Чего хочет статистик?

Схема математической статистики

Выборка: $x_1, \dots, x_n$

Параметр: θ

Как оценить:

- метод моментов
- метод максимального правдоподобия

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

Хорошие свойства:

- Несмещенная
- Состоятельная
- Эффективная

Точность оценки, прогнозов
(доверительные интервалы)

Ответы на вопросы
(проверка гипотез)

Союзники:

- Центральная предельная теорема (ЦПТ)
- Теорема Фишера
- $\chi^2_n, t_n, F_{n,k}$
- Ещё союзники!
- Дельта-метод

асимптотические (при большом n)

точные

Несмещённость

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

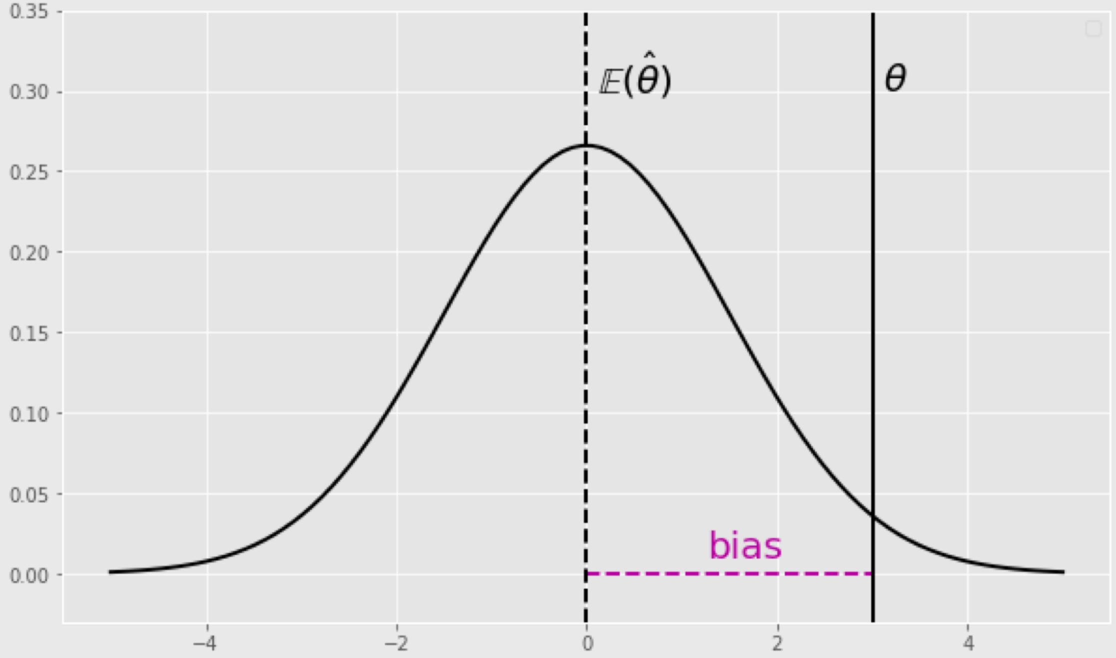
$$\mathbb{E}(\hat{\theta}) = \theta$$

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

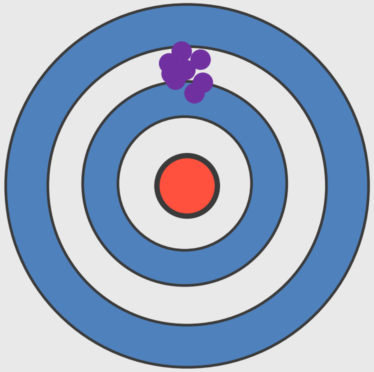
$$bias(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$$

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

Несмещённость



оценка 1



оценка 2

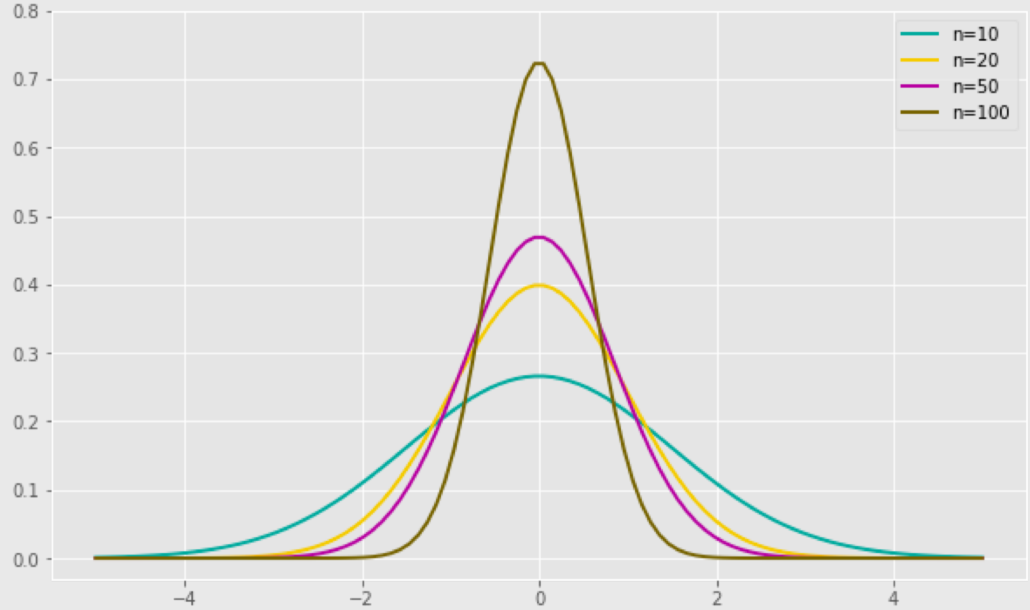
Состоятельность

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

$$\hat{\theta} \to \theta$$

- **Простым языком:** чем больше наблюдений тем мы ближе к истине

Состоятельность



оценки при n =10

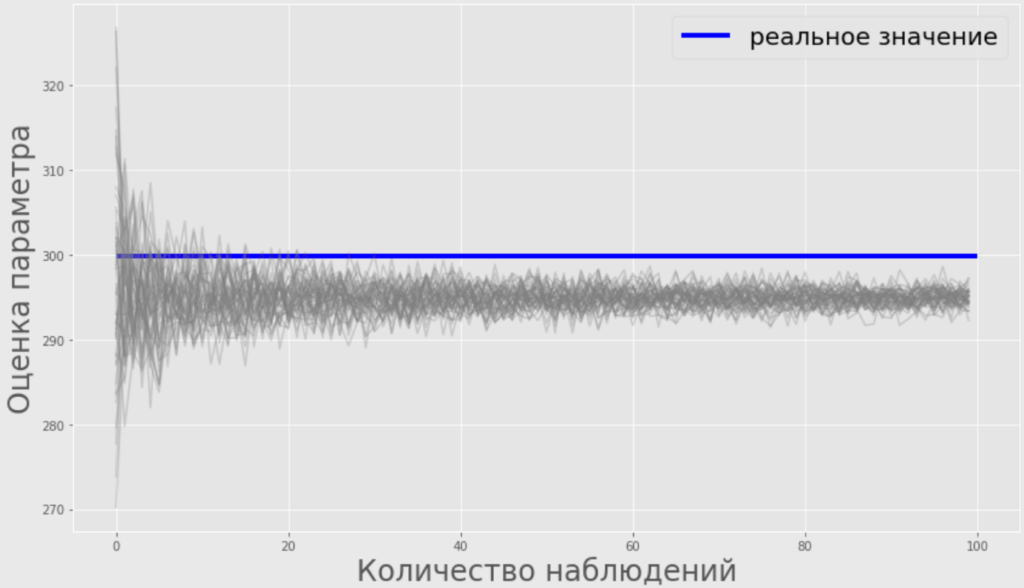
оценки при n =20

оценки при n =50

оценки при n =100

Состоятельность

Состоятельная оценка



Несостоятельная оценка

Асимптотическая несмещённость

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

$$\hat{\theta} \to \theta$$

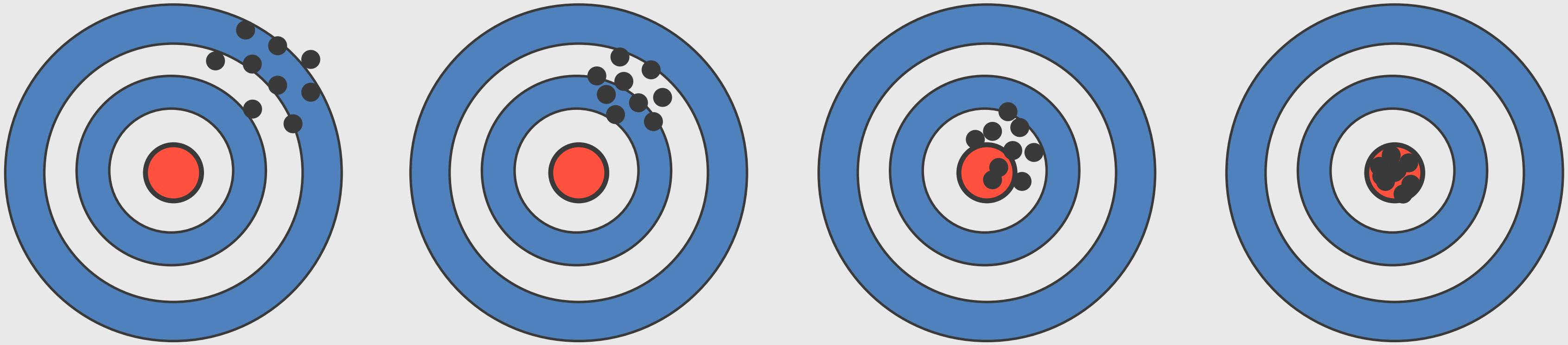
- Оценка называется **асимптотически несмещённой**, если её математическое ожидание сходится к оцениваемому параметру при $n \to \infty$:

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

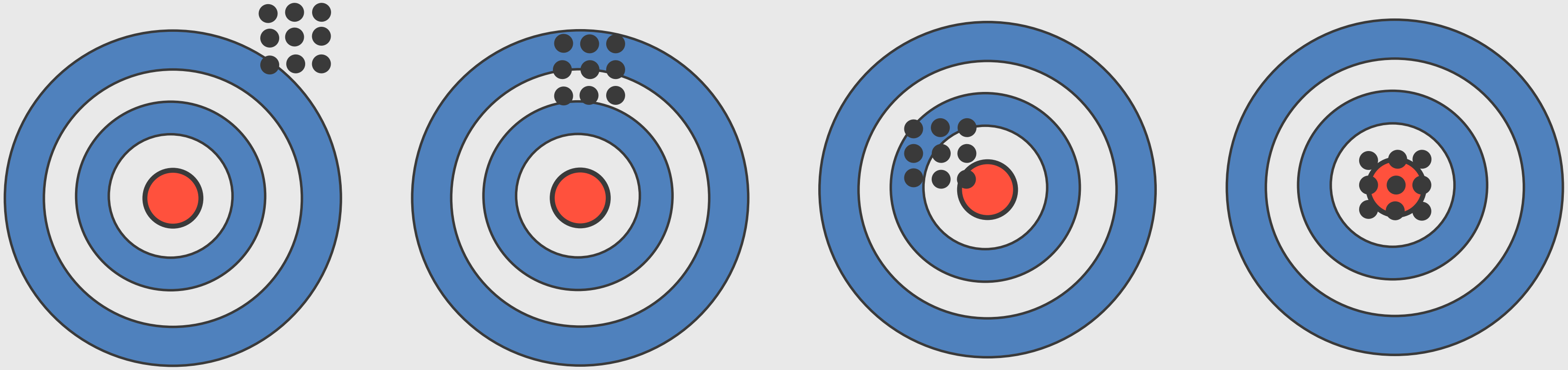
- **Простым языком:** если мы постоянно используем нашу оценку, в среднем, при очень больших n, мы не ошибаемся

Состоятельность vs асимпт. несмещенность

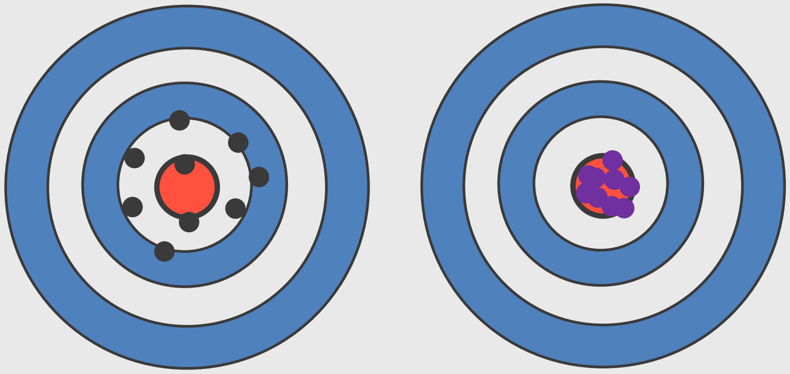
Асимптотически несмещённая и состоятельная оценка



Асимптотически несмещённая, но несостоятельная оценка



Сравнение оценок



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 предсказуемее оценка, тем точнее прогноз (уже доверительный интервал)

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 предсказуемее 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:

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

Bias-variance decomposition

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

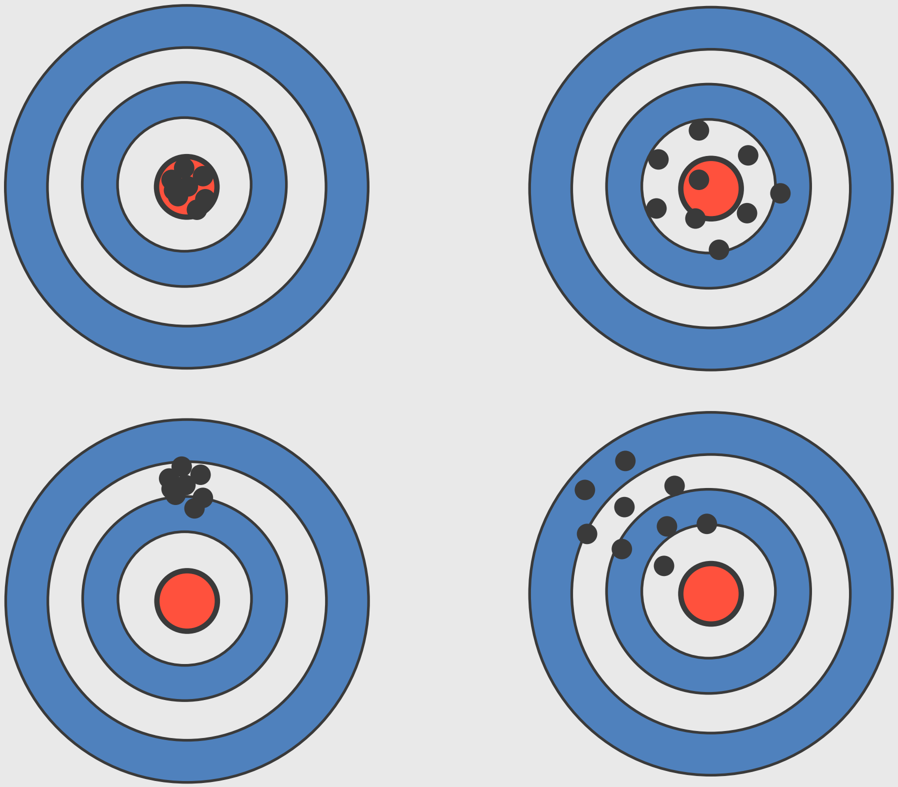
$$MSE = \mathbb{E}\left(\hat{\theta} - \theta\right)^2 = Var\left(\hat{\theta}\right) + \text{bias}^2(\hat{\theta})$$

Bias-variance decomposition

Низкий разброс | **Высокий разброс**

Низкое смещение

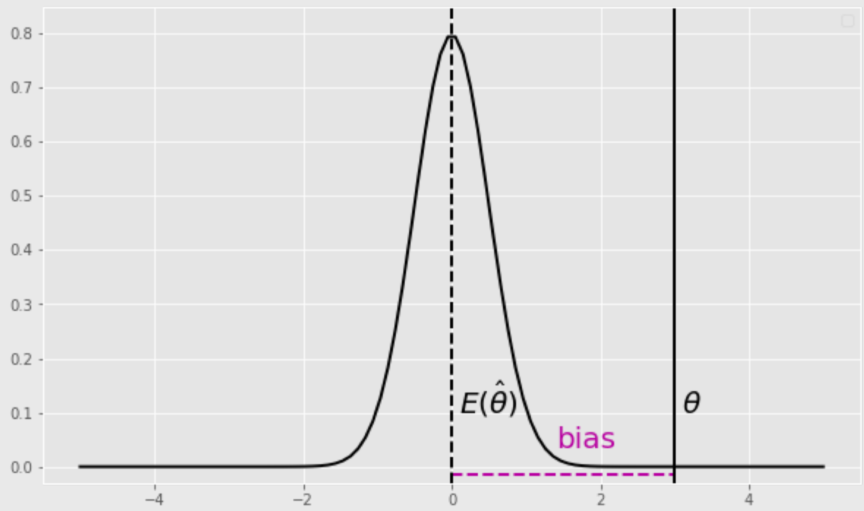
Высокое смещение



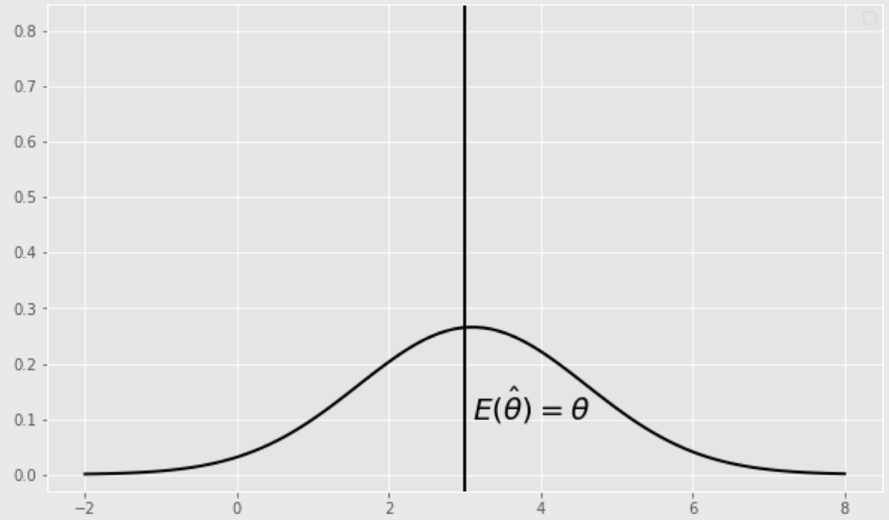
$$MSE = \mathbb{E}\left(\hat{\theta} - \theta\right)^2 = Var\left(\hat{\theta}\right) + \mathrm{bias}^2(\hat{\theta})$$

Bias-variance decomposition

Высокое смещение сильно сбило разброс ⇒ низкое MSE

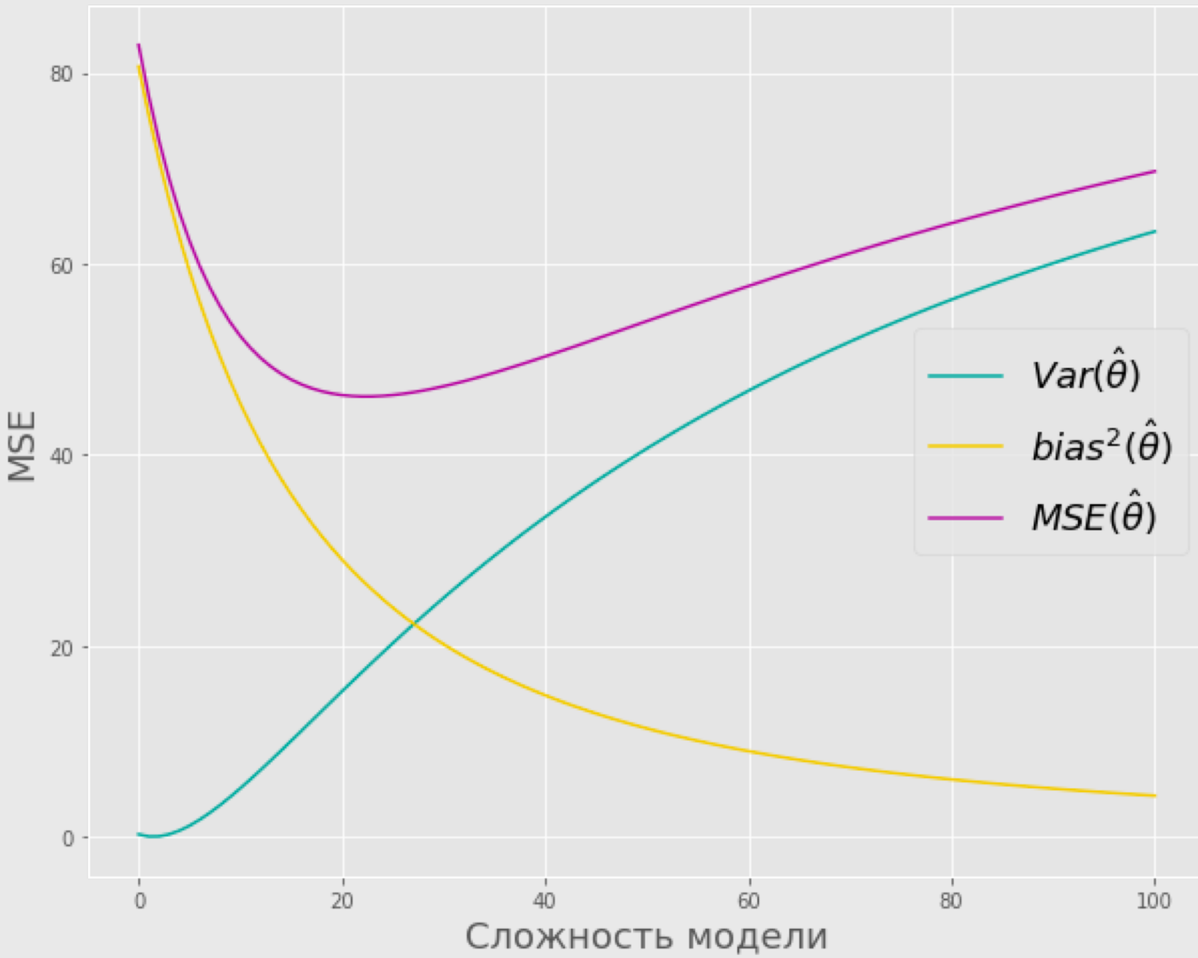


Несмещённая оценка с высокой дисперсией ⇒ высокое MSE



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

Bias-variance decomposition



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

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

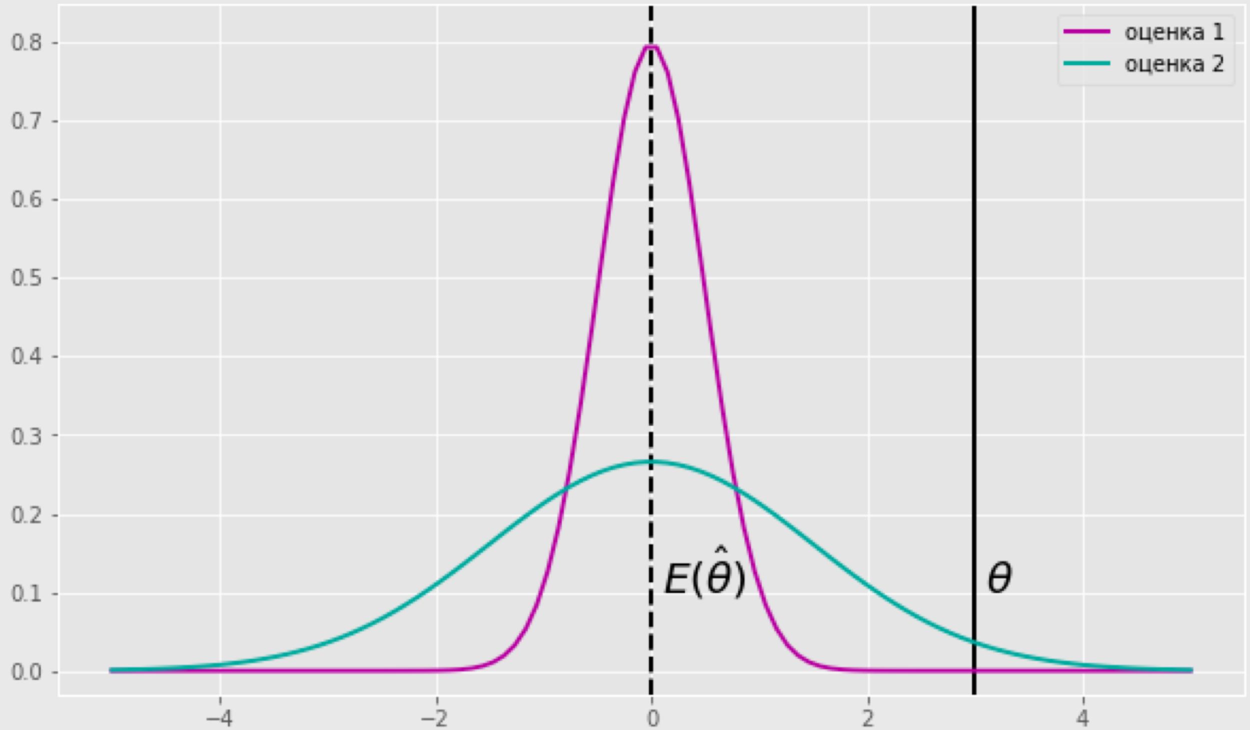
Эффективность

- Можно попробовать зафиксировать смещение и найти оценку с наименьшей дисперсией
- Такая оценка называется **эффективной** в классе $\mathrm{bias}(\hat{\theta})$
- Нас будут интересовать несмещённые эффективные оценки
- **Объяснение для бабушки:** эффективная оценка обладает самым узким доверительным интервалов в своём классе

Эффективность

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

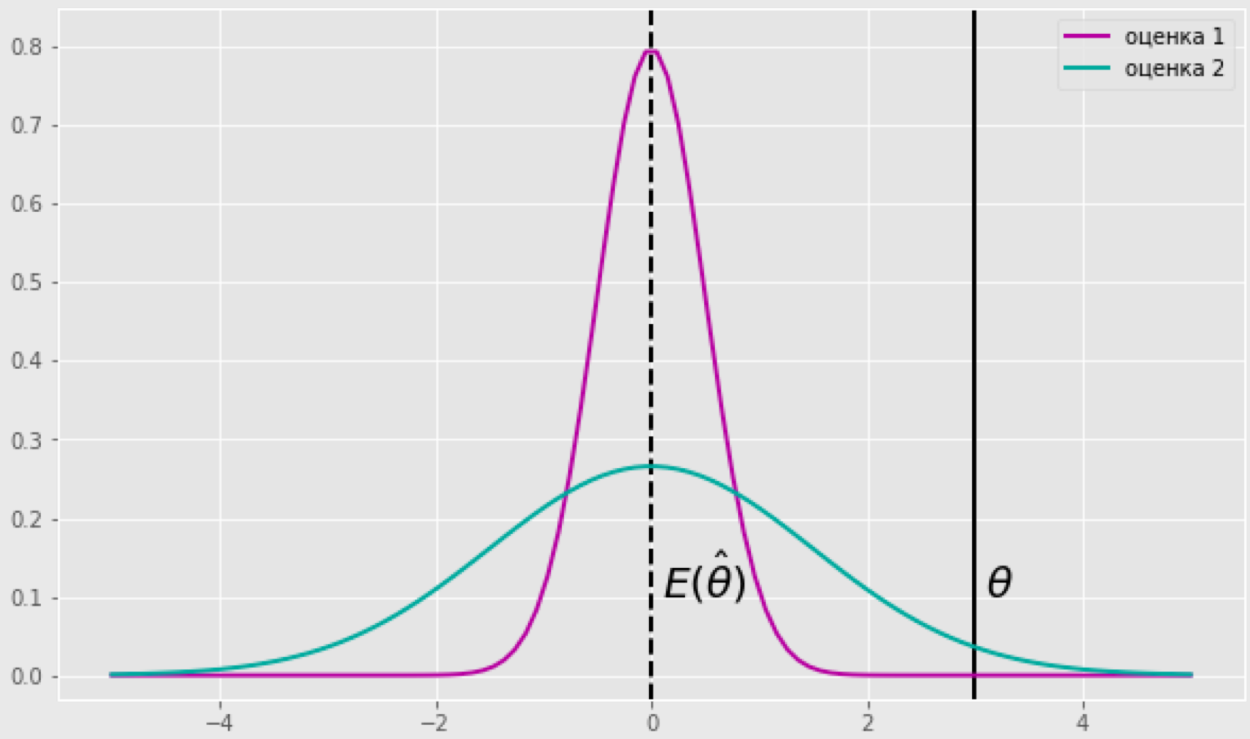
- Если у оценки 1 самая маленькая дисперсия из всех существующих ⇒ она для нас самая предпочтительная



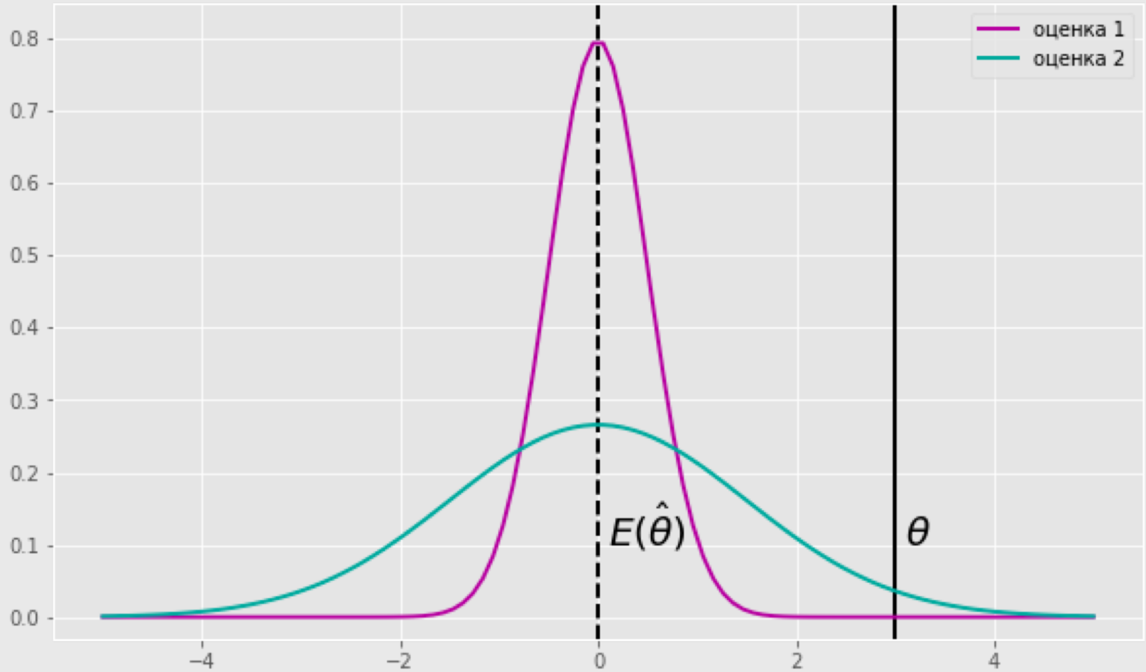
Эффективность

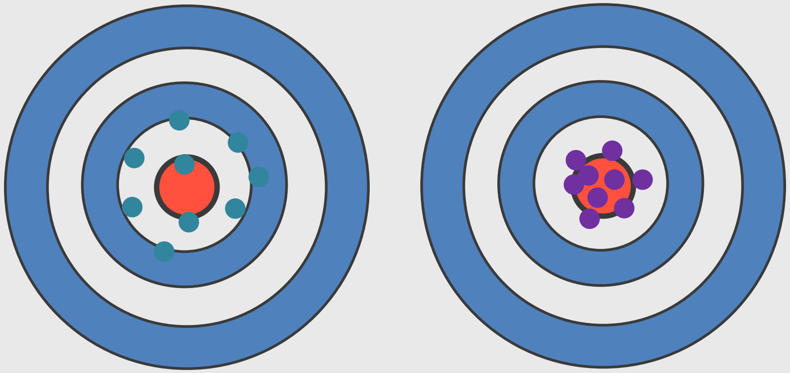
- То есть оценка 1 **эффективная** в классе с таким смещением

- 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 – используя оценку, при фиксированном размере выборки, много раз, в среднем 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 $\Rightarrow$ 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.