

Московский физико-технический институт

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА
II ЗАДАНИЕ

Автор: Яфаров Руслан,
Б13-202

весна 2024

1. Байесовские оценки

Напоминание (себе и для определенности (в Википедии, например, другая плотность)) Плотность распределения $\Gamma(\alpha, \lambda)$ с параметрами $\alpha > 0, \lambda > 0$ равна

$$p_{\alpha, \lambda}(x) = \frac{\alpha^\lambda x^{\lambda-1} e^{-\alpha x}}{\Gamma(\lambda)} I_{\mathbb{R}_+}(x)$$

1. $X_1 \sim \text{Exp}(\theta)$, θ имеет сопряженное априорное распределение $\Gamma(\alpha, \beta)$. Проверьте оценку $\theta^* = \frac{n+\beta}{\alpha+\sum_{i=1}^n X_i}$ на состоятельность.

Решение. По УЗБЧ $\bar{X} \xrightarrow{\text{п. н.}} \frac{1}{\theta}$. По теореме о наследовании сходимости $\theta^* = \frac{1+\frac{\beta}{n}}{\frac{\alpha}{n}+\bar{X}} \xrightarrow{\text{п. н.}} \theta \Rightarrow$ оценка сильно состоятельна.

2. По выборке X_1, \dots, X_n из пуассоновского распределения с параметром θ , где $\theta \sim \Gamma(\alpha, \lambda)$, постройте наилучшую оценку в байесовском подходе с квадратичной функцией потерь.

Решение. По теореме, лучшей оценкой в байесовском подходе с квадратичной функцией потерь является байесовская оценка.

$$q(t)p_t(x) = \frac{\alpha^\lambda t^{\lambda-1} e^{-\alpha t}}{\Gamma(\lambda)} I_{\mathbb{R}_+}(t) \frac{t^{\sum_{i=1}^n x_i}}{x_1! \dots x_n!} e^{-nt} = ct^{\sum_{i=1}^n x_i + \lambda - 1} e^{-t(n+\alpha)} I_{\mathbb{R}_+}(t) \Rightarrow$$

$$p_{\theta|X} \sim \Gamma\left(\alpha + n, \lambda + \sum_{i=1}^n X_i\right) \Rightarrow \hat{\theta} = \frac{\lambda + \sum_{i=1}^n X_i}{\alpha + n}$$

3. Пусть X_1, \dots, X_n - выборка из нормального распределения с параметрами $(\theta, 1)$. Найдите байесовскую оценку параметра θ , если априорное распределение θ есть $\text{Bin}(1, p)$. Будет ли полученная оценка состоятельной оценкой параметра θ ?

Решение.

$$\text{Bin}(1, p) = \text{Bern}(p), q(t)p_t(x) = p^t(1-p)^{1-t} \frac{1}{\sqrt{2\pi}} e^{-\frac{\sum_{i=1}^n (X_i - t)^2}{2}} \Rightarrow$$

$$\int_{\theta} q(t)p_t(x)dt = p \frac{1}{\sqrt{2\pi}} e^{-\frac{\sum_{i=1}^n (X_i - 1)^2}{2}} + (1-p) \frac{1}{\sqrt{2\pi}} e^{-\frac{\sum_{i=1}^n X_i^2}{2}} = \frac{e^{-\frac{\sum_{i=1}^n X_i^2}{2}}}{\sqrt{2\pi}} \left(p e^{\sum_{i=1}^n X_i - \frac{1}{2}} + 1 - p \right) \Rightarrow$$

$$p_{\theta|X}(t|x) = \frac{p^t(1-p)^{1-t} e^{\sum_{i=1}^n X_i t - \frac{t}{2}}}{p e^{\sum_{i=1}^n X_i - \frac{1}{2}} + 1 - p} \Rightarrow \hat{\theta} = \frac{p e^{\sum_{i=1}^n X_i - \frac{1}{2}}}{p e^{\sum_{i=1}^n X_i - \frac{1}{2}} + 1 - p} = 1 - \frac{1-p}{p(e^{\bar{X} - \frac{1}{2}})^n + 1 - p}$$

По УЗБЧ $\bar{X} \xrightarrow{\text{п. н.}} \theta \Rightarrow$ по теореме о наследовании сходимости $\hat{\theta} \xrightarrow{\text{п. н.}} \theta \Rightarrow$ оценка состоятельна.

4. Пусть X_1, \dots, X_n - выборка из равномерного распределения на отрезке $[0, \theta]$. Найдите байесовскую оценку параметра θ , если θ имеет априорное распределение а) равномерно

на отрезке $[0, 1]$, б) с плотностью $q(t) = 1/t^2$ при $t \geq 1$. Проверьте полученные оценки на состоятельность.

Решение.

1. При $n > 2$ имеем

$$q(t)p_t(x) = I_{[0,1]}(t) \frac{1}{t^n} I_{0 \leq x_{(1)} \leq x_{(n)} \leq t} \Rightarrow \int_{\Theta} q(t)p_t(x)dt = \int_{x_{(n)}}^1 \frac{dt}{t^n} = \frac{1}{n-1} \left(\frac{1}{x_{(n)}^{n-1}} - 1 \right) \Rightarrow$$

$$\hat{\theta} = (n-1) \frac{X_{(n)}^{n-1}}{1 - X_{(n)}^{n-1}} \int_{X_{(n)}}^1 \frac{dt}{t^{n-1}} = X_{(n)} \frac{n-1}{n-2} \frac{1 - X_{(n)}^{n-2}}{1 - X_{(n)}^{n-1}} = \frac{n-1}{n-2} \left(1 - \frac{1 - X_{(n)}}{1 - X_{(n)}^{n-1}} \right)$$

Так как $X_{(n)} \xrightarrow{\text{п. н.}} \theta < 1$ п.н., то $X_{(n)}^n \xrightarrow{\text{п. н.}} 0 \Rightarrow \hat{\theta} \xrightarrow{\text{п. н.}} \theta \Rightarrow$ оценка состоятельна.

2. Пусть $T(X) = \max(1, X_{(n)})$, тогда

$$q(t)p_t(x) = I_{[1,+\infty)}(t) \frac{1}{t^{n+2}} I_{0 \leq X_{(1)} \leq X_{(n)} \leq t} = \frac{I_{[T(X),+\infty)}(t)}{t^{n+2}} \Rightarrow \int_{\Theta} q(t)p_t(x)dt = \int_{T(X)}^{+\infty} \frac{dt}{t^{n+2}} =$$

$$\frac{1}{(n+1)T(X)^{n+1}} \Rightarrow \hat{\theta} = \frac{n+1}{n} T(X) \xrightarrow{\text{п. н.}} \theta \Rightarrow \text{оценка сильно состоятельна}$$

5. Пусть X_1, \dots, X_n - выборка из нормального распределения с параметрами $(\theta, 1)$. Найдите байесовскую оценку параметра θ , если априорное распределение θ есть $\mathcal{N}(b, \sigma^2)$.

Решение.

$$q(t)p_t(x) = c \exp \left(-\frac{(t-b)^2}{2\sigma^2} - \frac{\sum_{i=1}^n (X_i - t)^2}{2} \right) = c \exp \left(-\frac{1}{2\sigma^2} \left((t-b)^2 + \sigma^2 \sum_{i=1}^n (X_i - t)^2 \right) \right)$$

$$(t-b)^2 + \sigma^2 \sum_{i=1}^n (X_i - t)^2 = t^2(1 + \sigma^2 n) - 2t \left(b + \sigma^2 \sum_{i=1}^n X_i \right) + b^2 + \sigma^2 \sum_{i=1}^n X_i^2 =$$

$$(1 + \sigma^2 n) \left(t - \frac{b + \sigma^2 \sum_{i=1}^n X_i}{1 + \sigma^2 n} \right)^2 + b^2 + \sigma^2 \sum_{i=1}^n X_i^2 - \frac{(b + \sigma^2 \sum_{i=1}^n X_i)^2}{1 + \sigma^2 n} \Rightarrow$$

$$q(t)p_t(x) = \tilde{c} \exp \left(-\frac{1}{2} \frac{1 + \sigma^2 n}{\sigma^2} \left(t - \frac{b + \sigma^2 \sum_{i=1}^n X_i}{1 + \sigma^2 n} \right)^2 \right) \Rightarrow p_{\theta|X} \sim \mathcal{N} \left(\frac{b + \sigma^2 \sum_{i=1}^n X_i}{1 + \sigma^2 n}, \frac{\sigma^2}{1 + \sigma^2 n} \right) \Rightarrow$$

$$\hat{\theta} = \frac{b + \sigma^2 \sum_{i=1}^n X_i}{1 + \sigma^2 n}$$

2. Проверка гипотез и доверительное оценивание

6. X_1, \dots, X_n - выборка из распределения с плотностью

$$p_\theta(x) = \frac{3x^2}{8\theta^3} I_{[0, 2\theta]}(x)$$

С помощью статистики $X_{(1)}$ постройте точный доверительный интервал уровня доверия γ для параметра θ .

Решение. $F_{X_{(1)}}(x) = 1 - (1 - F_{X_1}(x))^n = 1 - (1 - \frac{x^3}{8\theta^3})^n, x \in [0, 2\theta] \Rightarrow F_{\frac{X_{(1)}}{2\theta}}(x) = F_{X_{(1)}}(2\theta x) = 1 - (1 - x^3)^n, x \in [0, 1]$.

$$P_\theta \left(z_{p_1} < \frac{X_{(1)}}{2\theta} < z_{p_2} \right) = P_\theta \left(\frac{X_{(1)}}{2z_{p_2}} < \theta < \frac{X_{(1)}}{2z_{p_1}} \right) = p_2 - p_1 = \gamma$$

Положим $p_2 = 1, p_1 = 1 - \gamma$, тогда $z_{p_2} = 1, z_{p_1} = \sqrt[3]{1 - \sqrt[n]{\gamma}}$

Ответ: $\left(\frac{X_{(1)}}{2}, \frac{X_{(1)}}{2\sqrt[3]{1 - \sqrt[n]{\gamma}}} \right)$

7. X_1, \dots, X_n - выборка, $X_1 = \xi + \eta$, где ξ, η - независимые случайные величины, $\xi \sim R[0, \theta], \eta \sim Bin(1, \theta)$. Постройте доверительный интервал для θ уровня доверия $1 - \alpha$ с помощью неравенства Чебышева.

Решение. $\mathbb{E}X_1 = \mathbb{E}\xi + \mathbb{E}\eta = \theta/2 + \theta = \frac{3\theta}{2}, \mathbb{D}X_1 = \mathbb{D}\xi + \mathbb{D}\eta = \frac{\theta^2}{12} + \theta(1 - \theta) = \theta - \frac{11}{12}\theta^2$

$$P_\theta \left(\left| \sum_{i=1}^n X_i - \frac{3\theta n}{2} \right| \geq \varepsilon n \right) \leq \frac{\theta - \frac{11}{12}\theta^2}{\varepsilon^2 n} \Rightarrow P_\theta \left(\bar{X} - \varepsilon < \frac{3\theta}{2} < \bar{X} + \varepsilon \right) \geq 1 - \frac{\theta - \frac{11}{12}\theta^2}{\varepsilon^2 n} = 1 - \alpha \Rightarrow$$

$$\varepsilon = \sqrt{\frac{\theta - \frac{11}{12}\theta^2}{\alpha n}}, \text{ Получаем } \begin{cases} \bar{X} - \frac{3\theta}{2} < \varepsilon, \\ \bar{X} - \frac{3\theta}{2} > -\varepsilon \end{cases} \Leftrightarrow \left(\bar{X} - \frac{3\theta}{2} \right)^2 < \varepsilon^2 \Leftrightarrow$$

$$\theta^2 \left(\frac{9}{4} + \frac{11}{12\alpha n} \right) - \theta \left(3\bar{X} + \frac{1}{\alpha n} \right) + \bar{X}^2 < 0 \Rightarrow$$

$$\text{Ответ: } \left(\frac{3\bar{X} + \frac{1}{\alpha n} - \sqrt{\left(3\bar{X} + \frac{1}{\alpha n} \right)^2 - 4 \left(\frac{9}{4} + \frac{11}{12\alpha n} \right) \bar{X}^2}}{2 \left(\frac{9}{4} + \frac{11}{12\alpha n} \right)}, \frac{3\bar{X} + \frac{1}{\alpha n} + \sqrt{\left(3\bar{X} + \frac{1}{\alpha n} \right)^2 - 4 \left(\frac{9}{4} + \frac{11}{12\alpha n} \right) \bar{X}^2}}{2 \left(\frac{9}{4} + \frac{11}{12\alpha n} \right)} \right)$$

8. Пусть X_1, \dots, X_n - выборка из гамма-распределения с параметрами (θ, λ) . Постройте асимптотический доверительный интервал для θ уровня доверия α , если а) λ известно, б) λ неизвестно.

Решение.

- а По ЦПТ $\sqrt{n}(\bar{X} - \frac{\lambda}{\theta}) \xrightarrow{d} \mathcal{N}(0, \frac{\lambda}{\theta^2})$ Пусть z_p - квантиль уровня p распределения $\mathcal{N}(0, 1)$.
Так как $\frac{\lambda}{\theta^2}$ непрерывна по θ на \mathbb{R}_+ , то

$$\lim_{n \rightarrow \infty} P\left(\sqrt{n} \frac{\bar{X} - \frac{\lambda}{\theta}}{\frac{\sqrt{\lambda}}{\bar{X}}} < z_\alpha\right) = \alpha$$

Ответ: $\left(0, \frac{\lambda}{\bar{X} - \frac{z_\alpha \sqrt{\lambda}}{\sqrt{n\bar{X}}}}\right)$

- б Найдем состоятельную оценку λ по методу моментов: $\begin{cases} \mathbb{E}X_1 = \frac{\lambda}{\theta} \\ \mathbb{E}X_1^2 = \frac{\lambda}{\theta^2} + \frac{\lambda^2}{\theta^2} \end{cases} \Rightarrow \hat{\lambda} =$
 $\frac{\bar{X}^2}{\bar{X}^2 - \bar{X}}$ Тогда по теореме о наследовании сходимости $\sqrt{n} \frac{\bar{X} - \frac{\lambda}{\theta}}{\frac{\sqrt{\lambda}}{\bar{X}}} \xrightarrow{d} \mathcal{N}(0, 1) \Rightarrow$

Ответ: $\left(0, \frac{\bar{X}^2}{\sqrt{\bar{X}^2 - \bar{X}} \left(\bar{X} \sqrt{\bar{X}^2 - \bar{X}} - \frac{z_\alpha}{\sqrt{n}}\right)}\right)$

9. Имеется X_1 - выборка объема 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.

Решение. Воспользуемся леммой Неймана-Пирсона и построим критерий вида $\{p_1(x) - \lambda p_0(x) \geq 0\}$. $p_1(x) = e^{-x} I_{[0, +\infty)}(x)$, $p_0(x) = I_{[0, 1]}(x)$ Так как мы хотим, чтобы $P_0(X \in S) \leq \alpha$, то

$$P_0(p_1(x) - \lambda p_0(x) \geq 0) = P_0(e^{-x} - \lambda \geq 0) = P_0(x \leq -\ln \lambda) = -\ln \lambda = \alpha \Rightarrow \lambda = e^{-\alpha}$$

$$\begin{aligned} \beta(1, S) &= P_1(X \in S) = P_1(X \in S | x > 1)P_1(x > 1) + P_1(X \in S | x \leq 1)P_1(x \leq 1) = \\ &= e^{-1}P_1(p_1(x) \geq 0 | x > 1) + (1 - e^{-1})P_1(x \leq \alpha | x \leq 1) = 1 + e^{-1} - e^{-\alpha} \end{aligned}$$

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

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

Решение.

$$P_0(X \in S) = P_0(X_{(n)} \leq c\theta_0) = F_{X_{(n)}}(c\theta_0) = c^n = \alpha \Rightarrow c = \sqrt[n]{\alpha}$$

Докажем, что данный критерий р.н.м.к уровня α Если $\theta \leq c\theta_0$, то $P_\theta(X \in S) = 1 \geq P_\theta X \in R) \forall R$ Если $\theta > \theta_0$, то $P_\theta(X \in S) = F_{X_{(n)}}(c\theta_0) + 1 - F_{X_{(n)}}(\theta_0) = 1 - (1 - \alpha) \left(\frac{\theta_0}{\theta}\right)^n$

$$P_\theta(X \notin R) = \frac{1}{\theta^n} \int_{[0, \theta]^n} I_{\bar{R}}(x_1, \dots, x_n) dx_1 \dots dx_n \geq \frac{1}{\theta^n} \int_{[0, \theta_0]^n} I_{\bar{R}}(x_1, \dots, x_n) dx_1 \dots dx_n =$$

$$\left(\frac{\theta_0}{\theta}\right)^n \frac{1}{\theta_0^n} \int_{[0, \theta_0]^n} I_{\bar{R}}(x_1, \dots, x_n) dx_1 \dots dx_n = \left(\frac{\theta_0}{\theta}\right)^n P_{\theta_0}(X \notin R) \geq (1 - \alpha) \left(\frac{\theta_0}{\theta}\right)^n \Rightarrow$$

$$P_\theta(X \in R) \leq 1 - (1 - \alpha) \left(\frac{\theta_0}{\theta}\right)^n = P_\theta(X \in S)$$

Для случая $c\theta_0 < \theta < \theta_0$ аналогичные рассуждения.

11. 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.к. уровня значимости α для проверки гипотезы H_0 против альтернативы 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 $\theta = \theta_0$), $H_1 : \theta < \theta_0$.

Решение.

а Если семейство невозрастает по статистике $T(X)$, то неубывает по статистике $-T(X)$.

Тогда из теоремы 10.1 получаем критерий $\{T(X) \leq -c\}$

б Делая замену $\tilde{\theta} = -\theta$, семейство $\{P_\theta\}$ будет неубывать по статистике $T(X) \Rightarrow$ по теореме критерий $\{T(X) \geq c\}$

12. * Показать, что любой равномерно наиболее мощный несмещённый (т.е. $\inf_{\theta \in \Theta_0} \beta(\theta, S) \geq \sup_{\theta \in \Theta_1} \beta(\theta, S)$) 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.

Решение.

13. Докажите, что в предположении гипотезы $H_0 : F = F_0$ для любого $x \in R$ выполнено

$$F_n^*(x) \xrightarrow[n \rightarrow \infty]{\text{п. н.}} F_0(x)$$

Решение. Пусть $B \in \mathcal{B}(\mathbb{R})$. Тогда $P^*(B) = \sum_{i=1}^n \frac{I_B(x)}{n}$. Функция I_B измеримая, тогда по УЗБЧ $\sum_{i=1}^n \frac{I_B(x)}{n} \xrightarrow{\text{п. н.}} \mathbb{E} I_B = P(x \in B) \Rightarrow$ положив $B = (-\infty, x)$ получаем то, что надо доказать

Теорема 1 (А. Колмогорова, 12.2 из C2). В предположении верности гипотезы $H_0 : F = F_0$ имеет место равенство

$$\lim_{n \rightarrow \infty} P(\sqrt{n} D_n \leq t) = K(t)$$

14. С помощью теоремы 1 докажите состоятельность критерия Колмогорова.

Решение.

15. Докажите, что при условии $0 \leq X_{(1)} \leq X_{(n)} \leq 1$ справедливо равенство

$$\int_0^1 (F_n^*(y) - y)^2 dy = \frac{1}{12n^2} + \frac{1}{n} \sum_{k=1}^n (X_{(k)} - (2k-1)/2n)^2$$

(с помощью этого представления часто вычисляется значение статистики ω^2 , которая используется в критерии Крамера–Мизеса–Смирнова).

Решение.

$$\begin{aligned} \int_0^1 (F_n^*(y) - y)^2 dy &= \int_0^{X_{(1)}} (0 - y)^2 dy + \int_{X_{(1)}}^{X_{(2)}} \left(\frac{1}{n} - y\right)^2 dy + \dots + \int_{X_{(n)}}^1 (1 - y)^2 dy = \\ &= \frac{1}{3} \left(X_{(1)}^3 + \sum_{i=1}^{n-1} \left(X_{(i+1)} - \frac{i}{n} \right)^3 - \left(X_{(i)} - \frac{i}{n} \right)^3 + (1-1)^3 - (X_{(n)} - 1)^3 \right) = \\ &= \frac{1}{3} \left(\sum_{i=1}^n \left(X_{(i)} - \frac{i-1}{n} \right)^3 - \sum_{i=1}^n \left(X_{(i)} - \frac{i}{n} \right)^3 \right) = \\ &= \frac{1}{3} \left(\sum_{i=1}^n \frac{1}{n} \left(X_{(i)}^2 - 2X_{(i)} \frac{i-1}{n} + \frac{i^2 - 2i + 1}{n^2} + X_{(i)}^2 - 2X_{(i)} \frac{i}{n} + \frac{i^2}{n^2} + X_{(i)}^2 - X_{(i)} \frac{2i-1}{n} + \frac{i^2 - i}{n^2} \right) \right) = \\ &= \frac{1}{n} \sum_{i=1}^n \left(X_{(i)}^2 - 2X_{(i)} \frac{2i-1}{2n} + \frac{4i^2 - 4i + 1}{4n^2} + \frac{4i^2 - 8i + 4 + 4i^2 + 4i^2 - 4i - 12i^2 + 12i - 3}{12n^2} \right) = \\ &= \frac{1}{n} \sum_{i=1}^n \left(X_{(i)} - \frac{2i-1}{2n} \right)^2 + \frac{1}{n} \sum_{i=1}^n \frac{1}{12n^2} = \frac{1}{12n^2} + \frac{1}{n} \sum_{k=1}^n (X_{(k)} - (2k-1)/2n)^2 \end{aligned}$$

16. Цифры $0, 1, 2, \dots, 9$ среди 800 первых десятичных знаков числа π появились 74, 92, 83, 79, 80, 73, 77, 75, 76, 91 раз соответственно. С помощью хи-квадрат критерия проверьте гипотезу о согласии этих данных с законом равномерного распределения на множестве $\{0, 1, \dots, 9\}$ на уровне значимости а) 0.05, б) 0.5, в) 0.8.

Решение.

$$\hat{\chi}^2 = \frac{1}{80} (6^2 + 12^2 + 3^2 + 1^2 + 0^2 + 7^2 + 3^2 + 5^2 + 4^2 + 11^2) = 5.125$$

а $\chi_{0.95,9}^2 = 16.9190 \geq \hat{\chi}^2 \Rightarrow$ мы не отвергаем H_0

б $\chi_{0.5,9}^2 = 8.3428 \geq \hat{\chi}^2 \Rightarrow$ мы не отвергаем H_0

в $\chi_{0.2,9}^2 = 5.3801 \geq \hat{\chi}^2 \Rightarrow$ мы не отвергаем H_0 .

17. Среди 5000 семей, имеющих трех детей, есть ровно 1010 семей с тремя мальчиками, 2200 семей с двумя мальчиками и одной девочкой, 950 семей с одним мальчиком и двумя девочками (во всех остальных семьях все дети - девочки). Можно ли с уровнем значимости $\alpha = 0.02$ считать, что количество мальчиков ξ в семье с тремя детьми имеет следующее распределение

$$P(\xi = 0) = \theta, P(\xi = 1) = \theta,$$

$$P(\xi = 2) = 2\theta, P(\xi = 3) = 1 - 4\theta,$$

где $\theta \in (0, 1/4)$?

Решение. Найдем ОМП

$$L(\theta) = \theta^{840} \theta^{950} (2\theta)^{2200} (1 - 4\theta)^{1010} = c\theta^{3990} (1 - 4\theta)^{1010} \Rightarrow l(\theta) = 3990 \ln \theta + 1010 \ln(1 - 4\theta) + \hat{c}$$

$$\frac{d}{d\theta} l(\theta) = \frac{3990}{\theta} - \frac{4 * 1010}{1 - 4\theta} = 0 \Rightarrow \hat{\theta} = \frac{3990}{4 * 5000} = \frac{1995}{2n} \Rightarrow$$

$$np_1(\hat{\theta}) = \frac{1995}{2}, np_2(\hat{\theta}) = \frac{1995}{2}, np_3(\hat{\theta}) = 1995, np_4(\hat{\theta}) = 1010 \Rightarrow$$

$$\hat{\chi}^2 = \frac{2}{1995} \left(\frac{315}{2} \right)^2 + \frac{2}{1995} \left(\frac{95}{2} \right)^2 + \frac{205^2}{1995} + 0 = \frac{6410}{133} \approx 48.2 > \chi_{0.98,3}^2 = 9.83741 \Rightarrow$$

ответ - нельзя.

18. Пусть X_1, \dots, X_n - выборка из распределения: а) $Bern(\theta)$, б) $Pois(\theta)$, в) $\mathcal{N}(\theta, 1)$, г) $\exp(\theta)$. Построить доверительный интервал для параметра θ .

Решение. Будем строить доверительные интервалы уровня доверия $\gamma = 1 - \alpha$

а

$$P_{\theta} \left(\left| \sum_{i=1}^n X_i - \theta n \right| \geq \varepsilon n \right) \leq \frac{\theta(1-\theta)}{\varepsilon^2 n} \Rightarrow P_{\theta} (\bar{X} - \varepsilon < \theta < \bar{X} + \varepsilon) \geq 1 - \frac{\theta(1-\theta)}{\varepsilon^2 n} = 1 - \alpha \Rightarrow$$

$$\varepsilon = \sqrt{\frac{\theta(1-\theta)}{\alpha n}}, \text{ Получаем } (\bar{X} - \theta)^2 < \varepsilon^2 \Leftrightarrow \theta^2 \left(1 + \frac{1}{\alpha n} \right) - \theta \left(2\bar{X} + \frac{1}{\alpha n} \right) + \bar{X}^2 < 0 \Rightarrow$$

$$\text{Ответ: } \left(\frac{2\bar{X} + \frac{1}{\alpha n} - \sqrt{(2\bar{X} + \frac{1}{\alpha n})^2 - 4(1 + \frac{1}{\alpha n})\bar{X}^2}}{2(1 + \frac{1}{\alpha n})}, \frac{2\bar{X} + \frac{1}{\alpha n} + \sqrt{(2\bar{X} + \frac{1}{\alpha n})^2 - 4(1 + \frac{1}{\alpha n})\bar{X}^2}}{2(1 + \frac{1}{\alpha n})} \right)$$

б По аналогии с предыдущей задачей, $\mathbb{E}X = \theta, \mathbb{D}X = \theta$, получаем неравенство $\theta^2 -$

$$\theta \left(2\bar{X} + \frac{1}{\alpha n} \right) + \bar{X}^2 < 0 \Rightarrow$$

$$\text{Ответ: } \left(\frac{2\bar{X} + \frac{1}{\alpha n} + \sqrt{\frac{4\bar{X}}{\alpha n} + \frac{1}{\alpha^2 n^2}}}{2}, \frac{2\bar{X} + \frac{1}{\alpha n} - \sqrt{\frac{4\bar{X}}{\alpha n} + \frac{1}{\alpha^2 n^2}}}{2} \right)$$

с $\sqrt{n}(\bar{X} - \theta) \sim \mathcal{N}(0, 1)$. Пусть z_p - квантиль уровня p распределения $\mathcal{N}(0, 1)$. Тогда $P(z_{\alpha/2} < \sqrt{n}(\bar{X} - \theta) < -z_{\alpha/2}) \Rightarrow \text{Ответ: } \left(\bar{X} + \frac{z_{\alpha/2}}{\sqrt{n}}, \bar{X} - \frac{z_{\alpha/2}}{\sqrt{n}} \right)$

d Заметим, что $nX_{(1)}\theta \sim \text{Exp}(1)$ Пусть z_p - p квантиль распределения $\text{Exp}(1)$. Тогда $P(z_{p_1} < nX_{(1)}\theta < z_{p_2}) = P\left(\frac{z_{p_1}}{nX_{(1)}} < \theta < \frac{z_{p_2}}{nX_{(1)}}\right) \Rightarrow l \sim z_{p_2} - z_{p_1} = \ln \frac{1-p_1}{1-p_2} = \ln \left(1 + \frac{\gamma}{1-p_2}\right) \rightarrow \min_{p_2 \in [\gamma, 1]} \Rightarrow p_2 = \gamma, p_1 = 0 \Rightarrow \text{Ответ: } \left(0, \frac{-\ln \alpha}{nX_{(1)}}\right)$

19. Рассмотрим распределения Коши с плотностью $p_\theta(x) = \frac{1}{\pi(1+(x-\theta)^2)}$. С помощью выборочной медианы построить доверительный интервал для θ^2 .

Решение.

20. Пусть X_1, \dots, X_n - выборка из экспоненциального распределения с параметром θ . Построить равномерно наиболее мощный критерий для проверки гипотезы $H_0 : \theta = \theta_0$ против альтернативы: а) $H_1 : \theta > \theta_0$; б) $H_1 : \theta < \theta_0$

Решение. Пусть $\theta_1 < \theta_2$.

$$\frac{p_{\theta_2}(X)}{p_{\theta_1}(X)} = \left(\frac{\theta_2}{\theta_1}\right)^n \exp\left\{(\theta_1 - \theta_2) \sum_{i=1}^n X_i\right\} \Rightarrow$$

То есть семейство с неубывающим отношением правдоподобия по статистике $T(X) = -\sum_{i=1}^n X_i$. Тогда можем воспользоваться теоремой 10.1

а Найдем $c : P_{\theta_0}(T(X) \geq c) = \alpha, \sum_{i=1}^n X_i \sim \Gamma(n, \theta) \Rightarrow$

$$P_{\theta_0}\left(\sum_{i=1}^n X_i \leq -c\right) = \alpha \Rightarrow c = -\Gamma_\alpha(n, \theta_0) \Rightarrow S = \left\{\sum_{i=1}^n X_i \leq \Gamma_\alpha(n, \theta_0)\right\}$$

б Аналогично получаем $S = \{\sum_{i=1}^n X_i \geq \Gamma_{1-\alpha}(n, \theta_0)\}$

21. Пусть X_1, \dots, X_n - выборка из $\text{Bern}(\theta)$. Проверить гипотезу $H_0 : \theta \leq \theta_0$ против альтернативы $H_1 : \theta > \theta_0$

Решение. Пусть $\theta_2 > \theta_1$.

$$\frac{p_{\theta_2}(X)}{p_{\theta_1}(X)} = \frac{\theta_2^{\sum_{i=1}^n X_i} (1-\theta_2)^{n-\sum_{i=1}^n X_i}}{\theta_1^{\sum_{i=1}^n X_i} (1-\theta_1)^{n-\sum_{i=1}^n X_i}} = c(\theta_1, \theta_2) \left(\frac{\theta_2}{\theta_1} \frac{1-\theta_1}{1-\theta_2}\right)^{\sum_{i=1}^n X_i} \Rightarrow$$

Семейство с возрастающим отношением правдоподобия по статистике $\sum_{i=1}^n X_i$ и критерий $S = \{\sum_{i=1}^n X_i \geq \text{Bin}_{1-\alpha}(n, \theta_0)\}$

3. Линейная регрессия. Проверка линейных гипотез.

22. Пусть

$$X_i = \beta_1 + i\beta_2 + \varepsilon_0 + \dots + \varepsilon_i,$$

$i = 0, 1, \dots, n$, где β_1, β_2 - неизвестные параметры, а $\varepsilon_0, \dots, \varepsilon_n$ - независимые, распределенные по закону $\mathcal{N}(0, \sigma^2)$ случайные величины. Сведите задачу к линейной модели и найдите оценки наименьших квадратов для β_1 и β_2 , а также несмещенную оценку для σ^2 .

Решение. Заметим, что $X_{i+1} - X_i = \beta_2 + \varepsilon_i$. Запишем модель линейной регрессии:

$$\begin{pmatrix} X_0 \\ X_1 - X_0 \\ \vdots \\ X_n - X_{n-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \varepsilon \Rightarrow$$

$$\begin{aligned} Z^T Z &= \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \Rightarrow (Z^T Z)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{n} \end{pmatrix} \Rightarrow \hat{\beta} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{n} \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} X_0 \\ X_1 - X_0 \\ \vdots \\ X_n - X_{n-1} \end{pmatrix} = \\ \begin{pmatrix} X_0 \\ \frac{X_n - X_0}{n} \end{pmatrix}, \hat{\sigma}^2 &= \frac{1}{n+1-2} \left\| \begin{pmatrix} X_0 \\ X_1 - X_0 \\ \vdots \\ X_n - X_{n-1} \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_0 \\ \frac{X_n - X_0}{n} \end{pmatrix} \right\|^2 = \frac{1}{n-1} \sum_{i=1}^n \left(X_i - X_{i-1} - \frac{X_n - X_0}{n} \right)^2 = \\ \frac{1}{n-1} \sum_{i=1}^n \left((X_i - X_{i-1})^2 - 2(X_i - X_{i-1}) \frac{X_n - X_0}{n} + \left(\frac{X_n - X_0}{n} \right)^2 \right) &= \\ \frac{1}{n-1} \left(\sum_{i=1}^n (X_i - X_{i-1})^2 - \frac{(X_n - X_0)^2}{n} \right) \end{aligned}$$

23. Пусть X_1, \dots, X_n - выборка из нормального распределения с параметрами (a, σ^2) . Докажите, что статистики \bar{X} и

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

независимы и вычислите распределение статистики nS^2 .

Решение. Построив модель линейной регрессии, мы обнаружим, что \bar{X} - оценка по МНК

$$\Rightarrow \begin{pmatrix} \bar{X} \\ \vdots \\ \bar{X} \end{pmatrix} = pr_L X, S^2 = \frac{1}{n} \left\| X - \begin{pmatrix} \bar{X} \\ \vdots \\ \bar{X} \end{pmatrix} \right\|^2 = \frac{1}{n} \|X - pr_L X\|^2 = \frac{1}{n} \|pr_{L^\perp} X\|^2, \text{ но по теореме об}$$

ортогональной проекции $pr_L X \perp pr_{L^\perp} X \Rightarrow \bar{X} \perp S^2$ Также по этой теореме $\frac{1}{\sigma^2} \|pr_{L^\perp} X\|^2 \sim \chi^2(n-1) \Rightarrow \frac{nS^2}{\sigma^2} \sim \chi^2(n-1) \Rightarrow nS^2 \sim \sigma^2 \chi^2(n-1)$

24. Пусть X_1, \dots, X_n - выборка из $\mathcal{N}(a, \sigma^2)$ (оба параметра неизвестны). Постройте точные доверительные интервалы для каждого из параметров a, σ^2 .

Решение. Из з. 23 известно, что $\frac{nS^2}{\sigma^2} \sim \chi^2(n-1) \Rightarrow$ доверительный интервал для $\sigma^2 \left(0, \frac{nS^2}{\chi_{1-\gamma}^2(n-1)}\right)$

Рассмотрим $\frac{\bar{X}-a}{\sqrt{S^2}} = \sqrt{n} \frac{\bar{X}-a}{\sigma} \frac{1}{\sqrt{\frac{nS^2}{\sigma^2}}} \sim t(n-1) \Rightarrow$

$$\text{Ответ: } \left(\bar{X} - t_{\frac{1+\gamma}{2}}(n-1) \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}, \bar{X} - t_{\frac{1-\gamma}{2}}(n-1) \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \right)$$

25. Взвешивание трех грузов массами a и b производится следующим образом: n_1 раз взвешивается первый груз (все ошибки измерения имеют распределение $\mathcal{N}(0, \sigma^2)$), n_2 раза взвешивается второй груз на тех же самых весах, затем n_3 раза на других весах взвешиваются первый и второй груз вместе, все ошибки измерения на которых имеют распределение $\mathcal{N}(0, 3\sigma^2)$. Сведите задачу к линейной модели и найдите оценки наименьших квадратов для a и b , а также оптимальную оценку для σ^2 .

Решение.

$$\tilde{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_{n_1} \\ X_{n_1+1} \\ \vdots \\ X_{n_1+n_2} \\ \frac{X_{n_1+n_2+1}}{\sqrt{3}} \\ \vdots \\ \frac{X_{n_1+n_2+n_3}}{\sqrt{3}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \vdots & \vdots \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \varepsilon \Rightarrow Z^T Z = \begin{pmatrix} n_1 + \frac{n_3}{3} & \frac{n_3}{3} \\ \frac{n_3}{3} & n_2 + \frac{n_3}{3} \end{pmatrix} \Rightarrow$$

$$\det Z^T Z = n_1 n_2 + \frac{n_3}{3}(n_1 + n_2) \Rightarrow (Z^T Z)^{-1} = \frac{1}{n_1 n_2 + \frac{n_3}{3}(n_1 + n_2)} \begin{pmatrix} n_2 + \frac{n_3}{3} & -\frac{n_3}{3} \\ -\frac{n_3}{3} & n_1 + \frac{n_3}{3} \end{pmatrix},$$

$$Z^T \tilde{X} = \begin{pmatrix} \sum_{i=1}^{n_1} X_i + \frac{1}{3} \sum_{j=n_1+n_2+1}^{n_1+n_2+n_3} X_j \\ \sum_{k=n_1+1}^{n_1+n_2} X_k + \frac{1}{3} \sum_{j=n_1+n_2+1}^{n_1+n_2+n_3} X_j \end{pmatrix}, \text{ Пусть } S_1 = \sum_{i=1}^{n_1} X_i, S_2 = \sum_{k=n_1+1}^{n_1+n_2} X_k, S_3 = \sum_{j=n_1+n_2+1}^{n_1+n_2+n_3} X_j \Rightarrow$$

$$\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \frac{1}{n_1 n_2 + \frac{n_3}{3}(n_1 + n_2)} \begin{pmatrix} n_2(S_1 + \frac{S_3}{3}) + \frac{n_3}{3}(S_1 - S_2) \\ n_1(S_2 + \frac{S_3}{3}) + \frac{n_3}{3}(S_2 - S_1) \end{pmatrix}, \hat{\sigma}^2 = \frac{1}{n_1 + n_2 + n_3 - 2} \|\tilde{X} - Z\hat{\theta}\|^2$$

Раскрывать я это не буду..

26. Пусть $X_i, i \in \{1, \dots, n\}$ - независимые случайные величины, распределенные по нормальному закону с параметрами $(a + bi, \sigma^2)$. Постройте точные доверительные интервалы для параметров a, b, σ^2 .

Решение. Запишем модель линейной регрессии

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \varepsilon, Z^T Z = \begin{pmatrix} n & \frac{n(n+1)}{2} \\ \frac{n(n+1)}{2} & \frac{n(n+1)(2n+1)}{6} \end{pmatrix} \Rightarrow (Z^T Z)^{-1} = \frac{2}{n(n-1)} \begin{pmatrix} 2n+1 & -3 \\ -3 & \frac{6}{n+1} \end{pmatrix}$$

$$Z^T X = \begin{pmatrix} \sum_{i=1}^n X_i \\ \sum_{i=1}^n i X_i \end{pmatrix} \Rightarrow \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} = \frac{2}{n(n-1)} \begin{pmatrix} \sum_{i=1}^n X_i (2n+1-3i) \\ 3 \sum_{i=1}^n X_i (\frac{2i}{n+1} - 1) \end{pmatrix}$$

Как известно из лекций,

$$\frac{1}{\sigma^2} \left\| X - Z \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \right\|^2 \sim \chi^2(n-2)$$

Тогда

$$P \left(\frac{1}{\sigma^2} \left\| X - Z \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \right\|^2 > \chi_{1-\gamma}^2(n-2) \right) = \gamma \Rightarrow \left(0, \frac{1}{\chi_{1-\gamma}^2(n-2)} \left\| X - Z \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \right\|^2 \right) -$$

точный доверительный интервал для σ^2 уровня доверия γ . Также известно, что

$$\sqrt{\frac{n-2}{\frac{2(2n+1)}{n(n-1)}}} \frac{\hat{a} - a}{\left\| X - Z \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \right\|} \sim t(n-2) \Rightarrow$$

Доверительный интервал для a равен

$$\left(\hat{a} - t_{\frac{1+\gamma}{2}}(n-2) \frac{\left\| X - Z \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \right\|}{\sqrt{\frac{2(2n+1)}{n(n-1)(n-2)}}}, \hat{a} - t_{\frac{1-\gamma}{2}}(n-2) \frac{\left\| X - Z \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \right\|}{\sqrt{\frac{2(2n+1)}{n(n-1)(n-2)}}}, \right)$$

Аналогично для b

$$\left(\begin{array}{c} \left\| X - Z \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \right\| \\ \hat{b} - t_{\frac{1+\gamma}{2}}(n-2) \frac{\left\| X - Z \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \right\|}{\sqrt{\frac{12(n-2)}{n(n-1)(n+1)}}}, \hat{b} - t_{\frac{1-\gamma}{2}}(n-2) \frac{\left\| X - Z \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \right\|}{\sqrt{\frac{12(n-2)}{n(n-1)(n+1)}}}, \end{array} \right)$$

27. X_1, \dots, X_n - выборка из распределения $\mathcal{N}(a_1, \sigma^2)$, Y_1, \dots, Y_m - выборка из распределения $\mathcal{N}(a_2, \sigma^2)$, Z_1, \dots, Z_k - выборка из распределения $\mathcal{N}(a_3, \sigma^2)$. Постройте F -критерий размера α для проверки гипотезы а) $H_0 : a_1 = a_2$ и $a_1 + a_2 = a_3$, б) $H_0 : a_1 = 2a_2$ и $a_1 + 3a_2 = a_3$.

Решение. Запишем модель линейной регрессии:

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \\ Y_1 \\ \vdots \\ Y_m \\ Z_1 \\ \vdots \\ Z_k \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ \vdots & & \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ \vdots & & \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \vdots & & \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \varepsilon, (Z^T Z)^{-1} = \begin{pmatrix} \frac{1}{n} & 0 & 0 \\ 0 & \frac{1}{m} & 0 \\ 0 & 0 & \frac{1}{k} \end{pmatrix}, \hat{a} = \begin{pmatrix} \bar{X} \\ \bar{Y} \\ \bar{Z} \end{pmatrix}$$

$$\hat{\sigma}^2 = \frac{1}{n+m+k-3} \left(\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2 + \sum_{i=1}^k (Z_i - \bar{Z})^2 \right)$$

$$\text{а) } A = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & -1 \end{pmatrix}, D = \begin{pmatrix} \frac{1}{n} + \frac{1}{m} & \frac{1}{n} - \frac{1}{m} \\ \frac{1}{n} - \frac{1}{m} & \frac{1}{n} + \frac{1}{m} + \frac{1}{k} \end{pmatrix}, \hat{\beta} = \begin{pmatrix} \hat{a}_1 - \hat{a}_2 \\ \hat{a}_1 + \hat{a}_2 - \hat{a}_3 \end{pmatrix}$$

$$(\det D)^{-1} = \frac{mnk}{3k+m+n} \Rightarrow D^{-1} = \frac{1}{3k+m+n} \begin{pmatrix} mk+nk+nm & nk-mk \\ nk-mk & mk+nk \end{pmatrix} \Rightarrow$$

$$S_F = \left\{ \begin{pmatrix} \bar{X} - \bar{Y} & \bar{X} + \bar{Y} - \bar{Z} \end{pmatrix} \begin{pmatrix} mk+nk+nm & nk-mk \\ nk-mk & mk+nk \end{pmatrix} \begin{pmatrix} \bar{X} - \bar{Y} \\ \bar{X} + \bar{Y} - \bar{Z} \end{pmatrix} \geq \right.$$

$$\left. \left(\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2 + \sum_{i=1}^k (Z_i - \bar{Z})^2 \right) \frac{2(3k+n+m)}{m+n+k-3} f_{1-\alpha}(2, n+m+k-3) \right\}$$

$$\text{b } A = \begin{pmatrix} 1 & -2 & 0 \\ 1 & 3 & -1 \end{pmatrix}, D = \begin{pmatrix} \frac{1}{n} + \frac{4}{m} & \frac{1}{n} - \frac{6}{m} \\ \frac{1}{n} - \frac{6}{m} & \frac{1}{n} + \frac{9}{m} + \frac{1}{k} \end{pmatrix}, \hat{\beta} - \beta_0 = \begin{pmatrix} \bar{X} - 2\bar{Y} \\ \bar{X} + 3\bar{Y} - \bar{Z} \end{pmatrix}$$

$$(\det D)^{-1} = \frac{mnk}{25k + m + 4n} \Rightarrow D^{-1} = \frac{1}{25k + m + 4n} \begin{pmatrix} mk + 9nk + nm & 6nk - mk \\ 6nk - mk & mk + 4nk \end{pmatrix} \Rightarrow$$

$$S_F = \left\{ \begin{pmatrix} \bar{X} - 2\bar{Y} & \bar{X} + 3\bar{Y} - \bar{Z} \end{pmatrix} \begin{pmatrix} mk + 9nk + nm & 6nk - mk \\ 6nk - mk & mk + 4nk \end{pmatrix} \begin{pmatrix} \bar{X} - 2\bar{Y} \\ \bar{X} + 3\bar{Y} - \bar{Z} \end{pmatrix} \geq \right.$$

$$\left. \left(\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2 + \sum_{i=1}^k (Z_i - \bar{Z})^2 \right) \frac{2(25k + m + 4n)}{m + n + k - 3} f_{1-\alpha}(2, n + m + k - 3) \right\}$$

28. Пусть $X_i \sim \mathcal{N}(a, i\sigma^2)$, $i = 1, \dots, n$, $Y_j \sim \mathcal{N}(jb, \sigma^2)$, $j = 1, \dots, m$, - независимые случайные величины, где a, b, σ^2 - неизвестные параметры. Сведите задачу к линейной модели и постройте F -критерий размера α для проверки гипотезы $H_0 : a + b = 1$.

Решение.

$$\begin{pmatrix} \frac{X_1}{\sqrt{1}} \\ \vdots \\ \frac{X_n}{\sqrt{n}} \\ Y_1 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{1}} & 0 \\ \vdots & \vdots \\ \frac{1}{\sqrt{n}} & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & m \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \varepsilon, (Z^T Z)^{-1} = \text{diag} \left(\frac{2}{n(n+1)}, \frac{6}{m(m+1)(2m+1)} \right),$$

$$\hat{\theta} = \begin{pmatrix} \frac{2}{n(n+1)} \sum_{i=1}^n \frac{X_i}{i} \\ \frac{6}{m(m+1)(2m+1)} \sum_{j=1}^m jY_j \end{pmatrix}$$

$$\hat{\sigma}^2 = \frac{1}{n+m-2} \left(\sum_{k=1}^n \frac{1}{k} \left(X_k - \frac{2}{n(n+1)} \sum_{i=1}^n \frac{X_i}{i} \right)^2 + \sum_{p=1}^m \left(Y_p - \frac{6p}{m(m+1)(2m+1)} \sum_{j=1}^m jY_j \right)^2 \right)$$

$$A = \begin{pmatrix} 1 & -1 \end{pmatrix} \Rightarrow D = \frac{2}{n(n+1)} + \frac{6}{m(m+1)(2m+1)} \Rightarrow$$

$$\frac{n+m-2}{\frac{2}{n(n+1)} + \frac{6}{m(m+1)(2m+1)}} \frac{\left(\frac{2}{n(n+1)} \sum_{i=1}^n \frac{X_i}{i} + \frac{6}{m(m+1)(2m+1)} \sum_{j=1}^m jY_j - 1 \right)^2}{\sum_{k=1}^n \frac{1}{k} \left(X_k - \frac{2}{n(n+1)} \sum_{i=1}^n \frac{X_i}{i} \right)^2 + \sum_{p=1}^m \left(Y_p - \frac{6p}{m(m+1)(2m+1)} \sum_{j=1}^m jY_j \right)^2} \sim$$

$$F(1, n+m-2) \Rightarrow S_F = \left\{ \frac{\left(\frac{2}{n(n+1)} \sum_{i=1}^n \frac{X_i}{i} + \frac{6}{m(m+1)(2m+1)} \sum_{j=1}^m jY_j - 1 \right)^2}{\sum_{k=1}^n \frac{1}{k} \left(X_k - \frac{2}{n(n+1)} \sum_{i=1}^n \frac{X_i}{i} \right)^2 + \sum_{p=1}^m \left(Y_p - \frac{6p}{m(m+1)(2m+1)} \sum_{j=1}^m jY_j \right)^2} \geq \right.$$

$$\left. \frac{\frac{2}{n(n+1)} + \frac{6}{m(m+1)(2m+1)}}{n+m-2} f_{1-\alpha}(1, n+m-2) \right\}$$

29. Используя метод линейной регрессии, постройте приближение функции $f(x)$ многочленом третьей степени по следующим данным:

$f(x_i)$	3.9	5.0	5.7	6.5	7.1	7.6	7.8	8.1	8.4
x_i	4.0	5.2	6.1	7.0	7.9	8.6	8.9	9.5	9.9

Решение. Пусть искомым многочлен равен $ax^3 + bx^2 + cx + d$. Тогда модель линейной регрессии выглядит следующим образом:

$$\begin{pmatrix} f(x_1) \\ \vdots \\ f(x_9) \end{pmatrix} = \begin{pmatrix} x_1^3 & x_1^2 & x_1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_9^3 & x_9^2 & x_9 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} + \varepsilon$$

Если считать это руками, то уйдет миллион лет и, кстати, если вычислять точно в виде дробей, оценивая $a/1000, b/100, c/10, d$, то получится такая матрица $(Z^T Z)^{-1}$

Input

$$\begin{pmatrix} 3014244005669 & 33620996231 & 381630821 & 4436327 \\ 33620996231 & 381630821 & 4436327 & 53309 \\ 381630821 & 4436327 & 53309 & 671 \\ 4436327 & 53309 & 671 & 9 \end{pmatrix}^{-1}$$

(matrix inverse)

Enlarge

Result

Enlarge

Data

Customize

Plain Text

$$\frac{1}{63889272833421192} \begin{pmatrix} 271735636 & -56787546343 & 3778438005837 & -79283636974068 \\ -56787546343 & 11939857168089 & -799801239296084 & 16899212322135336 \\ 3778438005837 & -799801239296084 & 53990447527625285 & -1150374729835829442 \\ -79283636974068 & 16899212322135336 & -1150374729835829442 & 24757040297579835354 \end{pmatrix}$$

Поэтому приведем программу на python:

```
import numpy as np

X = [4.0, 5.2, 6.1, 7.0, 7.9, 8.6, 8.9, 9.5, 9.9]
F = [3.9, 5.0, 5.7, 6.5, 7.1, 7.6, 7.8, 8.1, 8.4]
Z = list(map(lambda x: [x**3, x**2, x, 1], X))
Z_np = np.array(Z)
F_np = np.array(F)
theta = np.linalg.inv(Z_np.T @ Z_np) @ Z_np.T @ F_np
a, b, c, d = theta
print(a, b, c, d)
```

и получим

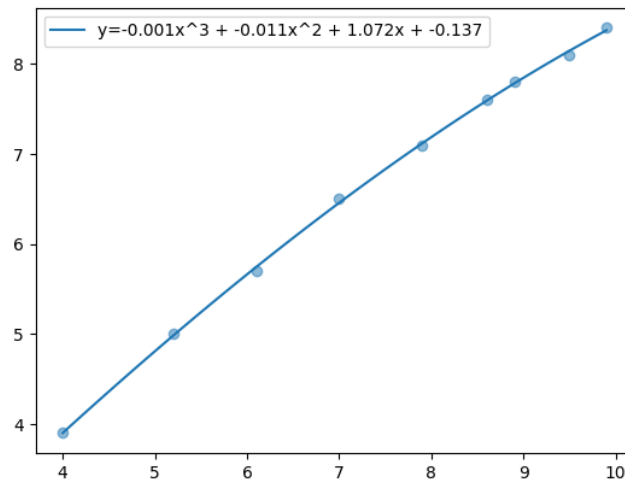
$$(a, b, c, d) = (-0.00100145912636040, -0.0114849959500949, 1.07154905548792, -0.137289441132785)$$

Вроде неплохо:

```
import matplotlib.pyplot as plt

plt.scatter(X, F, alpha=0.5)
xseq = np.linspace(min(X), max(X), num=100)
plt.plot(xseq, a * xseq**3 + b * xseq**2 + c * xseq + d,
         label=f'y={round(a, 3)}x^3 + {round(b, 3)}x^2 + {round(c, 3)}x + {round(d, 3)}')

plt.legend()
plt.show()
```



30. Убедиться в том, что наиболее мощный критерий для различения двух простых гипотез о симметричном относительно нуля распределении наблюдаемой случайной величины ξ $H_0 \mathcal{L}(\xi) = R[-a, a]$ и $H_1 \mathcal{L}(\xi) = \mathcal{N}(0, \sigma^2)$ (a и σ известны) имеет для больших выборок следующую асимптотическую форму

$$\mathfrak{X}_{1,a}^* = \left\{ \sum_{i=1}^n X_i^2 \leq \frac{n}{3}a^2 + \zeta_a \frac{2a^2}{3} \sqrt{\frac{n}{5}} \right\}, \Phi(\zeta_a) = a$$

Указание. Воспользоваться центральной предельной теоремой при отыскании распределения тестовой статистики.

Решение.

31. В последовательности независимых испытаний с двумя исходами вероятность “успеха” равна p . Построить критерий проверки гипотезы $H_0 p = 0$ против альтернативы $H_1 p = 0.01$

и определить наименьший объем выборки, при котором вероятности ошибок 1-го и 2-го родов не превышают 0.01.

Решение. По лемме Неймана-Пирсона наиболее мощный критерий $S_\lambda = \{p_1(X) - \lambda p_0(X) \geq 0\}$ Тогда вероятность ошибки первого рода равна $P_0((0.99)^n \geq \lambda) = I\{0.99^n \geq \lambda\} \Rightarrow \lambda$ должна быть такой, чтобы $0.99^n < \lambda$. Заметим, что тогда вероятность ошибки первого рода равна 0. Вероятность ошибки второго рода равна $P_1(0.01^{\sum_{i=1}^n X_i} 0.99^{n-\sum_{i=1}^n X_i} < \lambda 0^{\sum_{i=1}^n X_i})$. Если $\sum_{i=1}^n X_i = 0$, то $X \notin S(X)$ так как считаем, что выполнено $0.99^n < \lambda$, а если $\sum_{i=1}^n X_i \neq 0$, то $X \in S(X) \Rightarrow P_1(X \notin S(X)) = P_1(\sum_{i=1}^n X_i = 0) = 0.99^n \leq 0.01 \Rightarrow n = \log_{0.99}(0.01) = 459$. λ можно положить 1, итого критерий $\{p_1(X) - p_0(X) \geq 0\}$.

32. Имеется 2 гири с весами θ_1 и θ_2 . На одних и тех же весах сначала взвесили первую гирю, затем вторую, а потом обе сразу. Найти оценку наименьших квадратов для θ_1 и θ_2 и несмещенную оценку дисперсии ошибки измерений. Проверьте гипотезы: а) $H_0 : \theta_1 = \theta_2$; б) $H_0 : 2\theta_1 = 3\theta_2$.

Решение. Воспользуемся критерием Фишера для линейных гипотез, модель линейной регрессии:

$$X = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \theta + \varepsilon, (Z^T Z)^{-1} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \Rightarrow \hat{\theta} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} X_1 + X_3 \\ X_2 + X_3 \end{pmatrix} = \begin{pmatrix} \frac{2X_1 - X_2 + X_3}{3} \\ \frac{2X_2 - X_1 + X_3}{3} \end{pmatrix}$$

$$\hat{\sigma}^2 = \frac{1}{3-2} \|X - Z\hat{\theta}\|^2 = \frac{1}{9} ((2X_1 - X_2 + X_3)^2 + (2X_2 - X_1 + X_3)^2 + (X_1 + X_2 + 2X_3)^2) =$$

$$\frac{2}{3} (X_1^2 + X_2^2 + X_3^2 - X_1X_2 + X_1X_3 + X_2X_3)$$

а $A_1 = \begin{pmatrix} 1 & -1 \end{pmatrix}, \beta_0 = 0, \hat{\beta} = \hat{\theta}_1 - \hat{\theta}_2 = X_1 - X_2, D = A_1(Z^T Z)^{-1}A_1^T = 2 \Rightarrow D^{-1} = \frac{1}{2}$ Тогда $\frac{\frac{3}{4} \frac{X_1 - X_2}{X_1^2 + X_2^2 + X_3^2 - X_1X_2 + X_1X_3 + X_2X_3}}{\sim F(1, 1)}$ Тогда критерий уровня значимости α равен $\left\{ \frac{X_1 - X_2}{X_1^2 + X_2^2 + X_3^2 - X_1X_2 + X_1X_3 + X_2X_3} \geq \frac{4}{3} f_{1-\alpha}(1, 1) \right\}$

б $A_2 = \begin{pmatrix} 2 & -3 \end{pmatrix}, \hat{\beta} = 2\hat{\theta}_1 - 3\hat{\theta}_2 = \frac{7X_1 - 8X_2 - X_3}{3}, \beta_0 = 0, D = A_2(Z^T Z)^{-1}A_2^T = \frac{38}{3} \Rightarrow$ Критерий уровня значимости α равен $\left\{ \frac{7X_1 - 8X_2 - X_3}{X_1^2 + X_2^2 + X_3^2 - X_1X_2 + X_1X_3 + X_2X_3} \geq \frac{3}{76} f_{1-\alpha}(1, 1) \right\}$