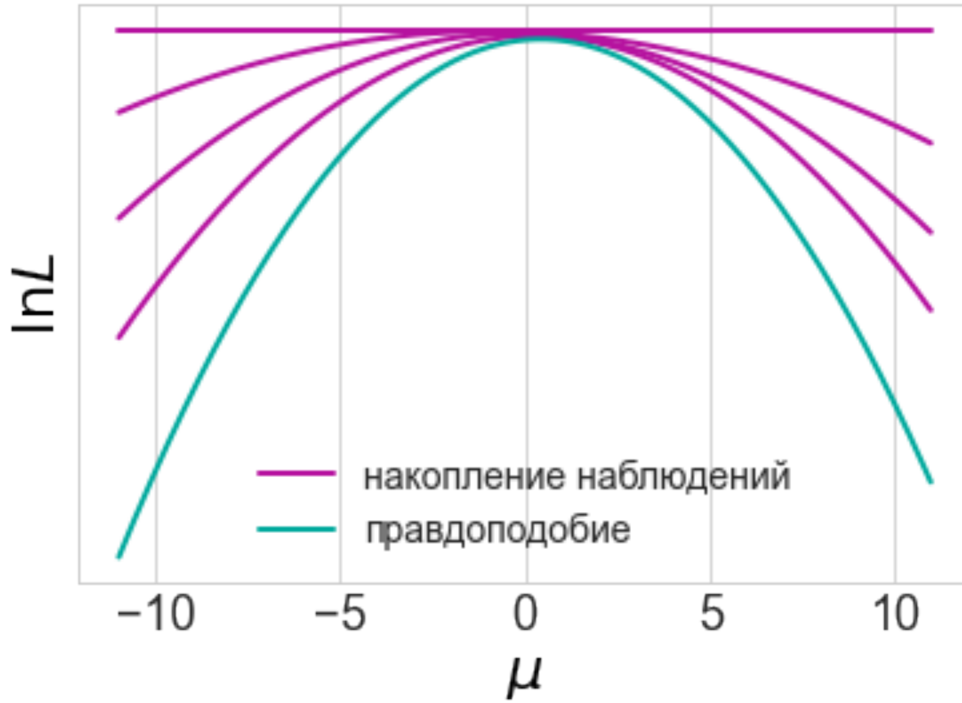
- Логарифмическая функция правдоподобия **для всей выборки** складывается как **сумма** логарифмических правдоподобий **отдельных наблюдений**

Пример: $N(\mu, 1)$



- Она имеет **более выраженный максимум,** чем больше выборка, тем ярче выражен максимум

Пример: $N(\mu, 1)$



- Каждая лиловая линия – добавление к сумме нового слагаемого
- С каждым слагаемым максимум становится более выраженным
- Каждое слагаемое добавляет нам **информацию**

Информация Фишера

- Чем выпуклее функция, тем чётче выражен максимум
- За выпуклость функции отвечает вторая производная, именно её, взятую со знаком минус, интерпретируют как **наблюденную информацию (observed information)**

$$J_o(\theta) = -\frac{\partial^2 \ln L}{\partial \theta^2}$$

Информация Фишера

- Если параметр векторный, то наблюденная информация описывается матрицей из вторых производных (матрица Гессе)

$$J_o(\theta) = -\left(\frac{\partial^2 \ln L}{\partial \theta_i \, \partial \theta_j}\right) = -H$$

Информация Фишера

- Математическое ожидание этой матрицы (по распределению наблюдений) называется **информационной матрицей Фишера**

$$J(\theta) = \mathbb{E}[J_o(\theta)] = -\mathbb{E}\left(\frac{\partial^2 \ln L}{\partial\theta_i\,\partial\theta_j}\right) = -\mathbb{E}(H)$$

Неравенство Рао-Крамера

Если функция плотности $f(x_i \mid \theta)$ удовлетворяет условиям регулярности, тогда для любой несмещённой оценки $\hat{\theta}$ выполняется неравенство Рао-Крамера:

$$Var\left(\hat{\theta}\right) \geq [J(\theta)]^{-1}$$

А также имеет место равенство:

$$J(\theta) = -\mathbb{E}\left(\frac{\partial^2 \ln L}{\partial\theta_i\, \partial\theta_j}\right) = -\mathbb{E}\left(\left[\frac{\partial \ln L}{\partial\theta}\right] \cdot \left[\frac{\partial \ln L}{\partial\theta}\right]^T\right)$$

Этим свойством можно пользоваться при проверке оценки на эффективность

Свойства ML-оценок

1. **Состоятельность:** $\operatorname*{plim}_{n\to\infty} \hat{\theta} = \theta$

2. **Асимптотическая эффективность:**

$$\lim_{n\to\infty} Var(\hat{\theta}) = [\, J(\theta)\,]^{-1}$$

3. **Асимптотическая нормальность:**

$$\hat{\theta} \overset{asy}{\sim} N(\theta, [\, J(\theta)\,]^{-1})$$

4. **Инвариантность:** если $\hat{\theta}$ – ML-оценка для θ, тогда если $\mathrm{g}(t)$ – непрерывная функция, то $\mathrm{g}(\hat{\theta})$ – ML-оценка для $\mathrm{g}(\theta)$

Асимптотическая нормальность

- ML-оценка асимптотически нормальна:

$$\hat{\theta} \sim N(\theta, [\,J(\theta)\,]^{-1})$$

- Это используют для строительства доверительных интервалов

- Нужно найти оценку для ковариационной матрицы (дисперсии) $[\,J(\theta)\,]^{-1}$

$$J\,(\theta) = -\,\mathbb{E}(\,H\,)$$

Пример: Нормальное распределение

Пример: нормальное распределение

Прологарифмируем:

$$\ln \mathrm{L} = -\frac{n}{2} \cdot \ln 2\,\pi - \frac{n}{2} \cdot \ln \sigma^2 - \frac{1}{2\,\sigma^2} \sum (x_i - \mu)^2 \to \max_{\mu, \sigma^2}$$

Первые производные:

$$\begin{cases} \dfrac{\partial \ln L}{\partial \mu} = \dfrac{1}{\sigma^2} \cdot \sum (x_i - \mu) \\ \dfrac{\partial \ln L}{\partial \sigma^2} = -\dfrac{n}{2\,\sigma^2} + \dfrac{1}{2\,\sigma^4} \sum (x_i - \mu)^2 \end{cases}$$

$$\boxed{\begin{aligned} \hat{\mu}_{ML} &= \bar{x} \\ \hat{\sigma}^2_{ML} &= \frac{1}{n} \sum (x_i - \bar{x})^2 \end{aligned}}$$

Пример: нормальное распределение

Первые производные:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \cdot \sum (x_i - \mu) \qquad \frac{\partial \ln L}{\partial \sigma^2} = \frac{\sum (x_i - \mu)^2}{2\,\sigma^4} - \frac{n}{2\sigma^2}$$

Вторые производные:

$$H = \begin{pmatrix} \dfrac{\partial^2 \ln L}{\partial \mu^2} & \dfrac{\partial^2 \ln L}{\partial \mu\, \partial \sigma^2} \\ \dfrac{\partial^2 \ln L}{\partial \sigma^2\, \partial \mu} & \dfrac{\partial^2 \ln L}{\partial (\sigma^2)^2} \end{pmatrix}$$

Пример: нормальное распределение

Первые производные:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \cdot \sum (x_i - \mu) \qquad \frac{\partial \ln L}{\partial \sigma^2} = \frac{\sum (x_i - \mu)^2}{2\,\sigma^4} - \frac{n}{2\sigma^2}$$

Вторые производные:

$$H = \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4}\sum (x_i - \mu) \\ -\frac{1}{\sigma^4}\sum (x_i - \mu) & -\frac{\sum (x_i - \mu)^2}{\sigma^6} + \frac{n}{2\,\sigma^4} \end{pmatrix}$$

$\mathbb{E}(H) - ?$

Пример: нормальное распределение

Первые производные:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \cdot \sum (x_i - \mu) \qquad \frac{\partial \ln L}{\partial \sigma^2} = \frac{\sum (x_i - \mu)^2}{2\,\sigma^4} - \frac{n}{2\sigma^2}$$

Вторые производные:

$$\mathbb{E}(H) = \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4} \sum (x_i - \mu) \\ -\frac{1}{\sigma^4} \sum (x_i - \mu) & -\frac{\sum (x_i - \mu)^2}{\sigma^6} + \frac{n}{2\,\sigma^4} \end{pmatrix}$$

$$\mathbb{E}\left(-\frac{n}{\sigma^2}\right) = -\frac{n}{\sigma^2}$$

Пример: нормальное распределение

Первые производные:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \cdot \sum (x_i - \mu) \qquad \frac{\partial \ln L}{\partial \sigma^2} = \frac{\sum (x_i - \mu)^2}{2\,\sigma^4} - \frac{n}{2\sigma^2}$$

Вторые производные:

$$\mathbb{E}(H) = \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4}\sum (x_i - \mu) \\ -\frac{1}{\sigma^4}\sum (x_i - \mu) & -\frac{\sum (x_i - \mu)^2}{\sigma^6} + \frac{n}{2\,\sigma^4} \end{pmatrix}$$

$$\mathbb{E}\left(-\frac{1}{\sigma^4}\sum (x_i - \mu)\right) = \frac{1}{\sigma^4}\sum (\mathbb{E}(x_i) - \mu) = \frac{1}{\sigma^4}\sum (\mu - \mu) = 0$$

Пример: нормальное распределение

Первые производные:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \cdot \sum (x_i - \mu) \qquad \frac{\partial \ln L}{\partial \sigma^2} = \frac{\sum (x_i - \mu)^2}{2\,\sigma^4} - \frac{n}{2\sigma^2}$$

Вторые производные:

$$\mathbb{E}(H) = \begin{pmatrix} -\frac{n}{\sigma^2} & 0 \\ 0 & -\frac{\sum (x_i - \mu)^2}{\sigma^6} + \frac{n}{2\,\sigma^4} \end{pmatrix}$$

$$\mathbb{E}\left(-\frac{\sum (x_i - \mu)^2}{\sigma^6} + \frac{n}{2\,\sigma^4}\right) = -\frac{n \cdot \sigma^2}{\sigma^6} + \frac{n}{2\,\sigma^4} = -\frac{n}{2\,\sigma^4}$$

Пример: нормальное распределение

Первые производные:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \cdot \sum (x_i - \mu) \qquad \frac{\partial \ln L}{\partial \sigma^2} = \frac{\sum (x_i - \mu)^2}{2\,\sigma^4} - \frac{n}{2\sigma^2}$$

Вторые производные:

$$\mathbb{E}(H) = \begin{pmatrix} -\frac{n}{\sigma^2} & 0 \\ 0 & -\frac{n}{2\,\sigma^4} \end{pmatrix}$$

Пример: нормальное распределение

Первые производные:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \cdot \sum (x_i - \mu) \qquad \frac{\partial \ln L}{\partial \sigma^2} = \frac{\sum (x_i - \mu)^2}{2\,\sigma^4} - \frac{n}{2\sigma^2}$$

Вторые производные:

$$J(\theta) = -\,\mathbb{E}(H) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\,\sigma^4} \end{pmatrix}$$

Пример: нормальное распределение

Первые производные:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \cdot \sum (x_i - \mu) \qquad \frac{\partial \ln L}{\partial \sigma^2} = \frac{\sum (x_i - \mu)^2}{2\,\sigma^4} - \frac{n}{2\sigma^2}$$

Вторые производные:

$$J(\theta) = -\,\mathbb{E}(H) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\,\sigma^4} \end{pmatrix}$$

$$[\,J(\theta)\,]^{-1} = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\,\sigma^4} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\,\sigma^4}{n} \end{pmatrix}$$

Пример: нормальное распределение

Первые производные:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \cdot \sum (x_i - \mu) \qquad \frac{\partial \ln L}{\partial \sigma^2} = \frac{\sum (x_i - \mu)^2}{2\,\sigma^4} - \frac{n}{2\sigma^2}$$

Вторые производные:

$$J(\theta) = -\,\mathbb{E}(H) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\,\sigma^4} \end{pmatrix}$$

$$\widehat{Var}\left(\hat{\theta}\right) = [\,\hat{j}(\theta)\,]^{-1} = \begin{pmatrix} \frac{\hat{\sigma}_{ML}^2}{n} & 0 \\ 0 & \frac{2\,\hat{\sigma}_{ML}^4}{n} \end{pmatrix}$$

Пример: нормальное распределение

$$\boxed{\begin{gathered}\hat{\mu}_{ML} = \bar{x} \\ \hat{\sigma}^2_{ML} = \frac{1}{n}\sum(x_i - \bar{x})^2\end{gathered}}$$

$$\begin{pmatrix}\hat{\mu}_{ML} \\ \hat{\sigma}^2_{ML}\end{pmatrix} \overset{asy}{\sim} N\left[\begin{pmatrix}\mu \\ \sigma^2\end{pmatrix}, \quad \begin{pmatrix}\frac{\hat{\sigma}^2_{ML}}{n} & 0 \\ 0 & \frac{2\,\hat{\sigma}^4_{ML}}{n}\end{pmatrix}\right]$$

Асимптотические доверительные интервалы:

$$\hat{\mu}_{ML} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\sigma}^2_{ML}}{n}} \qquad \hat{\sigma}^2_{ML} \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{2\,\hat{\sigma}^4_{ML}}{n}}$$

Резюме

- В анализе данных производные обычно используются для передачи информации
- Информация Фишера вычисляется как матрица из вторых производных и говорит о “крутости” максимума
- Если функция $f(x_i \mid \theta)$ удовлетворяет условиям регулярности, ML-оценка обладает оптимальными свойствами в асимптотическом плане

Резюме

- ML-оценка имеет асимптотически нормальное распределение, для поиска её дисперсии используют информацию Фишера
- В “сложных” (нерегулярных) случаях ML-оценка может терять эти свойства

Дельта-метод

Свойства ML-оценок

1. **Состоятельность:** $\operatorname*{plim}_{n\to\infty} \hat{\theta} = \theta$

2. **Асимптотическая эффективность:**

$$\lim_{n\to\infty} Var(\hat{\theta}) = [\,J(\theta)\,]^{-1}$$

3. **Асимптотическая нормальность:**

$$\hat{\theta} \overset{asy}{\sim} N(\theta, [\,J(\theta)\,]^{-1})$$

4. **Инвариантность:** если $\hat{\theta}$ – ML-оценка для θ, тогда если $\mathrm{g}(t)$ – непрерывная функция, то $\mathrm{g}(\hat{\theta})$ – ML-оценка для $\mathrm{g}(\theta)$

Дельта-метод

Если:

$$\hat{\theta} \overset{asy}{\sim} N(\theta, \hat{\sigma}^2)$$

$g(t)$ – дифференцируемая функция

Тогда:

$$g(\hat{\theta}) \overset{asy}{\sim} N\left(g(\theta),\ \hat{\sigma}^2 \cdot g'(\hat{\theta})^2\right)$$

- Удобно использовать для метода максимального правдоподобия, так как оценки распределены нормально и присутствует свойство инвариантности

Пример:

Если:

$$\hat{\theta} \overset{asy}{\sim} N(\theta, 9) \qquad \hat{\theta} = 2 \qquad g(t) = \frac{1}{t} \qquad g'(t) = -\frac{1}{t^2}$$

Тогда:

$$\frac{1}{\hat{\theta}} \overset{asy}{\sim} N\left(\frac{1}{\theta}, 9 \cdot \left(-\frac{1}{2^2}\right)^2\right) \quad g(\hat{\theta}) \overset{asy}{\sim} N\left(g(\theta),\ \hat{\sigma}^2 \cdot g'(\hat{\theta})^2\right)$$

❗ Важно понимать, что дельта-метод приближает распределение в окрестности математического ожидания

Двумерный дельта-метод

Если:

$$\hat{\theta} = \begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} \overset{asy}{\sim} N\left[\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \quad \begin{pmatrix} \hat{\sigma}_{11}^2 & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_{22}^2 \end{pmatrix}\right] \qquad \hat{\theta} \overset{asy}{\sim} \mathrm{N}(\theta, \hat{\Sigma})$$

$g(t_1, t_2)$ – дифференцируемая функция

Тогда:

$$g(\hat{\theta}) \overset{asy}{\sim} N\left(g(\theta), \nabla\hat{g}^T \, \hat{\Sigma} \, \nabla\hat{g}\right)$$

Где:

$$\nabla g = \begin{pmatrix} \dfrac{\partial g}{\partial t_1} \\ \dfrac{\partial g}{\partial t_2} \end{pmatrix} \qquad \nabla\hat{g} = \begin{pmatrix} \dfrac{\partial g}{\partial t_1}(\hat{\theta}) \\ \dfrac{\partial g}{\partial t_2}(\hat{\theta}) \end{pmatrix}$$

Пример

$$\begin{pmatrix}\hat{\theta}_1\\ \hat{\theta}_2\end{pmatrix} \overset{asy}{\sim} N\left[\begin{pmatrix}\theta_1\\ \theta_2\end{pmatrix}, \begin{pmatrix}8 & 0\\ 0 & 4\end{pmatrix}\right] \quad \hat{\theta} = \begin{pmatrix}1\\ 2\end{pmatrix} \quad g(t_1, t_2) = \frac{t_1}{t_2}$$

$$\nabla g = \begin{pmatrix}\dfrac{1}{t_2}\\ -\dfrac{t_1}{t_2^2}\end{pmatrix} \qquad \nabla \hat{g} = \begin{pmatrix}\dfrac{1}{2}\\ -\dfrac{1}{4}\end{pmatrix}$$

$$\nabla \hat{g}^T\, \hat{\Sigma}\, \nabla \hat{g} = (0.5, -0.25) \cdot \begin{pmatrix}8 & 0\\ 0 & 4\end{pmatrix} \cdot \begin{pmatrix}0.5\\ -0.25\end{pmatrix} =$$

$$= (4, -1) \cdot \begin{pmatrix}0.5\\ -0.25\end{pmatrix} = 2 + 0.25 = 2.25$$

$$\frac{\hat{\theta}_1}{\hat{\theta}_2} \overset{asy}{\sim} N\left(\frac{\theta_1}{\theta_2}, 2.25\right)$$

Дельта-метод

- ML-оценка обладает свойством инвариантности (функция от ML-оценки это ML оценка)
- Чтобы построить доверительный интервал для функции от ML-оценки, мы можем воспользоваться дельта-методом
- Дельта-метод хорошо приближает распределение в окрестности математического ожидания