Прикладная статистика в машинном обучении
Лекция 4. Часть 2.
Проверка гипотез

И. К. Козлов

(*Мехмат МГУ*)

2022

Критерий отношения правдоподобия

Likelihood

Statistics is all about Likelihood

Правдоподобие

Есть ряд тестов, основанных на правдоподобии.

Мы опишем один из них. Остальные обычно ему асимптотически эквивалентны:

Maximum Likelihood Estimation Theory Summary of Notation and Main Results

Статистика отношения правдоподобия

Тестируем гипотезы:

$$H_0 = \{\theta \in \Theta_0\}, \qquad H_1 = \{\theta \in \Theta_1\}$$

Статистика отношения правдоподобия:

$$\lambda = 2\log\left(\frac{\sup_{\theta\in\Theta}\mathcal{L}(\theta)}{\sup_{\theta\in\Theta_0}\mathcal{L}(\theta)}\right)$$

Статистика отношения правдоподобия

Проще говоря,

$$\lambda = 2\log\left(\frac{\mathcal{L}(\hat{\theta})}{\mathcal{L}(\hat{\theta}_0)}\right),$$

где $\hat{\theta}$ — MLE для Θ и $\hat{\theta}_0$ — MLE для Θ_0.

Статистика отношения правдоподобия

Проще говоря,

$$\lambda = 2\log\left(\frac{\mathcal{L}(\hat{\theta})}{\mathcal{L}(\hat{\theta}_0)}\right),$$

где $\hat{\theta}$ — MLE для Θ и $\hat{\theta}_0$ — MLE для Θ_0.

Замечание. Вместо λ можно взять отношение правдоподобий для Θ_0 и Θ_1

$$\lambda' = 2\log\left(\frac{\sup_{\theta\in\Theta_1}\mathcal{L}(\theta)}{\sup_{\theta\in\Theta_0}\mathcal{L}(\theta)}\right),$$

На практике это не принципиально, формулы проще для λ.

Тест отношения правдоподобия

Критерий отношения правдоподобия

Пусть

$$\theta = (\theta_1, \dots, \theta_q, \theta_{q+1}, \dots, \theta_r).$$

Пусть при нулевой гипотезе мы фиксируем последние $r - q$ параметров:

$$\Theta_0 = \{\theta : (\theta_{q+1}, \dots, \theta_r) = (\theta_{0,q+1}, \dots, \theta_{0,r})\}$$

Тест отношения правдоподобия

Критерий отношения правдоподобия

Пусть

$$\theta = (\theta_1, \dots, \theta_q, \theta_{q+1}, \dots, \theta_r).$$

Пусть при нулевой гипотезе мы фиксируем последние $r - q$ параметров:

$$\Theta_0 = \{\theta : (\theta_{q+1}, \dots, \theta_r) = (\theta_{0,q+1}, \dots, \theta_{0,r})\}$$

- При выполнении H_0 :

$$\lambda(x^n) \xrightarrow{d} \chi^2_{r-q},$$

  где $r - q = \dim\Theta - \dim\Theta_0$.

Тест отношения правдоподобия

Критерий отношения правдоподобия

Пусть

$$\theta = (\theta_1, \dots, \theta_q, \theta_{q+1}, \dots, \theta_r).$$

Пусть при нулевой гипотезе мы фиксируем последние $r - q$ параметров:

$$\Theta_0 = \{\theta : (\theta_{q+1}, \dots, \theta_r) = (\theta_{0,q+1}, \dots, \theta_{0,r})\}$$

- При выполнении H_0 :

$$\lambda(x^n) \xrightarrow{d} \chi^2_{r-q},$$

  где $r - q = \dim \Theta - \dim \Theta_0$.

- Примерное

$$p_{value} = \mathbb{P}(\chi^2_{r-q} > \lambda).$$

Критерий Неймана-Пирсона

Критерий Неймана-Пирсона

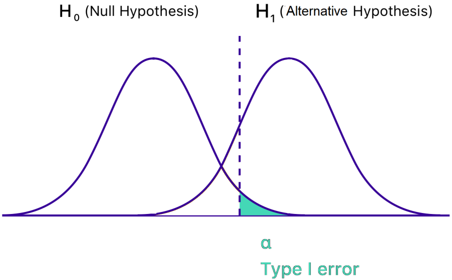
Критерий Неймана-Пирсона — самый мощный критерий.

Критерий Неймана-Пирсона

Сравниваем две простые гипотезы

$$H_0 : \theta = \theta_0, \qquad VS \qquad H_1 : \theta = \theta_1$$

Фиксируем силу критерия α (=вероятность ошибки I рода).



Т.е. вероятность критического множества равна α при выполнении H_0.

Критерий Неймана-Пирсона

Рассмотрим дискретный случай.

Есть M точек. Пусть при H_0 они равновероятны:

$$P_0(X_j) = \frac{1}{M}, \qquad P_1(X_j) = p_j$$

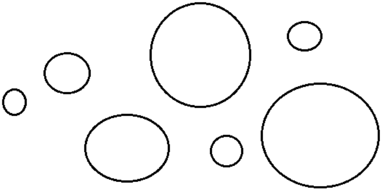
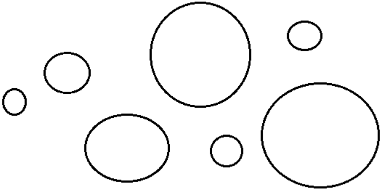


Рис.: Размер пропорционален $P_1(X_j) = p_j$

Критерий Неймана-Пирсона

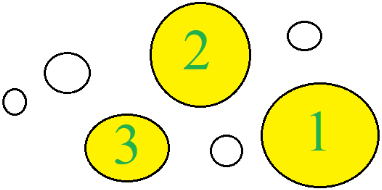


В каждой точке мы выбираем — H_0 или H_1.

Q: Какие $k = \alpha M$ точек взять за критическое множество (где мы выбираем H_1)?

Вероятность правильного выбора H_1 должна быть максимальна.

Критерий Неймана-Пирсона



A: Нужно брать точки с *наибольшим отношением правдоподобия*:

$$\frac{P_1(x_j)}{P_0(x_j)} \to \max$$

В общем случае аналогично.

Лемма Неймана-Пирсона

Сравниваем две простые гипотезы

$$H_0 : \theta = \theta_0, \qquad VS \qquad H_1 : \theta = \theta_1$$

Лемма Неймана-Пирсона

Рассмотрим отношение правдоподобий

$$T = \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} = \frac{\prod_{i=1}^{n} f(x_i;\theta_1)}{\prod_{i=1}^{n} f(x_i;\theta_0)}.$$

Пусть мы отклоняем H_0, если $T > k$. Если существует такое k, что

$$P_{\theta_0}(T > k) = \alpha,$$

то этот тест будет наиболее мощным тестом размера α.

Т.е. среди всех тестов размера α вероятность ошибки II рода $\beta(\theta_1)$ будет минимальна.

Pain

Применение леммы Неймана-Пирсона — сплошное удовольствие.

Критерий Неймана-Пирсона

Пример. $X_1, \ldots, X_n$ — i.i.d $\mathcal{N}(\mu, \sigma^2)$. Дисперсия σ^2 известна.

$$H_0 : \mu = \mu_0, \qquad VS \qquad H_1 : \mu = \mu_1 > \mu_0.$$

Критерий Неймана-Пирсона

Пример. $X_1, \dots, X_n$ — i.i.d $\mathcal{N}(\mu, \sigma^2)$. Дисперсия σ^2 известна.

$$H_0 : \mu = \mu_0, \qquad VS \qquad H_1 : \mu = \mu_1 > \mu_0.$$

- Отношение правдоподобий

$$T = \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} = \exp\left[\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i - \mu_0)^2 - \sum_{i=1}^{n}(x_i - \mu_1)^2\right)\right]$$

Критерий Неймана-Пирсона

Пример. $X_1, \dots, X_n$ — i.i.d $\mathcal{N}(\mu, \sigma^2)$. Дисперсия σ^2 известна.

$$H_0 : \mu = \mu_0, \qquad VS \qquad H_1 : \mu = \mu_1 > \mu_0.$$

- Отношение правдоподобий

$$T = \frac{\mathcal{L}(\theta_1)}{\mathcal{L}(\theta_0)} = \exp\left[\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i - \mu_0)^2 - \sum_{i=1}^{n}(x_i - \mu_1)^2\right)\right]$$

- Замечаем, что

$$\sum_{i=1}^{n}(x_i - \mu_0)^2 - \sum_{i=1}^{n}(x_i - \mu_1)^2 = n(\mu_0^2 - \mu_1^2) + 2n\bar{x}(\mu_1 - \mu_0).$$

Критерий Неймана-Пирсона

- Применяем Лемму Неймана-Пирсона. Критическое множество:

$$C = \left\{ x : \quad \exp\left[\frac{1}{2\sigma^2} \left(n(\mu_0^2 - \mu_1^2) + 2n\bar{x}(\mu_1 - \mu_0) \right) \right] \geq k \right\} =$$

$$= \left\{ x : \quad \bar{x} \geq \frac{\sigma^2}{n(\mu_1 - \mu_0)} \log k + \frac{\mu_0 + \mu_1}{2} \right\}$$

Критерий Неймана-Пирсона

- Применяем Лемму Неймана-Пирсона. Критическое множество:

$$C = \left\{x: \quad \exp\left[\frac{1}{2\sigma^2}\left(n(\mu_0^2 - \mu_1^2) + 2n\bar{x}(\mu_1 - \mu_0)\right)\right] \geq k\right\} =$$

$$= \left\{x: \quad \bar{x} \geq \frac{\sigma^2}{n(\mu_1 - \mu_0)} \log k + \frac{\mu_0 + \mu_1}{2}\right\}$$

- Обозначим правую часть за k^*.

При H_0 выполнено $\bar{x} \sim \mathcal{N}\left(\mu_0, \frac{\sigma}{\sqrt{n}}\right)$.

Находим k^* из условия $P_0(C) = \alpha$:

$$k^* = \mu_0 + z_{(1-\alpha)}\frac{\sigma}{\sqrt{n}}$$

Ответ: Тест задаётся описанным критическим множеством C.

Критерий Неймана-Пирсона

Доказательство в общем случае — см. Главу 12.

М. Б. Лагутин,
Наглядная математическая статистика.

Если не фиксировать размеры выборки, для последовательного тестирования[1] (см. Главу 12, §3), могут быть более мощные тесты.

[1] Это уже СлуПы

Критерий перестановок

Критерий перестановок

От перемены мест слагаемых сумма не меняется.

Критерий перестановок

Непараметрический критерий равенства распределений

Пусть $X_1, \ldots, X_n \sim F_X$ и $Y_1, \ldots, Y_m \sim F_Y$.

Тестируем гипотезы:

$$H_0 = \{F_X = F_Y\}, \qquad H_1 = \{F_X \neq F_Y\}$$

Q: Что здесь можно переставить (при H_0)?

Критерий перестановок

Непараметрический критерий равенства распределений

Пусть $X_1, \dots, X_n \sim F_X$ и $Y_1, \dots, Y_m \sim F_Y$.

Тестируем гипотезы:

$$H_0 = \{F_X = F_Y\}, \qquad H_1 = \{F_X \neq F_Y\}$$

Q: Что здесь можно переставить (при H_0)?

Идея. *Берём статистику* $T(X_1, \dots, X_n, Y_1, \dots, Y_m)$ *и переставляем аргументы.*

Критерий перестановок

- Вычисляем статистику

$$t = T(X_1, \dots, X_n, Y_1, \dots, Y_m)$$

- B раз случайным образом переставляем[2] данные $X_1, \dots, Y_m$ и вычисляем статистики $T_1, \dots, T_B$.

- Все T_i равновероятны, поэтому

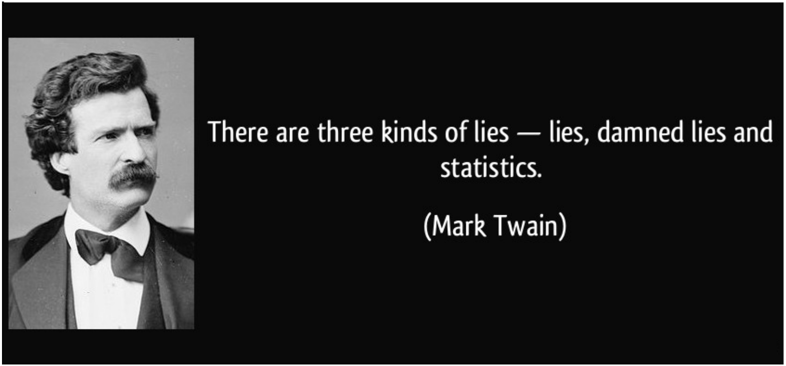
$$\text{p-value} = \frac{1}{B} \sum_{j=1}^{B} I(T_j > t).$$

[2] Не обязательно брать все перестановки

Заключение

3 kinds of lies

Никогда нет гарантий, что тест был составлен правильно, не был подогнан под ответ, или что не существует лучшей альтернативы.



Способ выстрелить себе в ногу

"Эффект Умного Ганса" = "Эффект Экспериментатора".

Экспериментатор влияет на результат.

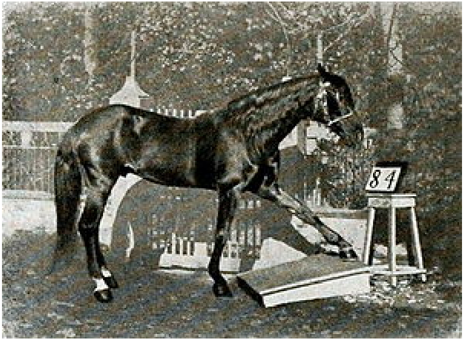
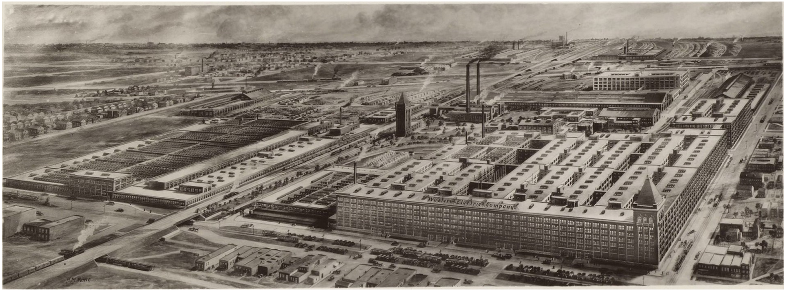


Рис.: Умный Ганс — не умел считать, но видел реакцию хозяина

Способ выстрелить себе в ногу

Насколько честно Вы ответите в соцопросе?

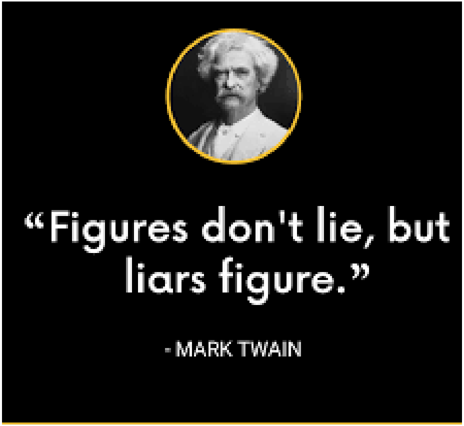
“Хоторнский эффект” — рабочие стали более эффективны из-за интереса к эксперименту.



Data never lies

Данные не врут. Врут люди.

К ML-статьям это тоже относится ☺.



Это всё теория, а как тестирование проходит на практике?

Обсудим A/B-тестирование на Лекции 7.