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. $X_1 \sim Exp(\theta)$, 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 $\overline{X} \xrightarrow{\text{п. н.}} \frac{1}{\theta}$. По теореме о наследовании сходимости $\theta^* = \frac{1+\frac{\beta}{n}}{\frac{\alpha}{n}+\overline{X}} \xrightarrow{\text{п. н.}} \theta \Rightarrow$ оценка сильно состоятельна.

2. По выборке $X_1, \ldots, X_n$ 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.

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

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

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

Решение.

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

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

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

По УЗБЧ $\overline{X} \xrightarrow{\text{п. н.}} \theta \Rightarrow$ по теореме о наследовании сходимости $\hat{\theta} \xrightarrow{\text{п. н.}} 1 \neq \theta \Rightarrow$ оценка несостоятельна.

4. Пусть $X_1, \ldots, X_n$ - 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 $X_{(n)} \xrightarrow{\text{п. н.}} \theta$, то $\hat{\theta} \xrightarrow{\text{п. н.}} 1 \Rightarrow$ оценка несостоятельна.

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

Решение.

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

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

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

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

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

2. Проверка гипотез и доверительное оценивание

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

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

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

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

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

8. Пусть $X_1, \ldots, X_n$ - 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.

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

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

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

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

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

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. Воспользуемя 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

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

$$\beta(1,S) = P_1(X \in S) = P_1(X \in S|x > 1)P_1(x > 1) + P_1(X \in S|x \leq 1)P_1(x \leq 1) =$$

$$e^{-1}P_1(p_1(x) \geq 0|x > 1) + (1 - e^{-1})P_1(x \leq \alpha|x \leq 1) = 1 + e^{-1} - e^{-\alpha}$$

10. Пусть $X_1, \ldots, X_n$ - 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

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

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 $P_\theta(X \in S) = F_{X_{(n)}}(c\theta_0) + 1 - F_{X_{(n)}}(\theta_0) = 1 - (1-\alpha)\left(\frac{\theta}{\theta_0}\right)^n$

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

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

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

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.

a Если семейство невозрастает по статистике $T(X)$, то неубывает по статистике $-T(X)$. Тогда из теоремы 10.1 получаем критерий $\{T(X) \leq -c\}$

b 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

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

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 С2)*.* В предположении верности гипотезы $H_0 : F = F_0$ имеет место равенство

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

14. С помощью теоремы 1 докажите состоятельность критерия Колмогорова.

Решение.

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

$$\int_0^1 (F_n^*(y) - y)^2 dy = \int_0^{X_{(1)}} (0-y)^2 dy + \int_{X_{(1)}}^{X_{(2)}} \left(\frac{1}{n} - y\right)^2 dy + \dots + \int_{X_{(n)}}^{1} (1-y)^2 dy =$$

$$\frac{1}{3}\left(X_{(1)}^3 + \sum_{i=1}^{n-1}\left(X_{(i+1)} - \frac{i}{n}\right)^3 - \left(X_{(i)} - \frac{i}{n}\right)^3 + (1-1)^3 - (X_{(n)} - 1)^3\right) =$$

$$\frac{1}{3}\left(\sum_{i=1}^{n}\left(X_{(i)} - \frac{i-1}{n}\right)^3 - \sum_{i=1}^{n}\left(X_{(i)} - \frac{i}{n}\right)^3\right) =$$

$$\frac{1}{3}\left(\sum_{i=1}^{n}\frac{1}{n}\left(X_{(i)}^2 - 2X_{(i)}\frac{i-1}{n} + \frac{i^2-2i+1}{n^2} + X_{(i)}^2 - 2X_{(i)}\frac{i}{n} + \frac{i^2}{n^2}\right)\right) =$$

$$\frac{2}{3n}\sum_{i=1}^{n}\left(X_{(i)}^2 - 2X_{(i)}\frac{2i-1}{2n} + \frac{4i^2-4i+1}{4n^2} + \frac{2i^2-4i+2+2i^2-4i^2+4i-1}{4n^2}\right) =$$

$$\frac{2}{3n}\sum_{i=1}^{n}\left(X_{(i)} - \frac{2i-1}{2n}\right)^2 + \frac{2}{3n}\sum_{i=1}^{n}\frac{1}{4n^2} = \frac{1}{6n^2} + \frac{2}{3n}\sum_{k=1}^{n}\left(X_{(k)} - (2k-1)/2n\right)^2$$

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.

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

a $\chi^2_{0.95,9} = 16.9190 \le \hat{\chi^2} \Rightarrow$ мы не отвергаем H_0

b $\chi^2_{0.5,9} = 8.3428 \le \hat{\chi^2} \Rightarrow$ мы не отвергаем H_0

c $\chi^2_{0.2,9} = 5.3801 \le \hat{\chi^2} \Rightarrow$ 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

$$L(\theta) = \theta^{800}\theta^{950}(2\theta)^{2200}(1-4\theta)^{1010} = c\theta^{3950}(1-4\theta)^{1010} \Rightarrow l(\theta) = 3950 \ln\theta + 1010 \ln(1-4\theta) + \hat{c}$$

$$\frac{d}{d\theta} l(\theta) = \frac{3950}{\theta} - \frac{4 * 1010}{1-4\theta} = 0 \Rightarrow \hat{\theta} = \frac{3950}{4 * 5000} = \frac{1975}{2n} \Rightarrow$$

$$np_1(\hat{\theta}) = \frac{1975}{2}, np_2(\hat{\theta}) = \frac{1975}{2}, np_3(\hat{\theta}) = 1975, np_4(\hat{\theta}) = 1010 \Rightarrow$$

$$\hat{\chi^2} = \frac{2}{1975}\left(\frac{375}{2}\right)^2 + \frac{2}{1975}\left(\frac{75}{2}\right)^2 + \frac{225^2}{1975} + 0 = \frac{4950}{79} \approx 62.66 > \chi^2_{0.99,3} = 11.34 \Rightarrow$$

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 $\gamma = 1 - \alpha$

a

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

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

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

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

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

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

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

d Заметим, что $nX_{(1)}\theta \sim Exp(1)$ Пусть z_p - p квантиль распределения $Exp(1)$. Тогда $P(z_{p_1}<nX_{(1)}\theta<z_{p_2})=P(\frac{z_{p_1}}{nX_{(1)}}<\theta<\frac{z_{p_2}}{nX_{(1)}}) \Rightarrow l \sim z_{p_2}-z_{p_1}=\ln\frac{1-p_1}{1-p_2}=\ln\left(1+\frac{\gamma}{1-p_2}\right) \to \min_{p_2\in[\gamma,1]} \Rightarrow p_2=\gamma, p_1=0 \Rightarrow$ 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 $X_1,\ldots,X_n$ - 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

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

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

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

21. Пусть $X_1,\ldots,X_n$ - выборка из $Bern(\theta)$. 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 $S = \{\sum_{i=1}^{n} X_i \geq Bin_{1-\alpha}(n, \theta_0)\}$

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:

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

$$Z^T Z = \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \Rightarrow (Z^T Z)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{n} \end{pmatrix} \Rightarrow \hat{\beta} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{n} \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} X_0 \\ X_1 - X_0 \\ \vdots \\ X_n - X_{n-1} \end{pmatrix} =$$

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

$$\frac{1}{n-1} \sum_{i=1}^{n} \left((X_i - X_{i-1})^2 - 2(X_i - X_{i-1}) \frac{X_n - X_0}{n} + \left(\frac{X_n - X_0}{n} \right)^2 \right) =$$

$$\frac{1}{n-1} \left(\sum_{i=1}^{n} (X_i - X_{i-1})^2 - \frac{(X_n - X_0)^2}{n} \right)$$

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

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

независимы и вычислите распределение статистики nS^2.

Решение. Построив модель линейной регрессии, мы обнаружим, что $\overline{X}$ - оценка по МНК $\Rightarrow \begin{pmatrix} \overline{X} \\ \vdots \\ \overline{X} \end{pmatrix} = pr_L X, S^2 = \frac{1}{n} \left\| X - \begin{pmatrix} \overline{X} \\ \vdots \\ \overline{X} \end{pmatrix} \right\|^2 = \frac{1}{n}\|X - pr_L X\|^2 = \frac{1}{n}\|pr_{L^\perp} X\|^2$, но по теореме об ортогональной проекции $pr_L X \perp\!\!\!\perp pr_{L^\perp} X \Rightarrow \overline{X} \perp\!\!\!\perp S^2$ 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 $X_1, \ldots, X_n$ - 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 σ^2 $\left(0, \frac{nS^2}{\chi^2_{1-\gamma}(n-1)}\right)$
Рассмотрим $\frac{\overline{X}-a}{\sqrt{nS^2}} = \frac{\overline{X}-a}{\sigma}\frac{1}{\sqrt{\frac{nS^2}{\sigma^2}}} \sim t(n-1) \Rightarrow$
Ответ: $\left(\overline{X} - t_{\frac{1+\gamma}{2}}(n-1)\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2}, \overline{X} - t_{\frac{1-\gamma}{2}}(n-1)\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2}\right)$

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.

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

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

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

$$\begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix} = \frac{1}{n_1n_2 + \frac{n_3}{3}(n_1+n_2)}\begin{pmatrix} n_2(S_1 + \frac{S_3}{3}) + \frac{n_3}{3}(S_1 - S_2)\\ n_1(S_2 + \frac{S_3}{3}) + \frac{n_3}{3}(S_2 - S_1)\end{pmatrix}$$

26. Пусть $X_i, i \in \{1, \ldots, n\}$ - 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

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

$$Z^TX = \begin{pmatrix}\sum_{i=1}^n X_i\\ \sum_{i=1}^n iX_i\end{pmatrix} \Rightarrow \begin{pmatrix}\hat{a}\\ \hat{b}\end{pmatrix} = \frac{2}{n(n-1)}\begin{pmatrix}\sum_{i=1}^n X_i\,(2n+1-3i)\\ 3\sum_{i=1}^n X_i\left(\frac{2i}{n+1} - 1\right)\end{pmatrix}$$

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

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

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

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

27. $X_1, \ldots, X_n$ - выборка из распределения $\mathcal{N}(a_1, \sigma^2), Y_1, \ldots, Y_m$ - выборка из распределения $\mathcal{N}(a_2, \sigma^2), Z_1, \ldots, Z_k$ - 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:

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

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

а $A = \begin{pmatrix} 1 & -1 & 0\\ 1 & 1 & -1 \end{pmatrix}, D = \begin{pmatrix} \frac{1}{n} + \frac{1}{m} & \frac{1}{n} - \frac{1}{m}\\ \frac{1}{n} - \frac{1}{m} & \frac{1}{n} + \frac{1}{m} + \frac{1}{k} \end{pmatrix}, \hat{\beta} = \begin{pmatrix} \hat{a_1} - \hat{a_2}\\ \hat{a_1} + \hat{a_2} - \hat{a_3} \end{pmatrix}$

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

$$S_F = \{\begin{pmatrix}\overline{X} - \overline{Y} & \overline{X} + \overline{Y} - \overline{Z}\end{pmatrix}\begin{pmatrix} mk + nk + nm & nk - mk\\ nk - mk & mk + nk \end{pmatrix}\begin{pmatrix}\overline{X} - \overline{Y}\\ \overline{X} + \overline{Y} - \overline{Z}\end{pmatrix} \geq$$

$$\frac{2(3k+n+m)}{m+n+k-3}f_{1-\alpha}(2, n+m+k-3)\}$$

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

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

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

$$\frac{2(25k + m + 4n)}{m + n + k - 3} f_{1-\alpha}(2, n + m + k - 3)\}$$

28. Пусть $X_i \sim \mathcal{N}(a, i\sigma^2)$, $i = 1, \ldots, n$, $Y_j \sim \mathcal{N}(jb, \sigma^2)$, $j == 1, \ldots, m$, - 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.

$$\begin{pmatrix} \frac{X_1}{\sqrt{1}} \\ \vdots \\ \frac{X_n}{\sqrt{n}} \\ \frac{Y_1}{\sqrt{1}} \\ \vdots \\ \frac{Y_m}{\sqrt{m}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{1}} & 0 \\ \vdots & \vdots \\ \frac{1}{\sqrt{n}} & 0 \\ 0 & \frac{1}{\sqrt{1}} \\ \vdots & \vdots \\ 0 & \frac{1}{\sqrt{m}} \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} + \varepsilon, (Z^T Z)^{-1} = diag\left(\frac{2}{n(n+1)}, \frac{2}{m(m+1)}\right), \hat{\theta} = \begin{pmatrix} \frac{2}{n(n+1)} \sum_{i=1}^n \frac{X_i}{i} \\ \frac{2}{m(m+1)} \sum_{j=1}^m \frac{Y_j}{j} \end{pmatrix}$$

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

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

$$\frac{n + m - 2}{\frac{2}{n(n+1)} + \frac{2}{m(m+1)}} \frac{\left(\frac{2}{n(n+1)} \sum_{i=1}^n \frac{X_i}{i} + \frac{2}{m(m+1)} \sum_{j=1}^m \frac{Y_j}{j} - 1 \right)^2}{\sum_{k=1}^n \frac{1}{k} \left(X_k - \sum_{i=1}^n \frac{X_i}{i} \right)^2 + \sum_{p=1}^m \frac{1}{p} \left(X_p - \sum_{j=1}^m \frac{Y_j}{j} \right)^2} \sim F(1, n + m - 2) \Rightarrow$$

$$S_F = \left\{ \frac{\left(\frac{2}{n(n+1)} \sum_{i=1}^n \frac{X_i}{i} + \frac{2}{m(m+1)} \sum_{j=1}^m \frac{Y_j}{j} - 1 \right)^2}{\sum_{k=1}^n \frac{1}{k} \left(X_k - \sum_{i=1}^n \frac{X_i}{i} \right)^2 + \sum_{p=1}^m \frac{1}{p} \left(X_p - \sum_{j=1}^m \frac{Y_j}{j} \right)^2} \geq \frac{\frac{2}{n(n+1)} + \frac{2}{m(m+1)}}{n + m - 2} f_{1-\alpha}(1, n + m - 2) \right\}$$

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 искомый 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 вычилять 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

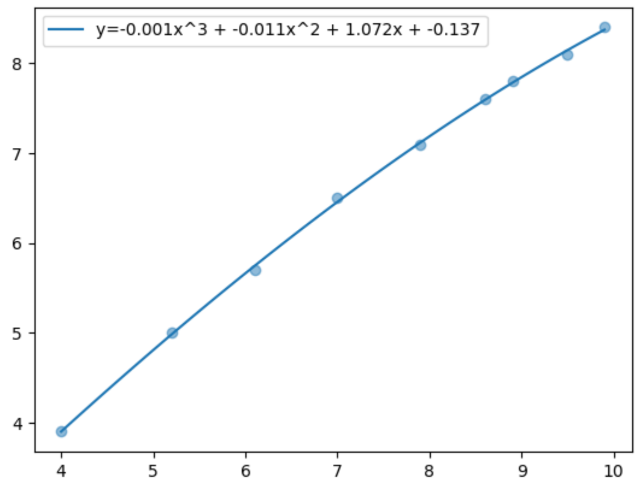
$$(a, b, c, d) = (-0.00100145912636040, -0.0114849959500949, 1.07154905548792, -0.137289441132785)$$

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 относитлтеьно нуля распределении набдюдаемой 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

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

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

Решение.

31. В последовательности независимых испытаний с двумя исходами вероятсноть "успеха" равна p. Построить критей 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 гирьки с весами θ_1 и θ_2. На одних и тех же весах сначала взвесили первую гирьку, 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:

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

$$\hat{\sigma^2} = \frac{1}{3-2} \left\| X - Z\hat{\theta} \right\|^2 = \frac{1}{9} \left((2X_1 - X_2 + X_3)^2 + (2X_2 - X_1 + X_3)^2 + (X_1 + X_2 + 2X_3)^2 \right) =$$
$$\frac{2}{3}(X_1^2 + X_2^2 + X_3^2 - X_1X_2 + X_1X_3 + X_2X_3)$$

a $A_1 = \begin{pmatrix} 1 & -1 \end{pmatrix}, \beta_0 = 0, \hat{\beta} = \hat{\theta_1} - \hat{\theta_2} = X_1 - X_2, D = A_1(Z^T Z)^{-1} A_1^T = 2 \Rightarrow D^{-1} = \frac{1}{2}$ Тогда $\frac{3}{4} \frac{X_1 - X_2}{X_1^2 + X_2^2 + X_3^2 - X_1X_2 + X_1X_3 + X_2X_3} \sim F(1,1)$ Тогда критерий уровня значимости α равен $\left\{ \frac{X_1 - X_2}{X_1^2 + X_2^2 + X_3^2 - X_1X_2 + X_1X_3 + X_2X_3} \geq \frac{4}{3} f_{1-\alpha}(1,1) \right\}$

b $A_2 = \begin{pmatrix} 2 & -3 \end{pmatrix}, \hat{\beta} = 2\hat{\theta_1} - 3\hat{\theta_2} = \frac{7X_1 - 8X_2 - X_3}{3}, \beta_0 = 0, D = A_2(Z^T Z)^{-1} A_2^T = \frac{38}{3} \Rightarrow$ Критерий уровня значимости α равен $\left\{ \frac{7X_1 - 8X_2 - X_3}{X_1^2 + X_2^2 + X_3^2 - X_1X_2 + X_1X_3 + X_2X_3} \geq \frac{3}{76} f_{1-\alpha}(1,1) \right\}$