Прикладная статистика в машинном обучении
Лекция 1

И. К. Козлов
(*Мехмат МГУ*)

2022

Прикладная статистика в машинном обучении

В 1ой половине курса обсуждаются различные способы оценивания параметров и проверки гипотез.

1. Введение. Точечные оценки. Метод максимального правдоподобия.
2. Вероятностный взгляд на машинное обучение.
3. Интервальные оценки. Бутстреп.
4. Проверка гипотез. Множественное тестирование гипотез.
5. Регрессионный анализ.
6. Непараметрическое оценивание.
7. A/B-тестирование.

Прикладная статистика в машинном обучении

2ая половина курса посвящена байесовскому подходу.

8. Байесовский подход к теории вероятности. Теорема Байеса. Аналитический байесовский вывод.
9. Байесовский взгляд на подбор моделей.
10. Метод релевантных векторов (RVM).
11. Модели с латентными переменными. EM-алгоритм.
12. Генеративные модели.

Сколько мучиться

1. 12 лекций и семинаров.
2. 3 теоретических задания.
3. 3 лабораторные работы.

Не очень жёстко

Жёсткость дедлайна.

Задание можно досдавать неделю после дедлайна.

$$\text{Штраф } = \min(50\%, 10\% * \text{число дней просрочки})$$

Оценки

Экзамена нет!

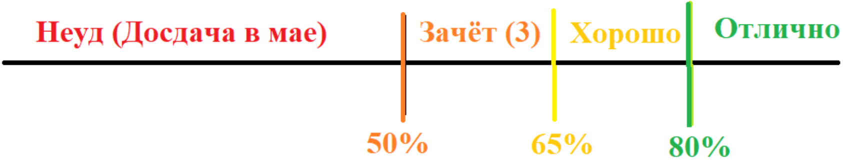
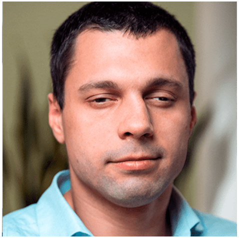


Рис.: Итоговая оценка

Standing on the shoulders of giants

Этот курс во многом основан на лекциях ШАД.

(a) А. М. Райгородский



(b) Е. В. Бурнаев

(c) Д. П. Ветров

Рис.: Лекторы, которым я благодарен

Литература

Некоторые полезные книжки:

- Wasserman L.
  All of Nonparametric Statistics.
- Wasserman L.
  All of Statistics.
- Bishop C.M.
  Pattern Recognition and Machine Learning.
- Murphy K.P.
  Machine Learning: A Probabilistic Perspective.
- М. Б. Лагутин,
  Наглядная математическая статистика.

Базовый тервер

Базовые факты по теории вероятностей см. в Приложении к

М. Б. Лагутин,
Наглядная математическая статистика.

или в начале книжки Вассермана

Wasserman L.
All of Nonparametric Statistics.

Стандартные обозначения

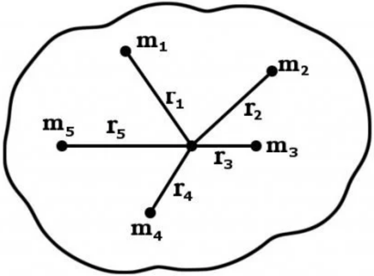
Распределения будут либо дискретными, либо абсолютно непрерывными.

Обозначения стандартные:

- $F(x)$ — функция распределения.
- $f(x)$ — плотность распределения (непр.) или функция вероятности (дискр.)

Матожидание и дисперсия

Матожидание и дисперсия — вероятностные аналоги центра масс и момента инерции относительно него.



$$I = \sum m_i r_i^2$$

Матожидание и дисперсия

Матожидание $a(x)$ — либо интеграл, либо конечная сумма:

$$\mathbb{E}\left(a(x)\right) = \int a(x) dF(x) = \begin{cases} \int a(x) f(x) dx, & \text{непрерывный случай,} \\ \sum_j a(x_j) f(x_j), & \text{дискретный случай.} \end{cases}$$

Дисперсия:

$$\mathbb{V}(X) = \mathbb{E}\left(X - \mathbb{E}(X)\right)^2.$$

Что такое статистика?

Основная задача статистики

Основной вопрос статистики:

$\boxed{\begin{array}{l}\text{Дана выборка } X_1, \ldots, X_n \sim F. \\ \text{Что можно сказать о } F?\end{array}}$

Статистическая модель

- Статистическая модель. Фиксируем семейство распределений

$$\mathcal{F} = \{F_\theta(x), \quad \theta \in \Theta\},$$

в котором лежит F.

- **Задача:** восстановить распределение F_θ (оценить параметры θ).

Основная задача статистики

Пример 1.

Курс будет по возможности наглядным и иметь отношение к ML.

Поговорим о деньгах.

Пример с деньгами

Курс будет по возможности наглядным и иметь отношение к ML.

Поговорим о ~~деньгах~~ монетках.

Орел и решка

Распределение Бернулли $\mathrm{Bernoulli}(p)$. Орёл выпадает с вероятностью p.

Реализация выборки — исходы бросков монетки.

Параметр $p = ?$

ЗБЧ

Пусть $X_1, \ldots, X_n$ — i.i.d. и $\mu = \mathbb{E}X_1 < \infty$.

Обозначим
$$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$

Закон больших чисел

$$\overline{X}_n \xrightarrow{\mathbb{P}} \mu.$$

Для тех, кто забыл

Случайные величины $X_1, \dots, X_n$ сходятся при $n \to \infty$ к случайной величине X

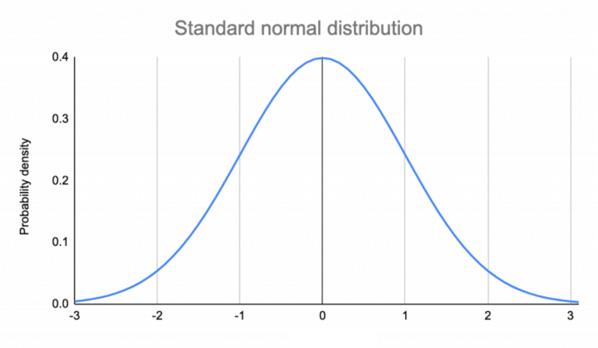
- по вероятности $X_n \xrightarrow{\mathbb{P}} X$, если $\forall \epsilon > 0$
$$\mathbb{P}\left(|X_n - X| > \epsilon\right) \to 0;$$
- почти наверное $X_n \xrightarrow{a.s.} X$, если
$$\mathbb{P}\left\{\omega : X_n(\omega) \to X(\omega)\right\} = 1;$$
- в среднем квадратичном $X_n \xrightarrow{L^2} X$, если
$$\mathbb{E}\left(X_n - X\right)^2 \to 0;$$
- по распределению $X_n \xrightarrow{d} X$, если
$$F_{X_n}(x) \to F_X(x)$$
для каждой точки непрерывности x функции распределения $F_X(x)$.

Нормальный пример

Пример 2.

Определить параметры μ и σ у стандартного нормального распределения

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$



ЦПТ

Нормальное распределение повсюду!

Центральная предельная теорема.

Если $\mathbb{V}X_i = \sigma^2 < \infty$, то

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\right).$$

Первая развилка

Рис.: Возникает 1ая развилка

Конечномерно ли пространство параметров Θ?

Два вида статистики

Статистика

$\dim \Theta < \infty$ — **Параметрическая**

$\dim \Theta = \infty$ — **Непараметрическая**

Непараметрическая статистика

Непараметрическая статистика

Задачи непараметрической статистики?

- Дана выборка $X_1, \dots, X_n \sim F$.

  Как **восстановить функцию распределения**

  $$F(x) = \mathbb{P}(X \le x)?$$

  Или **плотность** $f(x)$?

Непараметрическая статистика

Задачи непараметрической статистики?

- Дана выборка $X_1, \ldots, X_n \sim F$.

  Как **восстановить функцию распределения**

  $$F(x) = \mathbb{P}(X \leq x)?$$

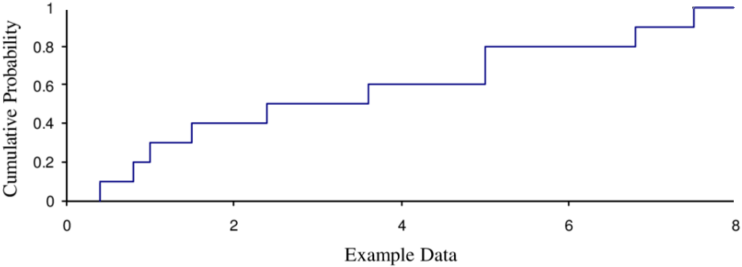
  Или **плотность** $f(x)$?

Q: Важнейший результат об этом?

Эмпирическая функция распределений

Эмпирическая функция распределения

$$\hat{F}(x) = \frac{\text{кол-во эл-тов в выборке} \leq \text{x}}{n} = \frac{1}{n}\sum_{i=1}^{n} I\left(X_i \leq x\right).$$



Теорема Гливенко-Кантелли

Теорема Гливенко-Кантелли

$$\sup_{x \in \mathbb{R}} \left| \hat{F}(x) - F(x) \right| \xrightarrow{a.s.} 0 \,.$$

Непараметрическая статистика

Другая задача непараметрической статистики:

- **Оценить статистики распределения** $T(F)$, например матожидание
$$T(F) = \int x dF(x).$$

Непараметрическая статистика

Другая задача непараметрической статистики:

- **Оценить статистики распределения** $T(F)$, например матожидание
$$T(F) = \int x dF(x).$$

Q: Простейший способ это сделать?

Лекция 1

Статистический функционал $T(F)$ — это функция от распределения F.

Примеры:

- Среднее значение
$$\mu = \int x dF(x),$$
- Дисперсия
$$\sigma^2 = \int (x - \mu)^2 dF(x),$$
- Медиана (квантиль распределения)
$$m = F^{-1}\left(\frac{1}{2}\right).$$

Plug-in оценивание

Выборочные оценки — подставить в интеграл эмпирическую функцию распределения $\hat{F}$.

Интеграл превращается в сумму

$$T\left(\hat{F}_n\right) = \int a(x) d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} a(X_i).$$

Регрессия

Ещё одна задача непараметрической статистики:

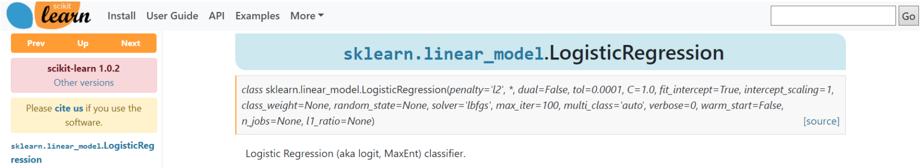
- **Регрессия**. Даны пары $(x_1, Y_1), \dots, (x_n, Y_n)$, где

$$Y_i = r(x_i) + \varepsilon_i, \qquad \text{и} \qquad \mathbb{E}\varepsilon_i = 0.$$

  Нужно восстановить функцию регрессии r.

Логистическая регрессия

Да, линейную и логистическую регрессию мы тоже обсудим.



Точечная оценка

Оценки параметров

3 важнейших способа оценивать параметры:

- Точечная оценка.
- Доверительные интервалы.
- Проверка гипотез.

Точечная оценка

Точечная оценка.

Оценка параметра θ — случайная величина от выборки

$$\hat{\theta}_n = g(X_1, \ldots, X_n).$$

По традиции обозначается $\hat{\theta}$ или $\hat{\theta}_n$.

Несмещённость

Что мы хотим от оценки?

- Смещение оценки:

$$\mathrm{Bias}_\theta(\hat{\theta}_n) = \mathbb{E}_\theta\left(\hat{\theta}_n\right) - \theta$$

Q. Почему написано $\mathbb{E}_\theta$?

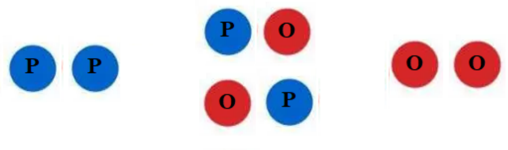
Несмещённость

- $\mathbb{E}_\theta\left(\hat{\theta}_n\right)$ — среднее значение статистики $\hat{\theta}$ по всем реализациям выборок $X_1, \ldots, X_n$ при фиксированном θ.

Несмещённость

- $\mathbb{E}_\theta\left(\hat{\theta}_n\right)$ — среднее значение статистики $\hat{\theta}$ по всем реализациям выборок $X_1, \ldots, X_n$ при фиксированном θ.

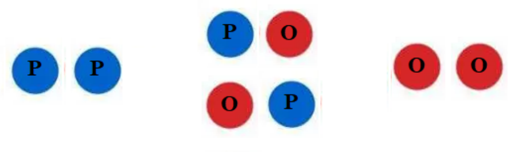
Пример. Кидаем монетку 2 раза. $\hat{\theta}$ — доля выпавших орлов.



Несмещённость

- $\mathbb{E}_\theta\left(\hat{\theta}_n\right)$ — среднее значение статистики $\hat{\theta}$ по всем реализациям выборок $X_1, \ldots, X_n$ при фиксированном θ.

Пример. Кидаем монетку 2 раза. $\hat{\theta}$ — доля выпавших орлов.



$$\mathbb{E}_\theta\left(\hat{\theta}\right) = 0 * (1-p)^2 + 2 * 0,5 * p(1-p) + 1 * p^2 = p.$$

Несмещённость

В общем случае

$$\mathbb{E}_\theta\left(\hat{\theta}_n\right) = \int g(x_1, \ldots, x_n) f_\theta(x_1) \ldots f_\theta(x_n) dx_1 \ldots dx_n.$$

Несмещённость

Оценка $\hat{\theta}_n$ **несмещенная**, если $\mathrm{Bias}_\theta(\hat{\theta}_n) = 0$ для любого θ.

Несмещённость — неплохо, но необязательно.

Стандартная ошибка

Несмещённость = в среднем угадываем параметр.

Но ~~какой ценой~~ каков разброс?

Стандартная ошибка — корень из дисперсии

$$\mathrm{se} = \mathrm{se}(\hat{\theta}_n) = \sqrt{\mathbb{V}_\theta(\hat{\theta}_n)}.$$

Состоятельность

Чем больше данных — тем меньше вероятность сильного отклонения оценки.

- Оценка $\hat{\theta}_n$ состоятельная, если
$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$$

Состоятельность

Чем больше данных — тем меньше вероятность сильного отклонения оценки.

- Оценка $\hat{\theta}_n$ состоятельная, если
$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta$$

Напомним, сходимость *по вероятности*:

$$\mathbb{P}\left(\left|\hat{\theta}_n - \theta\right| > \epsilon\right) \to 0, \qquad \forall \varepsilon > 0.$$

Зависимость между видами сходимости

Q: Как проверять состоятельность оценки?

Зависимость между видами сходимости

Q: Как проверять состоятельность оценки?

Идея. Вспомним зависимость между видами сходимости:

$$
\begin{array}{ccccc}
 & & \xrightarrow{L^2} & & \\
 & & \Downarrow & & \\
\xrightarrow{\text{a.s.}} & \Rightarrow & \xrightarrow{\mathbb{P}} & \Rightarrow & \xrightarrow{d}
\end{array}
$$

MSE

- Среднеквадратичная ошибка

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

Теорема

$$\text{MSE} = \left(\text{Bias}_\theta(\hat{\theta})\right)^2 + \mathbb{V}_\theta(\hat{\theta}).$$

Следствие. Если $\text{Bias}_\theta(\hat{\theta} \to 0$ и $\mathbb{V}_\theta(\hat{\theta}) \to 0$, то $\hat{\theta}$ — состоятельная оценка.

Bias-Variance trade-off

Теорема

$$\text{MSE} = \mathbb{V}_\theta(\hat{\theta}) + \left(\text{Bias}_\theta(\hat{\theta})\right)^2.$$

Доказательство.

Обозначим $\bar{\theta} = \mathbb{E}_\theta(\hat{\theta})$.

Далее просто раскрываем скобки:

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

Bias-Variance trade-off

Теорема

$$\mathrm{MSE} = \mathbb{V}_\theta(\hat{\theta}) + \left(\mathrm{Bias}_\theta(\hat{\theta})\right)^2.$$

Доказательство.

Обозначим $\bar{\theta} = \mathbb{E}_\theta(\hat{\theta})$.

Далее просто раскрываем скобки:

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

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

Bias-Variance trade-off

Теорема

$$\mathrm{MSE} = \mathbb{V}_\theta(\hat{\theta}) + \left(\mathrm{Bias}_\theta(\hat{\theta})\right)^2.$$

Доказательство.

Обозначим $\bar{\theta} = \mathbb{E}_\theta(\hat{\theta})$.

Далее просто раскрываем скобки:

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

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

$$= \mathbb{V}_\theta(\hat{\theta}) + \mathrm{Bias}(\hat{\theta})^2.$$

К теореме Гливенко-Кантелли

Пример. Эмпирическая функция распределения $\hat{F}_n(x)$ — несмещённая состоятельная оценка $F(x)$ в каждой точке x.

Теорема

Для любой фиксированной точки x:

$$\mathbb{E}\left(\hat{F}_n(x)\right) = F(x),$$

$$\mathbb{V}\left(\hat{F}_n(x)\right) = \frac{F(x)\left(1 - F(x)\right)}{n},$$

К теореме Гливенко-Кантелли

Пример. Эмпирическая функция распределения $\hat{F}_n(x)$ — несмещённая состоятельная оценка $F(x)$ в каждой точке x.

Теорема

Для любой фиксированной точки x:

$$\mathbb{E}\left(\hat{F}_n(x)\right) = F(x),$$

$$\mathbb{V}\left(\hat{F}_n(x)\right) = \frac{F(x)\,(1 - F(x))}{n},$$

$$\text{MSE} = \frac{F(x)\,(1 - F(x))}{n} \to 0,$$

$$\hat{F}_n(x) \xrightarrow{\mathbb{P}} F(x).$$

Асимптотическая нормальность

Зачастую нам достаточно выполнения статистических свойств в пределе.

Оценка $\hat{\theta}_n$ асимптотически нормальна, если

$$\frac{\hat{\theta}_n - \theta}{\mathsf{se}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Доверительные интервалы

Точечные оценки не идеальны

Монету подбросили 15 раз. Выпало

$$4 \text{ орла}, \qquad 11 \text{ решек}.$$

Готовы дать руку на отсечение, что $p = \dfrac{4}{15}$?

Доверительный интервал

Доверительный интервал с доверительной вероятностью $1 - \alpha$ для параметра θ — это интервал

$$C_n = (a, b), \qquad a = a(X_1, \ldots, X_n), \quad b = b(X_1, \ldots, X_n),$$

для которого

$$P_\theta(\theta \in C_n) \geq 1 - \alpha, \qquad \text{для всех } \theta \in \Theta.$$

Подробнее поговорим про доверительные интервалы на Лекции 3.

Brace yourself

Для построения точных доверительных интервалов нужно найти то или иное неравенство.

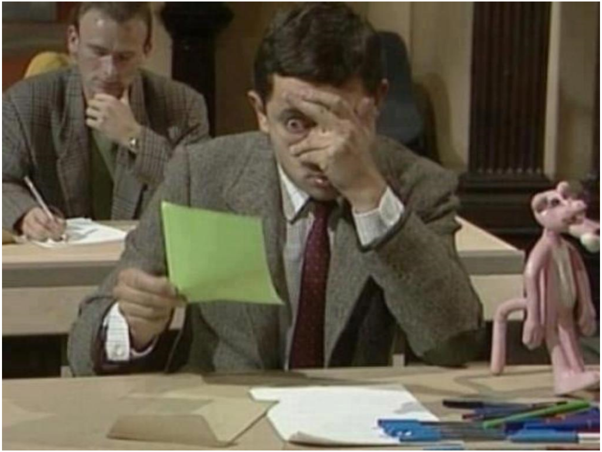


Рис.: Неравенства страшные

Неравенство Хёфдинга

Неравенство для отклонения от матожидания суммы ограниченных с.в.[1]:

Неравенство Хёфдинга

Пусть $X_1, \ldots, X_n$ — i.i.d. и $a_i \le X_i \le b_i$. Положим

$$S_n = X_1 + \cdots + X_n.$$

Тогда для любого $t > 0$

$$\mathbb{P}\left(S_n - \mathrm{E}\left[S_n\right] \ge t\right) \le \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

$$\mathbb{P}\left(\left|S_n - \mathrm{E}\left[S_n\right]\right| \ge t\right) \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

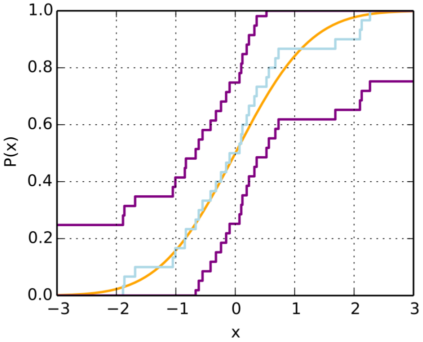
[1] Применимо для распределения Бернулли

Dvoretzky–Kiefer–Wolfowitz–Massart inequality

Для эмпирической функции распределения:

Dvoretzky–Kiefer–Wolfowitz–Massart inequality

$$\mathbb{P}\Big(\sup_{x\in\mathbb{R}}|F_n(x) - F(x)| > \varepsilon\Big) \leq 2e^{-2n\varepsilon^2} \qquad \text{для любого } \varepsilon > 0.$$



Проверка гипотез

Проверка гипотез

Частый вопрос — насколько маловероятно то или иное значение $\theta = \theta_0$?

Пример. Пусть
$$X_1, \dots, X_n \sim \mathrm{Bernoulli}(p).$$

- Нулевая гипотеза H_0: обмана нет $p = \frac{1}{2}$.
- Альтернативная гипотеза H_1: $p \neq \frac{1}{2}$.

Обсудим проверки гипотез в Лекции 4.

Метод максимального правдоподобия

Вспомним всё

Вспомним, как решаются статистические задачи.

Вспомним всё

Вспомним, как решаются статистические задачи.

Всё будет *нормально* и **максимально правдоподобно**.

MLE

- Известна плотность вероятности:

$$\mathcal{F} = \{f(x,\theta), \theta \in \Theta\}$$

- Функция правдоподобия = вероятность выборки

$$\mathcal{L}_n(\theta) = \prod_{i=1}^{n} f(X_i;\theta).$$

- Оценка максимального правдоподобия (MLE)

$$\hat{\theta}_{ML} = \arg\max_{\theta\in\Theta} \mathcal{L}_n(\theta).$$

Логарифмируем произведение

На практике **малые величины лучше суммировать, чем перемножать.**

Логарифм функции правдоподобия

$$\ell_n(\theta) = \log \mathcal{L}_n(\theta).$$

Заметём под ковёр

Свойства MLE верны при определённых условиях регулярности.

Рис.: По традиции опустим их

Строгие формулировки

Подробные формулировки и доказательство можно найти в

R. Hogg, J. McKean, A. Craig,
Introduction to Mathematical Statistics. Chapter 6.

Простейший случай

Для простоты[2] предположим, что

- MLE существует и единственна;
- $\hat{\theta}_{ML}$ — внутренняя точка Θ, и это единственное решение
$$\frac{\partial L(\theta)}{\partial \theta} = 0;$$
- функции $f(x;\theta)$ и $\int f(x;\theta)$ дважды дифференцируемы по θ (и производную можно вносить под знак интеграла).

[2] Это не все условия. Некоторые условия можно ослабить.

MLE инвариантна

- MLE **инвариантна.**

  Если $\tau = g(\theta)$ и $\hat{\theta}_{ML}$ — ML оценка для θ, то $\hat{\tau} = g(\hat{\theta}_{ML})$ — ML оценка для τ.

  Если g — биекция, то очевидно.

  $$\mathcal{L}(\theta) = \mathcal{L}(\tau),$$

  при замене координат максимум переходит в максимум.

  Можно доказать для любой g.

MLE состоятельна

- MLE **состоятельна**

$$\hat{\theta}_{ML} \xrightarrow{\mathbb{P}} \theta$$

MLE состоятельна

- MLE **состоятельна**

$$\hat{\theta}_{ML} \xrightarrow{\mathbb{P}} \theta$$

Далее для простоты $\dim \Theta = 1$ (параметр θ — число).

Следующее свойство — асимптотическая нормальность. Для оценки дисперсии нужно ввести несколько функций.

Информация Фишера

- Функция оценки

$$s(X;\theta) = \frac{\partial \log f(X;\theta)}{\partial \theta}.$$

- Информация Фишера

$$I_n(\theta) = \mathbb{V}(\sum_{i=1}^{n} s(X_i;\theta)) = \sum_{i=1}^{n} \mathbb{V}(s(X_i;\theta))$$

Часто обозначают $I_1(\theta) = I(\theta)$. Тогда $I_n(\theta) = n\, I(\theta)$.

Асимптотическая нормальность

- MLE **асимптотически нормальна**

$$\sqrt{n}\left(\hat{\theta}_{ML} - \theta\right) \quad \xrightarrow{d} \quad \mathcal{N}\left(0, \frac{1}{I(\theta)}\right).$$

Матрицу Фишера можно заменить её оценкой

$$\frac{\hat{\theta}_{ML} - \theta}{\hat{se}} \quad \xrightarrow{d} \quad \mathcal{N}(0, 1), \qquad \hat{se}^2 = \frac{1}{I_n(\hat{\theta}_{ML})}.$$

Асимптотическая нормальность

- MLE **асимптотически нормальна**

$$\sqrt{n}\left(\hat{\theta}_{ML} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right).$$

Матрицу Фишера можно заменить её оценкой

$$\frac{\hat{\theta}_{ML} - \theta}{\hat{\text{se}}} \xrightarrow{d} \mathcal{N}(0, 1), \qquad \hat{\text{se}}^2 = \frac{1}{I_n(\hat{\theta}_{ML})}.$$

Итак, MLE асимтотически несмещена и дисперсия примерно равна $\dfrac{1}{I_n(\theta)}$.

Неравенство Рао–Крамера

Неравенство Рао–Крамера

Пусть выполнены условия регулярности, а T — функция от выборки. Тогда

$$\mathbb{V}(T) \geq \frac{1}{I_n(\theta)} \left(\frac{\partial \mathbb{E}(T)}{\partial \theta} \right)^2 .$$

Неравенство Рао–Крамера

Неравенство Рао–Крамера

Пусть выполнены условия регулярности, а T — функция от выборки. Тогда

$$\mathbb{V}(T) \geq \frac{1}{I_n(\theta)} \left(\frac{\partial \mathbb{E}(T)}{\partial \theta} \right)^2.$$

Для несмещённых оценок $\mathbb{E}(\hat{\theta}) = \theta$ получаем

$$\mathbb{V}(\hat{\theta}) \geq \frac{1}{I_n(\theta)}.$$

MLE оптимальна

Оптимальная оценка— несмещённая оценка с минимальной дисперсией.

- MLE **асимптотически оптимальна**.
  - Она асимптотически несмещенна (т.к. асимпт. норм.)
  - Её дисперия стремится к $\mathbb{V}(\hat{\theta}) \geq \frac{1}{I_n(\theta)}$.
  - По неравенству Рао-Крамера дисперсия $\geq \mathbb{V}(\hat{\theta}) \geq \frac{1}{I_n(\theta)}$.

Байесовский подход

Допустимый уровень ошибок

Q: Какой доверительный интервал взять?

Допустимый уровень ошибок

Q: Какой доверительный интервал взять?

A: 95% или 99%.

Это традиция)

Многомодальность

Почемы мы ограничиваемся интервалом для θ?

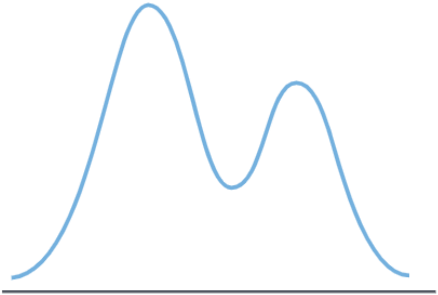


Рис.: У θ могут быть разные моды

Байесовский подход

Идея. Значение θ определяется распределением $p(\theta)$.

И треснул мир на пополам

Формула Байеса

При поступлении данных от априорного распределения $p(\theta)$ мы переходим к апостериорному $P(\theta \mid X)$.

Формула Байеса

$$P(\theta \mid X) = \frac{P(X \mid \theta)\, P(\theta)}{\int P(X \mid \theta)\, P(\theta) d\theta}.$$

Обсудим байесовский подход во 2ой части курса.

Резюме 1ой лекции

Дилемма смещения–дисперсии

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}_\theta(\hat{\theta}) + \mathrm{Bias}_\theta(\hat{\theta})^2.$$

Одна из ключевых проблем в обучении с учителем (**переобучение**).

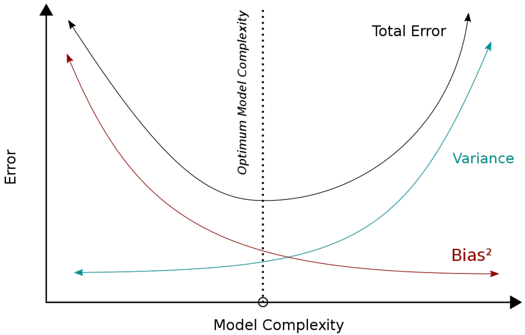


Рис.: Bias–variance tradeoff

Оценка максимального правдоподобия

Оценка максимального правдоподобия

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} \mathcal{L}_n(\theta)$$

- Состоятельна.
- Инвариантна.
- Асимптотически нормальна.
- Асимптотически оптимальна.

Разделы статистики

	Parametric	Nonparametric
Frequentist	I	II
Bayesian	III	Будет в ШАД (Гауссовские процессы)

Рис.: Разделы статистики, которые мы обсудим

To be continued

	Parametric	Nonparametric
Frequentist	I	II
Bayesian	III	Будет в ШАД (Гауссовские процессы)