Проверка гипотез

а) Основные понятия теории проверки гипотез

Износостойкость

$\boldsymbol{H_0}$ **(нулевая гипотеза): износостойкость одинаковая**

1 $\boldsymbol{H_1}$ **(альтернативная гипотеза): износостойкость в I группе больше, чем во II группе**

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

Пусть X **- случайная величина со значениями в** $\mathcal{X}$

$R \subset \mathcal{X}$ **- критическая область**

$\boldsymbol{x} \in \boldsymbol{R} \Rightarrow$ **гипотеза** $\boldsymbol{H_0}$ **отклоняется в пользу альтернативы**

$\boldsymbol{x} \notin \boldsymbol{R} \Rightarrow$ **гипотеза** $\boldsymbol{H_0}$ **экспериментальным данным не противоречит**

$$R = \left\{ x : \ T(x) > c \right\}$$

T **- статистика,** c **- критическое значение**

<u>Замечание</u>. На практике при тестировании гипотез зачастую лучше также строить и доверительный интервал – больше информации!

Тестирование гипотез ~ презумпция невиновности – предполагается невиновность до тех пор, пока факты не покажут обратное

	принять H_0	отклонить H_0
верна H_0	OK	Ошибка I рода
верна H_1	Ошибка II рода	OK

<u>Определение</u>. Функцией мощности критерия с критической областью R
3 **называют функцию**

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

Размер критерия определяется согласно формуле

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

Говорят, что критерий имеет уровень α если его размер меньше либо равен α

$\theta = \theta_0$ **- простая гипотеза**

$\theta > \theta_0$ **или** $\theta < \theta_0$ **- сложная гипотеза**

Двухсторонний критерий: $H_0 : \theta = \theta_0$ **vs.** $H_1 : \theta \neq \theta_0$

4 **Односторонний критерий:**

$$H_0 : \theta \leq \theta_0 \text{ vs. } H_1 : \theta > \theta_0$$

$$H_0 : \theta \geq \theta_0 \text{ vs. } H_1 : \theta < \theta_0$$

<u>Пример</u>. $X_1, \ldots, X_n \sim N(\mu, \sigma)$**,** σ **- известно.**

$$H_0 : \mu \leq 0 \text{ vs. } H_1 : \mu > 0$$

Значит, $\Theta_0 = (-\infty, 0]$ **и** $\Theta_1 = (0, \infty)$

Рассмотрим критерий:

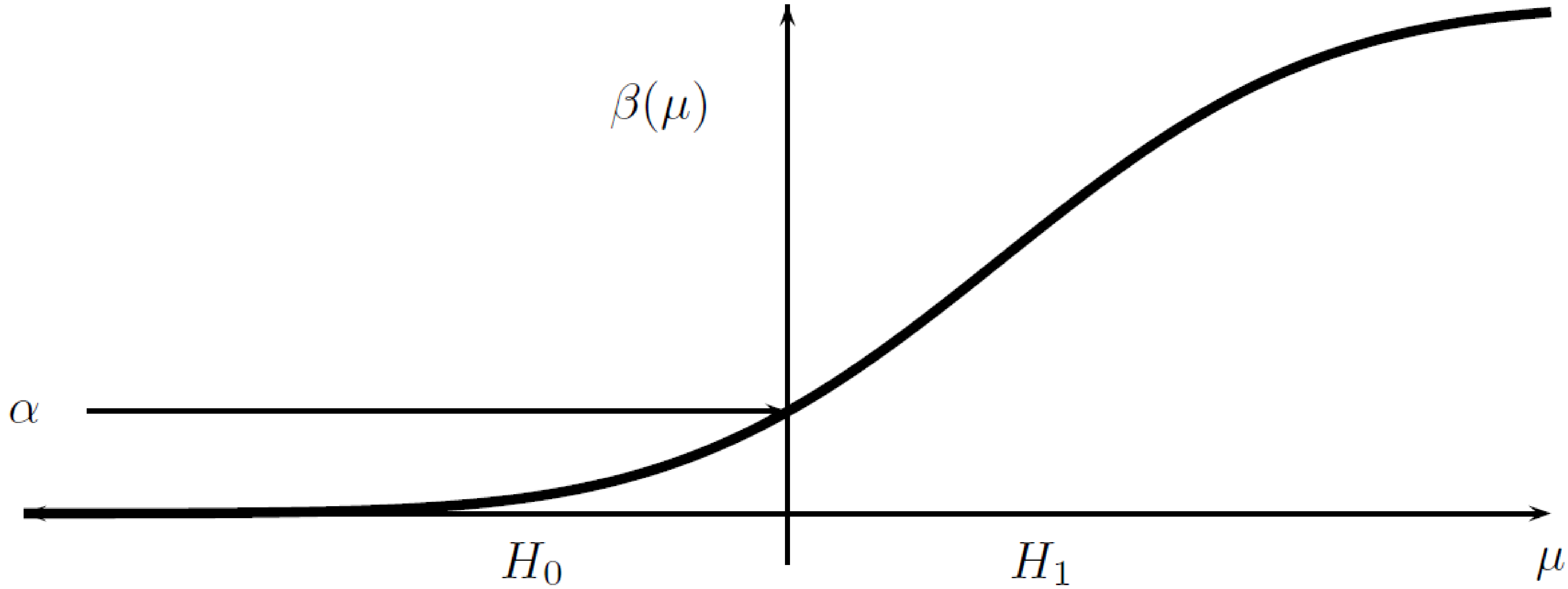
H_0 **отклоняется, если** $T > c$ **, где** $T = \overline{X}$

Критическая область:

$$R = \left\{(x_1, \dots, x_n) : \; T(x_1, \dots, x_n) > c\right\}$$

Обозначим через Z **- с.в., распределенную стандартно нормально**

$$\begin{aligned}
\beta(\mu) &= \mathbb{P}_\mu\left(\overline{X} > c\right) \\
&= \mathbb{P}_\mu\left(\frac{\sqrt{n}(\overline{X} - \mu)}{\sigma} > \frac{\sqrt{n}(c - \mu)}{\sigma}\right) \\
&= \mathbb{P}\left(Z > \frac{\sqrt{n}(c - \mu)}{\sigma}\right) \\
&= 1 - \Phi\left(\frac{\sqrt{n}(c - \mu)}{\sigma}\right)
\end{aligned}$$



6

Таким образом, размер критерия равен

$$\sup_{\mu \le 0} \beta(\mu) = \beta(0) = 1 - \Phi\left(\frac{\sqrt{n}c}{\sigma}\right)$$

Чтобы размер критерия равнялся α необходимо, чтобы

$$c = \frac{\sigma\,\Phi^{-1}(1-\alpha)}{\sqrt{n}}$$

Основная гипотеза отклоняется, если $\overline{X} > \sigma\,\Phi^{-1}(1-\alpha)/\sqrt{n}$**, то есть если**

7

$$\frac{\sqrt{n}\,(\overline{X}-0)}{\sigma} > z_\alpha,$$

где $z_\alpha = \Phi^{-1}(1-\alpha)$

<u>Наиболее мощный критерий</u> – критерий, который имеет максимальную мощность относительно гипотезы H_1 **среди критериев размера** α

b) Критерий Вальда

Пусть θ **- скалярный параметр,** $\widehat{\theta}$ **- его оценка,** $\widehat{\mathsf{se}}$ **- оценка стандартной ошибки оценки** $\widehat{\theta}$

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

Допустим, что $\widehat{\theta}$ **- асимптотически нормальная оценка, то есть**

$$\frac{(\widehat{\theta} - \theta_0)}{\widehat{\mathsf{se}}} \rightsquigarrow N(0,1)$$

Согласно критерию Вальда размера α **гипотеза** H_0 **отклоняется, если** $|W| > z_{\alpha/2}$**, где**

$$W = \frac{\widehat{\theta} - \theta_0}{\widehat{\mathsf{se}}}$$

<u>Теорема</u>. Асимптотически размер критерия Вальда равен α**, то есть** $\mathbb{P}_{\theta_0}\left(|W| > z_{\alpha/2}\right) \to \alpha$ **при** $n \to \infty$

<u>Доказательство</u>.

При условии, что $\theta = \theta_0$ **в силу асимптотической нормальности**
9 **выполнено, что** $(\widehat{\theta} - \theta_0)/\widehat{\mathsf{se}} \rightsquigarrow N(0,1)$**. Следовательно, вероятность отклонить основную гипотезу, когда она на самом деле верна, равняется**

$$
\begin{aligned}
\mathbb{P}_{\theta_0}\left(|W| > z_{\alpha/2}\right) &= \mathbb{P}_{\theta_0}\left(\frac{|\widehat{\theta} - \theta_0|}{\widehat{\mathsf{se}}} > z_{\alpha/2}\right) \\
&\to \mathbb{P}\left(|Z| > z_{\alpha/2}\right) \\
&= \alpha
\end{aligned}
$$

при $Z \sim N(0,1)$

<u>Замечание</u>. Можно использовать также статистику $W = (\widehat{\theta} - \theta_0)/\mathsf{se}_0$**, где** se_0 **- значение стандартной ошибки, подсчитанное при** $\theta = \theta_0$

<u>Теорема</u>. Допустим, что реальное значение θ **есть** $\theta_\star \neq \theta_0$**. В таком случае значение мощности критерия Вальда** $\beta(\theta_\star)$ **(вероятность** $P\left(H_1 \mid H_1\right)$

принять альтернативную гипотезу, когда она верна) в пределе равна

$$1 - \Phi\left(\frac{\theta_0 - \theta_\star}{\widehat{\mathsf{se}}} + z_{\alpha/2}\right) + \Phi\left(\frac{\theta_0 - \theta_\star}{\widehat{\mathsf{se}}} - z_{\alpha/2}\right)$$

<u>Замечание</u>. $\widehat{\mathsf{se}}$ **стремится к** 0 **при** $n \to \infty$ **=> мощность критерия Вальда большая, если а)** $\theta_\star$ **сильно отличается от** θ_0 **b) объем выборки большой**

Пример. **Сравнение двух алгоритмов прогнозирования**
Пусть алгоритм I тестируется на выборке объема m**, а другой алгоритм – на выборке объема** n**. Обозначим через** X **- неверных предсказаний для алгоритма I, а через** Y **- количество неверных предсказаний для алгоритма II. В таком случае** $X \sim \text{Binomial}(m, p_1)$ **и** $Y \sim \text{Binomial}(n, p_2)$**. Пусть также** $\delta = p_1 - p_2$

$$H_0 : \delta = 0 \text{ vs. } H_1 : \delta \neq 0$$

ОМП для δ **равно** $\widehat{\delta} = \widehat{p}_1 - \widehat{p}_2$**, оценка стандартной ошибки оценки равна**

$$\widehat{\mathsf{se}} = \sqrt{\frac{\widehat{p}_1(1 - \widehat{p}_1)}{m} + \frac{\widehat{p}_2(1 - \widehat{p}_2)}{n}}$$

Согласно критерия Вальда гипотеза H_0 **отвергается, если** $|W| > z_{\alpha/2}$**, где**

$$W = \frac{\widehat{\delta} - 0}{\widehat{\mathsf{se}}} = \frac{\widehat{p}_1 - \widehat{p}_2}{\sqrt{\frac{\widehat{p}_1(1-\widehat{p}_1)}{m} + \frac{\widehat{p}_2(1-\widehat{p}_2)}{n}}}$$

12

Допустим, что тестовое множество одинаково для обоих алгоритмов. В таком случае наблюдения зависимы.

Используем следующую стратегию (парное сравнение): пусть $X_i = 1$**, если алгоритм I выдал правильный ответ и** $X_i = 0$ **в противном случае. Аналогично,** $Y_i = 1$ **если алгоритм II выдал правильный ответ и** $Y_i = 0$ **в противном случае.**

Определим $D_i = X_i - Y_i$

номер	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
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	0	1	-1

$$\delta = \mathbb{E}(D_i) = \mathbb{E}(X_i) - \mathbb{E}(Y_i) = \mathbb{P}(X_i = 1) - \mathbb{P}(Y_i = 1)$$

Непараметрическая оценка для δ **равна** $\widehat{\delta} = \overline{D} = n^{-1}\sum_{i=1}^{n} D_i$ **и** $\widehat{\mathsf{se}}(\widehat{\delta}) = S/\sqrt{n}$**, где** $S^2 = n^{-1}\sum_{i=1}^{n}(D_i - \overline{D})^2$

$H_0 : \delta = 0$ **vs.** $H_1 : \delta \neq 0$

14 $$W = \widehat{\delta}/\widehat{\mathsf{se}}$$

гипотеза H_0 **отвергается, если** $|W| > z_{\alpha/2}$

<u>Пример</u>. Сравнение средних значений. Пусть $X_1, \ldots, X_m$ **и** $Y_1, \ldots, Y_n$ **- две независимые выборки из генеральных совокупностей, средние значения которых равны** μ_1 **и** μ_2 **соответственно,** s_1^2 **и** s_2^2 **- выборочные дисперсии. Положим** $\delta = \mu_1 - \mu_2$**.**

$$H_0 : \delta = 0 \quad \text{vs.} \quad H_1 : \delta \neq 0$$

$$\widehat{\delta} = \overline{X} - \overline{Y}$$

$$\widehat{\mathsf{se}} = \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

гипотеза H_0 **отвергается, если** $|W| > z_{\alpha/2}$**, где**

$$W = \frac{\widehat{\delta} - 0}{\widehat{\mathsf{se}}} = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

<u>Пример</u>. Сравнение двух медиан. Условия задачи аналогичны предыдущей с той лишь разницей, что теперь сравниваются значения двух медиан. Положим $\delta = \nu_1 - \nu_2$**, где** ν_1 **и** ν_2 **- медианы.**

$$H_0 : \delta = 0 \text{ vs. } H_1 : \delta \neq 0$$

Непараметрическая оценка δ **равна** $\widehat{\delta} = \widehat{\nu}_1 - \widehat{\nu}_2$**, где** $\widehat{\nu}_1$ **и** $\widehat{\nu}_2$ **- выборочные значения медиан. Оценка** $\widehat{\mathsf{se}}$ **стандартной ошибки оценки** $\widehat{\delta}$ **может быть получена с помощью бутстрепа. Статистика критерия Вальда имеет вид** $W = \widehat{\delta}/\widehat{\mathsf{se}}$

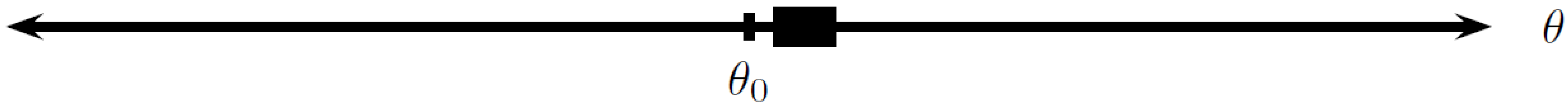
<u>Теорема</u>. Критерий Вальда размера α отклоняет $H_0 : \theta = \theta_0$ в пользу $H_1 : \theta \neq \theta_0$, если и только если $\theta_0 \notin C$, где

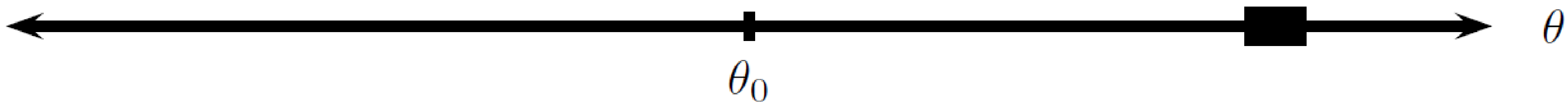
$$C = (\widehat{\theta} - \widehat{\mathsf{se}}\, z_{\alpha/2},\ \widehat{\theta} + \widehat{\mathsf{se}}\, z_{\alpha/2})$$

Таким образом, тестирование гипотезы эквивалентно проверке, попало ли значение θ_0 в доверительный интервал

<u>Замечание</u>. В случае отклонения гипотезы $\boldsymbol{H_0}$ говорят, что результат статистически значим. Однако «запас», с которым отклонили $\boldsymbol{H_0}$, может быть мал – результат мало значим с практической точки зрения.

Если доверительный интервал не включает θ_0, то $\boldsymbol{H_0}$ отклоняется. Однако значения в этом интервале могут быть близки к θ_0 (результат не значим с практической точки зрения), а могут быть и далеки от θ_0.





Таким образом, если даже результат статистически значим – это еще не означает, что он значим с практической точки зрения. При этом доверительны интервалы более информативны, чем просто проверка гипотез

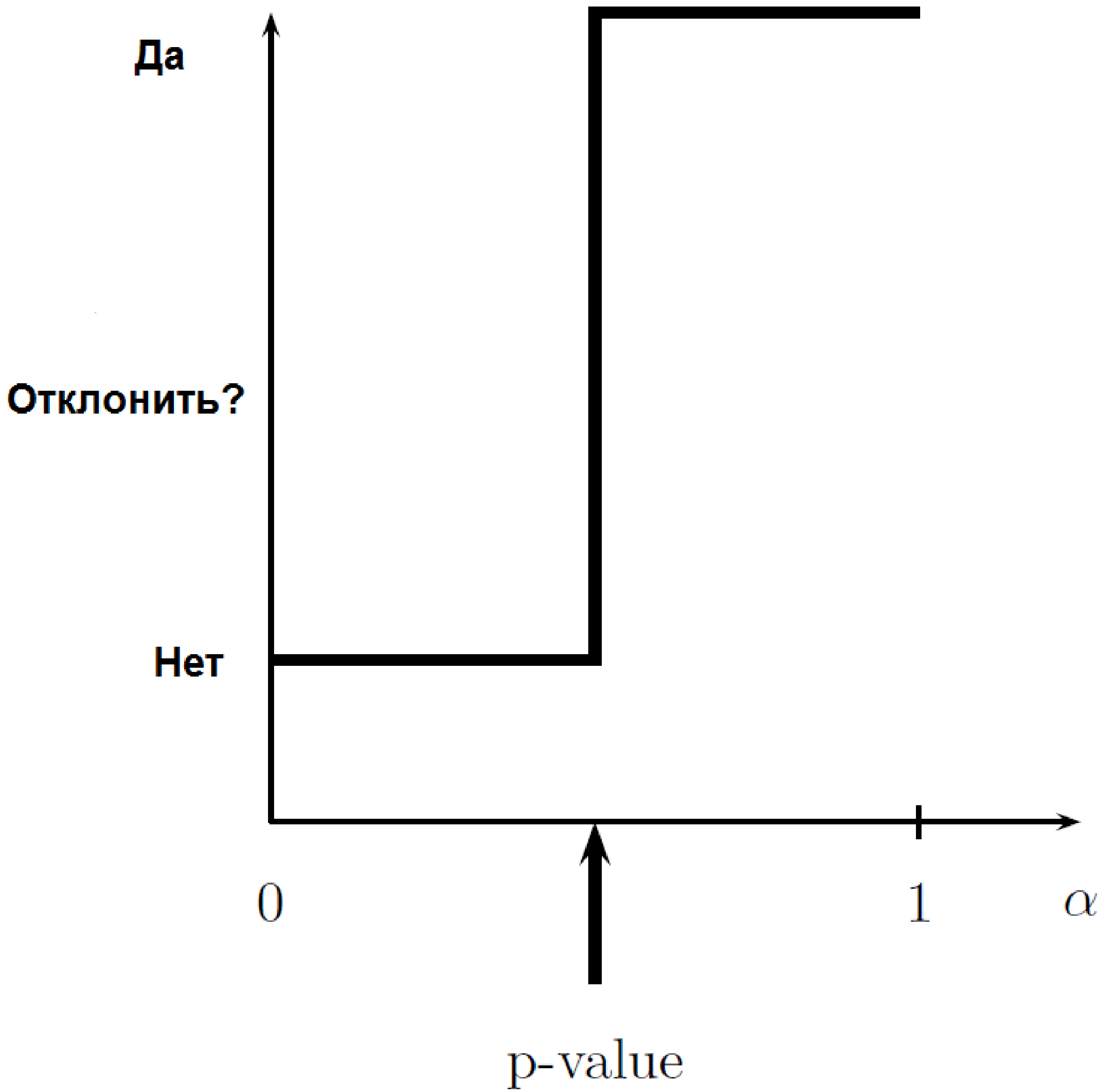
c) Квантили

Ответы типа “отклонить H_0” или “принять H_0” не очень информативны. Можно для каждого значения α узнать, отклоняется ли H_0 при этом значении α. Вообще говоря, если гипотеза была отклонена на уровне α, то и на уровне $\alpha' > \alpha$ она также будет отклонена

<u>Определение</u>. Допустим, что для каждого $\alpha \in (0, 1)$ имеется критерий размера $\alpha \in (0, 1)$ с критической областью R_α. Тогда

$$\text{p-value} = \inf\left\{\alpha : \; T(X^n) \in R_\alpha\right\}$$

p-value – наименьший уровень значимости, на котором еще можно отклонить H_0. Чем больше p-value – тем вероятнее, что H_0 надо отклонить.


Да
Отклонить?
Нет
0
1
α
p-value

p-value:

< 0.01 => H_0 заведомо не верна

0.01 – 0.05 => H_0 не верна

0.05 – 0.10 => H_0 скорее всего не верна

> 0.1 => ничего определенного о гипотезе H_0 сказать нельзя

Замечание. Большое p-value не является «подтверждением» H_0. Большое p–value появляется, если: a) H_0 верна b) H_0 неверна, но мощность теста невилика

Замечание. Не надо путать p-value и $\mathbb{P}(H_0|\mathrm{Data})$: p-value не является вероятностью того, что верна нулевая гипотеза

<u>Теорема</u>. Допустим, что критерий размера α имеет вид

$$\boldsymbol{H_0} \text{ **отвергается, если** } T(X^n) \geq c_\alpha$$

Тогда

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

22 **где x^n - реализация выборки X^n. Если $\Theta_0 = \{\theta_0\}$, то**

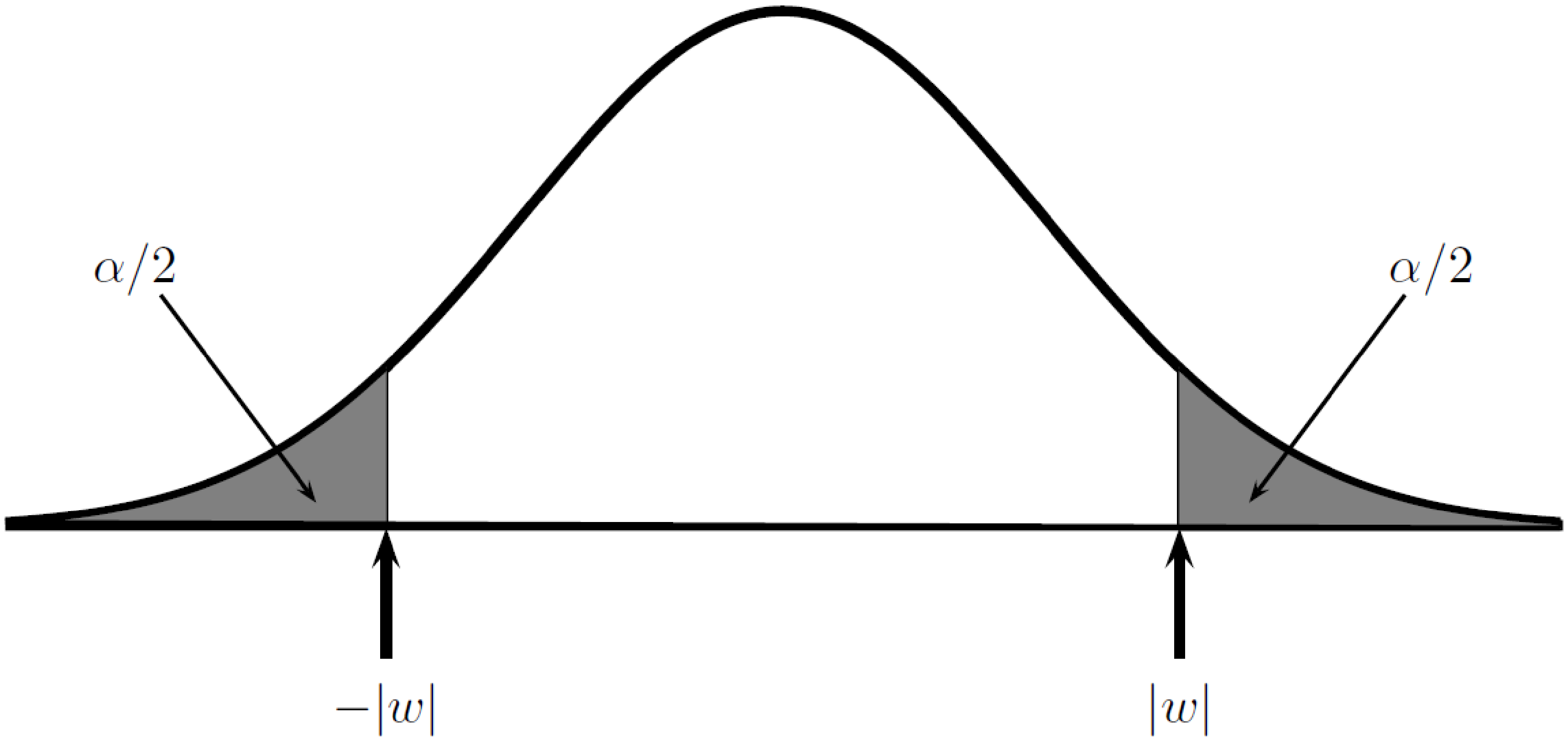
$$\text{p-value} = \mathbb{P}_{\theta_0}(T(X^n) \geq T(x^n))$$

p-value – это вероятность (при условии $\boldsymbol{H_0}$) того, что статистика $T(X^n)$ примет значение больше либо равное значения, которое реализовалось в опыте

<u>Теорема</u>. Пусть $w = (\widehat{\theta} - \theta_0)/\widehat{\mathsf{se}}$ **обозначает наблюденное значение статистики Вальда** W**. p-value в таком случае равно**

$$\text{p} - \text{value} = \mathbb{P}_{\theta_0}(|W| > |w|) \approx \mathbb{P}(|Z| > |w|) = 2\Phi(-|w|)$$ **, где**

$$Z \sim N(0, 1)$$



<u>Теорема</u>. Если распределение статистики критерия непрерывное, то тогда при гипотезе $H_0 : \theta = \theta_0$ **p-value ~ р.р. (0,1). Следовательно, если гипотеза** H_0 **отклоняется в случае, если p-value меньше** α**, то вероятность ошибки первого рода равна** α**.**

Другими словами, если H_0**, то p-value = случайной величине с р.р. на (0,1).**

Если верна H_1**, то распределение p-value будет концентрироваться около 0.**

<u>Пример</u>. Равенство значений холестерина в крови.

$$W = \frac{\widehat{\delta} - 0}{\widehat{\mathsf{se}}} = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}} = \frac{216.2 - 195.3}{\sqrt{5^2 + 2.4^2}} = 3.78$$

Пусть $Z \sim N(0,1)$**, тогда**

$$\text{p-value} = \mathbb{P}(|Z| > 3.78) = 2\mathbb{P}(Z < -3.78) = .0002$$

Чтобы протестировать равенство медиан, используем критерий

$$W = \frac{\widehat{\nu}_1 - \widehat{\nu}_2}{\widehat{\mathsf{se}}} = \frac{212.5 - 194}{7.7} = 2.4$$

d) Распределение хи-квадрат и критерий Пирсона

Пусть $Z_1, \ldots, Z_k$ **- независимые стандартно нормально распределенные случайные величины.** $V = \sum_{i=1}^{k} Z_i^2$, **тогда** $V \sim \chi_k^2$ **- хи-квадрат с** k **степенями свободы**

$$f(v) = \frac{v^{(k/2)-1} e^{-v/2}}{2^{k/2}\Gamma(k/2)}$$

$$\mathbb{E}(V) = k, \ \mathbb{V}(V) = 2k$$

$\chi_{k,\alpha}^2 = F^{-1}(1-\alpha)$ **- верхняя квантиль,** F **- функция распределения,**

т.е. $\mathbb{P}(\chi_k^2 > \chi_{k,\alpha}^2) = \alpha$

Пусть $X = (X_1, \ldots, X_k)$ **имеет мультиномиальное распределение с параметрами** (n, p)**. ОМП для** p **равно**

$$\widehat{p} = (\widehat{p}_1, \ldots, \widehat{p}_k) = (X_1/n, \ldots, X_k/n)$$

Обозначим через $p_0 = (p_{01}, \ldots, p_{0k})$ **некоторый фиксированный вектор**
 вероятностей

$$H_0 : p = p_0 \text{ vs. } H_1 : p \neq p_0$$

Положим $E_j = \mathbb{E}(X_j) = np_{0j}$ **- среднее значение** X_j **при гипотезе** H_0

<u>Определение</u>. Статистика χ^2 **Пирсона имеет вид**

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

Теорема. При гипотезе H_0 выполнено, что

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

Таким образом, гипотеза H_0 отвергается, если $T > \chi^2_{k-1,\alpha}$. Этот критерий имеет асимптотический уровень значимости α. p-value равняется $\mathbb{P}(\chi^2_{k-1} > t)$, где t - реализация значения статистики T

27

Пример. Горох Менделя. Два типа: круглые желтые зерна и сморщенные зеленые зерна. Имеется 4 типа потомков: круглые желтые, сморщенные желтые, круглые зеленые и сморщенные зеленые. Количество потомков каждого типа образуют мультиномиальное распределение с вероятностью $p = (p_1, p_2, p_3, p_4)$. Из теории следует, что

$$p_0 \equiv \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}\right)$$

В результате эксперимента: $n = 556$, $X = (315, 101, 108, 32)$

$$H_0 : p = p_0 \text{ vs. } H_1 : p \neq p_0$$

$$np_{01} = 312.75, \; np_{02} = np_{03} = 104.25, \; np_{04} = 34.75$$

$$\chi^2 = \frac{(315 - 312.75)^2}{312.75} + \frac{(101 - 104.25)^2}{104.25} + \frac{(108 - 104.25)^2}{104.25} + \frac{(32 - 34.75)^2}{34.75} = 0.47$$

$$\text{p-value} = \mathbb{P}(\chi^2_3 > .47) = .93$$

Таким образом, данные теории Менделя не противоречат

Теория тестирования гипотез не подходит для доказательства гипотезы H_0 Неспособность отклонить H_0 может быть как следствием того, что H_0 - истинная гипотеза, так и следствием того, что использованный критерий имеет небольшую мощность. М.б. лучше было построить доверительный интервал для расстояния между p и p_0

29 **е) Критерий перестановок – применяется для проверки того, отличаются ли распределения. Критерий перестановок – «точный» в том смысле, что он не использует предположения об асимптотической сходимости к нормальному распределению**

Допустим, что $X_1, \ldots, X_m \sim F_X$ и $Y_1, \ldots, Y_n \sim F_Y$ - две независимые выборки и H_0 - гипотеза, согласно которой распределения F_X и F_Y совпадают

$$H_0 : F_X = F_Y \text{ vs. } H_1 : F_X \neq F_Y$$

Обозначим через $T(x_1, \ldots, x_m, y_1, \ldots, y_n)$ - некоторую тестовую статистику, например,

$$T(X_1, \ldots, X_m, Y_1, \ldots, Y_n) = |\overline{X}_m - \overline{Y}_n|$$

Положим $N = m + n$ и рассмотрим все $N!$ перестановок объединенной выборки $X_1, \ldots, X_m, Y_1, \ldots, Y_n$. Для каждой из перестановок подсчитаем значение статистики T. Обозначим эти значения $T_1, \ldots, T_{N!}$. Если H_0 верна, то при фиксированных упорядоченных значениях $X_1, \ldots, X_m, Y_1, \ldots, Y_n$ значение статистики T распределены равномерно на множестве $T_1, \ldots, T_{N!}$.

Обозначим через распределение, согласно которому

$$\boldsymbol{P_0\left(T = T_i\right) = 1/N!, i = 1, \ldots, N!}$$

перестановочное распределение статистики T**. Пусть** t_{obs} **- значение статистики, которое было получено в опыте.**

$$\text{p-value} = \mathbb{P}_0(T > t_{obs}) = \frac{1}{N!}\sum_{j=1}^{N!} I(T_j > t_{obs})$$

<u>Пример</u>. **Допустим, что** $(X_1, X_2, Y_1) = (1, 9, 3)$**.** **Пусть** $T(X_1, X_2, Y_1) = |\overline{X} - \overline{Y}| = 2$**, тогда**

перестановка	значение T	вероятность
(1,9,3)	2	1/6
(9,1,3)	2	1/6
(1,3,9)	7	1/6
(3,1,9)	7	1/6
(3,9,1)	5	1/6
(9,3,1)	5	1/6

p-value равно $\mathbb{P}(T > 2) = 4/6$

Алгоритм (сокращение времени подсчетов):

1. Подсчитать $t_{\text{obs}} = T(X_1, \ldots, X_m, Y_1, \ldots, Y_n)$

2. Сделать перестановку. Подсчитать значение статистики

 3. Повторить предыдущий шаг B **раз. Пусть** $T_1, \ldots, T_B$ **- получившиеся значения**

4. Приближенное p-value

$$\frac{1}{B}\sum_{j=1}^{B} I(T_j > t_{\text{obs}})$$

<u>Пример</u>. Данные – уровни информационной РНА (вещество, несущее генетическую информацию) для каждого из генов, которые (как считается) характеризуют количество протеина, которое производит соответствующий ген (грубо говоря – чем больше значение характеристики – тем ген более активный). В таблице – значения уровней для генов 10 пациентов с двумя типами раковых клеток в печени. В таблице 2.638 генов.

	Type I					Type II				
Patient	1	2	3	4	5	6	7	8	9	10
Gene 1	230	-1,350	-1,580	-400	-760	970	110	-50	-190	-200
Gene 2	470	-850	-.8	-280	120	390	-1730	-1360	-1	-330
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Протестируем, отличается ли медиана уровня между группами. $T = |\widehat{\nu}_1 - \widehat{\nu}_2| = 710$**, при этом p-value = 0.045. Значит, если использовать уровень значимости** $\alpha =$**0.05, то окажется, что опытные данные гипотезе о равенстве медиан противоречат.**

f) Критерий на основе отношения правдоподобия

<u>Определение</u>. Рассмотрим гипотезы

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \notin \Theta_0$$

Пусть $\widehat{\theta}$ **- ОМП и** $\widehat{\theta}_0$ **- ОМП при** $\boldsymbol{\theta_0 \in \Theta_0}$
Статистика отношения правдоподобия имеет вил

$$\lambda = 2 \log \left(\frac{\sup_{\theta \in \Theta} \mathcal{L}(\theta)}{\sup_{\theta \in \Theta_0} \mathcal{L}(\theta)} \right) = 2 \log \left(\frac{\mathcal{L}(\widehat{\theta})}{\mathcal{L}(\widehat{\theta}_0)} \right)$$

<u>Замечание</u>. Замена $\boldsymbol{\Theta_0^c}$ **на** $\boldsymbol{\Theta}$ **фактически не влияет на критерий**

<u>Теорема</u>. Допустим, что $\theta = (\theta_1, \ldots, \theta_q, \theta_{q+1}, \ldots, \theta_r)$**. Пусть**

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

Пусть λ **- критерий на основе отношения правдоподобия. При гипотезе** $H_0 : \theta \in \Theta_0$

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

где $r - q$ **- размерность** Θ **за вычетом размерности** Θ_0**. p-value для критерия равно** $\mathbb{P}(\chi^2_{r-q} > \lambda)$

Например, если $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ **и необходимо проверить, что** $\theta_4 = \theta_5 = 0$**, тогда у предельного распределения имеется** $5 - 3 = 2$ **степеней свободы**

<u>Пример</u>. Горох Менделя. Статистика отношения правдоподобия для

$$H_0 : p = p_0 \text{ vs. } H_1 : p \neq p_0$$

принимает вид

$$\begin{aligned}
\lambda &= 2\log\left(\frac{\mathcal{L}(\widehat{p})}{\mathcal{L}(p_0)}\right) \\
&= 2\sum_{j=1}^{4} X_j \log\left(\frac{\widehat{p}_j}{p_{0j}}\right) \\
&= 2\left(315\log\left(\frac{\frac{315}{556}}{\frac{9}{16}}\right) + 101\log\left(\frac{\frac{101}{556}}{\frac{3}{16}}\right)\right. \\
&\quad \left. +108\log\left(\frac{\frac{108}{556}}{\frac{3}{16}}\right) + 32\log\left(\frac{\frac{32}{556}}{\frac{1}{16}}\right)\right) \\
&= 0.48
\end{aligned}$$

При гипотезе H_1 - 4 параметра. Так как сумма параметров должна равняться 1, то размерность пространства параметров равна 3. При гипотезе H_0 «свободных» параметров нет, значит количество степеней свободы равно 3 и χ_3^2 является предельным распределением.

$$\text{p-value} = \mathbb{P}(\chi_3^2 > .48) = .92$$

38 **Зачастую и критерий хи-квадрат и критерий отношения правдоподобий дают примерно одинаковые результаты при условии, что размер выборки достаточно большой**

g) Множественные тесты

В некоторых приложениях необходимо проверить много гипотез «за раз». См., например, пример по уровни информационной PHA – 2638 проверок.

Допустим, что каждый критерий проверяется на уровне α, то есть на во время каждой из проверок с вероятностью α основная гипотеза может быть неверно отвергнута. При этом вероятность допустить хотя бы одно

ложное отклонения основной гипотезы существенно больше => множественные тесты

Рассмотрим m различных случаев проверки гипотез

$$H_{0i} \text{ vs. } H_{1i},\ i = 1, \ldots, m$$

Обозначим через $P_1, \ldots, P_m$ величины m p-values для этих гипотез

<u>Метод Бонферрони</u>: Для заданных p-values $P_1, \ldots, P_m$ **основная гипотеза** H_{0i} **отклоняется, если**

$$P_i < \frac{\alpha}{m}$$

<u>Теорема</u>. При применении метода Бонферрони вероятность неправильно отклонить любую из основных гипотез меньше либо равна α

40

<u>Доказательство</u>. Обозначим через R **событие, что по крайней мере одна из основных гипотез была ложно отклонена,** R_i **- событие, что i-ая основная гипотеза была ложно отклонена.**

Напомним, что для любых событий $A_1, \ldots, A_k$ **выполнено, что** $\mathbb{P}(\bigcup_{i=1}^k A_i) \leq \sum_{i=1}^k \mathbb{P}(A_i)$

$$\mathbb{P}(R) = \mathbb{P}\left(\bigcup_{i=1}^{m} R_i\right) \leq \sum_{i=1}^{m} \mathbb{P}(R_i) = \sum_{i=1}^{m} \frac{\alpha}{m} = \alpha$$

Пример. **В случае с анализом уровней РНА для $\alpha = 0.05$ получаем, что $0.05 / 2.638 = 0.00001895375$**

Метод Бонферрони консервативен – суть метода в том, что он делает по сути невозможным, чтобы произошло даже одно ложное отклонение

Иногда лучше контролировать интенсивность ложных отклонений (False Discovery Rate, FDR) = среднее значение отношения ложных отклонений к числу отклонений вообще

Пусть m_0 **- количество «верных» нулевых гипотез,** $m_1 = m - m_0$**.**

	H_0 **не отклонена**	H_0 **отклонена**	Σ
H_0 **верна**	U	V	m_0
H_0 **неверна**	T	S	m_1
Σ	$m - R$	R	m

Доля ложных отклонений (False Discovery Proportion, FDP)

$$FDP = \begin{cases} V / R, R > 0 \\ 0, R = 0 \end{cases}$$

$$FDR = E\left(FDP\right)$$

<u>Метод Benjamini-Hochberg</u>:

1. Пусть $P_{(1)} < \cdots < P_{(m)}$ **- величины p-value, отсортированные по возрастанию**

2. Пусть C_m **в случае, если** $P_1, \ldots, P_m$ **независимы, в противном случае положим** $C_m = \sum_{i=1}^{m}(1/i)$**.**

3. Определим

$$\ell_i = \frac{i\alpha}{C_m m}$$

$$R = \max\left\{i : P_{(i)} < \ell_i\right\}$$

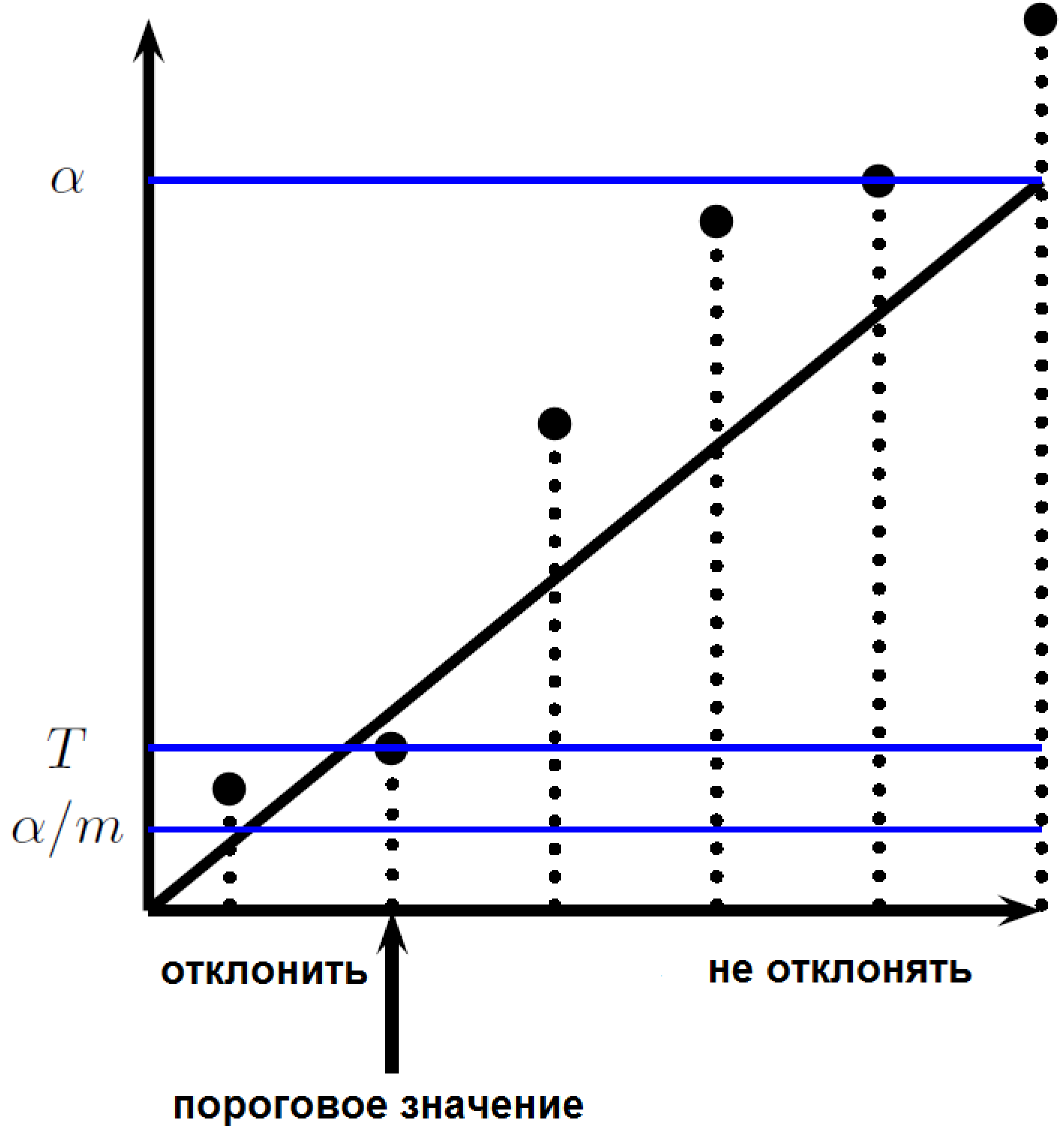
4. Пусть $T = P_{(R)}$ **- пороговое значение метода.**

5. Отклоняются такие H_{0i}**, для которых** $P_i \leq T$

<u>Теорема</u>. При применении метода Benjamini-Hochberg независимо от того, сколько основных гипотез на самом деле верно и независимо от того, какое распределение имеют $P_1, \ldots, P_m$ **в случае неверной основной гипотезы, выполнено, что**

$$FDR = E\left(FDP\right) \le \frac{m_0}{m}\alpha \le \alpha$$

<u>Пример</u>.



<u>Пример</u>. Допустим, что было протестировано 10 независимых гипотез, для которых были получены следующие p-value

0.00017 0.00448 0.00671 0.00907 0.01220

0.33626 0.39341 0.53882 0.58125 0.98617

При $\alpha = 0.05$ **метод Бонферрони отклоняет любую гипотезу, p-value которой меньше чем** $\alpha/10 = 0.005$ **=> будут отклонены только первые две гипотезы. Для метода Benjamini-Hochberg определяется наибольшее** i**, для которого** $P_{(i)} < i\alpha/m$**. В рассматриваемом случае это** $i = 5$**, то есть отклоняется первые 5 гипотез**

h) Критерий согласия

Необходимо проверить согласие параметрической модели распределения и полученных в эксперименте данных

Пусть $\mathfrak{F} = \{f(x;\theta) : \ \theta \in \Theta\}$ **- параметрическая модель,** $\theta = (\theta_1, \ldots, \theta_s)$**. Допустим, что область значений выборки** $= (-\infty, \infty)$**.**

Разделим $(-\infty, \infty)$ **на** k **непересекающихся интервалов** $I_1, \ldots, I_k$**. Для** $j = 1, \ldots, k$ **обозначим через**

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

вероятность попадания наблюдения в интервал I_j **при условии, что модель верна.**

Пусть N_j **- количество элементов выборки, которые попали в** I_j

Можно показать, что функция правдоподобия для параметра θ на основе наблюдений $N_1, \ldots, N_k$ (в предположении, что эти наблюдения имеют мультиномиальное распределение) имеет вид

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

Максимизируя $Q(\theta)$**, получаем оценку** $\widetilde{\theta} = (\widetilde{\theta}_1, \ldots, \widetilde{\theta}_s)$ **параметра** θ**.**

Критерий согласия имеет вид

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

<u>Теорема</u>. Пусть основная гипотеза H_0 состоит в том, что выборка – простая, элементы которой имеют распределение из параметрического класса $\mathfrak{F} = \{f(x;\theta) : \ \theta \in \Theta\}$. В условиях гипотезы H_0 статситика критерия Q сходится по распределению к χ^2_{k-1-s}. Таким образом, приближенное p-value критерия равно $\mathbb{P}(\chi^2_{k-1-s} > q)$, где q - наблюдаемое в эксперименте значение статистики Q

Если заменить $\widetilde{\theta}$ на ОМП $\widehat{\theta}$, то теорема перестанет выполняться.

Критерия согласия не позволяют проверить «правильность» модели. Если H_0 не была отклонена, то это не значит, что модель правильная. Модель могла быть не отклонена просто потому, что мощность критерия – низкая => лучше использовать непараметрические модели. Если H_0 была отклонена, то модель заведомо «плохая».

i) Критерий Неймана-Пирсона для случая двух простых гипотез, t-критерий

Лемма Неймана-Пирсона (наиболее мощный тест для случая двух простых гипотез):

$$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta = \theta_1$$

Пусть статистика

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

Допустим, что H_0 **отвергается при** $T > k$**. Если выбрать** k **так, что** $\mathbb{P}_{\theta_0}(T > k) = \alpha$**, то критерий на основе статистики (*) будет наиболее мощным размера** α**, то есть среди всех критериев размера** α **критерий Неймана-Пирсона имеет максимальную мощность** $\beta(\theta_1)$

j) t-тест. Для проверки гипотезы $H_0 : \mu = \mu_0$**, где** $\mu = \mathbb{E}(X_i)$ **, можно использовать критерий Вальда. В случае, если распределение данных близко к нормальному, а размер выборки не очень большой, имеет смысл использовать t-критерий.**

Случайная величина имеет распределение Стьюдента (t-распределение) с k **степенями свободы, если**

$$f(t) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)\left(1 + \frac{t^2}{k}\right)^{(k+1)/2}}$$

При $k \to \infty$ **t-распределение стремится к стандартному нормальному распределению. При** $k = 1$ **t-распределение совпадает с распределением Коши.**

Пусть $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$**, где параметры** $\theta = (\mu, \sigma^2)$ **неизвестны**

$$\boldsymbol{H_0 : \mu = \mu_0} \text{ vs. } \boldsymbol{H_0 : \mu \neq \mu_0}$$

Обозначим через S_n^2 **- выборочную дисперсию**

$$T = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S_n}$$

Основная гипотеза отвергается, если $|T| > t_{n-1,\alpha/2}$ **- квантиль распределения Стьюдента с** $n-1$ **степенью свободы**

При больших n **выполняется, что** $T \approx N(0,1)$**, то есть при больших** n **t-критерий эквивалентен критерию Вальда**