Ульянов В. В. Ушаков В Г.
Байрамов Н. Р. Нагапетян Т. А.

РЕШЕНИЯ ЗАДАЧ
МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Москва
2007

Аннотация. Данное методическое пособие предназначено для подготовки к экзамену по теории вероятности и математической статистике, который вот уже много лет проводится на факультете ВМиК МГУ после второго года обучения. Авторы постарались изложить наиболее стандартные решения задач, которые предлагались студентам на контрольных работах и экзаменах.

При решении задачи на экзамене студент не обязан вести изложение так, как предложено в данном пособии. Вместе с тем, по мнению авторов, предлагаемые здесь решения задач весьма подробно и полно раскрывают содержание решений.

Авторы благодарят Ульянова В. В., Шестакова О. В., которые научили их решать задачи по математической статистике. Также выражается благодарность Деревенцу Е., Дышкант Н., Ширяеву В. за ценные замечание касательно верстки и набора текста.

Ваши замечания, размышления, предложения, оценки и конструктивную критику направляйте по адресу nagapetyan@gmail.com.

В планы авторов входит дополнение данного пособия задачами по теории вероятности, которые предлагаются в третьем семестре обучения на втором курсе. И мы будем благодарны, если вы на тот же адрес будете присылать условия задач с семинарских, контрольных, зачетных работ.

Глава 1

Оценивание

Определения и теоремы

1. **Статистической структурой** называется совокупность $(\Omega, \mathcal{F}, \mathcal{P})$, где
   Ω — множество элементарных исходов,
   $\mathcal{F}$ — σ-алгебра *событий* — подмножеств Ω,
   $\mathcal{P}$ — семейство вероятностных мер на $\mathcal{F}$.
   Семейство $\mathcal{P}$ может быть параметрическим: $\mathcal{P}_\theta = \{P_\theta,\ \theta \in \Theta\}$.
   Как правило, рассматривают случайную величину X на Ω и индуцированную ею статистическую структуру $(\mathscr{X}, \mathcal{B}, \mathcal{P}_{X,\theta})$, где
   $\mathscr{X}$ — множество значений случайной величины X,
   $\mathcal{B}$ — борелевская σ-алгебра на прямой,
   $\forall B \in \mathcal{B} \quad P_{X,B} = P\{X \in B\}$.
2. **Повторная выборка** — это случайный вектор $(X_1, \ldots, X_n)$, в котором $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины.
3. **Статистикой** $T(X)$ называется любая измеримая функция $T : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ от выборки $X = (X_1, \ldots, X_n)$.
4. Пусть функция распределения случайных величин $X_1, \ldots, X_n$ зависит от параметра θ, и функция $T(x)$ возвращает приближенное значение функции $\tau(\theta)$ по заданному значению случайной выборки. Тогда можно рассматривать $T(x)$ как единичное наблюдение случайной величины $T(X) = T(X_1, \ldots, X_n)$. Случайная величина $T(X)$ — **оценка** функции $\tau(\theta)$.
5. $T(X)$ называется **несмещенной оценкой** функции $\tau(\theta)$, если $\mathbb{E}T(X) = \tau(\theta)$ для любого $\theta \in \Theta$.
   Пример. Пусть $X_1, \ldots, X_n$ — независимые одинаково распределенные случайные величины, $\mu = \mathbb{E}X_1$, $\sigma^2 = \mathbb{D}X_1$. Исследовать несмещенность оценки $\overline{X}^2$ для μ^2.
   Решение. Найдем математическое ожидание
   $$\begin{aligned}\mathbb{E}\overline{X}^2 &= \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^n X_i\Big)^2 = \\ &= \frac{1}{n^2}\mathbb{E}\Big(\sum_{i=1}^n X_i^2 + \sum_{i=1,j=1,i\neq j}^n X_i X_j\Big) = \\ &= \frac{1}{n^2}\big(n(\sigma^2+\mu^2) + (n^2-n)\mu^2\big) = \\ &= \mu^2 + \frac{\sigma^2}{n} > \mu^2,\end{aligned}$$
   поэтому $\overline{X}^2$ — смещенная оценка μ^2.
6. **Смещением** оценки $T(X)$ называется величина $B(T(X)) = \mathbb{E}T(X) - \tau(\theta)$.
7. **Среднеквадратичной погрешностью** оценки $T(X)$ называют $\mathbb{E}\big(T(X) - \tau(\theta)\big)^2 = \mathbb{D}T(X) + \big(B(T(X))\big)^2$.
8. **Погрешность оценки** $T(X)$ — это величина $e = |T(X) - \tau(\theta)|$.
9. Пусть $T_1(X)$ и $T_2(X)$ — несмещенные оценки $\tau(\theta)$.
   (a) Если $\mathbb{D}T_1(X) < \mathbb{D}T_2(X)$, оценка $T_1(X)$ **эффективнее** оценки $T_2(X)$.

(b) Эффективность $T_1(X)$ относительно $T_2(X)$, есть

$$\text{Эффективность} = \frac{\mathbb{D}T_1(X)}{\mathbb{D}T_2(X)}.$$

10. $T(X)$ называется **состоятельной оценкой** функции $\tau(\theta)$, если $T(X) \to \tau(\theta)$ по вероятности при $n \to \infty$ для любого $\theta \in \Theta$, то есть

$$\lim_{n\to\infty} P\big(|T(X)-\tau(\theta)| \le \varepsilon\big) = 1, \quad \text{или, что то же,}$$
$$\lim_{n\to\infty} P\big(|T(X)-\tau(\theta)| > \varepsilon\big) = 0.$$

11. $T(X)$ является *состоятельной* оценкой функции $\tau(\theta)$, если
    (a) $T(X)$ — несмещенная оценка, и
    (b) $\lim\limits_{n\to\infty} \mathbb{D}T(X) = 0$.
12. *Несмещенная* оценка:
    (a) может не существовать;
    (b) не единственна;
    (c) может быть бессмысленной;
    (d) не является, вообще говоря, состоятельной.
13. *Состоятельная* оценка:
    (a) не единственна;
    (b) может быть бессмысленной;
    (c) не является, вообще говоря, несмещенной.
14. Оценка $T(X)$ функции $\tau(\theta)$ называется **оптимальной**, если
    1) $T(X)$ — несмещенная, то есть $\mathbb{E}T(X) = \tau(\theta)$
    2) $T(X)$ имеет равномерно минимальную дисперсию, то есть для любой другой несмещенной оценки $T_1(X)$ функции $\tau(\theta)$ выполнено $\mathbb{D}_\theta T(X) \le \mathbb{D}_\theta T_1(X) \quad \forall \theta \in \Theta$.
15. Статистика $T(X) = \big(T_1(X), \ldots, T_k(X)\big)$ называется **достаточной**, если для любого борелевского множества A $P_\theta\big(X \in A \mid T(X)\big)$ не зависит от θ.
16. *Достаточная* статистика может не существовать.
17. **Функция прадоподобия.**
    Пусть $x_1, x_2, \ldots, x_n$ — значения повторной выборки из распределения $\mathcal{L}(X)$, зависящего от набора параметров $\theta = (\theta_1, \theta_2, \ldots, \theta_r)$. *Функция прадоподобия* выборки $L(x_1, \ldots, x_n; \theta)$ определяется следующим образом:
    (1) Если $\mathcal{L}(X)$ — дискретно,

    $$L(x_1, \ldots, x_n; \theta) = P\left(X_1 = x_1, \ldots, X_n = x_n; \theta\right).$$

    (2) Если $\mathcal{L}(X)$ — абсолютно непрерывно и $p(x;\theta)$ — плотность распределения случайной величины X,

    $$L(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} p(x_i; \theta).$$

18. **Критерий факторизации**
    Пусть $L(X,\theta)$ — функция правдоподобия выборки X, $T(X) = \big(T_1(X), \ldots, T_k(X)\big)$ — некоторая статистика.
    Тогда $T(X)$ — достаточная статистика тогда и только тогда, когда функцию правдоподобия можно представить в виде произведения $L(X,\theta) = g\big(T(X),\theta\big) \times h(X)$.
19. Статистика $T(X)$ называется **полной**, если из $\mathbb{E}\varphi(T(X)) = 0$ для любого θ следует равенство $\varphi(u) = 0$ почти всюду по распределению $T(X)$.
20. Семейство $\mathcal{P} = \big\{P_\theta,\ \theta \in \Theta \subset \mathbb{R}^k\big\}$, допускающее функцию правдоподобия вида

    $$L(X,\theta) = K(\theta) \times \exp\Big(\sum\nolimits_{i=1}^{k} a_i(\theta) T_i(X)\Big) \times h(X),$$

    называется **экспоненциальным семейством**.

21. **Теорема о полноте экспоненциальных семейств**
    Пусть
    - $\mathcal{P} = \{P_\theta,\ \theta \in \Theta \subset \mathbb{R}^k\}$ — экспоненциальное семейство, и
    - функции $a_1(\theta), \ldots, a_k(\theta)$ и параметрическое пространство Θ таковы, что $a(\theta) = (a_1(\theta), \ldots, a_k(\theta))$ зачерчивает некоторый k-мерный параллелепипед, когда θ пробегает Θ.

    Тогда $T(X) = (T_1(X), \ldots, T_k(X))$ является полной достаточной статистикой.
22. **Неравенство Рао-Крамера**
    Пусть $X_1, \ldots, X_n$ — выборка с функцией правдоподобия $L(X, \theta)$, а $T(X)$ — несмещенная оценка функции $\tau(\theta)$.
    Пусть $L(X, \theta)$, $T(X)$ и $\tau(\theta)$ удовлетворяют условию регулярности:
    1) Множество $X : L(X, \theta) > 0$ не зависит от θ;
    2) Функция $L(X, \theta)$ дифференцируема по θ и

    $$\frac{d}{d\theta} \int L(X, \theta)\, \mu(dX) = \int \frac{d}{d\theta} L(X, \theta)\, \mu(dX),$$
    $$\frac{d}{d\theta} \int T(X) L(X, \theta)\, \mu(dX) = \int T(X) \frac{d}{d\theta} L(X, \theta)\, \mu(dX);$$

    3) Функция $\tau(\theta)$ дифференцируема.
    Если существует конечный второй момент $T(X)$, то

    $$D_\theta T(X) \geq \frac{[\tau'(\theta)]^2}{\mathbb{E}_\theta \left(\frac{\partial}{\partial\theta} \ln L(X, \theta)\right)^2} \qquad \forall \theta \in \Theta.$$

    Это неравенсто превращается в равенство, если и только если существует такая фукция $a_n(\theta)$, что $T(X) - \tau(\theta) = a_n(\theta) \times \frac{\partial}{\partial\theta} \ln L(X, \theta)$.
    Оценка, для которой в неравенстве Рао-Крамера достигается равенство, называется **эффективной**.
    Эффективная оценка, если существует, является оптимальной.
    Пример. Плотность распределения случайной величины, распределенной по нормальному закону с неизвестным матожиданием θ и известной дисперсией σ^2, есть $f(x; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right)$. Используя неравенство Рао-Крамера, показать, что дисперсия любой несмещенной оценки $\widehat{\Theta}$ параметра θ не меньше, чем ${\sigma^2}/{n}$.
    Решение.
    (1) $\dfrac{\partial}{\partial\theta} \ln f(X; \theta) = \dfrac{X - \theta}{\sigma^2}$
    (2) $\left(\dfrac{\partial \ln f(X; \theta)}{\partial\theta}\right)^2 = \dfrac{(X - \theta)^2}{\sigma^4}$
    (3) $\mathbb{E}\, \dfrac{(X - \theta)^2}{\sigma^4} = \displaystyle\int_{-\infty}^{\infty} \dfrac{(X - \theta)^2}{\sigma^4} \left(\dfrac{1}{\sqrt{2\pi}\sigma} \exp\left(-\dfrac{(x - \theta)^2}{2\sigma^2}\right)\right) dx = \dfrac{1}{\sigma^2}$
    (4) $\mathbb{D}\widehat{\Theta} \geq \dfrac{1}{n\frac{1}{\sigma^2}} = \dfrac{\sigma^2}{n}$.
23. **Теорема Рао-Блекуэлла-Колмогорова**
    Пусть $T(X)$ — достаточная статистика выборки $X_1, \ldots, X_n$. Тогда если существует оптимальная оценка $T_1(X)$ для функции $\tau(\theta)$, то эта оценка является фукнцией от достаточной статистики $T(X)$: $T_1(X) = \varphi(T(X))$.
24. Измеримая функция от полной достаточной статистики является оптимальной оценкой своего математического ожидания.

25. **Метод моментов**.
Оценками методом моментов являются решения системы уравнений

$$\mu'_r = \mathbb{E}X^r = \frac{1}{n}\sum_{i=1}^{n} x_i^r = m'_r, \quad r = 1, 2, \ldots, k,$$

где k — число параметров.

26. Оценки *методом моментов*:
   (a) не единственны;
   (b) могут не быть функциями от достаточной или полной статистик.

27. **Оценкой максимального правдоподобия** (О.М.П.) $\hat{\theta}(X)$ параметра θ называется такое значение параметра, что

$$\max L(X, \theta) = L\big(X, \hat{\theta}(X)\big),$$

где $L(x, \theta)$ — функция правдоподобия выборки $X = (X_1, \ldots, X_n)$.
Часто оказывается проще максимизировать функцию $\ln L(x, \theta)$, что эквивалентно максимизации функции правдоподобия, поскольку $\ln L(x, \theta)$ есть монотонная функция от $L(x, \theta)$.
Пример. Пусть $X_1, \ldots, X_n$ — выборка из Пуассоновского распределения с параметром λ. Найти О.М.П. параметра λ.
Решение.
(1) $P_\lambda(X = x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$.
(2) Вычислим функцию правдоподобия:

$$L(\lambda) = \left(\frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\right)\left(\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\right)\cdots\left(\frac{e^{-\lambda}\lambda^{x_n}}{x_n!}\right) = \\ = \frac{e^{-n\lambda}\lambda^{x_1+x_2+\cdots+x_n}}{x_1!\, x_2! \cdots x_n!}, \\ \ln L(\lambda) = -n\lambda + (x_1 + x_2 + \cdots + x_n)\ln\lambda + \ln(x_1!\, x_2! \cdots x_n!).$$

(3) $\dfrac{\partial \ln L(\lambda)}{\partial \lambda} = -n + \dfrac{x_1 + x_2 + \cdots + x_n}{\lambda} = 0.$
(4) Решение относительно λ: $\overline{\lambda} = \dfrac{x_1 + x_2 + \cdots + x_n}{n} = \overline{x}$ — О.М.П. для λ.

28. Оценка *максимального правдоподобия* (О.М.П.):
   (a) не обязана быть состоятельной;
   (b) может не быть несмещенной;
   (c) не единственна.

29. Если существует единственная достаточная статистика T для параметра θ, то О.М.П. для θ является функцией от статистики T.

30. **Принцип инвариантности** для О.М.П.
Пусть $\hat{\theta}$ — оценка максимального правдоподобия для θ. Если $\tau(\cdot)$ — функция, обратная к которой однозначна, то О.М.П. для $\tau(\theta)$ есть $\tau(\hat{\theta})$.

31. **Различные оценки**.
Пусть $\{x_1, x_2, \ldots, x_n\}$ — множество наблюдений.
   (1) Нормальное распределение: $\mathcal{N}(\mu, \sigma^2)$.
      (a) Если σ известно:
         1. $\sum x_i$ — полная и достаточная статистика.
         2. Точечная оценка для μ: $\hat{\mu} = \bar{x} = \dfrac{1}{n}\sum x_i$ — О.М.П. и оптимальная оценка.
      (b) Если μ известно:
         1. $\sum (x_i - \mu)^2$ — полная и достаточная статистика.

2. Точечная оценка для σ^2: $\widehat{\sigma^2} = \dfrac{\sum (x_i - \mu)^2}{n}$ — О.М.П. и оптимальная оценка.

(c) Если μ и σ неизвестны:

1. $\{\sum x_i, \sum (x_i - \mu)^2\}$— полная и достаточная статистика.
2. Точечная оценка для μ: $\hat{\mu} = \bar{x} = \dfrac{1}{n}\sum x_i$ — О.М.П. и оптимальная оценка.
3. Точечная оценка для σ^2: $\widehat{\sigma^2} = \dfrac{\sum (x_i - \bar{x})^2}{n}$ — О.М.П.
4. Точечная оценка для σ^2: $\widehat{\sigma^2} = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$ — оптимальная оценка.
5. Точечная оценка для σ: $\widehat{\sigma} = \dfrac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{2}\Gamma\left(\frac{n}{2}\right)}\sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n-1}}$ — оптимальная оценка.

(2) Пуассоновское распределение с параметром λ:

(a) $\sum x_i$ — полная и достаточная статистика.

(b) Точечная оценка для λ: $\hat{\lambda} = \dfrac{1}{n}\sum x_i$ — О.М.П. и оптимальная оценка.

(3) Равномерное распределение на отрезке:

(a) Отрезок $[0, \theta]$.

1. $\max(x_i)$ — полная и достаточная статистика.
2. Точечная оценка для θ: $\widehat{\Theta} = \max(x_i)$ — О.М.П.
3. Точечная оценка для θ: $\widehat{\Theta} = \dfrac{n+1}{n}\max(x_i)$ — оптимальная оценка.

(b) Отрезок $[\alpha, \beta]$.

1. $\{\min(x_i), \max(x_i)\}$ — полная и достаточная статистика.
2. Точечная оценка для α: $\hat{\alpha} = \dfrac{n\min(x_i) - \max(x_i)}{n-1}$ — оптимальная оценка.
3. Точечная оценка для α: $\hat{\alpha} = \min(x_i)$ — О.М.П.
4. Точечная оценка для $\frac{\alpha+\beta}{2}$: $\widehat{\dfrac{\alpha+\beta}{2}} = \dfrac{\min(x_i) + \max(x_i)}{2}$ — оптимальная оценка.

(c) Отрезок $[\theta - \frac{1}{2}, \theta + \frac{1}{2}]$.

1. $\{\min(x_i), \max(x_i)\}$ — полная и достаточная статистика.
2. Точечная оценка для θ: $\widehat{\Theta} = \dfrac{\min(x_i) + \max(x_i)}{2}$ — О.М.П.

Задачи

Задача №1. Пусть $X_1, \ldots, X_n$ независимы и имеют биномиальное распределение $Bi(1, \theta)$, $0<\theta<1$. Доказать, что $T(X) = \sum_{i=1}^{n} X_i$ – полная и достаточная статистика.

Решение. Покажем, что $P_\theta(X{=}x \mid T(X){=}t)$ не зависит от θ:

$$P_\theta\big(X{=}x \mid T(X){=}t\big) = \frac{P\big(X_1{=}x_1, \ldots, X_n{=}x_n,\, T(X) = t\big)}{P\big(T(X){=}t\big)} = \Big\{\sum_{i=1}^{n} X_i \sim Bi(n, \theta)\Big\} =$$

$$= \frac{\theta^{\sum\limits_{i=1}^{n} x_i}\big(1{-}\theta\big)^{n-\sum\limits_{i=1}^{n} x_i}}{C_n^t \cdot \theta^t \big(1{-}\theta\big)^{n-t}} \times \mathbb{I}_{\left\{\sum\limits_{i=1}^{n} x_i = t\right\}} = \frac{1}{C_n^t} \times \mathbb{I}_{\left\{\sum\limits_{i=1}^{n} x_i = t\right\}}$$

Достаточность статистики $T(X)$ доказана.

Пусть функция $\varphi(\cdot)$ такова, что $\mathbb{E}_\theta \varphi\big(T(X)\big){=}0 \quad \forall \theta{\in}(0,1)$.

$$\mathbb{E}_\theta \varphi\big(T(X)\big) = \sum_{k=0}^{n} \varphi(k) C_n^k \theta^k (1{-}\theta)^{n-k} = \Big\{\frac{\theta}{1{-}\theta}{=}\tau,\ \tau{\in}(0,+\infty)\Big\} =$$

$$= (1{-}\theta)^n \sum_{k=0}^{n} \varphi(k) C_n^k \tau^k \equiv 0 \iff \sum_{k=0}^{n} \varphi(k) C_n^k \tau^k \equiv 0$$

$$\forall \tau{\in}(0,+\infty).$$

$\sum\limits_{k=0}^{n} \varphi(k) C_n^k \tau^k$ – многочлен степени не выше n – имеет континуум корней, следовательно, все его коэффициенты равны нулю: $\varphi(k){=}0,\ k = \overline{0,n}$.

Итак, из $\mathbb{E}_\theta \varphi\big(T(X)\big){=}0 \quad \forall \theta{\in}(0,1)$ следует $\varphi{\equiv}0$ по распределнию $T(X)$, что и означает полноту статистики $T(X)$. □

Задача №2. Пусть $X_1, \ldots, X_n$ независимы и имеют равномерное распределение на отрезке $[a; b]$. Найти оценку максимального правдоподобия для a и b.

Решение. Запишем функцию правдоподобия:

$$L(X; a, b) = \prod_{k=1}^{n} f(X_k) = \frac{1}{(b{-}a)^n} \mathbb{I}_{\left\{a \le X_{(1)} \le X_{(n)} \le b\right\}}$$

Зафиксируем b. $L(X; a, b) \to \max$, при $a \le X_{(1)}$ и $\frac{1}{(b-a)^n} \to \max$. Значит, оценкой максимального правдоподобия для a будет $\widehat{a} = X_{(1)}$.

Аналогично, оценка максимального правдоподобия для b: $\widehat{b} = X_{(n)}$. □

Задача №3. Пусть $X_1, \ldots, X_n$ независимы и имеют равномерное распределение на отрезке $[0; \theta]$. Исследовать несмещенность и состоятельность оценки $T(X){=}X_{(n)}$ параметра θ.

Решение. Поскольку функция распределения $X_{(n)}$ представляется в виде

$$F_{X_{(n)}}(x) = \begin{cases} 0, & y < 0, \\ \dfrac{y^n}{\theta^n}, & y \in [0, \theta], \\ 1, & y > 0, \end{cases}$$

то плотность $X_{(n)}$ имеет следующий вид:

$$f_{X_{(n)}}(x) = \begin{cases} \dfrac{ny^{n-1}}{\theta^n}, & y \in [0,\theta] \\ 0, & \text{иначе.} \end{cases}$$

значит,

$$\mathbb{E}\, X_{(n)} = \int_0^\theta y \cdot \frac{ny^{n-1}}{\theta^n} dy = \frac{n\theta}{n+1} \neq \theta$$

Таким образом, $T(X)$ несмещенной оценкой для θ не является[1].

Проверим теперь состоятельность оценки. Для любого достаточно малого $\varepsilon > 0$

$$P\big(X_{(n)} < \theta - \varepsilon\big) = \left(\frac{\theta-\varepsilon}{\theta}\right)^n \xrightarrow[n\to\infty]{} 0,$$

следовательно,

$$P\big(X_{(n)} \in [\theta-\varepsilon, \theta]\big) \xrightarrow[n\to\infty]{} 1.$$

То есть, $T(X) = X_{(n)} \xrightarrow{P} \theta$, что означает состоятельность оценки $T(X)$. □

Задача № 4. Пусть $X_1, \dots, X_n$ независимы и имеют пуассоновское распределение $\Pi(\theta)$, $\theta{>}0$. Доказать, что $T(X) = \sum_{i=1}^n X_i$ – достаточная и полная статистика.

Решение. Покажем, что $P\big(X{=}x \mid T(X){=}t\big)$ не зависит от θ:

$$P\big(X{=}x \mid T(X){=}t\big) = \frac{P\big(X{=}x, T(X){=}t\big)}{P\big(T(X){=}t\big)} = \Big\{T(X){=}\sum_{i=1}^n X_i \sim \Pi(n\theta)\Big\} =$$

$$= \frac{\prod\limits_{k=1}^n \exp(-\theta)\dfrac{\theta^{x_k}}{x_k!} \times \mathbb{I}_{\left\{\sum\limits_{k=1}^n x_k = t\right\}}}{e^{-n\theta}\dfrac{(n\theta)^t}{t!}} = \frac{t!}{n^t \cdot x_1! \dots x_n!} \times \mathbb{I}_{\left\{\sum\limits_{k=1}^n x_k = t\right\}}$$

Достаточность статистики $T(X)$ доказана. Покажем, что она является полной.

$$\mathbb{E}_\theta \varphi\big(T(X)\big) = \sum_{k=0}^\infty \varphi(k) \exp(-n\theta) \frac{(n\theta)^k}{k!} \equiv 0 \qquad \forall \theta \in \Theta. \tag{$*$}$$

При $\theta{=}0$ $\quad \mathbb{E}_\theta \varphi\big(T(X)\big) = \varphi(0)$, значит, $\varphi(0) = 0$. Разделив ($*$) на θ и устремив θ к нулю, получим $\varphi(1) = 0$.

Повторяя эту процедуру (разделить ($*$) на θ и перейти к пределу в нуле), придем к $\varphi(k){=}0$, $\quad k = 0, 1, 2, \dots,$ Таким образом, $\varphi \equiv 0$ по распределению $T(X)$, что означает полноту статистики $T(X) = \sum_{i=1}^n X_i$. □

Задача №5. Пусть $X_1, \dots, X_n$ независимы и имеют равномерное распределение на отрезке $[a, b]$. Найти оценку методом моментов для a и b по первым 2 моментам.

[1] Однако $T(X)$ – асимптотически несмещенная оценка θ.

Решение. Обозначим $M_1{=}\overline{X}$, $M_2{=}\overline{X^2}$ – эмпирические моменты первого и второго порядков соответственно.

Для равномерно распределенной на $[a,b]$ случайной величины X теоретические моменты первого и второго порядков следующие:

$$\begin{aligned}\mu_1 &= \mathbb{E}X = \frac{a{+}b}{2},\\ \mu_2 &= \mathbb{E}X^2 = \frac{1}{3}\left(a^2+ab+b^2\right).\end{aligned}$$

Из условия равенства теоретических моментов эмпирическим получим оценку методом моментов для a и b:

$$\begin{aligned}a &= M_1{-}\sqrt{3\left(M_2{-}M_1^2\right)},\\ b &= M_1{+}\sqrt{3\left(M_2{-}M_1^2\right)}.\end{aligned}$$

□

Задача № 6. Пусть слуыайные величины $X_1,\ldots,X_n$ независимы и имеют нормальное распределение $\mathcal{N}(\theta,1)$. Исследовать несмещенность и состоятельность оценки $T(X)=\overline{X}$ параметра θ.

Решение. $\mathbb{E}T(X){=}\mathbb{E}X_1{=}\theta$. $T(X)=\overline{X}$ – несмещенная оценка параметра θ. Воспользуемся неравенством Чебышева:

$$\forall\varepsilon>0\quad P\left(\left|\overline{X}-\mathbb{E}\overline{X}\right|<\varepsilon\right)>1-\frac{\mathbb{D}\overline{X}}{\varepsilon^2}\xrightarrow[n\to\infty]{}1, \tag{$*$}$$

поскольку дисперсия оценки

$$\mathbb{D}\overline{X}=\frac{1}{n}\mathbb{D}X_1=\frac{1}{n}\xrightarrow[n\to\infty]{}0.$$

$(*)$ означает, что

$$T(X)=\overline{X}\xrightarrow{P}\theta.$$

что, в свою очередь, означает состоятельность оценки. □

Задача №7. Пусть $X_1,\ldots,X_n$ независимы и имеют равномерное распределение на отрезке $[0,\theta]$. Доказать, что $T(X){=}\max\limits_{1\le i\le n}X_i$ - достаточная и полная статистика.

Решение. Проверим достаточности статистики $T(X)$. Функция правдоподобия выборки $X{=}(X_1,\ldots,X_n)$ имеет следующий вид:

$$L(X;\theta)=\prod_{k=1}^{n}\frac{1}{\theta}\mathbb{I}_{\left\{0\le X_k\le\theta\right\}}=\frac{1}{\theta^n}\mathbb{I}_{\left\{0\le X_{(1)}\right\}}\mathbb{I}_{\left\{X_{(n)}\le\theta\right\}}=g\big(T(X),\theta\big)\times h(X),\quad\text{где}$$

$$g\big(T(X),\theta\big)=\frac{1}{\theta^n}\mathbb{I}_{\left\{X_{(n)}\le\theta\right\}},\quad h(X)=\mathbb{I}_{\left\{0\le X_{(1)}\right\}}.$$

Выполнен критерий факторизации, значит, статистика $T(X){=}X_{(n)}$ – достаточная.

Пусть непрерывная функция $\varphi(\cdot)$ такова, что $\mathbb{E}_\theta\varphi\big(T(X)\big)\equiv 0 \quad \forall\theta\in(0,+\infty)$.

$$f_{X_{(n)}}(y)=\frac{d}{dy}F_{X_{(n)}}(y)=\begin{cases}\dfrac{ny^{n-1}}{\theta^n}, & y\in[0;\theta]\\ 0, & \text{иначе.}\end{cases}$$

$$\mathbb{E}_\theta\varphi(X_{(n)})=\int_0^\theta \varphi(y)\frac{ny^{n-1}}{\theta^n}\,dy\equiv 0 \quad\Longrightarrow\quad \int_0^\theta \varphi(y)y^{n-1}\,dy\equiv 0.$$

Дифференцируем по θ:

$$\varphi(\theta)\cdot\theta^{n-1}\equiv 0,$$

откуда следует

$$\varphi(\theta)\equiv 0 \text{ на } (0,+\infty).$$

Итак, из $\mathbb{E}_\theta\varphi\big(T(X)\big)\equiv 0$ следует $\varphi\equiv 0$ по распределению $T(X)$, что означает полноту статистики $T(X)=\max\limits_{1\le i\le n} X_i$. □

Задача №8. Пусть X имеет биномиальное распределение $Bi(n,\frac{1}{2})$. Найти оценку максимального правдоподобия для n.

Решение. Функция правдоподобия случайной величины X

$$L(x,\theta)=C_n^x\frac{1}{2^n}.$$

Найдем точки, в которых функция правдоподобия достигает своего максимума. Обозначим

$$a_n=C_n^x\frac{1}{2^n}.$$

Исследуем последовательность $\{a_n\}$ на монотонность.

$$\frac{a_{n+1}}{a_n}=\frac{1}{2}\cdot\frac{C_{n+1}^x}{C_n^x}=\frac{1}{2}\cdot\frac{(n+1)!\,x!\,(n-x)!}{k!\,(n+1-x)!\,n!}=\frac{n+1}{2(n-x+1)}\vee 1 \quad\Longleftrightarrow\quad 2x-1\vee n.$$

Получаем, $a_1<\ldots<a_{2x-2}<a_{2x-1}=a_{2x}>a_{2x+1}>\ldots>a_n$. Функция правдоподобия достигает максимума в точках $2x-1$ и $2x$, которые и будут оценками максимального правдоподобия параметра n. □

Задача №9. Пусть $X_1,\ldots,X_n$ независимы и $X_i=\begin{cases}1, & \theta,\\ 2, & \theta,\\ 3, & 1-2\theta.\end{cases}$ Найти одномерную достаточную статистику.

Решение. Функцию правдоподобия случайной величины X_1 можно записать, например, так[2]:

$$L_1(x;\theta)=P(X_1=x)=\theta^{\frac{(x-2)(x-3)}{2}}\theta^{\frac{(x-1)(x-3)}{-1}}(1-2\theta)^{\frac{(x-1)(x-2)}{2}}=\theta^{-\frac{x^2}{2}+\frac{3x}{2}}(1-2\theta)^{\frac{x^2}{2}-\frac{3x}{2}+1}.$$

[2]Разумеется, можно ее представить и в другом виде, важно лишь, чтобы в точках $1,2$ и 3 эта функция принимала значения θ,θ и $1-2\theta$ соответственно.

Функция правдоподобия выборки $X{=}(X_1,\dots,X_n)$:

$$L_n(X,\theta)=\theta^{-\sum\limits_{k=1}^{n}\frac{X_k^2-3X_k}{2}}\times(1{-}2\theta)^{\sum\limits_{k=1}^{n}\left(\frac{X_k^2-3X_k}{2}+1\right)}=\left(\frac{1{-}2\theta}{\theta}\right)^{\frac{1}{2}\sum\limits_{k=1}^{n}(X_k^2-3X_k)}\times(1{-}2\theta)^n=$$

$$=g\big(T(X),\theta\big)\times h(X),\qquad \text{где } T(X){=}\sum_{k=1}^{n}(X_k^2{-}3X_k),\quad h(X){=}1.$$

Для функции $T(X)$ выполнен критерий факторизации, значит, она и будет достаточной статистикой. $\square$

Задача №10. Пусть $X_1,\dots,X_n$ независимы и имеют равномерное распределение на отрезке $[\theta,\theta{+}1]$. Найти несмещенную оценку максимального правдоподобия для θ.

Решение. Запишем функцию правдоподобия:

$$L(X,\theta)=\mathbb{I}_{\left\{\theta\le X_{(1)}\le X_{(n)}\le\theta+1\right\}}=\mathbb{I}_{\left\{X_{(n)}-1\le\theta\le X_{(1)}\right\}}.$$

Оценка максимального правдоподобия для θ заключена на сегменте $[X_{(n)}{-}1,X_{(1)}]$. Для $x\in[\theta,\theta{+}1]$

$$F_{X_{(1)}}(x)=1{-}P\big(X_{(1)}\ge x\big)=1-(\theta{+}1{-}x)^n,\qquad F_{X_{(n)}}(x)=P\big(X_{(n)}\le x\big)=(x{-}\theta)^n,$$

$$f_{X_{(1)}}(x)=\frac{d}{dx}F_{X_{(1)}}(x)=n(\theta{+}1{-}x)^{n-1},\qquad f_{X_{(n)}}(x)=\frac{d}{dx}F_{X_{(n)}}(x)=n(x{-}\theta)^{n-1},$$

значит, математические ожидания случайных величин $X_{(1)}$ и $X_{(n)}$

$$\mathbb{E}X_{(1)}=\int_\theta^{\theta+1}x\cdot f_{X_{(1)}}(x)\,dx=\theta{+}\frac{1}{n{+}1},\qquad \mathbb{E}X_{(n)}=\int_\theta^{\theta+1}x\cdot f_{X_{(n)}}(x)\,dx=\theta{+}\frac{n}{n{+}1},$$

Следовательно, $\mathbb{E}\left(\dfrac{X_{(1)}{+}X_{(n)}{-}1}{2}\right){=}\theta$.

Наконец, $\dfrac{X_{(1)}{+}X_{(n)}-1}{2}\in\left[X_{(n)}{-}1,X_{(1)}\right]$, поэтому функция $T(X){=}\dfrac{X_{(1)}{+}X_{(n)}{-}1}{2}$ является несмещенной оценкой максимального правдоподобия для θ. $\square$

Задача №11. Пусть $X_1,\dots,X_n$ независимы и имеют гамма-распределение $\Gamma(\theta,2)$. Исследовать на несмещенность и состоятельность оценку $T(X)=\overline{X}$ для функции $\tau(\theta)={}^2\!/_\theta$.

Решение. $\mathbb{E}T(X)=\mathbb{E}X_1={}^2\!/_\theta=\tau(\theta)$. $T(X)$ – несмещенная оценка для $\tau(\theta)$.

Вычислим дисперсию оценки: $\mathbb{D}\overline{X}=\dfrac{1}{n}\,\mathbb{D}X_1=\dfrac{2}{n\theta^2}$.

Воспользовавшись неравенством Чебышева, получим

$$\forall\varepsilon>0\quad P\big(\left|\overline{X}-{}^2\!/_\theta\right|{<}\varepsilon\big)\ge 1{-}\frac{\mathbb{D}\overline{X}}{\varepsilon^2}=1{-}\frac{1}{\varepsilon^2}\cdot\frac{2}{n\theta^2}\xrightarrow[n\to\infty]{}1,$$

то есть $T(X)\xrightarrow{P}{}^2\!/_\theta$, что по определению доказывает состоятельность оценки $T(X)$. $\square$

Задача №12. Пусть $X_1, \ldots, X_n$ независимы и $X_i = \begin{cases} 1, & \theta_1, \\ 2, & \theta_2, \\ 3, & 1{-}\theta_1{-}\theta_2. \end{cases}$ Найти двумерную достаточную статистику.

Решение. Функцию правдоподобия случайной величины X_1:

$$L_1(x;\theta_1,\theta_2) = P(X_1{=}x) = \theta_1^{\frac{(x-2)(x-3)}{2}}\theta_2^{\frac{(x-1)(x-3)}{-1}}(1{-}\theta_1{-}\theta_2)^{\frac{(x-1)(x-2)}{2}} =$$

$$= \theta_1^{\frac{x^2}{2}-\frac{5x}{2}+3}\theta_2^{-x^2+4x+3}(1{-}\theta_1{-}\theta_2)^{\frac{x^2}{2}-\frac{3x}{2}+1}.$$

Функция правдоподобия выборки $X{=}(X_1,\ldots,X_n)$:

$$L_n(X;\theta_1,\theta_2) = \theta_1^{\sum\limits_{k=1}^{n}\frac{X_k^2}{2}-\frac{5X_k}{2}+3}\theta_2^{\sum\limits_{k=1}^{n}-X_k^2+4X_k+3}(1{-}\theta_1{-}\theta_2)^{\sum\limits_{k=1}^{n}\frac{X_k^2}{2}-\frac{3X_k}{2}+1} =$$

$$= g\big(T(X),\theta_1,\theta_2\big){\times}h(X),$$

где $T(X){=}\big(\sum_{k=1}^{n} X_k^2,\ \sum_{k=1}^{n} X_k\big)$, $\quad h(X){=}1.$

Для функции $T(X)$ выполнен критерий факторизации, значит, она и будет достаточной статистикой. □

Задача №13. Пусть $X_1, \ldots, X_n$ независимы и имеют гамма-распределение $\Gamma(\theta,\lambda)$, λ - известно. Найти оценку максимального правдоподобия для θ.

Решение.

$$L(X,\theta) = \prod_{k=1}^{n}\frac{\theta^\lambda}{\Gamma(\lambda)}\exp(-\theta X_k)\cdot X_k^{\lambda-1} = \frac{\theta^{n\lambda}}{(\Gamma(\lambda))^n}\exp\Big(-\theta\sum_{k=1}^{n}X_k\Big)\cdot\Big(\prod_{k=1}^{n}X_k\Big)^{\lambda-1},$$

$$\frac{\partial}{\partial\theta}\ln L(X,\theta) = \frac{\partial}{\partial\theta}\Big(n\lambda\ln\theta - n\ln\Gamma(\lambda) - \theta\sum_{k=1}^{n}X_k + (\lambda{-}1)\sum_{k=1}^{n}\ln X_k\Big) = \frac{n\lambda}{\theta} - \sum_{k=1}^{n}X_k.$$

В точке экстремума функции правдоподобия

$$\frac{\partial}{\partial\theta}\ln L(X,\widehat{\theta}) = 0 \quad\Longleftrightarrow\quad \widehat{\theta} = \frac{\lambda}{\overline{X}}.$$

В точке $\widehat{\theta}$ достигается максимум, поскольку

$$\frac{\partial^2}{\partial\theta^2}\ln L(X,\theta) = -\frac{n\lambda}{\theta^2} < 0.$$

Значит, $\widehat{\theta} = \lambda\big/\overline{X}$ – оценка максимального правдоподобия для θ. □

Задача №14. Пусть $X_1, \ldots, X_n$ независимы и имеют нормальное распределение $\mathcal{N}(0,\theta^2)$. Доказать, что $T(X) = \dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2$ – эффективная оценка функции $\tau(\theta){=}\theta^2$.

Решение. В неравенстве Рао – Крамера (22) равенство достигается, если и только если найдется фукция $a_n(\theta)$ такая, что

$$T(X) - \tau(\theta) = a_n(\theta)\times\frac{\partial}{\partial\theta}\ln L(X,\theta). \qquad (*)$$

Проверим это условие.

$$L(X,\theta) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi}\cdot\theta} \exp\Bigl(-\frac{X_k^2}{2\theta^2}\Bigr) = \frac{1}{(\sqrt{2\pi}\cdot\theta)^n} \exp\Bigl(-\frac{1}{2\theta^2}\sum_{k=1}^{n} X_k^2\Bigr)$$

$$\frac{\partial}{\partial\theta} \ln L(X,\theta) = -\frac{n}{\theta} + \frac{1}{\theta^3}\sum_{k=1}^{n} X_k^2 = \frac{n}{\theta^3}\Bigl(\sum_{k=1}^{n} X_k^2 - \theta^2\Bigr).$$

Полагая $a_n(\theta) = {}^{\theta^3}\!/_{n}$, получим равенство ($*$). Эффективность оценки доказана. □

Задача №15. Пусть $X_1,\ldots,X_n$ независимы и имеют гамма-распределение $\Gamma(\theta,\lambda)$. Найти оценку методом моментов для θ и λ по первым двум моментам.

Решение. Аналогично задаче №5, из системы уравнений

$$\begin{cases} \overline{X} = M_1 = \mu_1 = \mathbb{E}X = \dfrac{\lambda}{\theta}, \\ \overline{X^2} = M_2 = \mu_2 = \mathbb{E}X^2 = \dfrac{\lambda(\lambda+1)}{\theta^2}; \end{cases}$$

получим оценку методом моментов для θ и λ:

$$\theta = \frac{M_1}{M_2 - M_1^2}, \qquad \lambda = \frac{M_1^2}{M_2 - M_1^2}.$$ □

Задача № 16. Пусть $X_1,\ldots,X_n$ независимы и имеют биномиальное распределение $Bi(1,\theta)$. Требуется исследовать несмещенность и состоятельность оценки $T(X){=}\overline{X}$ для параметра θ.

Решение. $\mathbb{E}T(X){=}\mathbb{E}X_1{=}\,\theta$ – оценка является несмещенной.

Вычислим дисперсию оценки: $\mathbb{D}\overline{X}{=}\dfrac{1}{n}\mathbb{D}X_1{=}\dfrac{\theta(1{-}\theta)}{n}$.

Воспользовавшись неравенством Чебышева, получим

$$\forall\varepsilon>0 \qquad P\bigl(\bigl|\overline{X}-\theta\bigr|{<}\varepsilon\bigr) > 1-\frac{\mathbb{D}\overline{X}}{\varepsilon^2} = 1-\frac{\theta(1{-}\theta)}{n\varepsilon} \xrightarrow[n\to\infty]{} 1,$$

то есть $T(X){=}\overline{X} \xrightarrow{P} \theta$, что по определению доказывает состоятельность оценки $T(X)$. □

Задача №17. Пусть $X_1,\ldots,X_n$ независимы и имеют нормальное распределение $\mathcal{N}(\theta,1)$. Доказать, что $\overline{X}{-}\dfrac{1}{n}$ является оптимальной оценкой функции $\tau(\theta){=}\theta^2$.

Решение. Функция правдоподобия

$$L(X,\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(X_i{-}\theta)^2}{2}\right) = K(\theta)\exp\bigl(n\theta\overline{X}\bigr)h(X),$$

$$\text{где}\quad K(\theta){=}\exp\Bigl(-\frac{n\theta^2}{2}\Bigr),\quad h(X){=}\frac{1}{\bigl(\sqrt{2\pi}\bigr)^n}\exp\Bigl(-\frac{1}{2}\sum_{i=1}^{n} X_i^2\Bigr).$$

По теореме о полноте экспоненциальных семейств $T(X)=\overline{X}$ – полная достаточная статистика. Любая измеримая функция от полной достаточной статистики $\overline{X}$ является оптимальной оценкой своего математического ожидания.

В частности, $\overline{X}^2-{}^1\!/_n$ – оптимальная оценка для

$$\mathbb{E}\Big(\overline{X}^2-\frac{1}{n}\Big)=\mathbb{E}\,\overline{X}^2-\frac{1}{n}=\mathbb{D}\overline{X}+\big(\mathbb{E}\,\overline{X}\big)^2-\frac{1}{n}=\frac{1}{n}\mathbb{D}X_1+\theta^2-\frac{1}{n}=\theta^2,$$

что и требовалось доказать. □

Задача №18. Пусть $X_1,\ldots,X_n$ независимы и распределены с плотностью

$$f(x,\theta)=\begin{cases}\exp\{-(x-\theta)\}, & x>\theta,\\ 0, & x\le\theta.\end{cases}$$

Найти оценку максимального правдоподобия для θ.

Решение. Функция правдоподобия:

$$L(X,\theta)=\prod_{i=1}^{n}\Big(\exp\big(-(X_i-\theta)\big)\,\mathbb{I}_{\{X_i\ge\theta\}}\Big)=\exp\Big(n\theta-\sum_{i=1}^{n}X_i\Big)\mathbb{I}_{\{X_{(1)}\ge\theta\}}$$

Данная функция достигает максимума в точке $\widehat{\theta}=X_{(1)}$, которая и будет оценкой максимального правдоподобия для θ. □

Задача №19. Пусть $X_1,\ldots,X_n$ независимы и имеют гамма-распределение $\Gamma\big({}^1\!/_\theta,1\big)$. Доказать, что $T(X)=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$ является эффективной оценкой θ.

Решение. Для доказательства эффективности оценки снова воспользуемся критерием эффективности, а именно, покажем, что

$$T(X)-\tau(\theta)=a_n(\theta)\frac{\partial}{\partial\theta}\ln L(X,\theta). \qquad (*)$$

Функция правдоподобия

$$L(X,\theta)=\prod_{i=1}^{n}\frac{1}{\theta}\exp\Big(-\frac{X_i}{\theta}\Big),$$

$$\frac{\partial}{\partial\theta}\ln L(X,\theta)=-\frac{n}{\theta}+\frac{\sum_{i=1}^{n}X_i}{\theta}=\frac{n}{\theta^2}\big(\overline{X}-\theta\big).$$

Полагая $a_n(\theta)={}^{\theta^2}\!/_n$, получим равенство ($*$), значит, $T(X)$ – эффективная оценка θ. □

Задача №20. Пусть $X_1,\ldots,X_n$ независимы и имеют нормальное распределение $\mathcal{N}(\theta,2\theta)$. Найти оценку максимального правдоподобия для θ.

Решение. Функция правдоподобия

$$L(X;\theta)=\prod_{i=1}^{n}\frac{1}{\sqrt{4\pi\theta}}\exp\Big\{-\frac{(X_i-\theta)^2}{4\theta}\Big\};$$

$$\frac{\partial}{\partial\theta}\ln L(X,\theta)=-\frac{n}{2\theta}+\frac{1}{4\theta^2}\sum_{i=1}^{n}X_i^2-\frac{n}{4}.$$

В точке экстремума функции правдоподобия

$$\frac{\partial}{\partial\theta}\ln L(X,\theta)=0 \quad\iff\quad \theta^2+2\theta-\overline{X^2}=0.$$

В силу неотрицательности дисперсии $\theta>0$, значит, возможная точка максимума

$$\widehat{\theta}=-1+\sqrt{1+\overline{X^2}}.$$

В этой точке действительно достигается максимум функции правдоподобия, поскольку

$$\frac{\partial^2}{\partial\theta^2}\ln L(X,\theta)=\frac{n}{2\theta^3}\big(\theta-\overline{X^2}\big)\Big|_{\theta=\widehat{\theta}}=\frac{n}{2\theta^3}\Big(\sqrt{1+\overline{X^2}}-\big(1+\overline{X^2}\big)\Big)<0. \qquad \square$$

Задача №21. Пусть случайные величины $X_1,\ldots,X_n$ независимы и имеют гамма - распределение $\Gamma\big({}^1\!/_{\theta},1\big)$. Доказать, что $T(X)=\frac{n}{n+1}\overline{X}^2$ является оптимальной оценкой θ^2.

Решение. Функция правдоподобия

$$L(X;\theta)=\frac{1}{\theta^n}\exp\Big(-\frac{n}{\theta}\overline{X}\Big).$$

По теореме о полноте экспоненциальных семейств $\overline{X}$ – полная и достаточная статистика, поэтому любая измеримая функция от $\overline{X}$ является оптимальной оценкой своего математического ожидания. В частности, $T(X)=\dfrac{n}{n+1}\overline{X}^2$ – оптимальная оценка для

$$\mathbb{E}T(X)=\frac{n}{n+1}\mathbb{E}\overline{X}^2=\frac{n}{n+1}\Big(\mathbb{D}\overline{X}+\big(\mathbb{E}\overline{X}\big)^2\Big)=\frac{n}{n+1}\Big(\frac{1}{n}\mathbb{D}X_1+\big(\mathbb{E}X_1\big)^2\Big)=$$

$$=\frac{n}{n+1}\Big(\frac{\theta^2}{n}+\theta^2\Big)=\theta^2=\tau(\theta),$$

что и требовалось доказать. $\square$

Задача №22. Пусть случайные величины $X_1,\ldots,X_n$ независимы и распределены с плотностью $\exp\big\{-(x-\theta)-\exp[-(x-\theta)]\big\}$. Найти оценку максимального правдоподобия для θ.

Решение. Функция правдоподобия:

$$L(X,\theta)=\exp\Big\{n\theta-\sum_{i=1}^{n}X_i-\sum_{i=1}^{n}\exp(\theta-X_i)\Big\}.$$

В точках экстремума

$$\frac{\partial}{\partial\theta}\ln L(X,\theta)=n-\sum_{i=1}^{n}\exp(\theta-X_i)=0,$$

откуда

$$\widehat{\theta}=\ln n-\ln\sum_{i=1}^{n}\exp(-X_i).$$

Полученная оценка действительно является оценкой максимального правдоподобия, так как

$$\frac{\partial^2}{\partial \theta^2} \ln L(X,\theta) = -\sum_{i=1}^{n} \exp(\theta - X_i) < 0. \qquad \square$$

Задача №23. Пусть случайные величины $X_1, \ldots, X_n$ независимы и имеют пуассоновское распределение $\Pi(\theta)$, $\theta > 0$. Исследовать несмещенность и состоятельность оценки $T(X){=}\overline{X}$ параметра θ.

Решение. $\mathbb{E}T(X) = \mathbb{E}X_1 = \theta$. Несмещенность оценки доказана.

Дисперсия оценки $\mathbb{D}T(X) = \dfrac{1}{n}\mathbb{D}X_1 = \dfrac{\theta}{n} \xrightarrow[n\to\infty]{} 0$.

Воспользовавшись неравенством Чебышева, получим

$$\forall \varepsilon > 0 \quad P\big(\big|\overline{X} - \mathbb{E}\overline{X}\big| {<} \varepsilon\big) > 1 - \frac{\mathbb{D}\overline{X}}{\varepsilon^2} \xrightarrow[n\to\infty]{} 1,$$

то есть $T(X) \xrightarrow{P} \theta$, что по определению доказывает состоятельность оценки $T(X)$. $\square$

Задача №24. Пусть случайные величины $X_1, \ldots, X_n$ независимы и имеют биномиальное распределение $Bi(1,\theta)$, $0{<}\theta{<}1$. Доказать, что $T(X) = \dfrac{\overline{X}(1{-}\overline{X}) \cdot n}{n-1}$ является оптимальной оценкой $\tau(\theta) = \theta(1{-}\theta)$.

Решение. Функция правдоподобия выглядит следующим образом:

$$L(X,\theta) = \theta^{\sum\limits_{i=1}^{n} X_i} (1{-}\theta)^{n - \sum\limits_{i=1}^{n} X_i} = (1{-}\theta)^n \exp\left\{ n \ln \frac{\theta}{1{-}\theta} \overline{X} \right\}$$

По теореме о полноте экспоненциальных семейств $\overline{X}$ является полной и достаточной статистикой. Значит $\varphi(\overline{X})$ является оптимальной оценкой для $\mathbb{E}\varphi(\overline{X})$, где $\varphi(\cdot)$ – любая измеримая функция от $\overline{X}$.

$$\mathbb{E}T(X) = \frac{n}{n{-}1}\left(\mathbb{E}\overline{X} - \mathbb{E}\overline{X}^2\right) = \frac{n}{n{-}1}\left(\theta - \left(\mathbb{D}\overline{X} + \left(\mathbb{E}\overline{X}\right)^2\right)\right) =$$
$$= \frac{n}{n{-}1}\left(\theta - \frac{1}{n}\theta(1{-}\theta) - \theta^2\right) = \theta(1{-}\theta),$$

поэтому $T(X)$ является оптимальной оценкой для $\tau(\theta) = \theta(1{-}\theta)$. $\square$

Задача №25. Пусть случайные величины $X_1, \ldots, X_n$ независимы и имеют гамма-распределение $\Gamma(\theta, 2)$. Исследовать на несмещенность и состоятельность оценку $T(X) = \overline{X}$ для функции $\tau(\theta) = {}^2\!/_\theta$.

Решение. $\mathbb{E}T(X) = \mathbb{E}X_1 = {}^2\!/_\theta$. Несмещенность доказана.

Вычислим дисперсию оценки: $\mathbb{D}\overline{X} = \dfrac{1}{n}\mathbb{D}X_1 = {}^2\!/_{n\theta} \xrightarrow[n\to\infty]{} 0$.

Воспользовавшись неравенством Чебышева, получим

$$\forall \varepsilon > 0 \quad P\big(\big|\overline{X} - \mathbb{E}\overline{X}\big| {<} \varepsilon\big) > 1 - \frac{\mathbb{D}\overline{X}}{\varepsilon^2} \xrightarrow[n\to\infty]{} 1,$$

то есть $T(X) \xrightarrow{P} \theta$, что по определению доказывает состоятельность оценки $T(X)$. $\square$

Задача №26. Пусть $X_1, \ldots, X_n$ независимы и имеют биномиальное распределение $Bi(1,\theta)$. Доказать, что не существует оптимальной оценки для $\tau(\theta){=}\theta^{n+1}$.

Решение. Достаточно доказать, что не существует несмещенной оценки для $\tau(\theta)$. Предположим обратное: пусть найдется функция $T(X)$ такая, что

$$\mathbb{E}T(X) = \theta^{n+1}. \qquad (*)$$

Однако

$$\mathbb{E}T(X_1, \ldots, X_n) = \\ = T(0, \ldots, 0) \cdot (1{-}\theta)^n + T(1, 0, \ldots, 0) \cdot \theta(1{-}\theta)^{n-1} + \ldots + T(1, \ldots, 1) \cdot \theta^n$$

– многочлен степени не выше n, поэтому равенство $(*)$ невозможно.
Таким образом, не существует оптимальной оценки для $\tau(\theta){=}\theta^{n+1}$. $\square$

Задача №27. Пусть $X_1, \ldots, X_n$ независимы и имеют биномиальное распределение $Bi(1,\theta)$. Доказать, что не существует оптимальной оценки для $\tau(\theta) = \theta^{-1}$.

Решение. Решение аналогично решению задачи №26. $\square$

Задача № 28. Пусть $X_1, \ldots, X_n$ независимы и имеют пуассоновское распределение $\Pi(\theta)$. Доказать, что не существует оптимальной оценки для θ^{-2}.

Решение. Как и в задаче №26, доказав отсутствие несмещенной оценки, мы докажем отстутствие оптимальной.
Пусть существует функция $T(X) = T(X_1, \ldots, X_n)$ такая, что

$$\mathbb{E}T(X) = \theta^{-2} \qquad (*)$$

Её математическое ожидание

$$\mathbb{E}T(X_1, \ldots, X_n) = \sum_{\substack{(i_1,\ldots,i_n) \\ i_k \geq 0,\ k=0..n}} T(i_1, \ldots, i_n) \exp(-n\theta) \frac{\theta^{i_1+\ldots+i_n}}{i_1! \cdot \ldots \cdot i_n!} = \sum_{i=0}^{\infty} \exp(-n\theta) a_i \theta^i.$$

В силу несмещенности оценки

$$\mathbb{E}T(X) = \sum_{i=0}^{\infty} \exp(-n\theta) a_i \theta^i = \theta^{-2}.$$

Устремим θ к нулю:

$$\lim_{\theta \to +0} \mathbb{E}T(X) = a_0, \qquad \lim_{\theta \to +0} \theta^{-2} = +\infty,$$

Значит, равенство $(*)$ не выполнено ни для какой функции $T(X)$. $\square$

Задача №29. Пусть $X_1,\dots,X_n$ независимы и имеют равномерное распределение на отрезке $[a;b]$. Доказать, что $T(X){=}\dfrac{X_{(1)}{+}X_{(n)}}{2}$ является несмещенной и состоятельной оценкой функции $\tau(a,b)=\dfrac{a+b}{2}$.

Решение. Для доказательства несмещенности необходимо, чтобы математическое ожидание оценки было равно оцениваемой величине. Для вычисления матожидания необходимо знать плотность случайной величины, которую можно посчитать из функции распределения.

$$F_{X_{(1)}}(y)=1-\left(1-\frac{y-a}{b-a}\right)^n, \qquad F_{X_{(n)}}(y)=\left(\frac{y-a}{b-a}\right)^n,$$
$$f_{X_{(1)}}(y)=\frac{n(y-a)^{n-1}}{(b-a)^n}, \qquad f_{X_{(n)}}(y)=\frac{n(y-a)^{n-1}}{(b-a)^n},$$
$$\mathbb{E}X_{(1)}=\int_a^b y\cdot f_{X_{(1)}}y\,dy=\frac{na}{n{+}1}+\frac{b}{n{+}1}, \qquad \mathbb{E}X_{(n)}=\int_a^b y\cdot f_{X_{(n)}}(y)\,dy=\frac{nb}{n{+}1}+\frac{a}{n{+}1}.$$

Отсюда
$$\mathbb{E}T(X)=\frac{1}{2}(\mathbb{E}X_{(1)}{+}\mathbb{E}X_{(n)})=\frac{a{+}b}{2}.$$
Несмещенность оценки доказана.
Для произвольно малого $\varepsilon>0$
$$P\big(X_{(n)}>b{-}\varepsilon\big)=1-P\big(X_{(n)}\le b{-}\varepsilon\big)=1-\left(\frac{b{-}\varepsilon{-}a}{b{-}a}\right)^n\xrightarrow[n\to\infty]{}1, \quad \text{то есть } X_{(n)}\xrightarrow{P}b.$$

Аналогично,
$$X_{(1)}\xrightarrow{P}a.$$

Значит,
$$T(X){=}\frac{X_{(1)}{+}X_{(n)}}{2}\xrightarrow{P}\frac{a{+}b}{2},$$
что и означает состоятельность оценки. □

Задача №30. Пусть $X_1,\dots,X_n$ независимы и имеют равномерное распределение на отрезке $[a;b]$. Доказать, что $T(X){=}\dfrac{n{+}1}{n{-}1}\cdot\big(X_{(n)}{-}X_{(1)}\big)$ является несмещенной и состоятельной оценкой функции $\tau(a,b){=}b{-}a$.

Решение. Из предыдущей задачи:
$$\mathbb{E}X_{(1)}=\frac{na}{n{+}1}+\frac{b}{n{+}1}, \qquad \mathbb{E}X_{(n)}=\frac{nb}{n{+}1}+\frac{a}{n{+}1},$$
следовательно,
$$\mathbb{E}T(X)=b-a.$$
Несмещенность доказана.
$$X_{(1)}\xrightarrow{P}a, \qquad X_{(n)}\xrightarrow{P}b,$$

откуда

$$T(X)=\frac{n+1}{n-1}\cdot\left(X_{(n)}-X_{(1)}\right)\xrightarrow{P} b-a.$$

Состоятельность доказана. $\square$

Глава 2

Доверительные интервалы

Определения и теоремы

1. **Определения**.
Простая точечная оценка $\hat{\theta}$ параметра θ является лучшим предположением о значении θ, но ничего не сообщает об уверенности в правильности оценки. *Доверительный интервал* I, основанный на $\hat{\theta}$, используется для предположений о значении θ при известном размере выборке, распределении и *коэффициенте доверия* $1-\alpha$. Предположения имеют вид:

Вероятность того, что θ лежит в указанном интервале, равна $1-\alpha$.

Пусть доверительный интервал $I = \big(T_1(X), T_2(X)\big)$. Существует много способов задать $T_1(X)$ и $T_2(X)$ в зависимости от параметра θ и распределения случайной величины. Обычно границы доверительного интервала выбирают из следующих соображений:

$$\mathsf{P}\big(\theta < T_1(X)\big) = {}^{\alpha}\!/_2 \qquad \mathsf{P}\big(\theta > T_2(X)\big) = {}^{\alpha}\!/_2,$$

с тем, чтобы $\mathsf{P}\big(T_1(X) \le \theta \le T_2(X)\big) = 1-\alpha$.

Можно строить также односторонние доверительные интервалы, для коорых
(1) $T_1(X) = -\infty$ и $\mathsf{P}\big(\theta > T_2(X)\big) = \alpha$ (или $\mathsf{P}\big(\theta \le T_2(X)\big) = 1-\alpha$).
(2) $T_2(X) = +\infty$ и $\mathsf{P}\big(\theta < T_1(X)\big) = \alpha$ (или $\mathsf{P}\big(\theta \ge T_1(X)\big) = 1-\alpha$).

2. **Некоторые критические значения**.
Формулы для часто используемых доверительных интервалов обычно содержат критические значения для нормального распределения, t-распределения (распределения Стьюдента) или распределения χ^2 (см. таблицы 3 и 4). В таблице 1 собраны часто используемые для построения до верительных интервалов критические значения.

3. **Вычисление объема выборки**.
Для построения доверительного интервала заданной длины следует определить необходимый объем выборки, используя априорные параметры оценки и ограничения на погрешность оценки. Для доверительного интервала с уровнем доверия $1-\alpha$ положим $E =$ погрешность оценки (половина длины доверительного интервала). В таблице 2 представлены часто используемые формулы для вычисления объема выборки.
Пример. Исследователю нужно оценить вероятность успеха p в биномиальном эксперименте. Каков должен быть размер выборки (то есть сколько экспериментов нужно провести), чтобы оценить это величину с точностью 0.05 и уровнем доверия 0.99, т. е. найти n такое, что $\mathsf{P}\big(|p-\hat{p}| \le 0.05\big) \ge 0.99$.
Решение.
(1) Поскольку не задана никакая априорная оценка p, положим $p = 0,5$. Ограничение на погрешность оценки $E = 0.05$, $1-\alpha = 0.99$.
(2) Из таблицы 2, $n = \dfrac{z_{0.05}^2 \cdot pq}{E^2} = \dfrac{(2.5758)(0.5)(0.5)}{0.05^2} = 663.47$.
(3) Эта формула дает оценку размера выборки для наихудшего случая (поскольку неизвестна априорная оценка параметра p). Размер выборки должен быть не меньше, чем 664.

Распределение	α 0.10	0.05	0.01	0.001	0.0001
t-распределение					
$t_{\alpha/2,10}$	1.8125	2.2281	3.1693	4.5869	6.2111
$t_{\alpha/2,100}$	1.6602	1.9840	2.6259	3.3905	4.0533
$t_{\alpha/2,1000}$	1.6464	1.9623	2.5808	3.3003	3.9063
Нормальное распределение					
$z_{\alpha/2}$	1.6449	1.9600	2.5758	3.2905	3.8906
распределение χ^2					
$\chi^2_{1-\alpha/2,10}$	3.9403	3.2470	2.1559	1.2650	0.7660
$\chi^2_{\alpha/2,10}$	18.3070	20.4832	25.1882	31.4198	37.3107
$\chi^2_{1-\alpha/2,100}$	77.9295	74.2219	67.3276	65.8957	54.1129
$\chi^2_{\alpha/2,100}$	124.3421	129.5612	140.1695	153.1670	164.6591
$\chi^2_{1-\alpha/2,1000}$	927.5944	914.2572	888.5635	859.3615	835.3493
$\chi^2_{\alpha/2,1000}$	1074.6790	1089.5310	1118.9480	1153.7380	1183.4920
	0.90	0.95	0.99	0.999	0.9999
	$1-\alpha$				

Таблица 1: Критические значения

Параметр	Оценка	Размер выборки
μ	$\bar{x}$	$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$
p	$\hat{p}$	$n = \frac{(z_{\alpha/2})^2 \cdot pq}{E^2}$
$\mu_1 - \mu_2$	$\bar{x}_1 - \bar{x}_2$	$n_1 = n_2 = \frac{(z_{\alpha/2})^2(\sigma_1^2 + \sigma_2^2)^2}{E^2}$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$n_1 = n_2 = \frac{(z_{\alpha/2})^2(p_1 q_1 + p_2 q_2)}{E^2}$

Таблица 2: Вычисление объема выборки

4. **Часто встречающиеся доверительные интервалы**.
В таблице 3 представлен общий вид доверительных интервалов для одной выборки, в таблице 4 — для двух выборок. Для каждого параметра распределения даны формулы для вычисления доверительных интервалов с уровнем доверия $1-\alpha$.
Пример. Компания, разрабатывающая программное обеспечения, провела исследование среднего размера word processing файла. Для $n = 23$ произвольно выбранных файлов, $\bar{x} = 4822$ kb и $s = 127$. Определить доверительный интерва с уровнем доверия 0.95 для среднего размера word processing файлов.
Решение.
(1) Предполагается, что распределение размеров файлов — нормальное. Доверительный интервал для μ основан на t-распределении. Используем соответствующую формулу из таблицы 3.
(2) $1-\alpha = 0.95$; $\alpha = 0.05$; $\alpha/2 = 0.025$; $t_{\alpha/2,n-1} = t_{0.025,22} = 2.0739$.
(3) $k = \dfrac{2.0739 \cdot 127}{\sqrt{23}} = 54.92$.

Параметр	Предположение о распределении	Доверительный интервал с коэффициентом доверия $1-\alpha$
μ	n — большое, σ^2 — известно, или нормальное, σ^2 — известно	$\bar{x} \pm z_{\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$
μ	нормальное, σ^2 — неизвестно	$\bar{x} \pm t_{\alpha/2, n-1} \cdot \dfrac{s}{\sqrt{n}}$
σ^2	нормальное	$\left(\dfrac{(n-1)s^2}{\chi^2_{\alpha/2, n-1}}, \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2, n-1}}\right)$
p	биномиальное, n — большое	$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

Таблица 3: Часто встречающиеся доверительные интервалы: одна выборка

(4) Доверительный интервал с коэффициентом доверия 0.99 для μ — $(\bar{x}-k, \bar{x}+k) = (4767, 4877)$.

Параметр	Предположение о распределении	Доверительный интервал с коэффициентом доверия $1-\alpha$
$\mu_1 - \mu_2$	независимость, σ_1^2, σ_2^2 — известны; нормальное распределение или большое n	$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
$\mu_1 - \mu_2$	нормальность, независимость, $\sigma_1^2 = \sigma_2^2$ — неизвестны	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2, n_1+n_2-2} \cdot s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$
$\mu_1 - \mu_2$	нормальность, независимость, $\sigma_1^2 \neq \sigma_2^2$ — неизвестны	$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2, \nu} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $\nu \approx \dfrac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1-1} + \frac{\left(s_2^2/n_2\right)^2}{n_2-1}}$
$\mu_1 - \mu_2$	нормальность, n пар, зависимость	$\bar{d} \pm t_{\alpha/2, n-1} \cdot \frac{s_d}{\sqrt{n}}$
$\dfrac{\sigma_1^2}{\sigma_2^2}$	нормальность, независимость	$\left(\frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{\frac{\alpha}{2}, n_1-1, n_2-1}}, \frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{1-\frac{\alpha}{2}, n_1-1, n_2-1}}\right)$
$p_1 - p_2$	биномиальное распределение, n_1, n_2 — большие, независимость	$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

Таблица 4: Часто встречающиеся доверительные интервалы: две выборки

5. **Другие оценки**.
   1) **Доверительные интервалы для медиан**.
   Построить приближенный доверительный интервал с уровнем доверия $1-\alpha$ для медианы $\tilde{\mu}$, при больших n (основанный на порядковой статистике Wilcoxon'а).

(1) Построить порядковую статистику $\{w_{(1)}, w_{(2)}, \ldots, w_{(N)}\}$ для $N = \binom{n}{2} = \frac{n(n-1)}{2}$ средних $\frac{x_i+x_j}{2}$, для $1 \le i < j \le n$.
(2) Определить критическое значение $z_{\alpha/2}$ такое, что $\mathsf{P}(Z \ge z_{\alpha/2}) = \alpha/2$.
(3) Вычислить константы $k_1 = \left(\frac{N}{2} - \frac{z_{\alpha/2} N}{\sqrt{3n}}\right)$ и $k_2 = \left(\frac{N}{2} + \frac{z_{\alpha/2} N}{\sqrt{3n}}\right)$.
(4) Доверительный интервал с уровнем доверия $1-\alpha$ для медианы $\tilde{\mu}$ имеет вид $(w_{(k_1)}, w_{(k_2)})$ (см. таблицу 5).

	$\alpha = 0.05$		$\alpha = 0.01$	
n	k_1	k_2	k_1	k_2
7	1	20		
8	2	26		
9	4	32		
10	6	39	1	44
11	8	47	2	53
12	11	55	4	62
13	14	64	6	72
14	17	74	9	82
15	21	84	12	93
16	26	94	15	105
17	30	106	18	118
18	35	118	22	131
19	41	130	27	144
20	46	144	31	159

Таблица 5: Доверительные интервалы для медианы

2) **Разность медиан**.
Для построения доверительного интервала с уровнем доверия $1-\alpha$ для разности медиан $\tilde{\mu}_1 - \tilde{\mu}_2$ следующий алгоритм, основанный на процедуре Mann-Whitney-Wilcoxon'а. Предположим, размеры выборок досктаточно велики, и выборки независимы.
(1) Построить порядковую статистику $\{w_{(1)}, w_{(2)}, \ldots, w_{(N)}\}$ для $N = n_1 n_2$ разностей $x_i - y_j$, для $1 \le i \le n_1$, $1 \le j \le n_2$.
(2) Определить критическое значение $z_{\alpha/2}$ такое, что $\mathsf{P}(Z \ge z_{\alpha/2}) = \alpha/2$.
(3) Вычислить константы

$$k_1 = \left(\frac{n_1 n_2}{2} + 0.5 - z_{\alpha/2}\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}\right) \qquad \text{и}$$
$$k_2 = \left(\frac{n_1 n_2}{2} + 0.5 + z_{\alpha/2}\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}\right).$$

(4) Приближенный доверительный интервал с уровнем доверия $1-\alpha$ для $\tilde{\mu}_1 - \tilde{\mu}_2$ есть $(w_{(k_1)}, w_{(k_2)})$.

6. **Корректирующий множитель для конечных распределений**.
Пусть производится выборка без возвращения размера n из (конечного) распределения размера N. Если n — большая или существенная часть распределения, то интуитивно понятно, что точечная оценка, основанная на этой выборке, должна быть точнее, чем если бы распределение было бесконечным. В таких случаях, поэтому,

стандартное отклонение выборочного среднего и стандартное отклонение вероятности успеха в испытаниях Бернулли умножается на *корректирующий множитель для конечных распределений*:

$$\sqrt{\frac{N-n}{N-1}}$$

При посторении доверительного интервала для достижения большей точности оценки на эту функцию от n и N умножается *критическое расстояние*. Если размер выборки составляет *менее* 5% от всего распределения, корректирующий множитель для конечных распределений, как правило, не используется.

Доверительные интервалы, построенные с учетом корректирующего множителя:

(1) Пусть производится выборка без возвращения размера n из распределения размера N. Если предполагается нормальное распределение, граничные точки для доверительного интервала для среднего значения распределения μ выбираются следующим образом:

$$\bar{x} \pm z_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}.$$

(2) В случае биномиального распределения, граничные точки для доверительного интервала для вероятности успеха в единичном эксперименте p выбираются следующим образом:

$$\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \cdot \sqrt{\frac{N-n}{N-1}}.$$

Задачи

Задача №31. Пусть $X_1, \dots, X_n$ независимы и имеют равномерное распределение на отрезке $[0;\theta]$. Построить кратчайший доверительный интервал для θ с коэффициентом доверия α, основанный на центральной статистике $G(X,\theta) = \dfrac{\max\limits_{1\le i\le n} X_i}{\theta}$.

Решение. Разрешив неравенство $g_1 < G(X,\theta) < g_2$ относительно θ, получим доверительный интервал (θ_1, θ_2):

$$g_1 < \frac{X_{(n)}}{\theta} < g_2 \quad\Longleftrightarrow\quad \frac{X_{(n)}}{g_2} < \theta < \frac{X_{(n)}}{g_1}: \quad \theta_1 = \frac{X_{(n)}}{g_2},\ \theta_2 = \frac{X_{(n)}}{g_1}.$$

Длину его $X_{(n)}\times\Big(\dfrac{1}{g_1} - \dfrac{1}{g_2}\Big)$ нужно минимизировать при заданном уровне доверия

$$\alpha = \mathsf{P}(\theta_1 < \theta < \theta_2) = \mathsf{P}\Big(g_1 < \frac{X_{(n)}}{\theta} < g_2\Big) = \mathsf{P}\Big(\frac{X_{(n)}}{\theta} < g_2\Big) - \mathsf{P}\Big(\frac{X_{(n)}}{\theta} < g_1\Big) = g_2^n - g_1^n,$$
$$0 \le g_1 < g_2 \le 1.$$

Поскольку длина интервала должна быть минимальной, иные значения g_1 и g_2 рассматривать не имеет смысла.

Минимум выражения $X_{(n)}\times\Big(\dfrac{1}{g_1} - \dfrac{1}{g_2}\Big)$ при условии $g_2^n - g_1^n = \alpha,\ 0 \le g_1 < g_2 \le 1$ достигаетя на $g_2 = 1,\ g_1 = \sqrt[n]{1-\alpha}$.

Итак, кратчайший доверительный интервал, основанный на центральной статистике $G(X,\theta)$, имеет вид $\left(X_{(n)}, \dfrac{X_{(n)}}{\sqrt[n]{1-\alpha}}\right)$. □

Задача №32. Пусть $X_1, \dots, X_n$ независимы и имеют биномиальное распределение $Bi(1,\theta)$, $0<\theta<1$. Построить равномерено наиболее мощный критерий размера α для проверки гипотезы $H_0: \theta = \theta_0$ при альтернативе $H_1: \theta < \theta_0$. Найти функцию мощности.

Решение. Построим наиболее мощный критерий для проверки H_0 при простой альтернативе $H_1: \theta = \theta_1$, $\theta_1 < \theta_0$, используя лемму Неймана–Пирсона:

$$\frac{L_1}{L_0} = \frac{\theta_1^{\sum\limits_{i=1}^n X_i}(1-\theta_1)^{n-\sum\limits_{i=1}^n X_i}}{\theta_0^{\sum\limits_{i=1}^n X_i}(1-\theta_0)^{n-\sum\limits_{i=1}^n X_i}} = \left(\frac{1-\theta_1}{1-\theta_0}\right)^n \left(\frac{\frac{\theta_1}{1-\theta_1}}{\frac{\theta_0}{1-\theta_0}}\right)^{\sum\limits_{i=1}^n X_i} \ge c_\alpha \Longleftrightarrow T(X) = \sum_{i=1}^n X_i \le c'_\alpha.$$

Существует c'_α для которого выполняется следующее неравенство:

$$\alpha'' = \sum_{i=0}^{c'_\alpha - 1} C_n^i \theta_0^i (1-\theta_0)^{n-i} < \alpha \le \sum_{i=0}^{c'_\alpha - 1} C_n^i \theta_0^i (1-\theta_0)^{n-i} = \alpha'.$$

При $\alpha = \alpha'$ критическая функцмя имеет вид:

$$\varphi(X) = \begin{cases} 1, & T(X) \le c'_\alpha \\ 0, & T(X) > c'_\alpha \end{cases}.$$

В случае $\alpha < \alpha'$, критерий является рандомизированным и из

$$\alpha = \mathbb{E}_{\theta_0}\varphi(X) = \mathsf{P}_{\theta_0}\big(T(X) < c'_\alpha\big) + \varepsilon_\alpha \mathsf{P}_{\theta_0}\big(T(X) = c'_\alpha\big) \quad\Longrightarrow\quad \varepsilon_\alpha = \frac{\alpha - \alpha''}{\alpha' - \alpha''}$$

получаем

$$\varphi(X) = \begin{cases} 1, & T(X) < c'_\alpha, \\ \varepsilon_\alpha, & T(X) = c'_\alpha, \\ 0, & T(X) > c'_\alpha; \end{cases}$$

где

$$\varepsilon_\alpha = \frac{\alpha - \sum_{i=0}^{c'_\alpha-1} C_n^i \theta_0^i (1-\theta_0)^{n-i}}{C_n^{c'_\alpha} \theta_0^{c'_\alpha} (1-\theta_0)^{n-c'_\alpha}}.$$

Функция мощности

$$W(\theta) = \mathbb{E}_\theta \varphi(X) = \mathsf{P}_\theta\big(T(X){<}c'_\alpha\big) + \varepsilon_\alpha \mathsf{P}_\theta\big(T(X){=}c'_\alpha\big) = \mathsf{P}_\theta\big(T(X){<}c'_\alpha\big) + \varepsilon_\alpha \theta^{c'_\alpha}(1-\theta)^{n-c'_\alpha} =$$
$$= \sum_{i=0}^{c'_\alpha-1} C_n^i \theta^i (1-\theta)^{n-i} + \Big(\alpha - \sum_{i=0}^{c'_\alpha-1} C_n^i \theta_0^i (1-\theta_0)^{n-i}\Big) \frac{\theta^{c'_\alpha}(1-\theta)^{n-c'_\alpha}}{\theta_0^{c'_\alpha}(1-\theta_0)^{n-c'_\alpha}}.$$

Построенный критерий – наиболее мощный, если гипотеза H_1 – простая (то есть $H_1 : \theta{=}\theta_1$). При построении критерия значение θ_1 используется неявно: важно лишь, что $\theta_1 < \theta_0$. Значит, построенный критерий – равномерно наиболее мощный. □

Задача № 33. Пусть $X_1, \dots, X_n$ независимы и имеют нормальное распределение $\mathcal{N}(\theta, 1)$. Построить кратчайший доверительный интервал для θ с коэффициентом доверия α, основанный на центральной статистике $\sqrt{n}(\overline{X}{-}\theta)$.

Решение.

$$G(X,\theta) = \sqrt{n}\Big(\overline{X} - \theta\Big) = \frac{\sum_{i=1}^n X_i - n\theta}{\sqrt{n}} \sim \mathcal{N}(0,1).$$

$$\mathsf{P}\big(\tau_1 < G(X,\theta) < \tau_2\big) = \mathsf{P}\Big(\overline{X} - \frac{\tau_1}{\sqrt{n}} < \theta < \overline{X} - \frac{\tau_2}{\sqrt{n}}\Big).$$

Доверительный интервал имеет вид (θ_1, θ_2), где

$$\theta_1 = \overline{X} - \frac{\tau_1}{\sqrt{n}}, \qquad \theta_2 = \overline{X} - \frac{\tau_2}{\sqrt{n}}.$$

Длину его $\theta_2{-}\theta_1 = \dfrac{\tau_2{-}\tau_1}{\sqrt{n}}$ нужно минимизировать при условии

$$\alpha = \mathsf{P}\big(\tau_1 < G(X,\theta) < \tau_2\big) = \Phi(\tau_2) - \Phi(\tau_1),$$

где $\Phi(x)$ – функция распределения стандартного нормального закона.

Применив метод множителей Лагранжа, получим $\tau_2 = -\tau_1 = \tau_{\frac{1+\alpha}{2}}$, где $\tau_{\frac{1+\alpha}{2}}$ – квантиль порядка $\frac{1+\alpha}{2}$ функции $\Phi(x)$.

Итак, кратчайший доверительный интервал, основанный на центральной статистике $G(X,\theta)$, имеет вид $\Big(\overline{X} - \dfrac{\tau_{\frac{1+\alpha}{2}}}{\sqrt{n}}, \overline{X} + \dfrac{\tau_{\frac{1+\alpha}{2}}}{\sqrt{n}}\Big)$. □

Задача №34. Пусть $X_1, \dots, X_n$ независимы и имеют равномерное распределение на отрезке $[0;\theta]$. Построить наиболее мощный критерий размера α для проверки гипотезы $H_0 : \theta = \theta_0$ при альтернативе $H_1 : \theta = \theta_1 > \theta_0$. Найти мощность критерия.

Решение. Воспользуемся леммой Неймана-Пирсона:

$$\frac{L_1}{L_0} = \frac{\prod\limits_{i=1}^{n} \frac{1}{\theta_1} I_{\{0 \leqslant X_i \leqslant \theta_1\}}}{\prod\limits_{i=1}^{n} \frac{1}{\theta_0} I_{\{0 \leqslant X_i \leqslant \theta_0\}}} = \left(\frac{\theta_0}{\theta_1}\right)^n \frac{I_{\{X_{(n)} \leqslant \theta_1\}}}{I_{\{X_{(n)} \leqslant \theta_0\}}} = \begin{cases} \infty, & X_{(n)} > \theta_0 \\ \left(\frac{\theta_0}{\theta_1}\right)^n, & X_{(n)} \leq \theta_0 \end{cases}$$

Критическая функция $\varphi(X) = \begin{cases} 1, & X_{(n)} > \theta_0, \\ \varepsilon_\alpha, & X_{(n)} \leq \theta_0. \end{cases}$

$\alpha = \mathbb{E}_{\theta_0}\varphi(X) = 1 \cdot \mathsf{P}_{\theta_0}(X_{(n)} > \theta_0) + \varepsilon_\alpha \cdot \mathsf{P}_{\theta_0}(X_{(n)} \leq \theta_0) = 1 \cdot 0 + \varepsilon_\alpha \cdot 1$, откуда $\alpha = \varepsilon_\alpha$.

Мощность критерия

$$\begin{aligned} W(\varphi, \theta_1) = \mathbb{E}_{\theta_1}\varphi(X) = 1 \cdot \mathsf{P}_{\theta_1}\big(X_{(n)} > \theta_0\big) + \alpha \cdot \mathsf{P}_{\theta_1}\big(X_{(n)} \leq \theta_0\big) = \\ = 1 - \mathsf{P}_{\theta_1}\big(X_{(n)} \leq \theta_0\big) + \alpha \cdot \mathsf{P}_{\theta_1}\big(X_{(n)} \leq \theta_0\big) = \\ = 1 - (1-\alpha) \cdot \prod_{k=1}^{n} \mathsf{P}_{\theta_1}\big(X_k \leq \theta_0\big) = 1 - (1-\alpha)\Big(\frac{\theta_0}{\theta_1}\Big)^n. \end{aligned}$$

□

Задача №35. Пусть $X_1, \ldots, X_n$ независимы и имеют пуассоновское распределение $\Pi(\theta)$. Построить центральный доверительный интервал с коэффициентом доверия α, используя точечную оценку $T(X) = \overline{X}$.

Решение.

$$X_i \sim \Pi(\theta), \quad n\overline{X} \sim \Pi(n\theta), \quad \mathsf{P}_\theta\Big(\overline{X} = \frac{k}{n}\Big) = \mathsf{P}_\theta\big(x\overline{X} = k\big) = \exp(-n\theta)\frac{(n\theta)^k}{k!}.$$

Функция распределения $T(X)$

$$F_{T(X)}(t;\theta) = \begin{cases} \sum\limits_{k=0}^{[nt]} \exp(-n\theta)\dfrac{(n\theta)^n}{k!}, & t \geq 0, \\ 0, & t < 0; \end{cases}$$

монотонна по θ при $\theta \geq 0$, поскольку

$$\frac{\partial}{\partial\theta} F_{T(X)}(t;\theta) = -n\exp(-n\theta)\frac{(n\theta)^{[nt]}}{[nt]!} < 0.$$

Границы доверительного интервала (θ_1, θ_2) однозначно задаются уравнениями

$$F_{T(X)}(t;\theta_1) = \frac{1+\alpha}{2}, \qquad F_{T(X)}(t;\theta_2) = \frac{1-\alpha}{2}$$

при $t = T(X)$.

□

Глава 3

Проверка гипотез

Определения и теоремы

1. **Введение**.
Проверка гипотез — это формальная процедура проверки правильности некоторого утверждения об одном или нескольких поараметрах распределения. Используя информацию, полученную из выборки, гипотезу либо отклоняют, либо принимают.
В каждой проверке гипотезы участвуют 4 составляющие:
(1) **Нулевая гипотеза** H_0 — утверждение об одном или нескольких поараметрах распределения. Предполагается, что она верна.
(2) **Альтернативная гипотеза** H_1 — противоположное утверждение. Альтернативная гипотеза верна, если неверна нулевая гипотеза.
(3) **Статистический критерий** (или просто **критерий**) — правило, согласно которому гипотеза H_0 принимается или отвергается.
(4) **Критическая область** RR — множество чисел, выбранное таким образом, что в случае попадания значения выборки в это множество нулевая гипотеза отвергается. Одно или более **критических значений** отделяют критическую область от остальных значений выборки.
При проверке гипотез учитывают две вероятности ошибок, показанных в таблице 1 и описанных ниже.
(1) **Ошибка первого рода** – H_0 отвергается, в то время как она верна. Вероятность ошибки I рода обычно обозначают α: P(ошибка I рода) $= \alpha$. Типичные значения α: 0.05, 0.01, 0.001.
(2) **Ошибка второго рода** — H_0 принимается, в то время как она неверна. Вероятность ошибки II рода зависит от реального значения параметров распределения и обозначается обычно через β (или $\beta(\theta)$): P(ошибка II рода) $= \beta$. **Мощность критерия** равна $1 - \beta$.

	H_0 принимается	H_0 отвергается
H_0 верна	Верное решение	Ошибка I рода: α
H_0 неверна	Ошибка II рода: β	Верное решение

Таблица 1: Ошибки проверки гипотез

Замечания.
(a) α — это **уровень значимости** критерия. Выборка **значима**, если ее значение лежит в критической области.
(b) Значения α и β имеют обратную зависимость: при увеличении α β уменьшается и наоборот.
(c) Для одновременного уменьшения α и β, следует увеличить объем выборки.

Таблицы.
В таблицах 2 и 3 представлены некоторые гипотезы и критерии для их проверки.

Пример. Производитель сухих завтраков заявляет, что масса содержимого каждой упаковки — 24 унции. Для проверки этого утверждения группа покупателей выбрала

Нулевая гипотеза, свойства распределения	Альтернативная гипотеза	Статистика	Критическая область
$\mu = \mu_0$, большое n или нормальность, σ^2 — известно	$\mu > \mu_0$	$Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$	$Z \geq z_\alpha$
	$\mu < \mu_0$		$Z \leq -z_\alpha$
	$\mu \neq \mu_0$		$\lvert Z\rvert \geq z_{\alpha/2}$
$\mu = \mu_0$, нормальность, σ^2 — неизвестно	$\mu > \mu_0$	$T = \frac{\overline{X}-\mu_0}{S/\sqrt{n}}$	$T \geq t_{\alpha,\, n-1}$
	$\mu < \mu_0$		$T \leq -t_{\alpha,\, n-1}$
	$\mu \neq \mu_0$		$\lvert T\rvert \geq t_{\alpha/2,\, n-1}$
$\sigma = \sigma_0^2$, нормальность	$\sigma^2 > \sigma_0^2$	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$\chi^2 \geq \chi^2_{\alpha,\, n-1}$
	$\sigma^2 < \sigma_0^2$		$\chi^2 \leq \chi^2_{1-\alpha,\, n-1}$
	$\sigma^2 \neq \sigma_0^2$		$\chi^2 \leq \chi^2_{1-\alpha/2,\, n-1}$, или $\chi^2 \geq \chi^2_{\alpha/2,\, n-1}$
$p = p_0$, биномиальный эксперимент, большое n	$p > p_0$	$Z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$	$Z \geq z_\alpha$
	$p < p_0$		$Z \leq -z_\alpha$
	$p \neq p_0$		$\lvert Z\rvert \geq z_{\alpha/2}$

Таблица 2: Проверка гипотез: одна выборка

случайным образом 17 упаковок сухих завтраков и взвесила их содержимое. Среднее занчение: $\bar{x} = 23.55$, выборочная дисперсия $s = 1.5$. Свидетельствует ли это о о том, что производитель вводит покупателей в заблуждение? Использовать коэффициент значимости $\alpha = 0.05$.

Решение.

(1) Задача заключается в оценке математического ожидания распределения μ. Распределение масс содержимого упаковок предполагается нормальным, дисперсия неизвестна. Альтернативная гипотеза: $\mu < \mu_0 = 24$.

(2) Четыре составные части проверки гипотезы таковы:

$H_0 : \mu = 24 = \mu_0$

$H_1 : \mu < 24$

$$TS : T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

$RR : T \leq -t_{\alpha,\, n-1} = -t_{0.05,\, 16} = -1.7459$

(3) $$T = \frac{23.55 - 24}{1.5/\sqrt{17}} = -1.2369$$

(4) Вывод: Значение статистики $\bar{x}$ не лежит в критической области (то есть $p = 0.1170 > 0.05$). Поэтому нет оснований предполагать, что математическое ожидание выборки меньше 24.

Пример. Производитель деталей автомобиля заявляет, что новый продукт, при установке его на фильтр двигателя, уменьшает расход газа. Были измерены расстояния, пройденные автомобилями на единицу расхода газа, с испольхованием этого механизма и без него. Среднее значение разностей (до–после): $\bar{d} = -1.2$, $s_D = 3.5$. Свидетельствует ли это о том, что новый механизм действительно уменьшает расход газа? $\alpha = 0.01$.

Решение.

(1) Задача заключается в оценке математического ожидания $\mu_D = \mu_1 - \mu_2$ распределения разности двух случайных величин. Распределение значений предполагается нормальным, случайные величины зависимы. Альтернативная гипотеза: $\mu_D < \Delta_0 = 0$.

(2) Четыре составные части проверки гипотезы таковы:

Свойства распределения			
Нулевая гипотеза	Альтернативная гипотеза	Статистика	Критическая область
n_1, n_2 – большие, независимость, σ_1^2, σ_2^2 — известны, или нормальность, независимость, σ_1^2, σ_2^2 — известны			
$\mu_1 - \mu_2 = \Delta_0$	$\mu_1 - \mu_2 > \Delta_0$	$Z = \frac{(\overline{X}_1 - \overline{X}_2) - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$Z \geq z_\alpha$
	$\mu_1 - \mu_2 < \Delta_0$		$Z \leq -z_\alpha$
	$\mu_1 - \mu_2 \neq \Delta_0$		$\lvert Z \rvert \geq z_{\alpha/2}$
нормальность, независимость, $\sigma_1^2 = \sigma_2^2$ — неизвестны			
$\mu_1 - \mu_2 = \Delta_0$	$\mu_1 - \mu_2 > \Delta_0$	$T = \frac{(\overline{X}_1 - \overline{X}_2) - \Delta_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$	$T \geq t_{\alpha,\, n_1+n_2-2}$
	$\mu_1 - \mu_2 < \Delta_0$		$T \leq -t_{\alpha,\, n_1+n_2-2}$
	$\mu_1 - \mu_2 \neq \Delta_0$	$S_p = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$	$\lvert T \rvert \geq t_{\alpha/2,\, n_1+n_2-2}$
нормальность, независимость, σ_1^2, σ_2^2 — неизвестны, $\sigma_1^2 \neq \sigma_2^2$			
$\mu_1 - \mu_2 = \Delta_0$	$\mu_1 - \mu_2 > \Delta_0$	$T' = \frac{(\overline{X}_1 - \overline{X}_2) - \Delta_0}{S_p \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$	$T' \geq t_{\alpha,\, \nu}$
	$\mu_1 - \mu_2 < \Delta_0$		$T' \leq -t_{\alpha,\, \nu}$
	$\mu_1 - \mu_2 \neq \Delta_0$	$\nu \approx \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{\left(S_1^2/n_1\right)^2}{n_1-1} + \frac{\left(S_2^2/n_2\right)^2}{n_2-1}}$	$\lvert T' \rvert \geq t_{\alpha/2,\, \nu}$
нормальность, n пар, зависимость			
$\mu_D = \Delta_0$	$\mu_D > \Delta_0$	$T = \frac{\overline{D} - \Delta_0}{S_D/\sqrt{n}}$	$T \geq t_{\alpha,\, n-1}$
	$\mu_D < \Delta_0$		$T \leq -t_{\alpha,\, n-1}$
	$\mu_D \neq \Delta_0$		$\lvert T \rvert \geq t_{\alpha/2,\, n-1}$
нормальность, независимость			
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$F = \frac{S_1^2}{S_2^2}$	$F \geq F_{\alpha,\, n_1-1, n_2-1}$
	$\sigma_1^2 < \sigma_2^2$		$F \leq F_{1-\alpha,\, n_1-1, n_2-1}$
	$\sigma_1^2 \neq \sigma_2^2$		$F \leq F_{1-\alpha/2,\, n_1-1, n_2-1}$, или $F \geq F_{\alpha/2,\, n_1-1, n_2-1}$
биномиальный эксперимент (испытания Бернулли), n_1,n_2 — большие, независимость			
$p_1 - p_2 = 0$	$p_1 - p_2 > 0$	$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$	$Z \geq z_\alpha$
	$p_1 - p_2 < 0$		$Z \leq z_\alpha$
	$p_1 - p_2 \neq 0$	$\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}, \quad \hat{q} = 1 - \hat{p}$	$\lvert Z \rvert \geq z_{\alpha/2}$
биномиальный эксперимент, n_1,n_2 — большие, независимость			
$p_1 - p_2 = \Delta_0$	$p_1 - p_2 > \Delta_0$	$Z = \frac{(\hat{p}_1 - \hat{p}_2) - \Delta_0}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$	$Z \geq z_\alpha$
	$p_1 - p_2 < \Delta_0$		$Z \leq z_\alpha$
	$p_1 - p_2 \neq \Delta_0$		$\lvert Z \rvert \geq z_{\alpha/2}$

Таблица 3: Проверка гипотез: две выборки

$$
\begin{aligned}
&H_0 : \mu_D = \Delta_0 = 0 \\
&H_1 : \mu_D < 0 \\
&TS : T = \frac{\bar{D} - \Delta_0}{S_d/\sqrt{n}} \\
&RR : T \leq -t_{\alpha,\, n-1} = -t_{0.01,\, 9} = -2.8214
\end{aligned}
$$

(3) $T = \frac{-1.2 - 0}{3.5/\sqrt{10}} = -1.0842$

(4) Вывод: Значение статистики $\bar{x}$ не лежит в критической области (то есть $p = 0.1532 > 0.01$). Поэтому нет оснований предполагать, что новый продукт уменьшает расход топлива.

2. **Лемма Немана-Пирсона**.
Пусть нулевая гипотеза $H_0 : \theta = \theta_0$, альтернатива $H_1 : \theta = \theta_1$, $L(\theta)$ — функция правдоподобия, вычисленная в точке θ. Для заданного α критическая область наиболее мощного критерия задается неравенством

$$\frac{L(\theta_0)}{L(\theta_1)} < k.$$

3. **Проверка гипотез путем сравнения функций правдоподобия**.
Пусть нулевая гипотеза $H_0 : \theta \in \Omega_0$, альтернатива $H_1 : \theta \in \Omega_1$, $\Omega_0 \cap \Omega_1 = \emptyset$ и $\Omega_0 \cup \Omega_1 = \Omega$. Через $L(\hat{\Omega}_0)$ обозначим функцию правдоподобия, где все неизвестные параметры заменены максимальными своими значениями, удовлетворяющими условию $\theta \in \Omega_0$, $L(\hat{\Omega})$ — то же для $\theta \in \Omega$. Определим

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})}.$$

Для проверки гипотезы рассматриваем λ как тестовую статистику; критическая область определяется из $\lambda \leq k$ (для $0 < k < 1$).

При выполнении некоторых условий и для больших n, $-2\ln\lambda$ имеет приближенно распределение χ^2 с количеством степеней свободы, равным числу параметров или функций от параметров.

4. **Критерий χ^2 Пирсона**.
Пусть n_i — число наблюдений i-й категории ($i = 1, 2, \ldots, k$) и $n = n-1+n-2+\ldots+n_k$.
$H_0 : p_1 = p_{10}$, $p_2 = p_{20}$,…,$p_k = p_{k0}$
$H_1 : p_i \neq p_{i0}$ хотя бы для одного i

$$TS : \chi^2 = \sum_{i=1}^{k} \frac{(\text{наблюдаемое} - \text{ожидаемое})^2}{\text{ожидаемое}} = \sum_{i=1}^{k} \frac{(n_i - np_{i0})^2}{np_{i0}}$$

Если верна нулевая гипотеза, χ^2 имеет приближенно распределени хи-квадрат с $k-1$ степенями свободы. Это приближение удовлетворительно при $np_{i0} \geq 5$ для всех i.

$RR : \chi^2 \geq \chi^2_{\alpha,\,k-1}$

5. **Критерий независимости χ^2. Таблицы независимости.**
Рассматриваются 2 признака A и B, принимающие значения $\{a_i\}_{i=1}^I$ и $\{b_j\}_{j=1}^J$ соответственно. Прверяется предположение о независимости этих признаков. Таблица независимости размера $I \times J$ имеет вид

	b_1	b_2	$\ldots$	b_J	Всего
a_1	n_{11}	n_{12}	$\cdots$	n_{1J}	$n_{1\cdot}$
a_2	n_{21}	n_{22}	$\cdots$	n_{2J}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
a_I	n_{I1}	n_{I2}	$\cdots$	n_{IJ}	$n_{I\cdot}$
Всего	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot J}$	n

где $n_{k\cdot} = \sum_{j=1}^J n_{kj}$ и $n_{\cdot k} = \sum_{i=1}^I n_{ik}$. Если имеет место независимость признаков, вероятность получить определенный набор значений признаков при известных $\{n_{\cdot k}, n_{k\cdot}\}$

есть

$$\mathsf{P}\big(n_{11},\ldots,n_{IJ}\mid n_{1\cdot},\ldots,n_{\cdot J}\big)=\frac{\left(\prod_i^I n_{i\cdot}!\right)\left(\prod_j^J n_{\cdot j}!\right)}{n!\prod_i^I\prod_j^J n_{ij}!}.$$

Пусть таблица независимости содержит I строк и J столбцов, n_{ij} — количество экземпляров со значениями признаков $\{a_i,b_j\}$, и $\hat{e}_{ij}$ — оценка ожидаемого числа таких экземпляров. Тестовая статистика есть

$$\chi^2=\sum_{\substack{\text{по всем}\\ \text{ячейкам}}}\frac{(\text{наблюдаемое}-\text{ожидаемое})^2}{\text{ожидаемое}}=\sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(n_{ij}-\hat{e}_{ij})^2}{\hat{e}_{ij}},$$

где

$$\hat{e}_{ij}=\frac{(\text{всего в } i\text{-й строке})(\text{всего в } j\text{-м столбце})}{\text{всего экземпляров}}=\frac{n_{i\cdot}n_{\cdot j}}{n}.$$

Если верна нулевая гипотеза, χ^2 имеет приближенно распределени хи-квадрат с $(I-1)(J-1)$ степенями свободы. Это приближение удовлетворительно при $\hat{e}_{ij}\geq 5$ для всех i и j.

6. **Таблицы независимости размера 2×2.**
Таблица независимости размера 2×2 — часто встречаюшийся частный случай таблиц независимости. Каждый из n элементов обладает (взаимоисключающими) свойствами 1 или 2 и I или II.

	I	II	Всего
1	a	$A-a$	A
2	b	$B-b$	B
Всего	r	$n-r$	n

Если r, A и B фиксированы, вероятность заданной конфигурации

$$f(a\mid r,A,B)=\frac{\binom{A}{a}\binom{B}{b}}{\binom{n}{r}}=\frac{A!B!r!(n-r)!}{n!a!b!(A-a)!(B-b)!}.$$

При заданных a, A и B можно определить критическое значение r так, чтобы вероятность $f(a\mid r,A,B)$ принимала нужное нам значение. Тогда число элементов, обладающих свойствами 1 и I определяется как разность $b=r-a$.
Пример. Для сравнения вероятностей успеха двух распределений составлена таблица независимости размера 2×2:

	Успех	Неуспех	Всего
Выборка из распределения 1	7	2	9
Выборка из распределения 2	3	3	6
Всего	10	5	15

Есть ли основания предполагать, что вероятности успеха в распределениях различны? Уровень значимости $\alpha=0.05$.
Решение.
(1) Для данных значений a, A и B при неизвестном r таблица независимости выглядит следующим образом:

	Успех	Неуспех	Всего
Выборка из распределения 1	7	2	9
Выборка из распределения 2	b	$6-b$	6
Всего	r	$15-r$	15

(2) Для $7 \le r \le 13$ условная вероятность $f(a \mid r, A, B)$ равна:

r	7	8	9	10	11	12	13
$f(a \mid r, A, B)$	0.0056	0.034	0.11	0.24	0.40	0.47	0.34

(3) Из этой таблицы получим, что наибольшее значение r, для которого вероятность наблюдать заданное значение a менее, чем $\alpha = 0.05$, есть $r = 8$. В этом случае вероятность наблюдать a равна 0.034.

(4) Следовательно, критическое значение $b = r - a = 8 - 7 = 1$. Если в полученных данных $b \le 1$, нулевая гипотеза $H_0 : p_1 = p_2$ отвергается. В нашем случае $b = 3$.

(5) Вывод: Значение тестовой статистики не лежит в критической области. Нет оснований предполагать, что вероятности успеха в распределениях различны.

7. **Критические значения для проверки возмущений**.
Для проверки на возмущения можно рассматривать $\max\limits_{i=1,2,\ldots} (x_i - \bar{x})$ — макисмальное отклонение наблюдаемых значений от их среднего, которое средует нормализовать с учетом стандартного отклонения или его оценки. Другой способ — используя отношения элементов вариационного ряда.

(a) Чтобы определить, является ли наименьший элемент вариационного ряда возмущением, вычислим

$$r_{10} = \frac{x_{(2)} - x_{(1)}}{x_{(n)} - x_{(1)}}.$$

Аналогично, чтобы определить, является ли наибольший элемент вариационного ряда возмущением, вычислим

$$r_{10} = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}}.$$

(b) Чтобы определить, является ли наименьший элемент вариационного ряда возмущением, не используя $x_{(n)}$ вычислим

$$r_{11} = \frac{x_{(2)} - x_{(1)}}{x_{(n-1)} - x_{(1)}}.$$

Аналогично, чтобы определить, является ли наибольший элемент вариационного ряда возмущением, не используя $x_{(1)}$ вычислим

$$r_{11} = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(2)}}.$$

Таблицы 4 и 5 содержат критические значения для r_{10} и r_{11}.

Таблица 4: Критические значения для r_{10} ($\mathsf{P}[r_{10} > R] = \alpha$)

n	$\alpha = 0.005$	0.01	0.02	0.05	0.10	0.50	0.90	0.95
3	0.994	0.988	0.976	0.941	0.886	0.500	0.114	0.059
4	0.926	0.889	0.846	0.745	0.679	0.324	0.065	0.033
5	0.821	0.780	0.729	0.642	0.557	0.250	0.048	0.023
6	0.740	0.698	0.644	0.560	0.482	0.210	0.038	0.018
7	0.680	0.637	0.586	0.507	0.434	0.184	0.032	0.016
8	0.634	0.590	0.543	0.468	0.399	0.166	0.029	0.014
9	0.598	0.555	0.510	0.437	0.370	0.152	0.026	0.013
10	0.568	0.527	0.483	0.412	0.142	0.142	0.025	0.012
11	0.475	0.438	0.399	0.338	0.111	0.111	0.019	0.010
12	0.425	0.391	0.356	0.300	0.096	0.096	0.017	0.008
13	0.393	0.362	0.329	0.277	0.088	0.088	0.015	0.008
14	0.372	0.341	0.309	0.260	0.082	0.082	0.014	0.007

Таблица 5: Критические значения для r_{11} ($\mathsf{P}[r_{11} > R] = \alpha$)

n	$\alpha = 0.005$	0.01	0.02	0.05	0.10	0.50	0.90	0.95
4	0.995	0.991	0.981	0.955	0.910	0.554	0.131	0.069
5	0.937	0.916	0.876	0.807	0.728	0.369	0.078	0.039
6	0.839	0.805	0.763	0.689	0.609	0.288	0.056	0.028
7	0.782	0.740	0.689	0.610	0.530	0.241	0.045	0.022
8	0.725	0.683	0.631	0.554	0.479	0.210	0.037	0.019
9	0.677	0.635	0.587	0.512	0.441	0.189	0.033	0.016
10	0.639	0.597	0.551	0.477	0.409	0.173	0.030	0.014
15	0.522	0.486	0.445	0.381	0.323	0.129	0.023	0.011
20	0.464	0.430	0.392	0.334	0.282	0.110	0.019	0.010
25	0.426	0.394	0.359	0.294	0.255	0.098	0.027	0.009
30	0.399	0.369	0.336	0.283	0.236	0.090	0.016	0.008

Задачи

Задача №36. Пусть $X_1, \ldots, X_n$ независимы и имеют гамма-распределение $\Gamma(\theta, 1)$. Построить равномерено наиболее мощный критерий размера α для проверки гипотезы $H_0 : \theta{=}\theta_0$ при альтернативе $H_1 : \theta{>}\theta_0$. Найти функцию мощности.

Решение. Построим наиболее мощный критерий для проверки H_0 при простой альтернативе $H_1 : \theta{=}\theta_1$, $\theta_1{>}\theta_0$, воспользовавшись леммой Неймана-Пирсона:

$$\frac{L_1(X,\theta)}{L_0(X,\theta)} = \frac{\prod\limits_{i=1}^{n} \theta_1 \exp(-\theta_1 X_i)}{\prod\limits_{i=1}^{n} \theta_0 \exp(-\theta_0 X_i)} = \left(\frac{\theta_1}{\theta_0}\right)^n \exp\Big\{\big(\theta_0{-}\theta_1\big) \sum_{i=1}^{n} X_i\Big\} \geqslant c_\alpha \iff T(X){=}\sum_{i=1}^{n} X_i \le c'_\alpha.$$

c'_α найдем из условия

$$\alpha = P_{\theta_0}\left(T(X) \le c'_\alpha\right) = P_{\theta_0}\Big(2\theta_0 \sum_{i=1}^{n} X_i \le 2\theta_0 c'_\alpha\Big) = \left\{ \begin{array}{c} 2\theta_0 X_i \sim \Gamma\left(\frac{1}{2}, 1\right) = \chi^2_2 \\ 2\theta_0 \sum_{i=1}^{n} X_i \sim \chi^2_{2n} \end{array} \right\} = F_{2n}(2\theta_0 c'_\alpha),$$

где $F_m(a) = \int\limits_{-\infty}^{a} k_m(x)\, dx$, а $k_m(x)$ – плотность распределения случайной величины χ^2_m.

Отсюда $2\theta_0 c'_\alpha = \chi^2_{\alpha,2n} \implies c'_\alpha = \dfrac{\chi^2_{\alpha,2n}}{2\theta_0}$, где $\chi^2_{\alpha,2n}$ – квантиль порядка α функции $F_{2n}(y)$.

Критическая функция $\varphi(X) = \begin{cases} 1, & T(X) \le \chi^2_{\alpha,2n}/2\theta_0\,, \\ 0, & T(X) > \chi^2_{\alpha,2n}/2\theta_1\,. \end{cases}$

Функция мощности будет представляться в следующем виде:

$$W(\theta) = \mathbb{E}_\theta \varphi(X) = P\big(T(X) \le c'_\alpha\big) = P_\theta\big(2\theta T(X) \le 2\theta c'_\alpha\big) = F_{2n}\Big(\frac{\theta}{\theta_0}\chi^2_{\alpha,2n}\Big).$$

Построенный критерий – наиболее мощный, если гипотеза H_1 – простая, то есть $H_1 : \theta{=}\theta_1$. При построении критерия значение θ_1 используется неявно, важно лишь, что $\theta_1{>}\theta_0$. Значит, построенный критерий – равномерно наиболее мощный. $\square$

Задача №37. Пусть $X_1, \ldots, X_n$ независимы и имеют гамма-распределение $\Gamma(\theta, 2)$. Построить кратчайший доверительный интервал для θ с коэффициентом доверия α, основанный на центральной статистике $G(X,\theta) = \theta \cdot \sum\limits_{i=1}^{n} X_i$

Решение. Прежде всего заметим, что

$$2\theta X_i \sim \Gamma\Big(\frac{1}{2}, 2\Big) \quad \implies \quad 2\theta \sum_{i=1}^{n} X_i \sim \Gamma\Big(\frac{1}{2}, 2n\Big) = \chi^2_{4n}.$$

Написав определение доверительного интервала с уровнем доверия α и воспользовавшись описанными выше соотношениями получим следующую цепочку равенств:

$$\alpha = P\big(g_1 < 2G(X,\theta) < g_2\big) = P\Big(\frac{g_1}{2n\overline{\overline{X}}} < \theta < \frac{g_2}{2n\overline{\overline{X}}}\Big) = F_{4n}(g_2) - F_{4n}(g_1),$$

где $F_{4n}(y)$ – функция распределения χ^2_{4n}.

Для построения наименьшего интервала необходимо минимизировать $g_2 - g_1$ при условии $F_{4n}(g_2) - F_{4n}(g_1) = \alpha$.

Запишем функцию Лагранжа:

$$F(\lambda) = g_2 - g_1 - \lambda(F_{4n}(g_2) - F_{4n}(g_1) - \alpha).$$

Продифференцируем ее по переменным g_1, g_2 и λ и приравняем полученные выражения к нулю. В результате получим систему уравнений

$$\begin{cases} F_{4n}(g_2) - F_{4n}(g_1) = \alpha, \\ F'_{4n}(g_2) = F'_{4n}(g_2). \end{cases} \tag{$*$}$$

Кратчайший доверительный интервал для θ с коэффициентом доверия α, основанный на центральной статистике $G(X,\theta)$, имеет вид $\left(\dfrac{g_1}{2n\overline{\overline{X}}}, \dfrac{g_2}{2n\overline{\overline{X}}}\right)$, где g_1 и g_2 - решения системы ($*$). $\square$

Задача № 38. Пусть $X_1, \ldots, X_n$ независимы и имеют пуассоновское распределение $\Pi(\theta)$. Построить равномерно наиболее мощный критерий размера α для проверки гипотезы $H_0 : \theta = \theta_0$ при альтернативе $H_1 : \theta < \theta_0$. Найти функцию мощности.

Решение. Построим наиболее мощный критерий для простой альтернативы H_1 : $\theta = \theta_1,\ \theta_1 < \theta_0$. Для этого воспользуемся леммой Неймана–Пирсона:

$$\frac{L_1}{L_0} = \frac{\prod\limits_{i=1}^{n} e^{-\theta_1}\,\dfrac{\theta_1^{X_i}}{X_i!}}{\prod\limits_{i=1}^{n} e^{-\theta_0}\,\dfrac{\theta_0^{X_i}}{X_i!}} = \left(\frac{\theta_1}{\theta_0}\right)^{\sum\limits_{i=1}^{n} X_i} \exp\bigl(-n(\theta_1-\theta_0)\bigr) \geq c_\alpha \quad\Longleftrightarrow\quad T(X) = \sum_{i=1}^{n} X_i \leq c'_\alpha.$$

Учтем, что $T(X) \sim \Pi(\theta n)$. c'_α найдем из условия

$$F_{\theta_0}(c'_\alpha - 1) < \alpha \leq F_{\theta_0}(c'_\alpha), \quad c'_\alpha \in \mathbb{Z},$$

где $F_{\theta_0}(y)$ – функция распределения $T(X)$ при условии, что гипотеза H_0 верна:

$$F_{\theta_0}(y) = \sum_{k=0}^{[y]} \exp(-n\theta_0)\frac{(n\theta_0)^k}{k!}.$$

Критическая функция

$$\varphi(X) = \begin{cases} 1, & T(X) < c'_\alpha, \\ \varepsilon_\alpha, & T(X) = c'_\alpha, \\ 0, & T(X) > c'_\alpha; \end{cases} \qquad \text{где } \varepsilon_\alpha = \frac{\alpha - F_{\theta_0}(c'_\alpha - 1)}{F_{\theta_0}(c'_\alpha) - F_{\theta_0}(c'_\alpha - 1)}.$$

Если найдется $c'_\alpha \in \mathbb{Z}$ такое, что $F_{\theta_0}(c'_\alpha) = \alpha$, критерий будет нерандомизированным:

$$\varphi(X) = \begin{cases} 1, & T(X) \leq c'_\alpha, \\ 0, & T(X) > c'_\alpha. \end{cases}$$

Построенный критерий – наиболее мощный, если гипотеза H_1 - простая ($H_1 : \theta{=}\theta_1$). При построении критерия значение θ_1 используется неявно: важно лишь, что $\theta_1{<}\theta_0$. Значит, построенный критерий – равномерно наиболее мощный.

Функция мощности:

$$W(\varphi,\theta) = \mathbb{E}_\theta \varphi(X) = 1 \cdot P_\theta\bigl(T(X) < c'_\alpha\bigr) + \varepsilon_\alpha \cdot P_\theta\bigl(T(X) = c'_\alpha\bigr) = F_\theta(c'_\alpha{-}1) + \varepsilon_\alpha F_\theta(c'_\alpha),$$

$$\text{где} \quad F_\theta(y) = \sum_{k=0}^{[y]} \exp(-n\theta)\frac{(n\theta)^k}{k!}. \qquad \square$$

Задача №39. Пусть $X_1, \ldots, X_n$ независимы и имеют плотность распределения

$$f(x,\theta) = \begin{cases} \exp\bigl\{-(x-\theta)\bigr\}, & x > \theta, \\ 0, & x \le \theta. \end{cases}$$

Построить наиболее мощный критерий размера α для проверки гипотезы $H_0 : \theta{=}\theta_0$ при альтернативе $H_1 : \theta{=}\theta_1 < \theta_0$. Найти мощность критерия.

Решение. Для решения данной задачи воспользуемся леммой Неймана–Пирсона:

$$\frac{L_1}{L_0} = \frac{\prod\limits_{i=1}^{n} \exp(\theta_1 - X_i)\mathbb{I}_{\left\{X_i > \theta_1\right\}}}{\prod\limits_{i=1}^{n} \exp(\theta_0 - X_i)\mathbb{I}_{\left\{X_i > \theta_0\right\}}} = \exp\bigl(n(\theta_1 - \theta_0)\bigr)\frac{\mathbb{I}_{\left\{X_{(1)} > \theta_1\right\}}}{\mathbb{I}_{\left\{X_{(1)} > \theta_0\right\}}} > c_\alpha.$$

Поскольку

$$\frac{L_1}{L_0} = \begin{cases} \infty, & X_{(1)} \le \theta_0, \\ \exp\bigl(n(\theta_1 - \theta_0)\bigr), & X_{(1)} > \theta_0; \end{cases}$$

то при $X_{(1)} \le \theta_0$ и $\dfrac{L_1}{L_0} > c_\alpha$ для любого α критическая функция принимает вид

$$\varphi(X) = \begin{cases} 1, & X_{(1)} \le \theta_0, \\ \varepsilon_\alpha, & X_{(1)} > \theta_0. \end{cases}$$

Отыщем ε_α:

$$\alpha = \mathbb{E}_{\theta_0}\varphi(X) = 1 \cdot P_{\theta_0}\bigl(X_{(1)} \le \theta_0\bigr) + \varepsilon_\alpha \cdot P_{\theta_0}\bigl(X_{(1)} > \theta_0\bigr) = 1 \cdot 0 + \varepsilon_\alpha \cdot 1,$$

откуда $\alpha = \varepsilon_\alpha$.
Мощность критерия

$$W(\varphi, \theta_1) = \mathbb{E}_{\theta_1}\varphi(X) = P_{\theta_1}\bigl(X_{(1)} \le \theta_0\bigr) + \alpha P_{\theta_1}\bigl(X_{(1)} > \theta_0\bigr) =$$
$$= \left(1 - \Bigl(\int_{\theta_0}^{+\infty} \exp(\theta_1 - x)\, dx\Bigr)^n\right) + \alpha\Bigl(\int_{\theta_0}^{+\infty} \exp(\theta_1 - x)\, dx\Bigr)^n = 1 - (1-\alpha)\exp\bigl(n(\theta_1 - \theta_0)\bigr).$$

□

Задача №40. Пусть $X_1, \ldots, X_n$ независимы и имеют равномерное распределение на отрезке $[0, \theta]$. Построить наиболее мощный критерий размера α для проверки гипотезы $H_0 : \theta{=}\theta_0$ при альтернативе $H_1 : \theta{=}\theta_1 < \theta_0$. Найти мощность критерия.

Решение. Как и в предыдущих задачах, воспользуемся леммой Неймана–Пирсона (следует отметить, что поведение при $X_{(1)}{<}0$ и $X_{(n)} > \theta_0$ нас не интересует):

$$\frac{L_1}{L_0} = \frac{\prod\limits_{i=1}^{n} \frac{1}{\theta_1}\mathbb{I}_{\left\{0 \le X_i \le \theta_1\right\}}}{\prod\limits_{i=1}^{n} \frac{1}{\theta_0}\mathbb{I}_{\left\{0 \le X_i \le \theta_0\right\}}} = \left(\frac{\theta_0}{\theta_1}\right)^n \frac{\mathbb{I}_{\left\{X_{(n)} \le \theta_1\right\}}}{\mathbb{I}_{\left\{X_{(n)} \le \theta_0\right\}}} = \begin{cases} 0, & X_{(n)} > \theta_1, \\ \left(\dfrac{\theta_0}{\theta_1}\right)^n, & X_{(n)} \le \theta_1. \end{cases}$$

Следовательно, критическая функция имеет вид

$$\varphi(X) = \begin{cases} \varepsilon_\alpha, & X_{(n)} \le \theta_1, \\ 0, & X_{(n)} > \theta_1. \end{cases}$$

Рассмотрим следующее выражения для α:

$$\alpha = \mathbb{E}_{\theta_0}\varphi(X) = \varepsilon_\alpha P_{\theta_0}\left(\frac{X_{(n)}}{\theta_0} \le \frac{\theta_1}{\theta_0}\right) = \varepsilon_\alpha \left(\frac{\theta_1}{\theta_0}\right)^n,$$

откуда $\varepsilon_\alpha = \min\left\{\alpha\cdot\left(\frac{\theta_0}{\theta_1}\right)^n; 1\right\}$.

Возможны два случая:

1. $\varepsilon_\alpha = 1$, мощность критерия $W(\varphi, \theta_1) = \mathbb{E}_{\theta_1}\varphi(X) = 1$;
2. $\varepsilon_\alpha = \alpha\cdot\left(\dfrac{\theta_0}{\theta_1}\right)^n$, мощность критерия $W(\varphi, \theta_1) = \mathbb{E}_{\theta_1}\varphi(X) = \varepsilon_\alpha P_{\theta_1}(X_{(n)} \le \theta_1) = \varepsilon_\alpha$. $\square$

Литература

[1] *Д. М. Чибисов, В. И. Пагурова.* Задачи по математической статистике. М.: Издательство Московского университета, 1990.

[2] *Г. И. Ивченко, Ю. И. Медведев, А. В. Чистяков.* Сборник задач по математической статистике. М.: "Высшая школа", 1989.

Оглавление