



ФКН

Департамент больших данных и
информационного поиска

Москва 2025

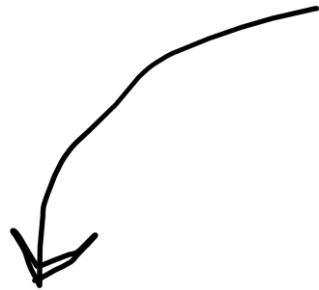
Лекция 4

Доверительные интервалы и проверка гипотез

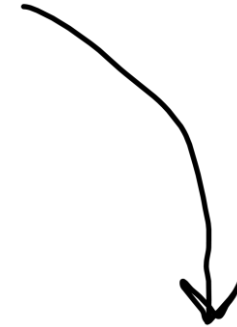
Машинное обучение в цифровом продукте

Полякова И.Ю.

Оценки параметров/статистик



Точечные



Интервальные

Доверительный интервал

- Доверительным интервалом уровня α для оцениваемого параметра θ является такой интервал со случайными концами L и R , для которого

1. $L < R$ почти наверное;
2. $P(L < \theta < R) = 1 - \alpha$



L и R – это случайные величины/функции от выборки!

Чем шире доверительный интервал – тем менее устойчивая оценка

Пример

$$X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$$

Наб. дана $E(X) = ?$

- Записываем дов. интервал по определению, пользуясь ЦПТ
- Раскрываем скобки
- Получаем левую и правую границы оценки

$$P\left(z_{\alpha/2} < \frac{\frac{1}{n} \sum X_i - \mu}{\sigma/\sqrt{n}} < z_{1-\alpha/2}\right) = 1 - \alpha$$

$$L = \frac{\sum X_i}{n} - \frac{\sigma}{\sqrt{n}} \cdot z_{\alpha/2}; \quad R = \frac{\sum X_i}{n} + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha/2}$$

$$L \leq \mu \leq R$$

Пример

???

$$L = \frac{\sum x_i}{n} - \frac{6}{\sqrt{n}} \cdot z_{\alpha/2} ; \quad R = \frac{\sum x_i}{n} + \frac{6}{\sqrt{n}} \cdot z_{1-\alpha/2}$$
$$L \leq \mu \leq R$$

- Теоретическое СКО нам неизвестно практически никогда!
- Однако, по лемме Слуцкого имеем право заменять теоретическое СКО на выборочное СКО и считать интервал
- При маленьких выборках рекомендуется использовать t-распределение вместо z-распределения

Про лемму Слуцкого:

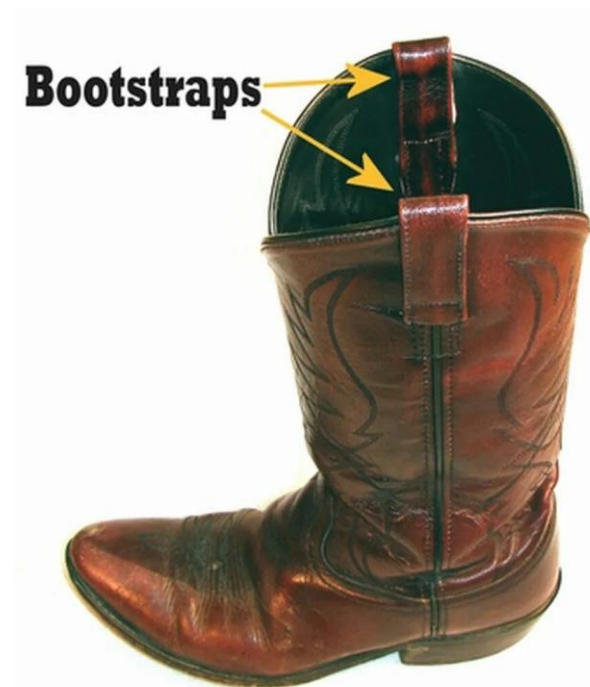
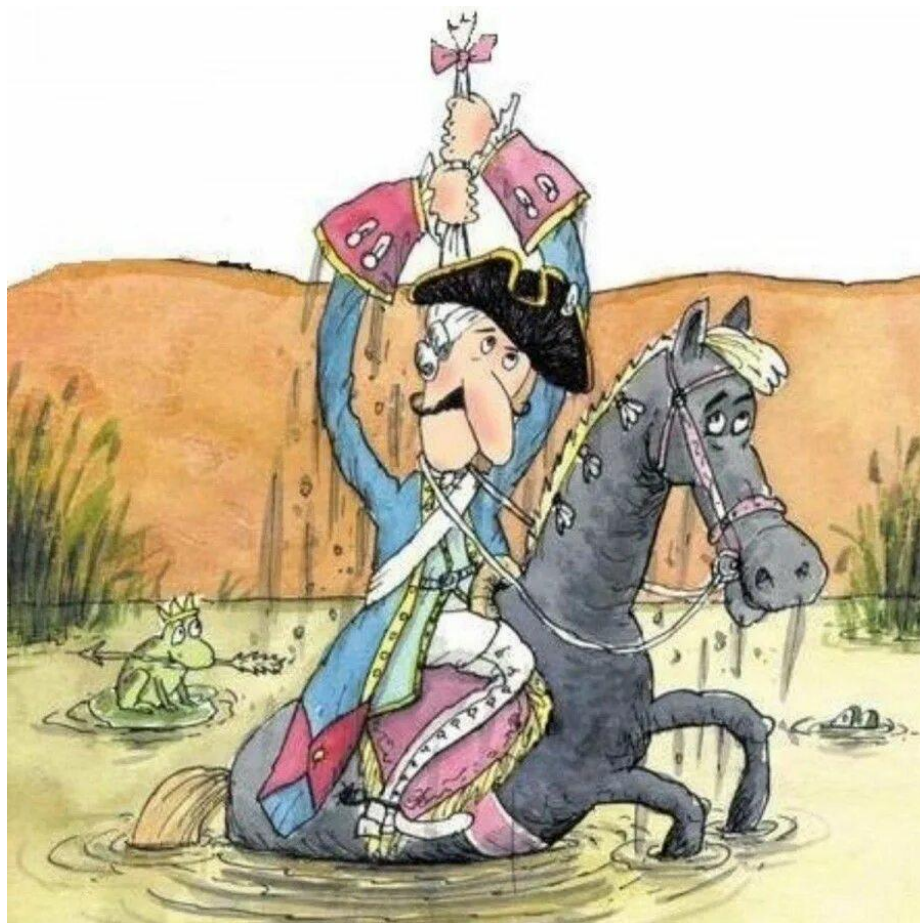
https://ru.wikipedia.org/wiki/%D0%A2%D0%B5%D0%BE%D1%80%D0%B5%D0%BC%D0%B0_%D0%A1%D0%BB%D1%83%D1%86%D0%BA%D0%BE%D0%B3%D0%BE

Пример

- При $n \rightarrow \infty$ статистики L и R будут сходиться к мат. ожиданию, что является следствием ЗБЧ

Как еще считать доверительный интервал?

Бутстреп



Бутстреп

Бутстреп – набор техник, позволяющих искусственно генерировать большие данные из уже имеющихся, при этом получается распределение, похожее на исходное распределение выборки.

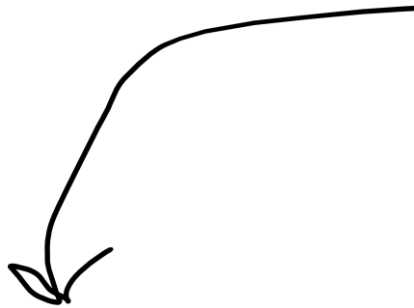
Этими данными можно манипулировать, добиваясь нужных для ЦПТ и других полезных конструкций предположений.

Можно понимать как одну из разновидностей метода Монте-Карло

Бутстреп

- Бутстреп НЕ позволяет сделать оценку точнее (доверительный интервал уже)
- Можно сказать, что бутстреп из данных «непонятного» распределения приводит данные к чуть более «понятному» распределению
- **Главное:** бутстреп позволяет дать интервальную оценку параметра, тогда когда совсем не понятно, как ее построить более формальными методами

Бутстреп



Параметрический



Непараметрический

Непараметрический бутстреп

Предпосылка использования: нет уверенности в том, что данные подчиняются какому-либо известному параметрическому распределению

Непараметрический бутстреп

Идея:

- Имеется исходная выборка
- Из этой выборки **случайно, с возвращением** извлекается новая выборка такого же размера n : такая выборка называется **бутстреп-выборка**
- Так как выборка с возвращением, некоторые исходные точки могут попасть в новую выборку несколько раз, а некоторые - ни разу

Непараметрический бутстреп

Идея:

- Для каждой бутстреп-выборки вычисляется интересующая нас статистика θ^*
- Предыдущие шаги повторяются большое количество раз. В результате мы получаем бутстреп-распределение статистики θ^* .
- Бутстреп-распределение используется для построения доверительных интервалов, оценки стандартной ошибки и т.д.

Параметрический бутстреп

Предпосылка использования: предполагаем, что данные пришли из распределения определенного типа. Однако нам неизвестны параметры этого распределения

Параметрический бутстреп

Идея:

- Используем исходную выборку, чтобы оценить эти параметры. Затем генерируем новые выборки уже из этого подобранного параметрического распределения

Параметрический бутстреп

Идея:

- Генерируем новую выборку размера n не из исходных данных, а из этого подобранного параметрического распределения
- Для каждой такой сгенерированной выборки вычисляется статистика θ^*
- Получаем бутстреп-распределение статистики θ^*
- Так же используем бутстреп-распределение для построения доверительных интервалов, оценки стандартной ошибки и т.д.

Доверительный интервал на основе бутстрепа

- Для каждой бутстреп выборки вычисляем оцениваемую величину
- Сортируем вычисленные статистики по возрастанию, получаем эмпирическое распределение
- Из этого распределения вычисляем левосторонний и/или правосторонний квантиль в соответствие с назначенным уровнем значимости
- Значения квантилей и будут границами доверительного интервала на основе бутстрепа

Jackknife



- Зачастую более экономный подвид бутстрепа (но массово используется реже)
- Идея в том, чтобы сгенерировать n выборок из исходной, где каждая выборка получается путем удаления одного единственного i -го элемента
- Такой метод не очень хорошо показывает себя в вычислении порядковых статистик, медианы и квантилей

Проверка статистических гипотез

Проверка гипотез

- **Статистическая гипотеза:** предположение о свойствах генеральной совокупности (о виде распределения, значении параметра и т.д.), которое можно проверить на основе выборочных данных

Гипотеза это «догадка» исследователя, она может быть сформулирована на основе анализа литературы, EDA, априорных знаний и тд...

Проверка гипотез

$$H_0 \quad \text{vs} \quad H_1$$

- Выдвигаются две гипотезы: нулевая (H_0) и альтернативная (H_1)
- Гипотезы обязаны быть взаимоисключающими!
- Но при этом не обязаны покрывать все «множество допустимых значений»

Например: постановка $H_0: \theta \neq 5$ и $H_1: \theta < 5$ является валидной

Проверка гипотез

$$H_0 \quad \text{vs} \quad H_1$$

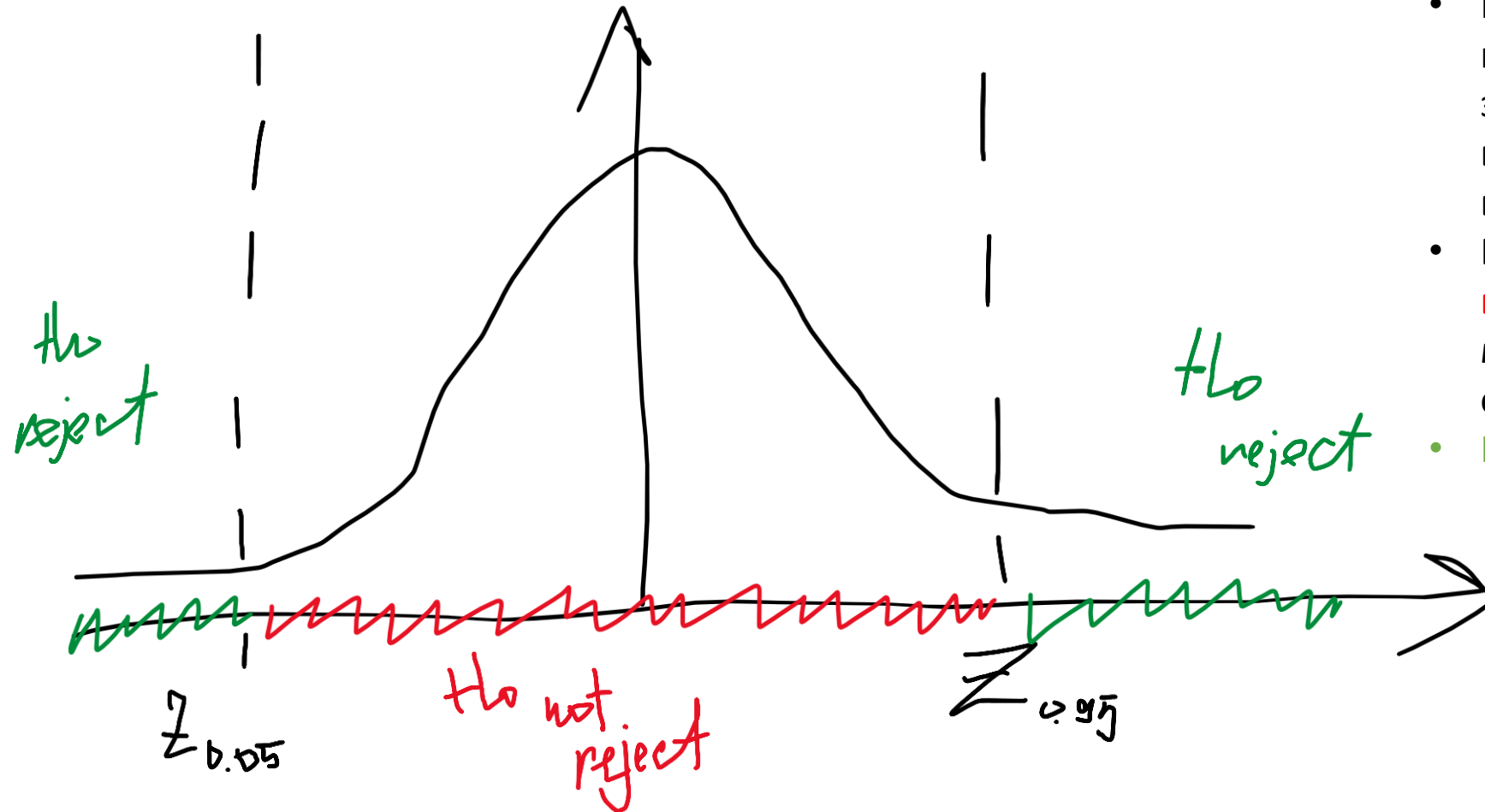
- Как правило, гипотеза H_0 говорит о состоянии «по умолчанию», «об отсутствии эффекта/различия/связи»
- H_1 , в свою очередь, гипотеза, противоречащая нулевой: то, что мы хотим «доказать»

Пример H_1 : разработанный препарат лучше плацебо, спрос мужчин и женщин на предметы роскоши различен и тд...

Процедура проверки гипотез

- Для проверки гипотез используется идея доверительных интервалов
- Применяем статистический критерий: правило, по которому на основании выборки принимается решение — отклонить или не отклонить нулевую гипотезу

Пример

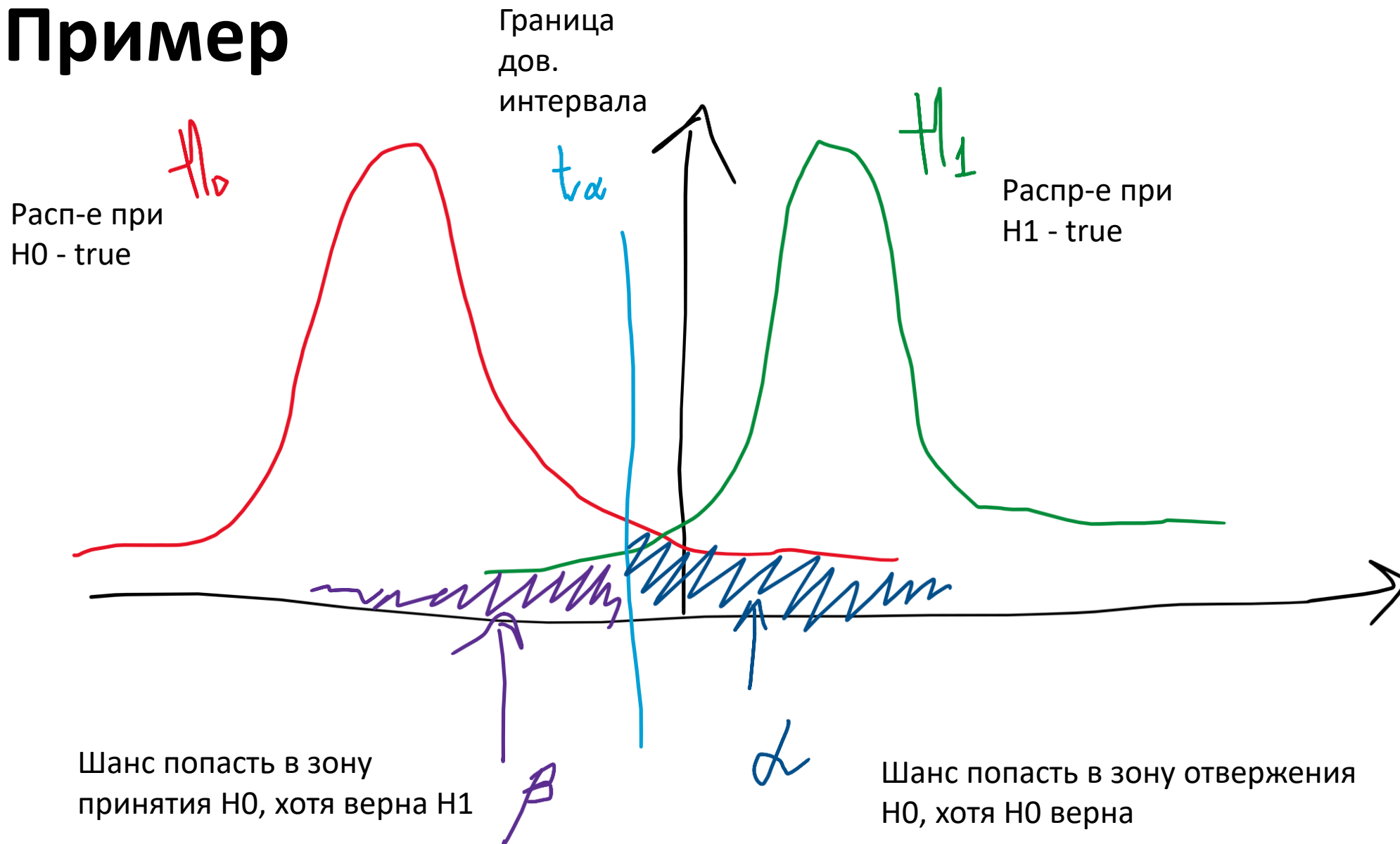


- Мы знаем теор. распр-е статистики при верной H_0
- Можем проверить, насколько вычисленное значение статистики вероятно получить при выполненной H_0
- Если попадает в область **высокой вероятности**, то мы не имеем оснований отклонять H_0
- **Иначе**, H_0 отклоняется

Проверка гипотез: уровень значимости

- При проверке гипотез α обозначается **«уровень значимости»**
- Задается исследователем
- Обозначает в-ть ошибки первого рода, которую мы считаем «приемлемой» совершить
- Ошибка первого рода – отвержение H_0 при том, что H_0 верна
- Ошибка второго рода – НЕотвержение H_0 при том, что H_0 неверна
- Чем меньше ошибку первого рода мы готовы «терпеть» – тем шире будет наш доверительный интервал и тем выше в-ть совершить ошибку второго рода

Пример



«Золотое правило» проверки гипотез

$$\begin{aligned} p - value < \alpha &\Rightarrow H_0 \text{ reject} \\ p - value \geq \alpha &\Rightarrow H_0 \text{ not reject} \end{aligned}$$

- P-value - это шанс получить такую же или более «экстремальную» статистику при истинности H_0
- P-value – это минимальный уровень α , при котором мы отклонили бы H_0

Распределения, произведенные от нормального

Распределение хи-квадрат

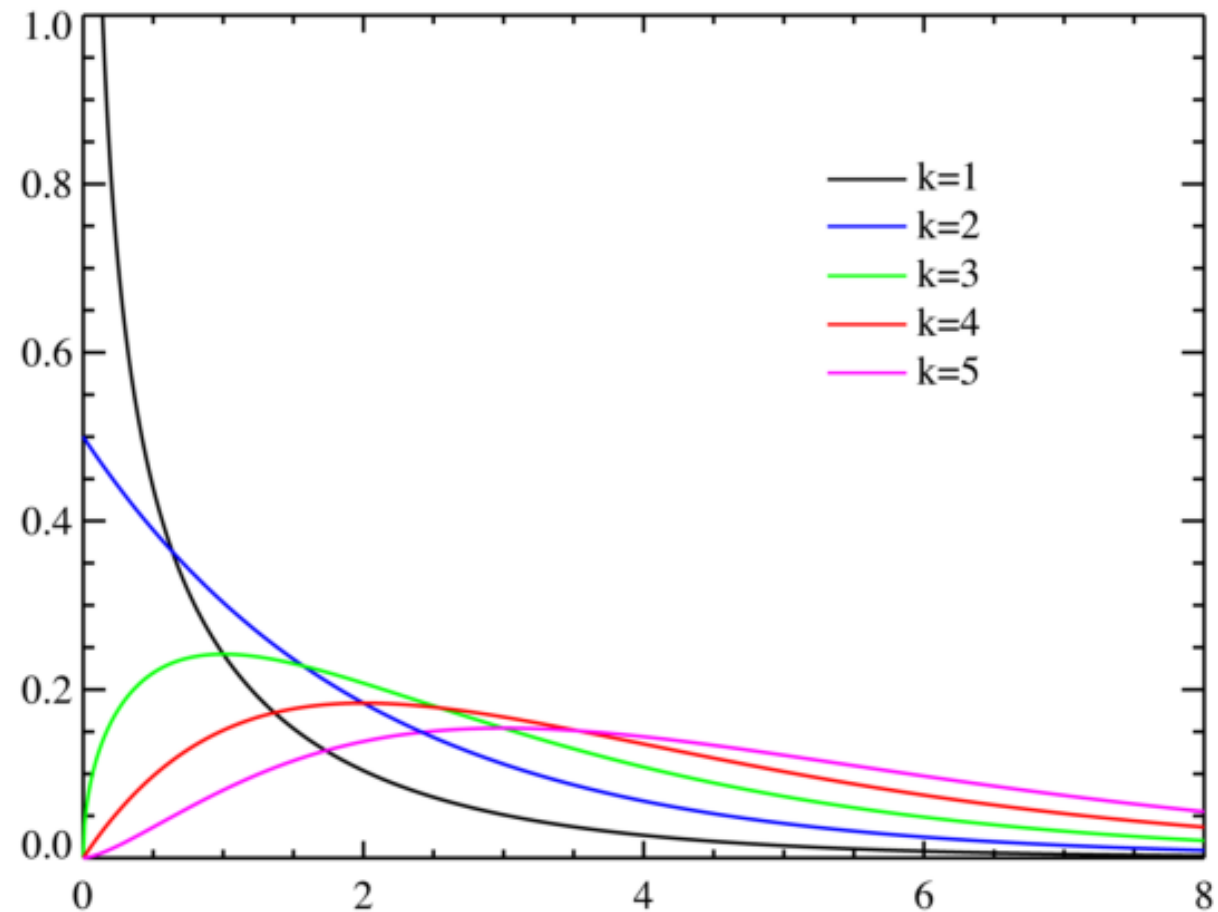
Пусть X_1, \dots, X_k независимы и имеют стандартное нормальное распределение $\mathcal{N}(0, 1)$.

Распределением χ^2 (хи-квадрат) с k степенями свободы называется распределение случайной величины

$$Y = X_1^2 + \dots + X_k^2.$$

Обозначение: χ_k^2 или H_k .

Распределение хи-квадрат



Распределение Стьюдента

Пусть X_0, X_1, \dots, X_k независимы и имеют стандартное нормальное распределение $\mathcal{N}(0, 1)$.

Распределением Стьюдента с k степенями свободы называется распределение случайной величины

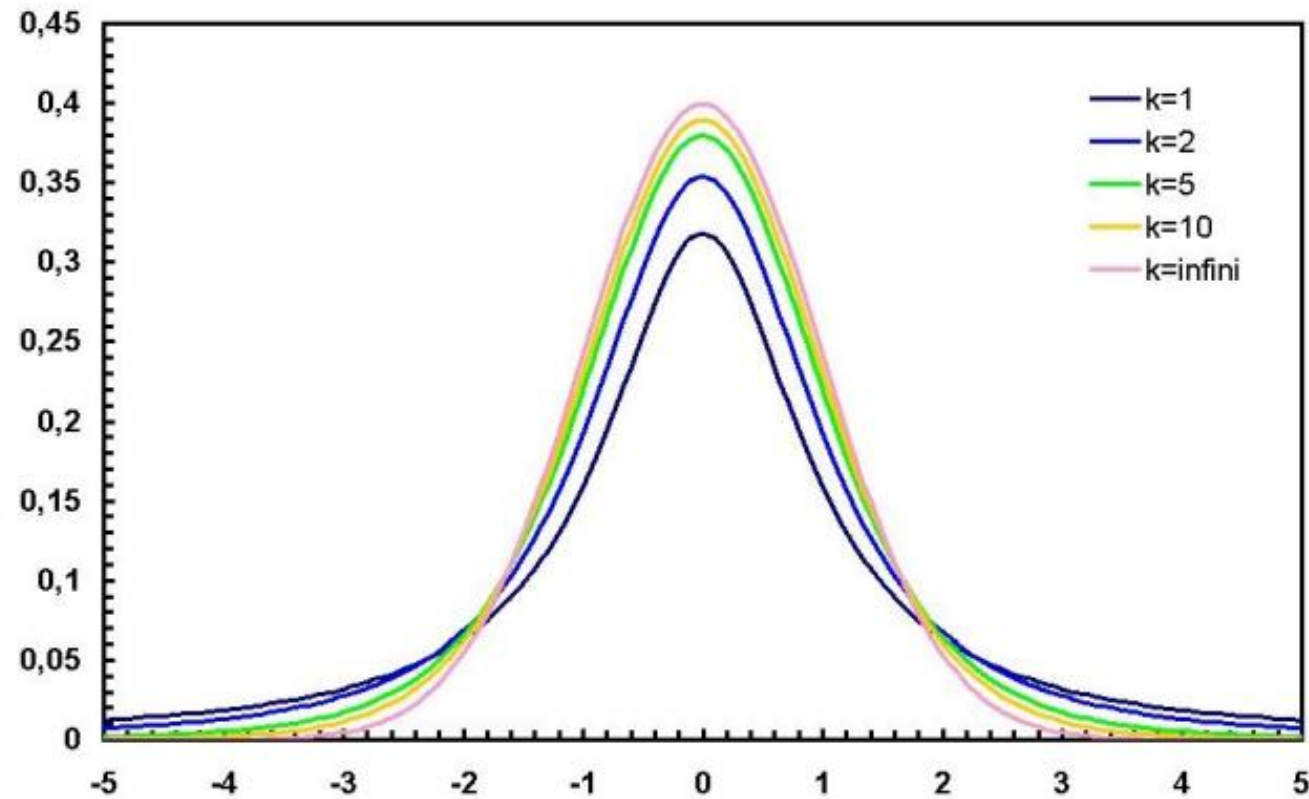
$$Y = \frac{X_0}{\sqrt{\frac{X_1^2 + \dots + X_k^2}{k}}}.$$

← $\mathcal{N}(0, 1)$

← $\chi^2(k)$

Обозначение: t_k или T_k .

Распределение Стьюдента



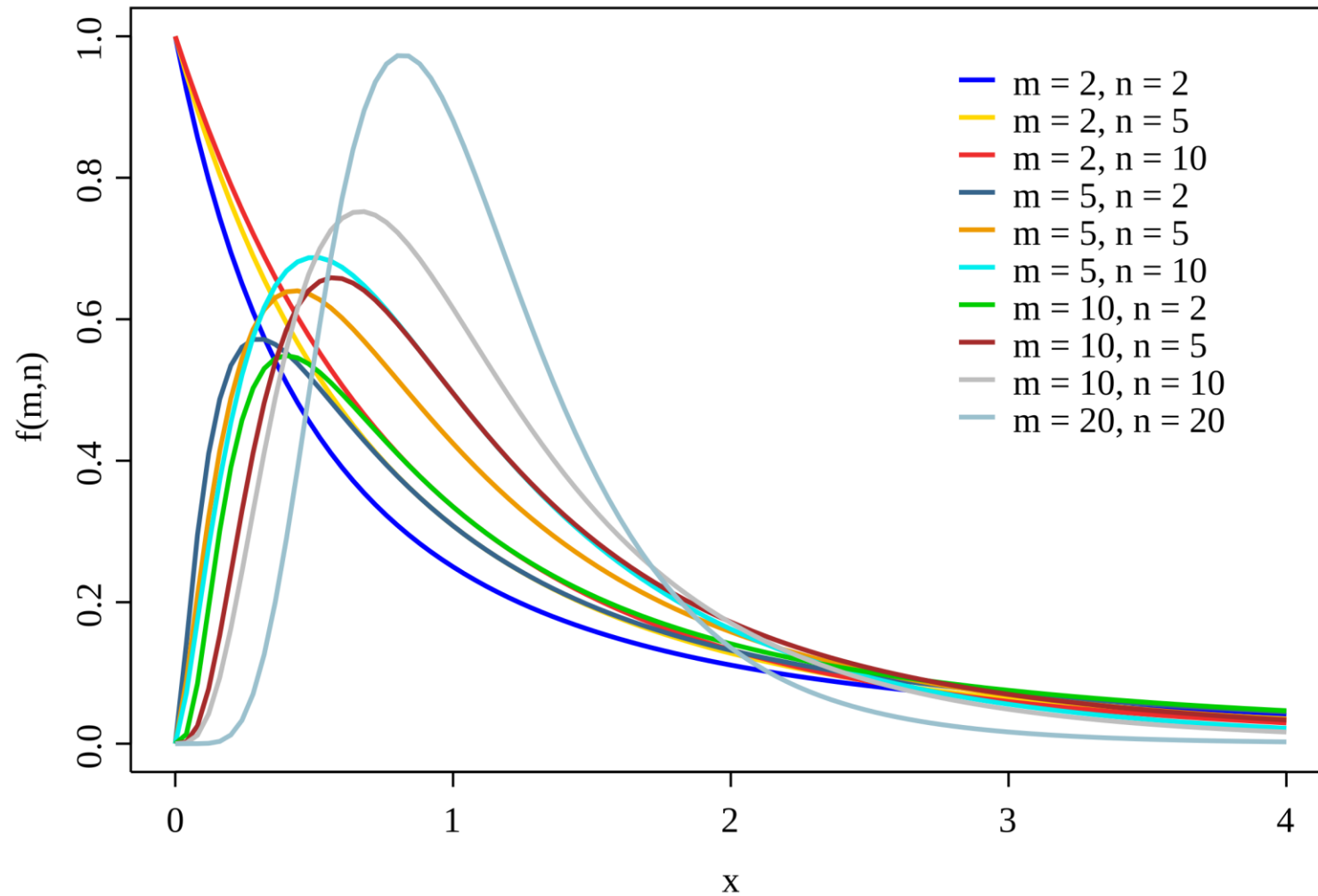
Распределение Фишера

Распределение Фишера показывает распределение отношения двух независимых СВ, которые распределены по закону хи-квадрат и имеют степени свободы k_1 и k_2 соответственно:

$$F = \frac{X / k_1}{Y / k_2}$$

$$F \sim F(k_1, k_2)$$

Распределение Фишера



Гипотезы об однородности/критерии согласия

Пример

- В какой из кофеев X и Y качество кофе более непредсказуемо?

$$H_0: \sigma_x^2 = \sigma_y^2$$

$$H_1: \sigma_x^2 > \sigma_y^2$$

$$T = \frac{\sigma_x^2}{\sigma_y^2} \sim F(n-1; m-1)$$

В пользу отклонения H_0 говорит очень большая статистика T (очень высокое отношение двух дисперсий)

Областью H_0 reject здесь является полуинтервал:

$$(F_{1-\alpha}; +\infty)$$

Пример

- Процедура, сконструированная в предыдущем примере хорошо подходит для небольших выборок ($n < 1000$)
- Однако при крупных n более состоятельными можно считать тест Колмогорова-Смирнова (для непрерывных распр-й) или тест хи-квадрат (для дискретных)

Критерий Колмогорова

- Критерий о виде неизвестного распределения

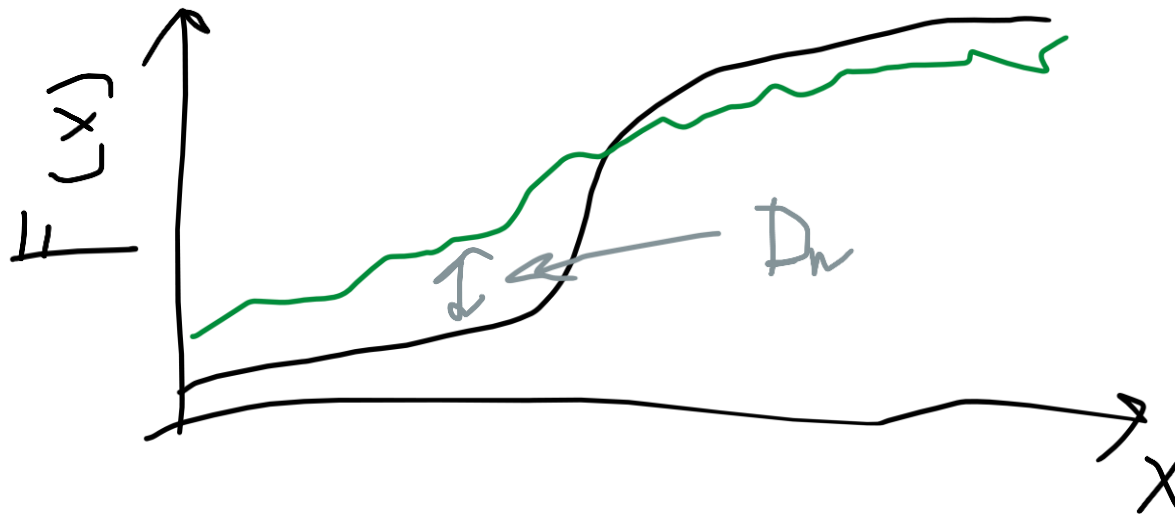
$$H_0: F_x(x) = F_0(x)$$

$$H_1: F_x(x) \neq F_0(x)$$

Критерий Колмогорова

- Критерий о виде неизвестного распределения
- Нужно ввести меру расстояния между двумя распределениями

$$D_n(X_1, \dots, X_n) = \sup_x |F_0(x) - \hat{F}_n(x)|$$



Теорема Колмогорова

- При справедливости H_0 , распределения статистики D_n одинаково для любых непрерывных распределений, при этом его функция распр-я:

$$\lim_{n \rightarrow \infty} P(\sqrt{n} \cdot D_n \leq z) = 1 + 2 \sum_{k=1}^{\infty} (-1)^k \exp^{-2k^2 z^2}$$

- Критерий для проверки:

$$K_n = \sqrt{n} \cdot \sup_x |F_0(x) - \hat{F}_n(x)|$$

$K_n \sim_{th}$ распр-е
Колмогорова

Критерий Пирсона

- «Аналог» теста Колмогорова для дискретных величин;
- Необходимо сравнить все теоретические частоты с эмпирическими;

X	z_1	z_2	\dots	z_s	Возможные значения СВ
$P(X=z_i)$	$p_1(\theta)$	$p_2(\theta)$	\dots	$p_s(\theta)$	Теоретические вероятности
$V(X=z_i)$	v_1	v_2	\dots	v_s	Эмпирические частоты

$$T = \sum_{j=1}^s \frac{(v_j - n \cdot p_j(\hat{\theta}))^2}{n \cdot p_j(\hat{\theta})} \sim \chi^2_{s-k-1}$$

Тестовая статистика представляет собой нормированную разность квадратов наблюдаемых частот и теоретических

Как проверить распределение на нормальность?

- **Тест Шапиро-Уилка** – один из самых мощных тестов для малых выборок (H_0 : данные распределены нормально)
- **Тест Жарка-Бера** – применяется для больших выборок (H_0 : данные распределены нормально)
- Тест Жарка-Бера делает вывод на основе асимметрии и эксцесса
- **QQ-plot**: график квантиль-квантиль
- Позволяет увидеть тяжелые хвосты, асимметрию и тд...

Дополнительно

- Подробнее при критерии согласия:
https://github.com/pileyan/applied_statistics_2024/blob/master/lect04/S4.pdf
- Про бутстрап у Максима Каледина:
<https://github.com/XuMuK1/psmo/blob/master/lectures/Lec12-Boots.pdf>
- Оригинальная статья: B.Efron. Bootstraps methods: Another look at the jackknife. *The annals of Statistics*, 7(1): 1-26, 1979
- ВШЭ: прикладная статистика (week10):
https://www.youtube.com/watch?v=2p24KPez62U&list=PLCf-cQCe1FRyCcf47wwBIDObNcz4ud2_L

