Прикладная статистика в машинном обучении
Лекция 4.
Проверка гипотез

И. К. Козлов
(*Мехмат МГУ*)

2022

Тестирование гипотез

Тестирование гипотез — огромная область.

Мы рассмотрим простейшие тесты, в основном из

Wasserman L.
All of Statistics.

Много других тестов описано в Главах 12, 13

М. Б. Лагутин,
Наглядная математическая статистика.

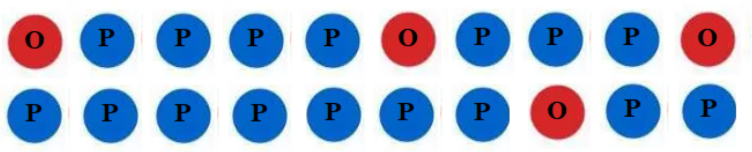
Статистика. Ликбез

1. Статистические гипотезы
   - 2 типа ошибок
   - p-value
   - Пайплайн
2. Тест Вальда
   - Свойства теста Вальда
   - Примеры
3. Тест Колмогорова
4. χ^2 критерий Пирсона
   - Распределение хи-квадрат
   - Мультиномиальное распределение
   - Критерий согласия
   - Тест на однородность
   - Точный тест Фишера

Проверка гипотез

Статистическая гипотеза — любое предположение о распределении и свойствах случайной величины.

Скажем, подбросили монетку 20 раз. Орёл выпал 4 раза.



Возникает сомнение — *а не бракованная ли монетка*?

Точно ли это выборка из Bernoulli(p), где $p = \frac{1}{2}$?

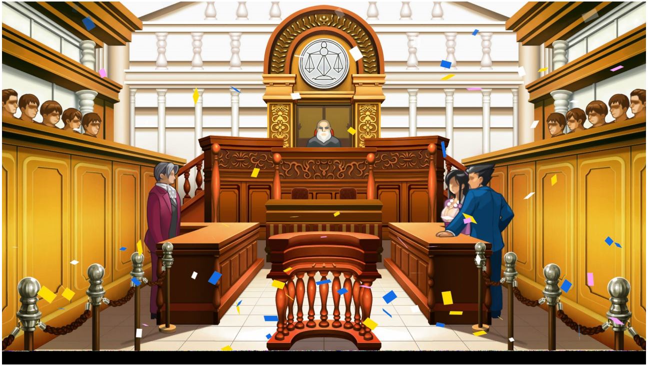
Проверка гипотез

Проверка гипотез похожа на судебное разбирательство.

- Нулевая гипотеза H_0 (обмана нет, $p = \frac{1}{2}$)

сталкивается с

- альтернативной гипотезой H_1 (есть обман, $p \neq \frac{1}{2}$).



Проверка гипотез

В общем случае:

- Разбиваем пространство параметров на 2 непересекающихся подмножества

$$\Theta = \Theta_0 \cup \Theta_1.$$

- Сравниваем две гипотезы:

$$H_0 : \theta \in \Theta_0 \qquad \text{VS} \qquad H_1 : \theta \in \Theta_1.$$

  - H_0 — нулевая (основная) гипотеза.
  - H_1 — альтернативная гипотеза.

Простые и сложные гипотезы

- Гипотеза вида $\theta = \theta_0$ называется простой гипотезой. (Иначе гипотеза сложная).
- Обычно тест двусторонний
$$H_0 : \theta = \theta_0 \qquad \text{VS} \qquad H_1 : \theta \neq \theta_0.$$
- Бывают односторонние тесты
$$H_0 : \theta \leq \theta_0 \qquad \text{VS} \qquad H_1 : \theta \geq \theta_0$$
или
$$H_0 : \theta \geq \theta_0 \qquad \text{VS} \qquad H_1 : \theta \leq \theta_0$$

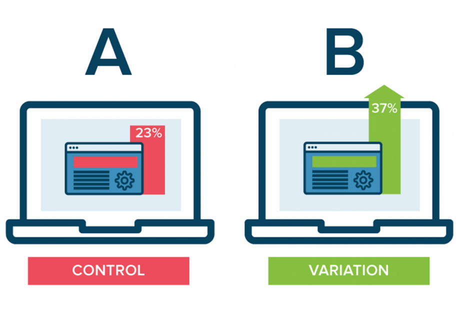
Равенство средних

Могут быть самые разные гипотезы. Мы рассмотрим простейшие примеры.

Частый вопрос — *равны ли средние значения*:

$$H_0 = \{\mathbb{E}X = \mathbb{E}Y\}, \qquad H_1 = \{\mathbb{E}X \neq \mathbb{E}Y\}$$

Вопрос из серии — *если поменять цвет кнопки в приложении, то будем ли больше зарабатывать?*



Статистический критерий

Статистический критерий — математическое правило, по которому отвергают/не отвергают нулевую гипотезу (с заданным уровнем значимости).

Q: Как “доказать”, что статистическая гипотеза не верна?

Критическая область

- Пытаемся найти статистику X, что
  - H_0 "маловероятно" (значения X велики),
  - H_1 "достаточно вероятно" (значения X невелики).
- Формально, строим критическую область R.
  - $X \in R \quad \Rightarrow \quad$ отклоняем H_0.
  - $X \notin R \quad \Rightarrow \quad$ гипотеза H_0 не противоречит экспериментальным данным.

Проверка гипотез

Обычно критическая область имеет вид

$$R = \{x : T(x) > c\} \qquad \text{или} \qquad \{x : |T(x)| > c\}$$

Презумпция невиновности

Презумпция невиновности.

Основная гипотеза верна, пока не доказано обратное.

Рис.: Гипотезу нельзя “доказать”. Только не найти достаточно “улик” против неё.

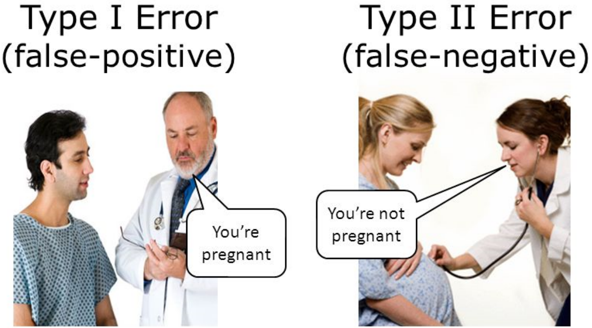
2 типа ошибок

Ошибки тестирования

Q: Какие ошибки мы можем допустить при тестировании гипотез?

Ошибки тестирования

Q: Какие ошибки мы можем допустить при тестировании гипотез?



Два типа ошибок

		reality	
		H_0 = true	H_0 = false
conclusion	H_0 is not rejected	OK	type II error
	H_0 is rejected	type I error	OK

Рис.: Два типа ошибок.

Стандартные обозначения ошибок

Стандартные обозначения:

- α — вероятность ошибки I рода (False Positive).
- β — вероятность ошибки II рода (False Negative).

Мощность теста

Более формально,

- Функция мощности[1] = вероятность отклонить H_0:

$$\pi(\theta) = \mathbb{P}_\theta(X \in R).$$

  Это вероятность попасть в критическую область.

  Если $\theta \in \Theta_0$, то $\alpha = \pi(\theta)$ — вероятность *ошибки I рода*.

  Если $\theta \in \Theta_1$, то $\beta = 1 - \pi(\theta)$ — вероятность *ошибки II рода*.

[1] У Вассермана обозначена $\beta(\theta)$.

Размер критерия

- Размер критерия = вероятность *ошибки I рода*

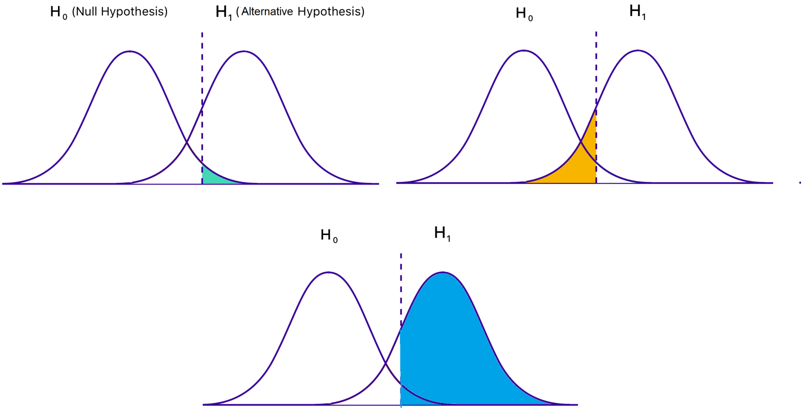
$$\alpha = \sup_{\theta \in \Theta_0} \pi(\theta).$$

Критерий имеет уровень значимости α, если его размер <= α.

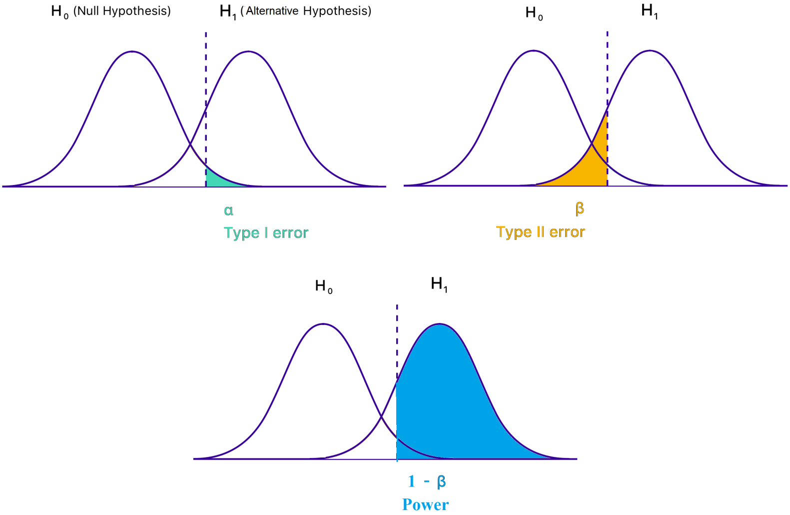
Повторение - мать учения

Изобразим гипотезы как два нормальных распределения.

Q: Каким ошибкам/понятиям соответствуют закрашенные на рис. области?



Ошибки и мощность теста

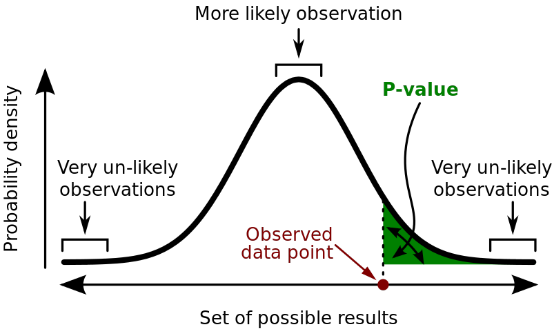


p-value

p-value

Ответ “гипотеза отклоняется/не отклоняется” неинформативен.

- p-value — наименьший уровень значимости, при котором принимается H_0.



p-value

p-value = вероятность наблюдать

ещё большие значения тестовой статистики,

если верна нулевая гипотеза H_0.

p-value

Формально, если

- статистический критерий $T(X) > c$,
- x^n — наблюдаемая реализация выборки X

то

$$\text{p-value} = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta \left(T(X^n) \geq T(x^n) \right).$$

p-value на практике

С какой вероятностью мы готовы ошибиться?

Обычно на практике p-value:

- $< 0.01 \Rightarrow H_0$ заведомо не верна,
- $0.01 - 0.05 \Rightarrow H_0$ не верна,
- $0.05 - 0.1 \Rightarrow H_0$ скорее всего не верна,
- $> 0.1 \Rightarrow$ не найдено свидетельств против H_0.

p-value

WARNING! *Большой* p-value *НЕ означает, что нулевая гипотеза верна.*

Возможно: H_0 неверно, а у теста маленькая мощность.

p-value

WARNING! *Большой* p-value *НЕ означает, что нулевая гипотеза верна.*

Возможно: H_0 неверно, а у теста маленькая мощность.

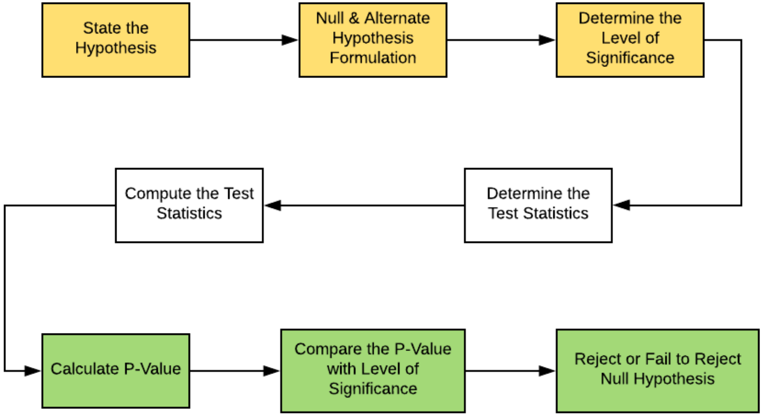
WARNING-2! p-value *это НЕ* $\mathbb{P}(H_0|Data)$ и вообще

$$\mathbb{P}(Data|H_0) \neq \mathbb{P}(H_0|Data).$$

Пайплайн

1. Статистические гипотезы
   - 2 типа ошибок
   - p-value
   - Пайплайн
2. Тест Вальда
   - Свойства теста Вальда
   - Примеры
3. Тест Колмогорова
4. χ^2 критерий Пирсона
   - Распределение хи-квадрат
   - Мультиномиальное распределение
   - Критерий согласия
   - Тест на однородность
   - Точный тест Фишера

Пайплайн для проверки гипотез:

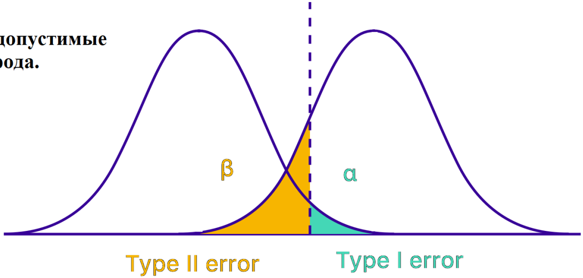


Повторим пайплайн

Схематичный пайплайн

1. Фиксируем гипотезы: $H_0 : \mu_A = \mu_B$ $H_1 : \mu_A \neq \mu_B$

2. Фиксируем допустимые ошибки I и II рода.



3. Выбираем тест и считаем p-value

4. Если p-value < α, то отклоняем H_0

Перерыв

Перерыв

Тест Вальда

Тестирование гипотез

Q: Как построить стат-тест?

A: Нужно знать распределение для $T(X)$.

Q-2: В каких известных теоремах возникает распределение?

Тестирование гипотез

Q: Как построить стат-тест?

A: Нужно знать распределение для $T(X)$.

Q-2: В каких известных теоремах возникает распределение?

A: Ну конечно ЦПТ!

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

Тест Вальда

Тест Вальда

- Тестируем гипотезы:

$$H_0 : \theta = \theta_0 \qquad \text{VS} \qquad H_1 : \theta \neq \theta_0.$$

- $\hat{\theta}$ — оценка параметра θ, $\hat{\text{se}}$ — оценка стандартной ошибки $\hat{\theta}$.

- Пусть $\hat{\theta}$ асимптотически нормально:

$$\frac{\hat{\theta} - \theta_0}{\hat{\text{se}}} \xrightarrow{d} \mathcal{N}(0, 1).$$

- Тест Вальда размера α: отклоняем H_0 тогда и только тогда, когда

$$|W| > z_{\alpha/2}, \qquad W = \frac{\hat{\theta} - \theta_0}{\hat{\text{se}}}.$$

Тест Вальда

Замечание. Можно использовать другой критерий — заменить в

$$W = \frac{\hat{\theta} - \theta_0}{\widehat{\mathrm{se}}},$$

знаменатель $\widehat{\mathrm{se}}$ на se_0 — значение стандартной ошибки при $\theta = \theta_0$.

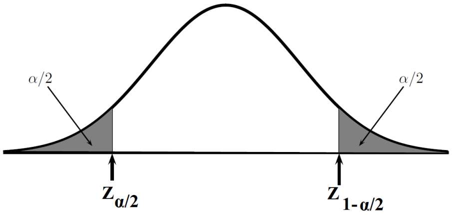
Особой разницы не будет.

Свойства теста Вальда

Свойства теста Вальда

- Поскольку $W \xrightarrow{d} \mathcal{N}(0,1)$, асимптотически размер критерия Вальда равен α:

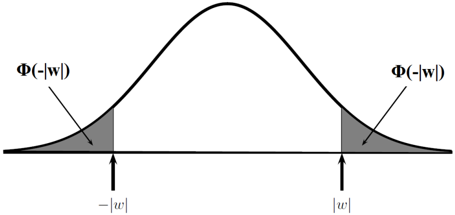
$$P_\theta(|W| > z_{\alpha/2}) \to \alpha.$$



Свойства теста Вальда

- Пусть $w = \dfrac{\hat{\theta} - \theta_0}{\hat{\text{se}}}$ — наблюдаемое значение статистики Вальда.

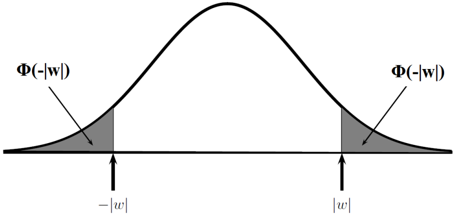
Q: Чему равно p-value?



Свойства теста Вальда

- Пусть $w = \dfrac{\hat{\theta} - \theta_0}{\hat{\text{se}}}$ — наблюдаемое значение статистики Вальда.

Q: Чему равно p-value?



A: Пусть $Z \sim \mathcal{N}(0, 1)$, тогда

$$\text{p-value} = P_{\theta_0}(|W| > |w|) \approx P(|Z| > |w|) = 2\Phi(-|w|).$$

Теста Вальда и доверительный интервал

- Критерий Вальда $\Leftrightarrow$ попадание в доверительный интервал.

Гипотеза H_0 отклоняется $\Leftrightarrow$ θ_0 не попадает в доверительный интервал

$$C_n = \left(\hat{\theta} - \hat{\text{se}}\ z_{\alpha/2}, \quad \hat{\theta} + \hat{\text{se}}\ z_{\alpha/2}, \right)$$

Доказательство:

$$|W| = \left|\frac{\hat{\theta} - \theta_0}{\hat{\text{se}}}\right| > z_{\alpha/2}, \qquad \Leftrightarrow \qquad \hat{\theta} \in C_n.$$

Практическая значимость

Доверительный интервал зачастую несёт больше информации.

Примеры

Сравнение средних

Пример 1. Сравнение двух средних.

- Пусть $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$ — две независимые выборки со средними μ_X и μ_Y соответственно.

Сравнение средних

Пример 1. Сравнение двух средних.

- Пусть $X_1, \ldots, X_n$ и $Y_1, \ldots, Y_m$ — две независимые выборки со средними μ_X и μ_Y соответственно.

- Обозначим $\delta = \mu_X - \mu_Y$. Гипотезы:

$$H_0 : \quad \delta = 0 \qquad VS \qquad H_1 : \delta \neq 0.$$

Сравнение средних

- **Q**: К чему согласно ЦПТ стремится $\bar{X} - \bar{Y}$?

Сравнение средних

- **Q**: К чему согласно ЦПТ стремится $\bar{X} - \bar{Y}$?
- **A**: $\bar{X}$ и $\bar{Y}$ независимы и асимпт. нормальны, поэтому

$$\bar{X} - \bar{Y} \approx \mathcal{N}(\delta, \mathbb{V}(\bar{X}) + \mathbb{V}(\bar{Y})).$$

Сравнение средних

- **Q**: К чему согласно ЦПТ стремится $\bar{X} - \bar{Y}$?
- **A**: $\bar{X}$ и $\bar{Y}$ независимы и асимпт. нормальны, поэтому

$$\bar{X} - \bar{Y} \approx \mathcal{N}(\delta, \mathbb{V}(\bar{X}) + \mathbb{V}(\bar{Y})).$$

- Получаем, оценка стандартной ошибки

$$\hat{\mathrm{se}} = \sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}},$$

где S_X^2 и S_Y^2 — выборочные дисперсии.

Сравнение средних

- **Тест Вальда размера** α отклоняет H_0, если

$$|W| > z_{\alpha/2},$$

где

$$W = \frac{\hat{\delta} - 0}{\hat{\mathrm{se}}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}}.$$

Сравнение медиан

Пример 2. Сравнение медиан.

Вместо средних в Примере 1 берём медианы ν_X и ν_Y.

Берём разницу выборочных оценок $\hat{\delta} = \hat{\nu}_X - \hat{\nu}_Y$.

Сравнение медиан

Тест Вальда по сути тот же

$$W = \frac{\hat{\delta}}{\widehat{\mathrm{se}}}$$

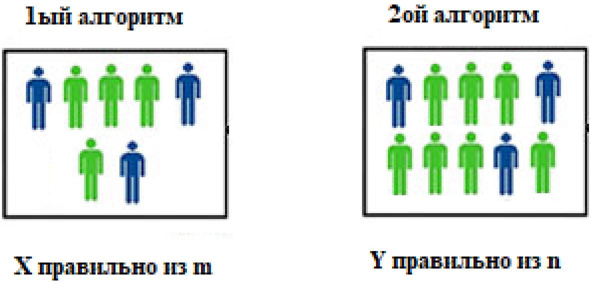
Q: Как оценить $\widehat{\mathrm{se}}$?

Нет аналитической формулы — используем бутстреп.

Рис.: Bootstrap that I used to know

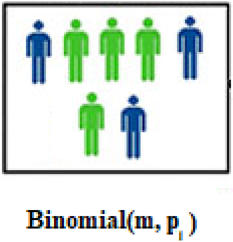
Сравнение алгоритмов

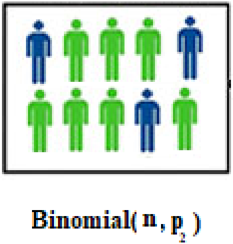
Пример 3. Сравниваем предсказания 2 алгоритмов на 2 множествах.



Q: Какие распределения у X и Y?

Сравнение алгоритмов





Q: Как протестировать $p_1 = p_2$?

Сравнение алгоритмов

A: Биномиальные распределения — суммы распределений Бернулли.

Если $X = X_1 + \dots + X_m$ и $Y = Y_1 + \dots + Y_n$, то

$$p_1 - p_2 = \frac{X_1 + \dots + X_m}{m} - \frac{Y_1 + \dots + Y_n}{n} \xrightarrow{d}$$

$$\xrightarrow{d} \mathcal{N}\left(p_1 - p_2, \frac{p_1(1 - p_1)}{m} + \frac{p_2(1 - p_2)}{n}\right).$$

Зависимые алгоритмы

Тест Вальда:

$$W = \frac{\hat{\delta}}{\hat{\mathrm{se}}}.$$

- Оценка разности матожиданий:

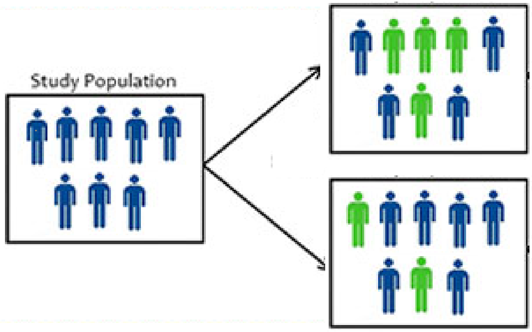
$$\hat{\delta} = \hat{p}_1 - \hat{p}_2,$$

- Оценка стандартной ошибки:

$$\hat{\mathrm{se}} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{m} + \frac{\hat{p}_2(1-\hat{p}_2)}{n}}.$$

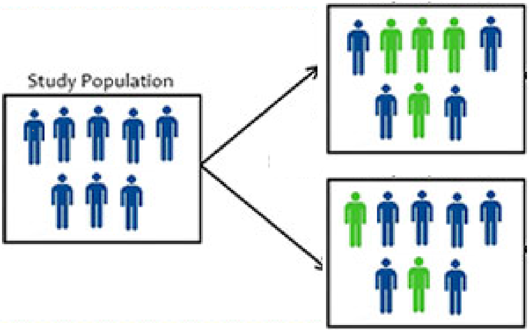
Зависимые алгоритмы

Пример 4. А если оба алгоритма — на одном множестве. Есть разница?



Зависимые алгоритмы

Пример 4. А если оба алгоритма — на одном множестве. Есть разница?



А: Да, X_i и Y_j теперь зависимы.

Зависимые алгоритмы

Идея. Вместо X_i и Y_i рассматриваем $X_i - Y_i$.

Test Case	X_i	Y_i	$D_i = X_i - Y_i$
1	1	0	1
2	1	1	0
3	1	1	0
4	0	1	-1
5	0	0	0
⋮	⋮	⋮	⋮
n	0	1	-1

Зависимые алгоритмы

Тест Вальда:

$$W = \frac{\hat{\delta}}{\hat{\text{se}}}.$$

- Оценка разности матожиданий:

$$\hat{\delta} = \frac{1}{n}\sum_{i=1}^{n} D_i,$$

- Оценка стандартной ошибки:

$$\hat{\text{se}} = \frac{S}{\sqrt{n}}, \qquad S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(D_i - \bar{D}\right)^2.$$

Тест Колмогорова

Колмогоров

Раз лекции на мехмате, нужно упомянуть Колмогорова)

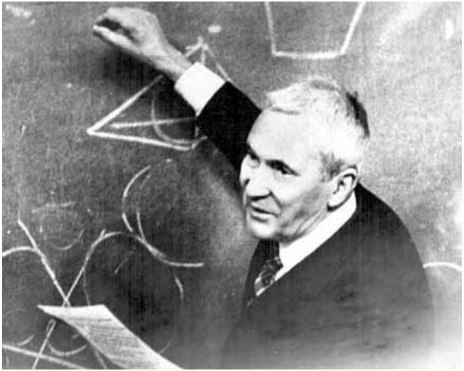


Рис.: А.Н. Колмогоров

Не будем использовать этот тест.

Теорема Колмогорова

Пусть $F(x)$ — *непрерывная* функция распределения.

Теорема Колмогорова

$$\sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{d} K$$

Где K — распределение Колмогорова

$$K(x) = \begin{cases} \sum\limits_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2}, & x > 0; \\ 0, & x \leqslant 0. \end{cases}$$

Тест Колмогорова

Проверка принадлежности выборки распределению.

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

Отвергаем гипотезу, что $F = F_0$, если

$$t(X^n) = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \geq C.$$

Перерыв

χ^2 критерий Пирсона

Распределение хи-квадрат

Три “Богатыря” Статистики

Сейчас мы познакомимся с 3 очень важными распределениями.

Рис.: Три “Богатыря” Статистики — Стьюдент, Фишер и Хи-квадрат

Три важных распределения

Есть 3 важных распределения:

1. Хи-квадрат χ^2.
2. t-распределение Стьюдента.
3. Распределение Фишера-Снедекора F_{d_1,d_2} (F-распределение).

Три важных распределения

$$\chi_k^2 = Z_1^2 + \dots + Z_k^2$$

$$Z_i \sim \mathcal{N}(0,1)$$

Хи-квадрат = сумма квадратов нормальных распределений

χ^2-критерий Пирсона

(Goodness of fit)

$$\text{Student} = \frac{\mathcal{N}(0,1)}{\sqrt{\chi_k^2/k}}$$

расп. Стьюдента

Критерий Стьюдента

$$\mathbb{E}X = \mathbb{E}Y$$

$$F_{d_1,d_2} = \frac{\chi_{d_1}^2/d_1}{\chi_{d_2}^2/d_2}$$

расп. Фишера

Критерий Фишера

$$\mathbb{V}X = \mathbb{V}Y$$

2 популярных критерия

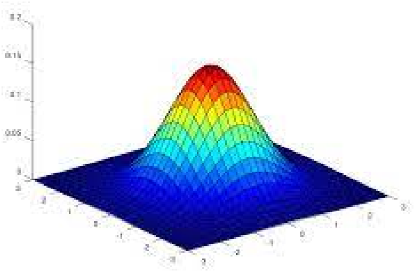
(Будет на 7 Занятии)

Контрольный вопрос

Q: Рассмотрим многомерное нормальное распределение

$$\xi = (\xi_1, \ldots, \xi_k) \sim \mathcal{N}(0, I_k).$$

Где распределение хи-квадрат?

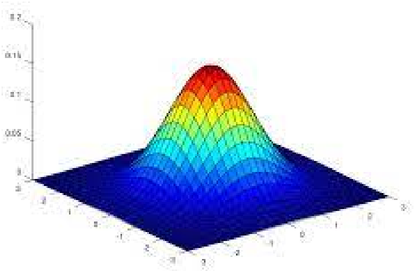


Контрольный вопрос

Q: Рассмотрим многомерное нормальное распределение

$$\xi = (\xi_1, \ldots, \xi_k) \sim \mathcal{N}(0, I_k).$$

Где распределение хи-квадрат?



A: Квадрат длины вектора

$$\|\xi\|^2 = \xi_1^2 + \cdots + \xi_k^2.$$

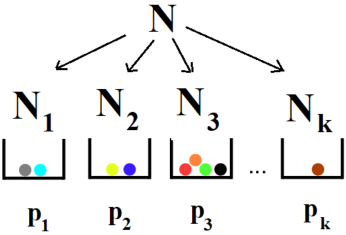
Мультиномиальное распределение

1. Статистические гипотезы
   - 2 типа ошибок
   - p-value
   - Пайплайн
2. Тест Вальда
   - Свойства теста Вальда
   - Примеры
3. Тест Колмогорова
4. χ^2 критерий Пирсона
   - Распределение хи-квадрат
   - **Мультиномиальное распределение**
   - Критерий согласия
   - Тест на однородность
   - Точный тест Фишера

χ^2-критерий Пирсона

χ^2-критерий Пирсона используется, когда данные "раскладываются по коробочкам".

Мы проверяем — *совпадают ли эмпирические частоты с теоретическими.*



Мультиномиальное распределение

Q: Какое самое известное распределение с возможными исходами $1, \ldots, k$.

A: Пусть $X = (X_1, \ldots, X_k)$ имеют *мультиномиальное* (n, p) *распределение*.

$$p(\mathbf{x}) = \frac{n!}{x_1! \ldots x_k!} p_1^{x_1} \ldots p_k^{x_k}, \qquad \sum_{j=1}^{k} x_j = n$$

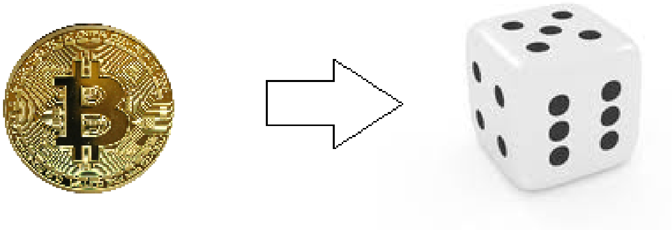


Рис.: От монеток переходим к кубикам

χ^2-статистика Пирсона

Проверяем гипотезу

$$H_0 : p = p_0 \qquad VS \qquad H_1 : p \neq p_0.$$

χ^2 статистика Пирсона

$$T = \sum_{j=1}^{k} \frac{(X_j - np_{0j})^2}{np_{0j}} = \sum_{j=1}^{k} \frac{(X_j - E_j)^2}{E_j}.$$

Здесь $E_j = \mathbb{E}X_j = np_{0j}$ — матожидание при H_0.

χ^2-статистика Пирсона

- При выполнении H_0:

$$T \xrightarrow{d} \chi^2_{k-1}$$

- Тест: отвергаем H_0, если $\chi^2_{k-1} > \chi^2_{k-1,\alpha}$ соответствующей квантили.

- $p - value = \mathbb{P}(\chi^2_{k-1} > t)$, где t — наблюдаемое значение статистики.

Критерий согласия

Тест параметрической модели

Фиксируем параметрическую модель

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

Принадлежит ли выборка этой модели?

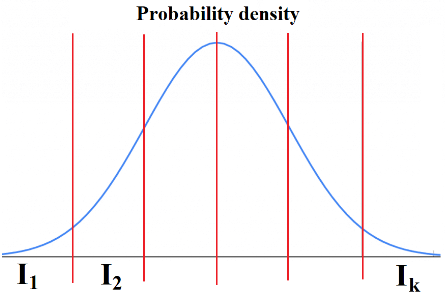
Создаём мультиномиальное распределение

Идея! “Разложим выборку X_i по коробочкам”,
чтобы получить мультиномиальное распределение.

Создаём мультиномиальное распределение

Делим прямую на k непересекающихся отрезков $I_1, \ldots, I_k$

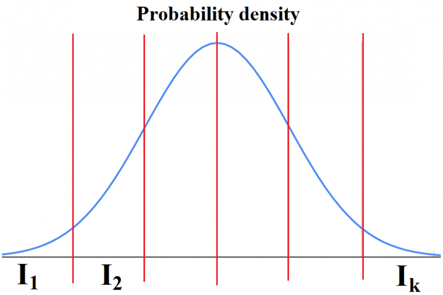
Q: Вероятность попасть в отрезок?



Создаём мультиномиальное распределение

Делим прямую на k непересекающихся отрезков $I_1, \ldots, I_k$

Q: Вероятность попасть в отрезок?



A: Вероятность:

$$p_j(\theta) = \int_{I_j} f(x; \theta) dx$$

Создаём мультиномиальное распределение

Вероятность зависит от параматра:

$$p_j(\theta) = \int_{I_j} f(x;\theta)dx$$

Q: Какой параметр θ взять?

Создаём мультиномиальное распределение

Вероятность зависит от параматра:

$$p_j(\theta) = \int_{I_j} f(x;\theta)dx$$

Q: Какой параметр θ взять?

A: Это Статистика — поэтому самый правдоподобный (в описанном ниже смысле).

Правдоподобие

- Пусть в отрезок I_j попало N_j наблюдений.
- **Q**: Какова вероятность этого?

Правдоподобие

- Пусть в отрезок I_j попало N_j наблюдений.
- **Q**: Какова вероятность этого?
- **A**: *Правдоподобие:*

$$Q(\theta) = \prod_{j=1}^{k} p_j(\theta)^{N_j}.$$

Выбираем самое правдоподобное

- *Выбираем самое правдоподобное*:

$$(\tilde{\theta}_1, \ldots, \tilde{\theta}_s) = \mathrm{argmax}_\theta\, Q(\theta)$$

Замечание. $\tilde{\theta}$ — это НЕ MLE $\hat{\theta}$, не перепутайте.

Goodness-of-fit Test

Критерий согласия

Рассмотрим статистику

$$Q = \sum_{j=1}^{k} \frac{(N_j - np_j(\tilde{\theta}))^2}{np_j(\tilde{\theta})}.$$

Goodness-of-fit Test

Критерий согласия

Рассмотрим статистику

$$Q = \sum_{j=1}^{k} \frac{(N_j - np_j(\tilde{\theta}))^2}{np_j(\tilde{\theta})}.$$

- При выполнении H_0

$$Q \xrightarrow{d} \chi^2_{k-1-s}.$$

- Поэтому (приближённо)

$$\text{p-value} = \mathbb{P}(\chi^2_{k-1-s} > q),$$

  где q — наблюдаемое значение статистики.

Goodness-of-fit Test

Критерий согласия

Рассмотрим статистику

$$Q = \sum_{j=1}^{k} \frac{(N_j - np_j(\tilde{\theta}))^2}{np_j(\tilde{\theta})}.$$

- При выполнении H_0

$$Q \xrightarrow{d} \chi^2_{k-1-s}.$$

- Поэтому (приближённо)

$$\text{p-value} = \mathbb{P}(\chi^2_{k-1-s} > q),$$

  где q — наблюдаемое значение статистики.

Здесь k — число отрезков $I_1, \ldots, I_k$,

s — число параметров $\theta_1, \ldots, \theta_s$.

Тест на однородность

1. Статистические гипотезы
   - 2 типа ошибок
   - p-value
   - Пайплайн
2. Тест Вальда
   - Свойства теста Вальда
   - Примеры
3. Тест Колмогорова
4. χ^2 критерий Пирсона
   - Распределение хи-квадрат
   - Мультиномиальное распределение
   - Критерий согласия
   - Тест на однородность
   - Точный тест Фишера

Читайте хорошие книжки

Книжка так себе, но местами иллюстрации наглядные.

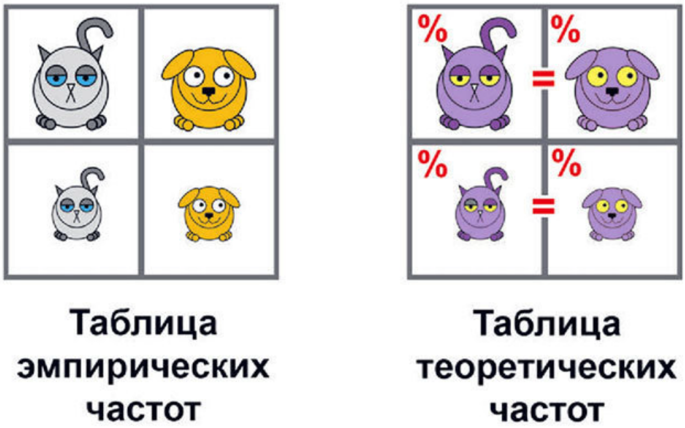
Тест на однородность

Кроме “коробочек” аналогичный критерий можно применять к “табличкам”.

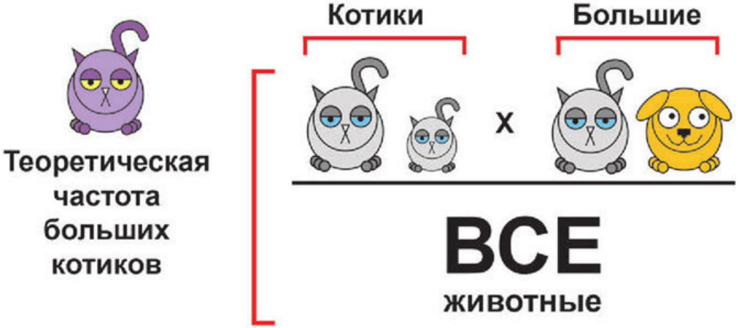
Проверяем — *одинаковые ли распределения в столбиках*?

Тест на однородность

Составляем 2 таблицы

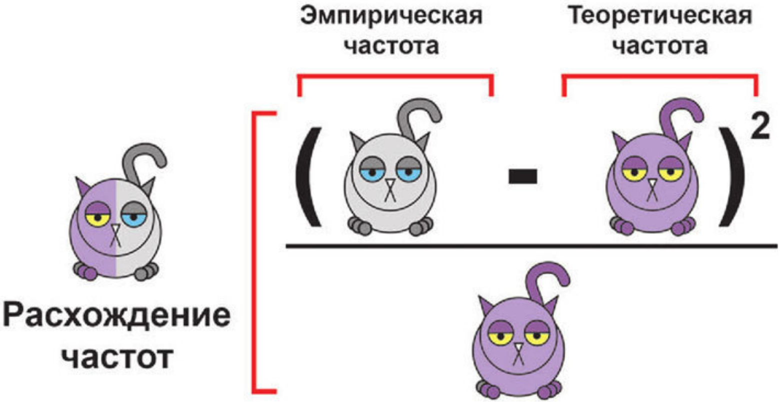


Теоретические частоты вычисляем так:



Тест на однородность

Вычисляем расхождения частот:



Тест на однородность

Сумма расхождений частот имеет распределение χ^2.

В общем случае число степеней свободы $df = (Rows - 1) \times (Cols - 1)$.

Точный тест Фишера

Дополнительный материал

Критерий хи-квадрат асимптотический (НЕ точный).

Rule of Thumb. Тест не годится, если значение в таблице сопряжённости меньше 5 (или 10 для 1 степени свободы).

Точный тест Фишера

Если хочется точного теста, можно фиксировать суммы в строках и/или столбцах таблицы.

Пример: Точный тест Фишера.

Рис.: см. также “Lady tasting tea” experiment

Точный тест Фишера

Суммы в строках и столбцах фиксированы.

	Tea	**Milk**	*Row Total*
Correct	a	b	$a + b$
Wrong	c	d	$c + d$
Column Total	$a + c$	$b + d$	$a + b + c + d$ (=n)

Вероятность набора задаётся гипергеометрическим распределением:

$$p = \binom{a+b}{a}\binom{c+d}{c} \Big/ \binom{n}{a+c} = \frac{(a+b)!\,(c+d)!\,(a+c)!\,(b+d)!}{n!\,a!\,b!\,c!\,d!}$$

$p - value$ = сумма вероятностей для наблюдаемого a и более экстремальных.

Конец 1ой части лекции

Конец 1ой части лекции.

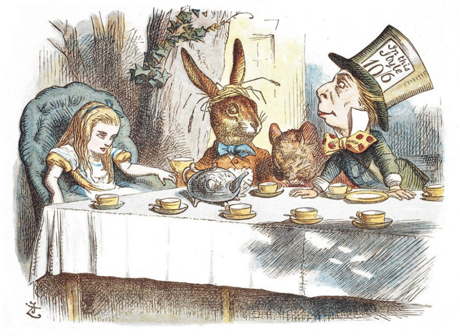


Рис.: Time for Tea Break