

Seminar 9: Проверка статистических гипотез

Carlos Buitrago

Table of contents

1	1. Постановка задачи	1
2	2. Функция мощности, размер и уровень	2
2.1	2.1. Почему уровень и мощность — разные вещи	2
3	3. Лемма Неймана—Пирсона	3
3.1	3.1. Бернулли: почему статистика $\sum X_i$?	3
4	4. Монотонное отношение правдоподобия (MLR)	4
5	5. Равномерное распределение и зависимость носителя от параметра	4
6	6. Экспоненциальное семейство и точные критические значения	5
7	7. Двусторонние альтернативы и отсутствие UMP	6
8	8. Дисперсия нормального распределения: простая против простой	7
9	9. Особая задача 1: равномерное против показательного	8
10	10. Допустимость и несмещённость критериев	9
11	11. Практикум: проверяем теорию методом Монте-Карло	9
11.1	11.1. Эмпирическая ошибка I рода и мощность	9
11.2	11.2. Размер выборки и вероятность ошибки II рода	10
12	12. Задачи для самостоятельного решения (стр. 75–76)	10
13	13. Контрольные вопросы	12

1 1. Постановка задачи

Пусть $\mathcal{P}$ — семейство распределений наблюдения X . **Основная гипотеза** $H_0 : P \in \mathcal{P}_0$ проверяется против **альтернативы** $H_1 : P \in \mathcal{P}_1$, где множества распределений не пересекаются. Для параметрического семейства $\{P_\theta : \theta \in \Theta\}$ записываем $H_0 : \theta \in \Theta_0$, $H_1 : \theta \in \Theta_1$.

Критерий задаётся критической областью S в выборочном пространстве: если $X \in S$, отвергаем H_0 ; иначе не отвергаем. Неотвержение **не доказывает** истинность H_0 .

! Две ошибки

- Ошибка I рода: отвергнуть верную H_0 .
- Ошибка II рода: не отвергнуть H_0 , когда верна H_1 .
- При фиксированном ограничении ошибки I рода стараемся максимизировать вероятность обнаружения альтернативы.

2. Функция мощности, размер и уровень

Для критической области S функция мощности определяется формулой

$$\beta(\theta, S) = P_\theta(X \in S).$$

При $\theta \in \Theta_0$ это вероятность ошибки I рода, при $\theta \in \Theta_1$ — мощность. Вероятность ошибки II рода при альтернативе θ равна $1 - \beta(\theta, S)$.

Размер критерия: $\alpha_0(S) = \sup_{\theta \in \Theta_0} \beta(\theta, S)$. Критерий имеет **уровень значимости** α , если $\alpha_0(S) \leq \alpha$ (равенство не обязательно).

Равномерно наиболее мощный (р.н.м., UMP) критерий уровня α : для любого другого критерия R уровня α и каждого $\theta \in \Theta_1$ выполнено $\beta(\theta, S) \geq \beta(\theta, R)$.

💡 Стратегия решения

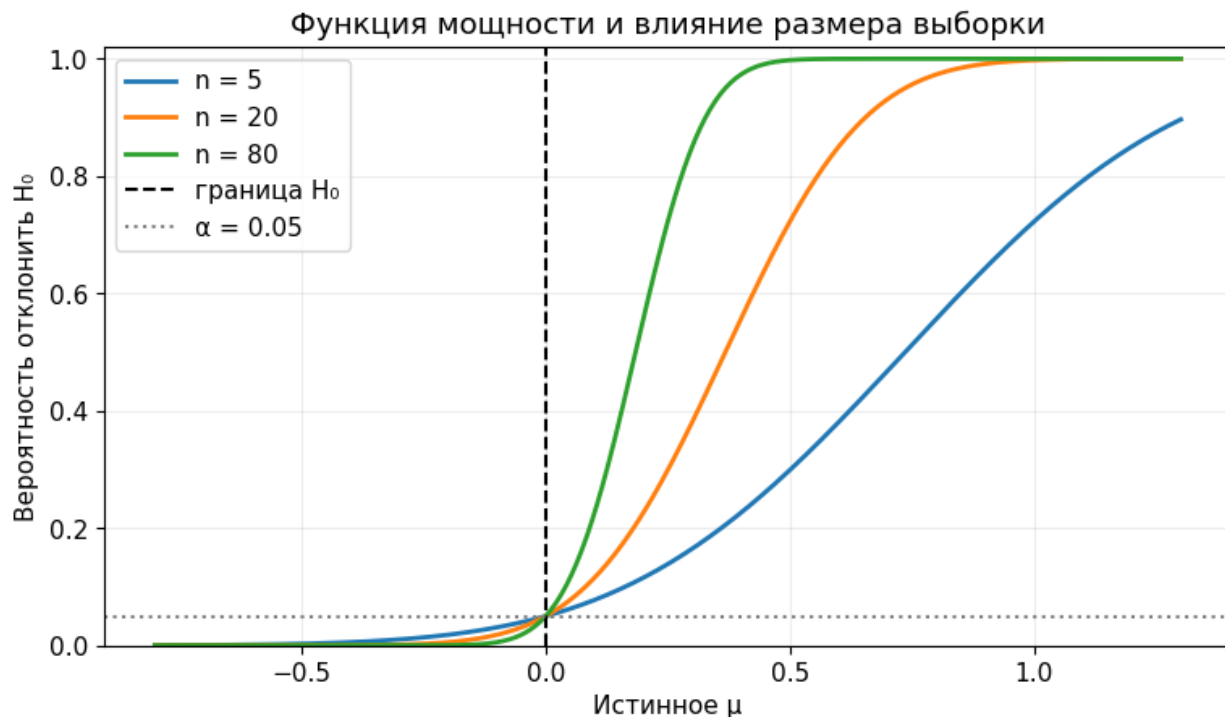
1. Выписать H_0 , H_1 и плотность/вероятности выборки.
2. Найти статистику, монотонную по отношению правдоподобия.
3. Определить направление критической области.
4. Выбрать порог из распределения **при** H_0 .
5. Вычислить мощность при произвольном параметре альтернативы.
6. Обосновать оптимальность леммой Неймана—Пирсона или теоремой о монотонном отношении правдоподобия.

2.1. Почему уровень и мощность — разные вещи

Для нормальной модели $X_i \sim N(\mu, 1)$ проверяем $H_0 : \mu \leq 0$ против $H_1 : \mu > 0$. Правосторонний критерий имеет вид $\sqrt{n} \bar{X} > z_{1-\alpha}$, где $z_q = \Phi^{-1}(q)$. Его функция мощности:

$$\beta(\mu) = 1 - \Phi(z_{1-\alpha} - \sqrt{n} \mu).$$

При $\mu = 0$ она равна α ; при $\mu > 0$ растёт к единице.



3. Лемма Неймана—Пирсона

! Лемма (простая гипотеза против простой альтернативы)

Пусть $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$ и существуют плотности p_0, p_1 относительно общей меры. Критерий, отвергающий H_0 при большом отношении $p_1(x)/p_0(x)$, является наиболее мощным среди критериев данного уровня α . В общем случае на границе $p_1 = cp_0$ допускается рандомизация, чтобы получить точный уровень.

$$\varphi(x) = \begin{cases} 1, & p_1(x) > cp_0(x), \\ \gamma, & p_1(x) = cp_0(x), \\ 0, & p_1(x) < cp_0(x). \end{cases}$$

Идея доказательства. Для произвольного конкурирующего критерия ψ уровня не выше α имеем $(\varphi - \psi)(p_1 - cp_0) \geq 0$ поточечно. Интегрируя, получаем

$$E_1 \varphi - E_1 \psi \geq c(E_0 \varphi - E_0 \psi) \geq 0.$$

Последнее неравенство использует $E_0 \varphi = \alpha$ (при точной калибровке). Для нерандомизированных дискретных критериев может получиться уровень строго ниже α .

Важная оговорка: если $p_0(x) = 0$ и $p_1(x) > 0$, такие точки всегда выгодно включить в критическую область: это даёт мощность без ошибки I рода.

3.1. Бернулли: почему статистика $\sum X_i$?

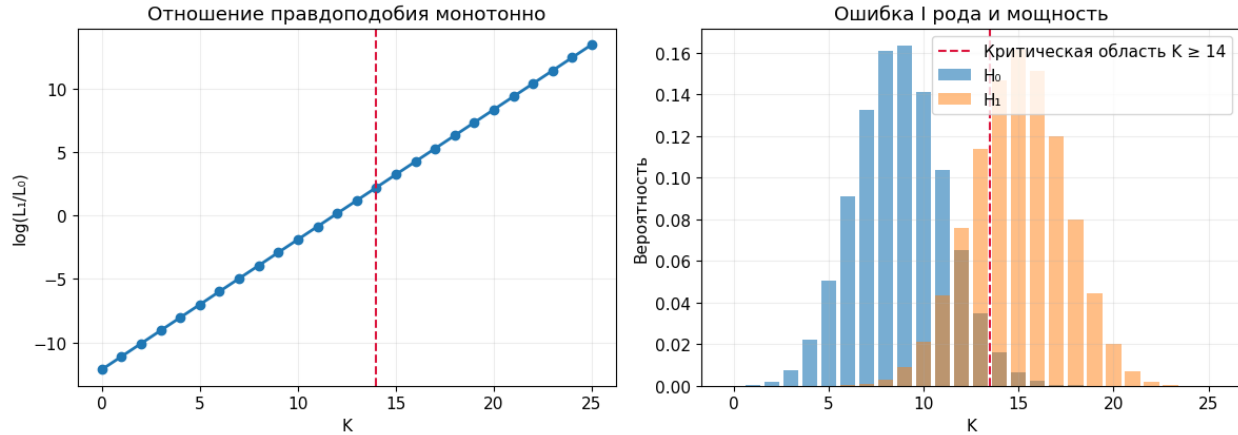
Для $X_i \stackrel{iid}{\sim} \text{Bern}(p)$ положим $K = \sum_{i=1}^n X_i$. Тогда

$$L(p; x) = p^K (1 - p)^{n-K},$$

$$\frac{L(p_1; x)}{L(p_0; x)} = \left(\frac{p_1(1-p_0)}{p_0(1-p_1)} \right)^K \left(\frac{1-p_1}{1-p_0} \right)^n.$$

Если $p_1 > p_0$, отношение растёт по K : отвергаем при **большом** K . Если $p_1 < p_0$, при **маленьком** K .

При дискретном распределении порог без рандомизации обычно даёт консервативный уровень. Для точного уровня можно рандомизировать на одном граничном значении K .



Фактический уровень: 0.02546060824372137

Мощность при p_1 : 0.732282173361202

4. Монотонное отношение правдоподобия (MLR)

Семейство $\{P_\theta\}$ имеет **неубывающее отношение правдоподобия по статистике** $T(X)$, если при любых $\theta_1 > \theta_0$ отношение $p_{\theta_1}(x)/p_{\theta_0}(x)$ является неубывающей функцией $T(x)$ (на общей области определения плотностей). Если отношение невозрастает, направление неравенств меняется.

! Теорема о р.н.м. критерии (Карлин—Рубин)

При неубывающем отношении правдоподобия по T критерий $H_0 : \theta \leq \theta_0$ против $H_1 : \theta > \theta_0$ отвергает при $T > c$ (при необходимости с рандомизацией на $T = c$). Порог выбирается по граничному распределению при θ_0 .

Для альтернативы $\theta < \theta_0$ критерий отвергает при малых T . Это можно получить заменой параметра $\eta = -\theta$ и статистики $-T$.

Почему оптимален для всех альтернатив сразу? Лемма Неймана—Пирсона для каждой фиксированной $\theta_1 > \theta_0$ приводит к **одной и той же форме** критической области $\{T > c\}$; критический порог определяется только θ_0 и α . Монотонность мощности по параметру гарантирует контроль ошибки на всём составном H_0 .

Пример: в $N(\mu, \sigma^2)$ с известной σ^2 статистика $T = \sum X_i$ имеет MLR по μ . Для $\text{Exp}(\text{rate } \lambda)$ отношение правдоподобия по $T = \sum X_i$ **невозрастает** по T при росте λ .

5. Равномерное распределение и зависимость носителя от параметра

Если $X_i \sim U[0, \theta]$, то совместная плотность

$$p_\theta(x) = \theta^{-n} \mathbf{1}\{0 \leq X_{(1)}, X_{(n)} \leq \theta\}, \quad X_{(n)} = \max_i X_i.$$

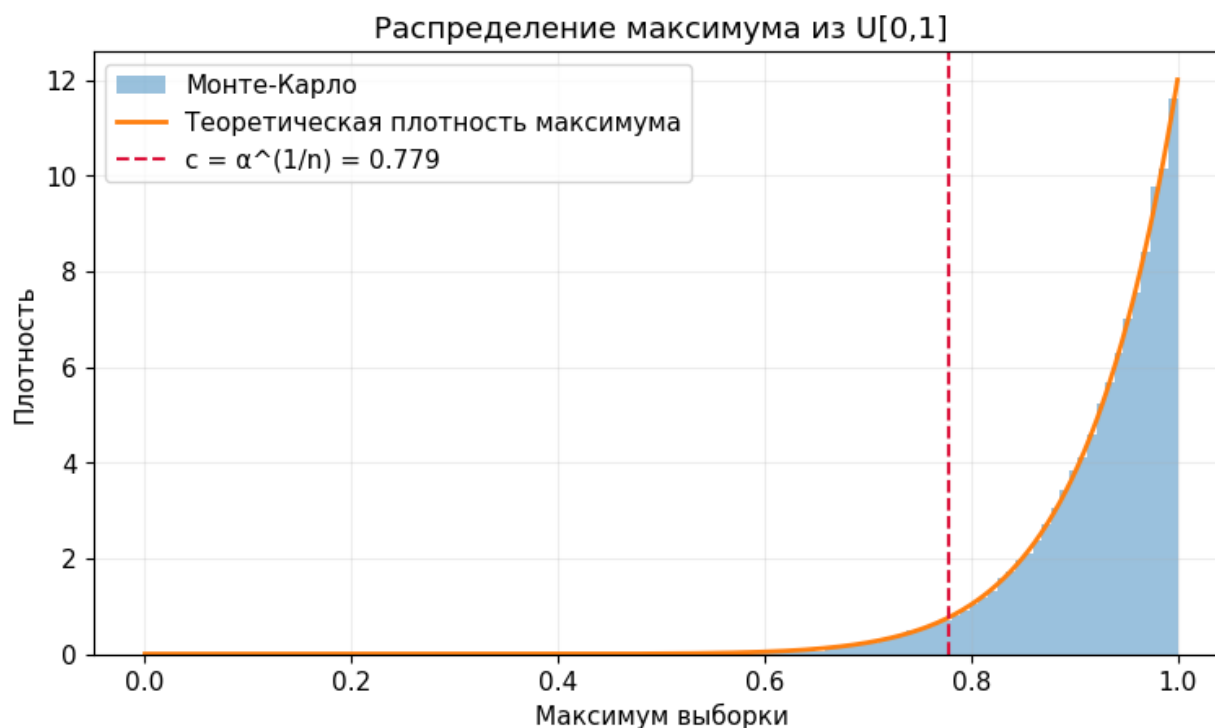
Для $0 \leq t \leq \theta$ имеем

$$P_\theta(X_{(n)} \leq t) = \left(\frac{t}{\theta} \right)^n.$$

Для проверки $H_0 : \theta = \theta_0$ против $H_1 : \theta < \theta_0$ естественна область $X_{(n)} \leq c\theta_0$, где $c = \alpha^{1/n}$.

⚠ Поддержка зависит от параметра

При $\theta > \theta_0$ наблюдение $X_{(n)} > \theta_0$ невозможно при H_0 и возможно при альтернативе. Такие точки следует включать в критическую область без увеличения уровня. Это особенно важно для двусторонней альтернативы в задаче 6.



Эмпирический уровень: 0.049825 Теоретический: 0.05

6. Экспоненциальное семейство и точные критические значения

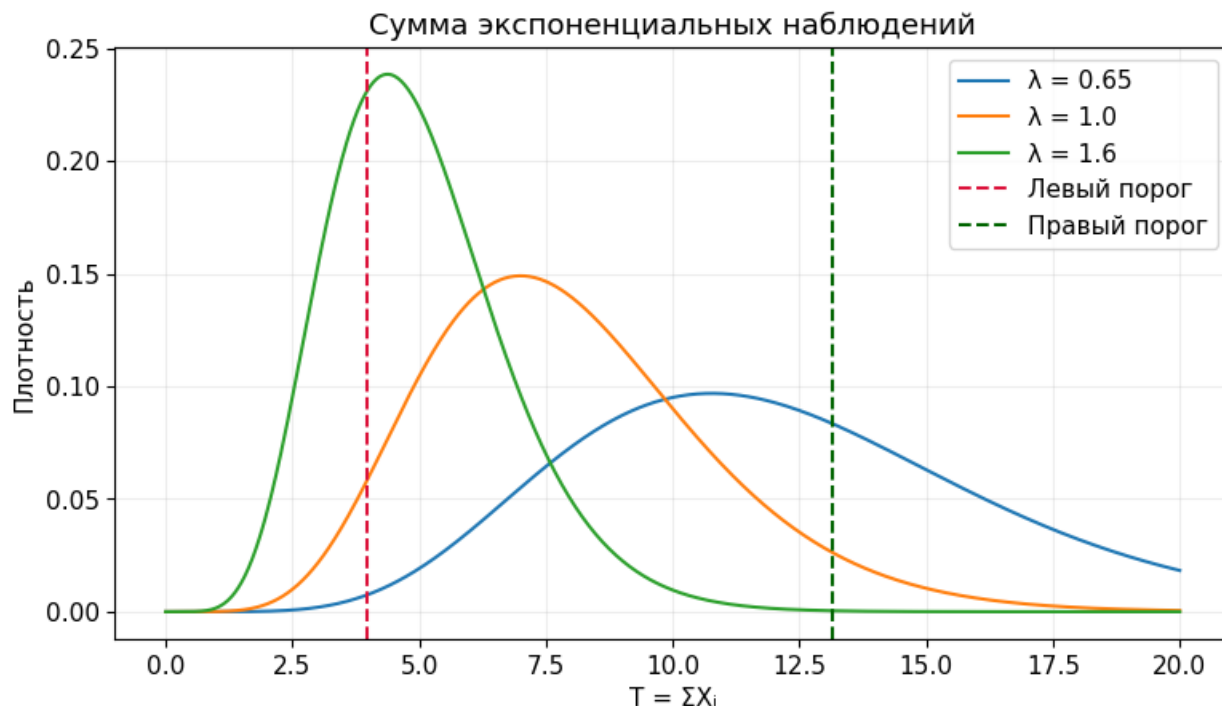
Если $X_i \sim \text{Exp}(\lambda)$, где λ — **интенсивность** (rate), то

$$L(\lambda; x) = \lambda^n \exp\left(-\lambda \sum_i x_i\right), \quad T = \sum_i X_i \sim \text{Gamma}(n, \text{scale} = 1/\lambda).$$

Эквивалентно, $2\lambda T \sim \chi_{2n}^2$.

- Для $H_1 : \lambda > \lambda_0$ критическая область $T \leq \chi_{2n;\alpha}^2 / (2\lambda_0)$.
- Для $H_1 : \lambda < \lambda_0$ критическая область $T \geq \chi_{2n;1-\alpha}^2 / (2\lambda_0)$.

Для $N(\mu, 1)$ статистика $\sqrt{n}(\bar{X} - \mu_0) \sim N(0, 1)$ при границе H_0 ; для правосторонней альтернативы сравниваем с $z_{1-\alpha}$, для левосторонней — с z_α .

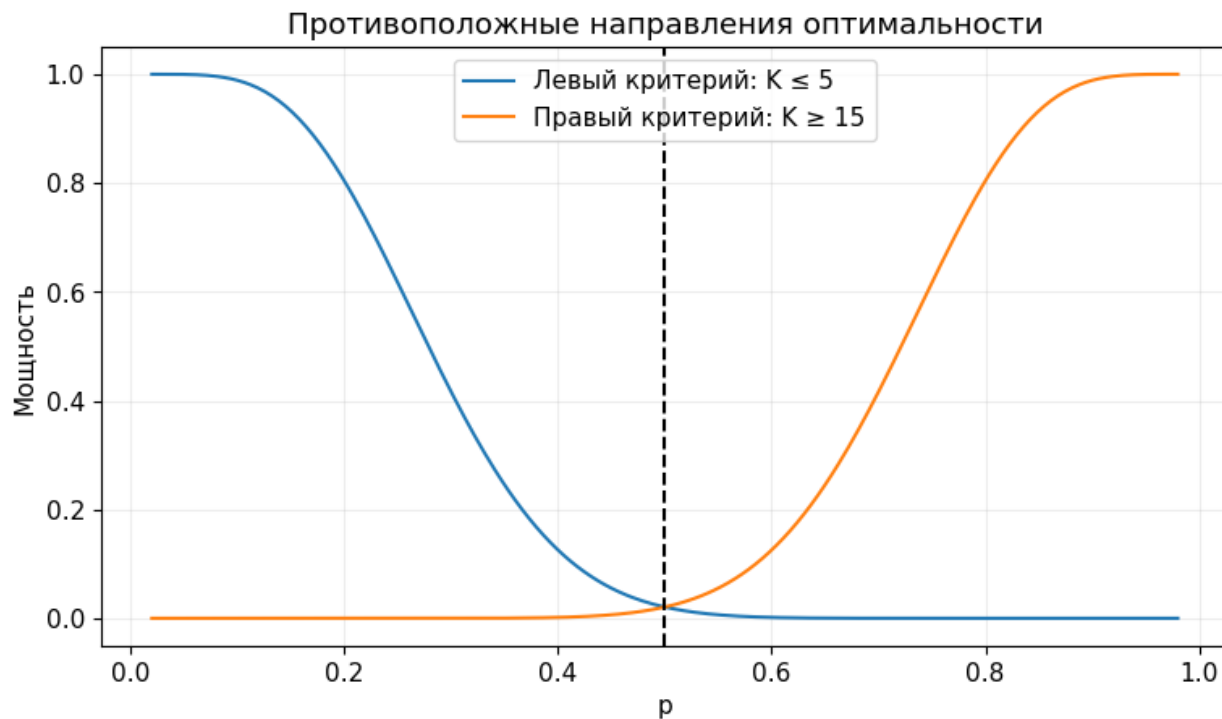


7 7. Двусторонние альтернативы и отсутствие UMP

Для $H_1 : \theta \neq \theta_0$ наиболее мощные критерии против альтернатив справа и слева могут требовать **противоположных** критических областей. Поэтому UMP-критерий для двусторонней альтернативы часто не существует. Это нужно **доказывать**, а не просто утверждать.

Для Бернулли можно сравнить единственные (или вынужденно совпадающие с ними) наиболее мощные области для $p_1 > p_0$ и $p_2 < p_0$ по лемме Неймана—Пирсона. При дискретности нужно аккуратно учитывать граничные уровни и возможность рандомизации.

Рандомизированный критерий — функция $\varphi(x) \in [0, 1]$, задающая вероятность отвергнуть H_0 при наблюдении x . Его мощность $\beta(\theta, \varphi) = E_\theta \varphi(X)$. Это расширение полезно для точного уровня в дискретных моделях.



8. Дисперсия нормального распределения: простая против простой

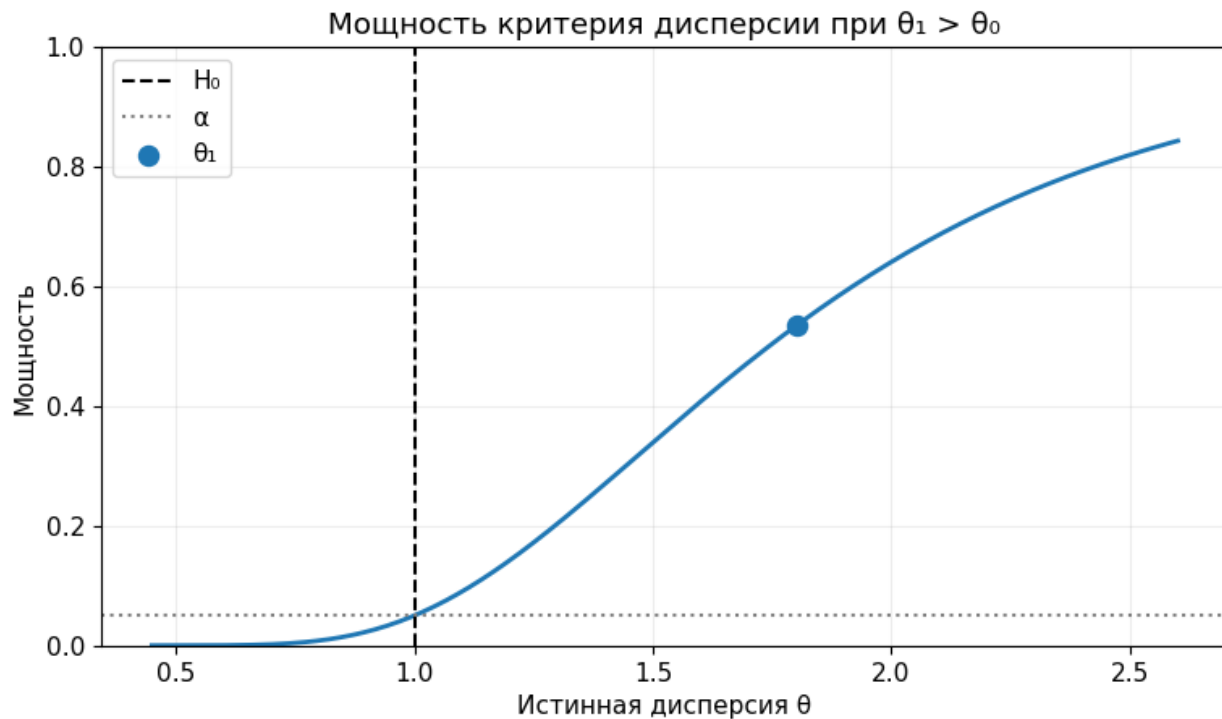
В задаче 5 рассматривается $X_i \sim N(0, \theta)$, где $\theta > 0$ — **дисперсия**, а не стандартное отклонение. Плотность выборки

$$L(\theta; x) = (2\pi\theta)^{-n/2} \exp\left(-\frac{\sum_i x_i^2}{2\theta}\right).$$

Для $H_0 : \theta = \theta_0$ против $H_1 : \theta = \theta_1$ логарифм отношения правдоподобия:

$$\log \frac{L(\theta_1; x)}{L(\theta_0; x)} = -\frac{n}{2} \log \frac{\theta_1}{\theta_0} + \frac{1}{2} \left(\frac{1}{\theta_0} - \frac{1}{\theta_1} \right) \sum_i x_i^2.$$

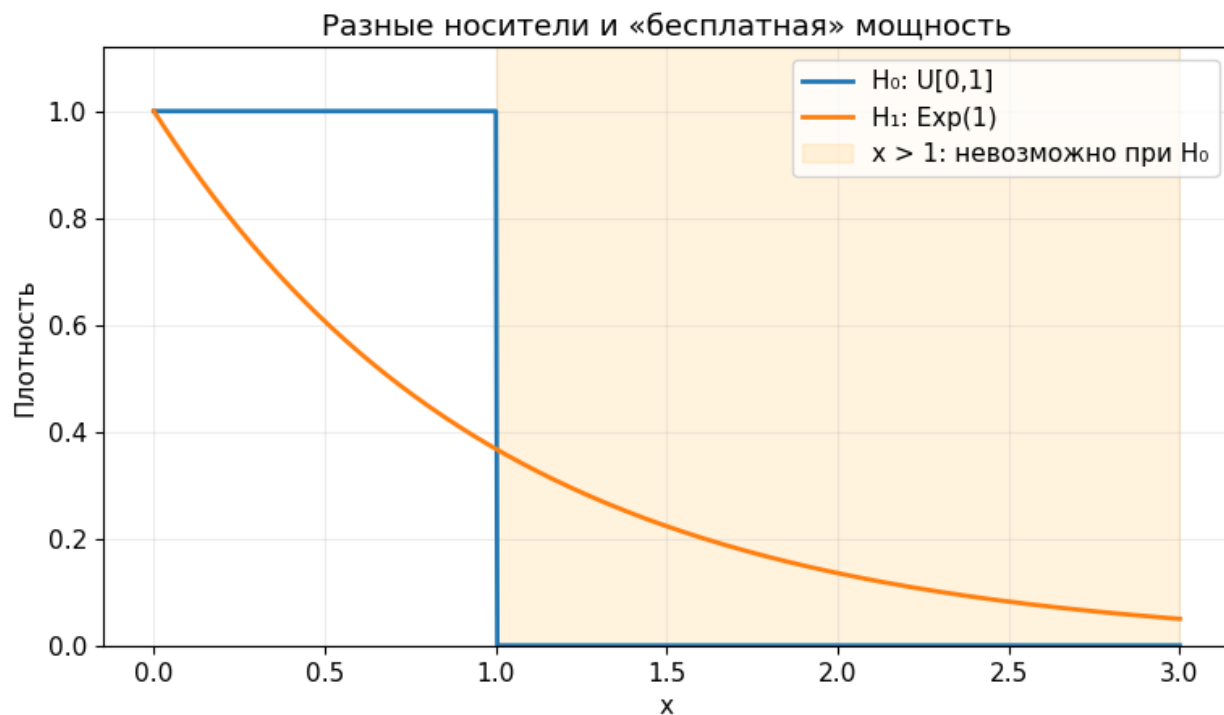
Следовательно, при $\theta_1 > \theta_0$ отвергаем для **больших** $T = \sum X_i^2$, при $\theta_1 < \theta_0$ — для **малых**. При H_0 $T/\theta_0 \sim \chi_n^2$, поэтому порог вычисляется точно. Мощность находится из распределения $T/\theta_1 \sim \chi_n^2$.



9 9. Особая задача 1: равномерное против показательного

Для одного наблюдения X гипотезы $H_0 : U[0, 1]$ и $H_1 : \text{Exp}(1)$ — **простые**. На $[0, 1]$ отношение плотностей равно e^{-x} , а при $x > 1$ плотность при H_0 равна нулю, при H_1 положительна.

Поэтому при построении критерия Неймана—Пирсона обязательно рассмотрите две части критической области: $x > 1$ (бесплатная мощность) и участок внутри $[0, 1]$, выбираемый по отношению правдоподобия. Мощность вычисляется интегрированием экспоненциальной плотности по обеим частям.



10. Допустимость и несмещённость критериев

Критерий S называется **несмещённым** (в смысле мощности), если

$$\inf_{\theta \in \Theta_1} \beta(\theta, S) \geq \sup_{\theta \in \Theta_0} \beta(\theta, S).$$

Это понятие **не связано** с несмещённостью оценок вида $E\hat{\theta} = \theta$.

Критерий S **допустим**, если не существует другого критерия R с $\beta(\theta, R) \geq \beta(\theta, S)$ для всех параметров альтернативы (и при сопоставимом контроле ошибки I рода), причём строгое неравенство выполняется хотя бы при одной альтернативе.

💡 Подсказка к задаче 8

Предположите, что другой критерий R уровня α не менее мощен на всей альтернативе и строго лучше где-то. Сопоставьте нижнюю грань мощности S на альтернативе с верхней гранью мощности S на нулевой гипотезе. Для строгого доказательства в формулировке задачи важно учитывать **равномерную наиболее мощность** S : она сразу запрещает строгое улучшение среди критериев уровня α .

11. Практикум: проверяем теорию методом Монте-Карло

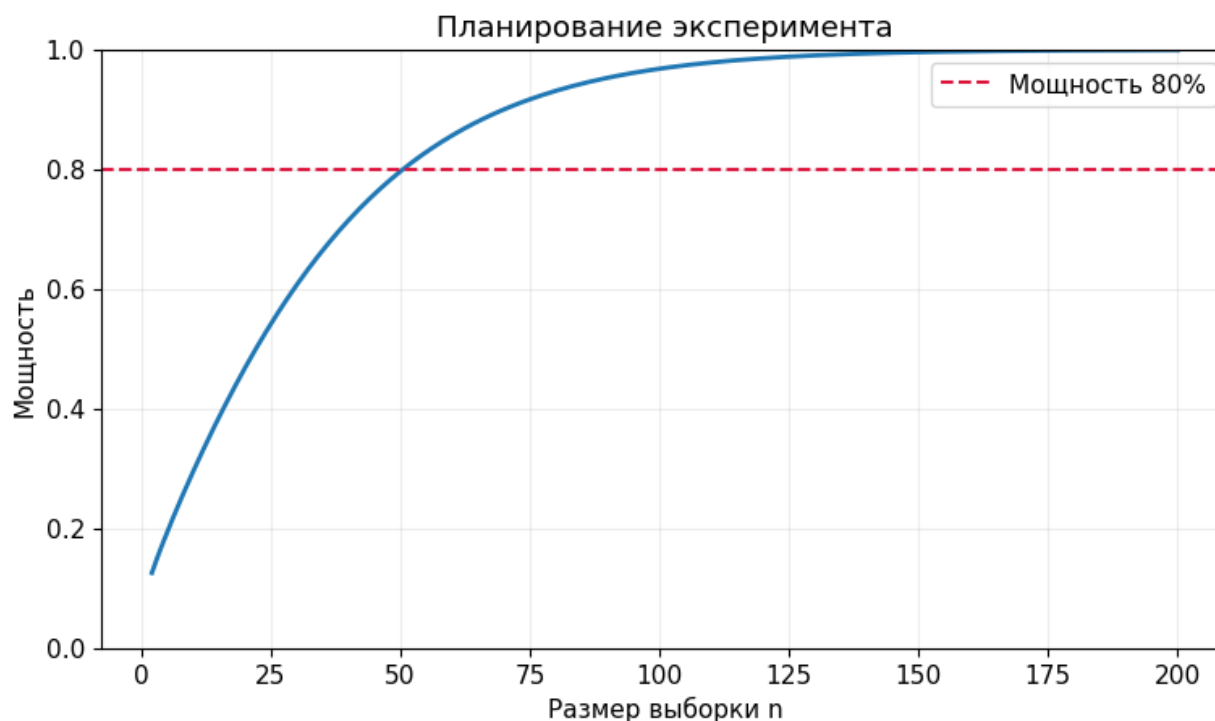
11.1. Эмпирическая ошибка I рода и мощность

Смоделируем правосторонний критерий нормального среднего и сравним с точной формулой.

$\mu=0.00$	эмпирическая мощность=0.0495	теория=0.0500
$\mu=0.15$	эмпирическая мощность=0.2079	теория=0.2052
$\mu=0.30$	эмпирическая мощность=0.4971	теория=0.4993
$\mu=0.50$	эмпирическая мощность=0.8653	теория=0.8630

11.2. Размер выборки и вероятность ошибки II рода

При фиксированной альтернативе $\mu_1 > \mu_0$ вероятность ошибки II рода равна $\Phi(z_{1-\alpha} - \sqrt{n}(\mu_1 - \mu_0))$. Это позволяет планировать объём выборки для заданной мощности.



Минимальное n для мощности $\geq 80\%$: 51

12. Задачи для самостоятельного решения (стр. 75–76)

Ниже приведены все восемь задач из пособия. Полные решения намеренно не включены. Под каждой задачей — раскрываемая подсказка с нужным инструментом.

💡 Задача 1

Имеется X_1 — выборка объёма 1. Основная гипотеза H_0 состоит в том, что X_1 имеет равномерное распределение на отрезке $[0, 1]$, альтернатива — в том, что X_1 имеет показательное распределение с параметром 1. Постройте наиболее мощный критерий уровня значимости α для различения этих гипотез и вычислите его мощность.

Подсказка и необходимые результаты

Лемма Неймана—Пирсона; разделите $x \in [0, 1]$ и $x > 1$; используйте $p_1/p_0 = e^{-x}$ на $[0, 1]$.

💡 Задача 2

$X_1, \dots, X_n$ — выборка из экспоненциального распределения с параметром θ . Постройте равномерно наиболее мощный критерий уровня значимости α проверки гипотезы $H_0 : \theta = \theta_0$ против альтернативы: (а) $H_1 : \theta > \theta_0$; (б) $H_1 : \theta < \theta_0$.

Подсказка и необходимые результаты

Используйте $T = \sum X_i$, гамма-распределение и монотонное отношение правдоподобия. Здесь θ — интенсивность.

💡 Задача 3

$X_1, \dots, X_n$ — выборка из нормального распределения с параметрами $(\theta, 1)$. Постройте равномерно наиболее мощный критерий уровня значимости α проверки: (а) $H_0 : \theta \geq \theta_0$ против $H_1 : \theta < \theta_0$; (б) $H_0 : \theta \leq \theta_0$ против $H_1 : \theta > \theta_0$.

Подсказка и необходимые результаты

Статистика $\sqrt{n}(\bar{X} - \theta_0)$, нормальные квантили; максимальная ошибка I рода достигается на границе.

💡 Задача 4

$X_1, \dots, X_n$ — выборка из распределения $\text{Bern}(\theta)$. Докажите, что не существует равномерно наиболее мощного критерия произвольного уровня значимости α для проверки $H_0 : \theta = \theta_0$ против $H_1 : \theta \neq \theta_0$.

Подсказка и необходимые результаты

Сопоставьте оптимальные области для $\theta_1 > \theta_0$ и $\theta_2 < \theta_0$. Уточните роль нерандомизированных критериев и крайних уровней.

💡 Задача 5

$X_1, \dots, X_n$ — выборка из распределения $N(0, \theta)$. Постройте равномерно наиболее мощный критерий уровня значимости α проверки $H_0 : \theta = \theta_0$ против $H_1 : \theta = \theta_1$ и найдите его мощность.

Подсказка и необходимые результаты

Примените лемму Неймана—Пирсона к $T = \sum X_i^2$ и распределение $T/\theta \sim \chi_n^2$. Разделите случаи $\theta_1 > \theta_0$ и $\theta_1 < \theta_0$.

💡 Задача 6

Пусть $X_1, \dots, X_n$ — выборка из равномерного распределения на отрезке $[0, \theta]$, $\theta > 0$. Постройте р.н.м. критерий уровня значимости α для проверки $H_0 : \theta = \theta_0$ против $H_1 : \theta \neq \theta_0$ в виде

$$S(X_1, \dots, X_n) = \{X_{(n)} \leq c\theta_0\} \cup \{X_{(n)} > \theta_0\}.$$

Подсказка и необходимые результаты

Вычислите $P_{\theta_0}(X_{(n)} \leq c\theta_0) = c^n$; объясните добавление области $X_{(n)} > \theta_0$.

💡 Задача 7

Пусть $\{P_\theta, \theta \in \Theta\}$ — семейство с невозрастающим отношением правдоподобия по статистике $T(X)$, а $\alpha < 1$ — некоторое положительное число. Постройте р.н.м. критерий уровня значимости α для проверки гипотез H_0 против H_1 , где: (а) $H_0 : \theta \leq \theta_0$ (или $\theta = \theta_0$), $H_1 : \theta > \theta_0$; (б) $H_0 : \theta \geq \theta_0$ (или $\theta = \theta_0$), $H_1 : \theta < \theta_0$.

Подсказка и необходимые результаты

Сделайте замену параметра $\eta = -\theta$ и статистики $U = -T$, после чего примените теорему о несмещённом отношении правдоподобия.

💡 Задача 8

Показать, что любой равномерно наиболее мощный несмещённый (т.е. $\inf_{\theta \in \Theta_1} \beta(\theta, S) \geq \sup_{\theta \in \Theta_0} \beta(\theta, S)$) критерий S является допустимым, т.е. не существует другого критерия R , который был

бы не менее мощен, чем S , при всех альтернативах и более мощен хотя бы при одной из альтернатив.

Подсказка и необходимые результаты

Разберите определения UMP и допустимости; проверьте, какое множество конкурирующих критериев предполагается (одинаковый уровень/ограничение на ошибку I рода).

13. Контрольные вопросы

1. Чем размер критерия отличается от номинального уровня?
2. Почему вероятность ошибки II рода зависит от конкретной альтернативы?
3. Почему для дискретной статистики точный уровень может требовать рандомизации?
4. Почему для двусторонних альтернатив UMP может не существовать?
5. Что меняется, если носитель распределения зависит от параметра?
6. Почему у экспоненциальной интенсивности правосторонняя альтернатива ведёт к **левому** хвосту суммы?
7. В чём отличие несмещённого критерия от несмещённой оценки?

i Замечание о строгости формулировок

Задачи 4 и 8 требуют аккуратной интерпретации класса критериев (рандомизированные или нет), диапазона α и допустимого множества конкурентов. В пособии формулировки краткие; при доказательствах следует явно указать эти соглашения.