

# 1 1

## 1.1

Če zapišemo gostote v eksponentni obliki, dobimo

$$f(x; a, b) = \frac{b^a}{\Gamma(a)} e^{-\langle (a+1, b), (\ln(x), 1/x) \rangle} \cdot \mathbb{1}_{(0, \infty)}(x),$$

torej dobimo po Neuman-Fischerjevem faktorizacijskem izreku, da je

$$T(x) = (\ln(x), 1/x)$$

zadostna statistika za ta model.

## 1.2

Ker smo v parametričnem modelu z zvezno odvedljivimi gostotami, lahko iščemo stacionarne točke gostot z logaritmčno enačbo verjetja. Računamo

$$l(a, b) := \ln(f(x; a, b)) = a \ln(b) - \ln(\Gamma(a)) - (a+1) \ln(x) - \frac{b}{x}$$

$$\frac{\partial l}{\partial a}(a, b) = \ln(b) - \frac{\Gamma'(a)}{\Gamma(a)} - \ln(x)$$

$$\frac{\partial l}{\partial b}(a, b) = \frac{a}{b} - \frac{1}{x}.$$

Zapišemo lahko logaritemsko enačbo verjetja

$$\ln(b) - \frac{\Gamma'(a)}{\Gamma(a)} - \ln(x) = \frac{a}{b} - \frac{1}{x} = 0,$$

iz česar takoj dobimo  $b = ax$ , za drugo enačbo pa potem sledi, da rešujemo

$$\Psi(a) := \ln(a) - \frac{\Gamma'(a)}{\Gamma(a)} = 0.$$

Po posledici Bernsteinovega izreka o monotonih funkcijah velja, da je za  $x > 0$

$$\Psi(x) - \frac{1}{2x} > 0,$$

torej naša funkcija nima ničel in zato cenilka po metodi največjega verjetja ne obstaja (digamma wiki).

## 1.3

Za  $k \in \mathbb{N}$  velja

$$m_k := P_{(a,b)}[\text{id}^k] = \int_0^\infty \frac{b^a}{\Gamma(a)} x^{-a-1} e^{-b/x} x^k dx = \frac{b^a}{\Gamma(a)} \frac{\Gamma(a-k)}{b^{a-k}} \int_0^\infty \frac{b^{a-k}}{\Gamma(a-k)} x^{-(a-k)-1} e^{-b/x} dx = b^k \frac{\Gamma(a-k)}{\Gamma(a)}.$$

Označimo

$$\hat{b}(x_1, x_2, x_3) := \left( \frac{x_1}{x_2} - \frac{x_2}{x_3} \right)^{-1}$$

$$\hat{a}(x_1, x_2, x_3) := \frac{\hat{b}(x_1, x_2, x_3)}{x_1} + 1$$

in opazimo, da je  $\hat{b}(m_1, m_2, m_3) = b$  in  $\hat{a}(m_1, m_2, m_3) = a$ , torej dobimo zaporedje cenilk po metodi momentov

$$(\hat{a}, \hat{b}) \circ (\hat{m}_1^n, \hat{m}_2^n, \hat{m}_3^n),$$

kjer je

$$\hat{m}_k^n(x) = \frac{1}{n} \sum_{i=1}^n x_i^k.$$

Ker je  $g = (\hat{a}, \hat{b})$  zvezna na  $\mathbb{R}_+^3$ , je ta cenilka dosledna. Sicer pa  $g$  ni trivialna preslikava in zato ne moremo enostavno analizirati porazdelitve, niti nujno izračunati povprečne vrednosti.

## 1.4

Naj bodo torej  $X_i \sim P_\vartheta$  NEP. Definirajmo NEP vektorje  $Y_i := (X_i, X_i^2, X_i^3)$  in označimo

$$M_n := (\hat{m}_1^n, \hat{m}_2^n, \hat{m}_3^n) \quad \text{in} \quad \hat{\vartheta}_n := g \circ M_n.$$

Po centralnem limitnem izreku velja

$$\sqrt{n} \cdot (M_n(X_1, \dots, X_n) - \mu) = \sqrt{n} \cdot \left( \frac{1}{n} \sum_{i=1}^n Y_i - \mu \right) \xrightarrow[n \rightarrow \infty]{D} N(0, \Sigma),$$

kjer je  $\mu$  pričakovana vrednost,  $\Sigma$  pa variančno-kovariančna matrika vektorjev  $Y_i$ . Potem nam metoda delta skupaj s Cramer-Woldovim pravilom zagotovi

$$\sqrt{n} \cdot (\hat{\vartheta}_n(X_1, \dots, X_n) - \vartheta) = \sqrt{n} \cdot (g(M_n(X_1, \dots, X_n)) - g(\mu)) \xrightarrow[n \rightarrow \infty]{D} N(0, \mathcal{J}_g(\mu) \Sigma \mathcal{J}_g(\mu)^T),$$

saj je očitno  $\mu = [m_1, m_2, m_3]^T$ .

## 1.5

Enostavno izračunamo

$$\Sigma = \begin{bmatrix} m_2 - m_1^2 & m_3 - m_1 m_2 & m_4 - m_1 m_3 \\ m_3 - m_1 m_2 & m_4 - m_2^2 & m_5 - m_2 m_3 \\ m_4 - m_1 m_3 & m_5 - m_2 m_3 & m_6 - m_3^2 \end{bmatrix}$$

in označimo

$$A := \mathcal{J}_g(\mu) \Sigma \mathcal{J}_g(\mu)^T.$$

Definirajmo sedaj zaporedje cenilk  $\hat{A}_n$  tako, da zamenjamo v  $A$  elemente  $m_i$  z  $\hat{m}_i^n$ . Potem uporabimo Slutskyjev izrek in zveznost korena ter inverza spd matrike, da dobimo

$$\pi_n(X_1, \dots, X_n, \vartheta) := \sqrt{n} \cdot \hat{A}_n^{-1/2} \left( \hat{\vartheta}_n(X_1, \dots, X_n) - \vartheta \right) \xrightarrow[n \rightarrow \infty]{D} N(0, I).$$

Naj bo  $B \subset \mathbb{R}^2$  neka množica, ki zadošča

$$P(\pi_n(X_1, \dots, X_n, \vartheta) \in B) \approx N(0, I)(B) = 1 - \alpha,$$

Sledi

$$P(\pi_n(X_1, \dots, X_n, \vartheta) \in B) = P(\hat{\vartheta}_n(X_1, \dots, X_n) - \vartheta \in \hat{A}_n^{1/2}(B)/\sqrt{n}) = P(\vartheta \in \hat{\vartheta}_n(X_1, \dots, X_n) - \hat{A}_n^{1/2}(B)/\sqrt{n}),$$

torej je območje zaupanja za  $\vartheta$  enako

$$\hat{\vartheta}_n(X_1, \dots, X_n) - \hat{A}_n^{1/2}(B)/\sqrt{n}.$$

Vprašanje izbire množice  $B$  lahko rešimo s predpostavko, da je to zaprta krogla s središčem v 0 v neki normirani metriki, s čimer se verjetno da dokazati enoličnost njenega radija.

## 2 2

Naj bodo  $X_i \sim P_\mu = N(\mu, \sigma_0)$  NEP za  $i = 1, \dots, n$  in fiksna  $\sigma_0 \in \mathbb{R}^+$ ,  $n \in \mathbb{N}$ . Po Rao-Blackwellevem izreku lahko najdemo NCEND kot pogojno upanje nepristranske cenilke  $\mathbb{1}_{\{X_1 \leq 5\}}$  pogojno na kompletno in zadostno statistiko

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Označimo  $Y = \bar{X} - \mu$  in definirajmo

$$\tilde{X} = X_1 - \zeta Y, \text{ kjer je } \zeta = P[X_1 Y] / \sigma(Y)^2.$$

Zanjo velja

$$P[\tilde{X}Y] = P[X_1Y] - \zeta P[Y^2] = 0,$$

torej sta  $\tilde{X}$  in  $Y$  nekorelirani, ker pa sta normalni, sta tudi neodvisni. Sedaj lahko zapišemo

$$\begin{aligned} P[\mathbb{1}_{\{X_1 \leq 5\}} \mid \bar{X} = t] &= P(X_1 \leq 5 \mid \bar{X} = t) = P(\tilde{X} + \zeta Y \leq 5 \mid Y + \mu = t) = P(\tilde{X} + \zeta(t - \mu) \leq 5) \\ &= P(X_1 - \zeta(\bar{X} - \mu) + \zeta(t - \mu) \leq 5) = P(X_1 - \zeta\bar{X} \leq 5 + \zeta t). \end{aligned}$$

Konkretno:

$$P[X_1Y] = P[(\sigma Z_1 + \mu)Y] = P[\sigma Z_1Y] = P[\sigma Z_1 \sum_{i=1}^n \frac{\sigma}{n} Z_i] = \frac{\sigma^2}{n} P[Z_1^2] = \frac{\sigma^2}{n},$$

torej skupaj z dejstvom  $\sigma(Y) = \sigma/\sqrt{n}$  dobimo  $\zeta = 1$  in NCEND

$$U(\bar{x}) = P(\sigma(1 + 1/\sqrt{n})Z \leq 5 + \bar{x}) = \Phi\left(\frac{5 + \bar{x}}{\sigma(1 + 1/\sqrt{n})}\right)$$

## 3 3

### 3.1

Fiksirajmo  $n \in \mathbb{N}$  in označimo  $P_\vartheta = \left(\int_- f(k; \vartheta) dc(k)\right)^{\times n}$ . Ker je

$$f(k; \vartheta) = -\frac{\vartheta^k}{k \ln(1 - \vartheta)} = -\frac{e^{k \ln(\vartheta)}}{k \ln(1 - \vartheta)} \implies f(k_1, \dots, k_n; \vartheta) = (-\ln(1 - \vartheta))^{-n} \prod_{i=1}^n k_i^{-1} e^{\ln(\vartheta) \sum_{i=1}^n k_i}$$

in  $\ln((0, 1)) = (-\infty, 0) \in \mathcal{T}_{EVK} \setminus \{\emptyset\}$ , je to eksponentni model s polnim rangom, torej dobimo kompletno zadostno statistiko

$$T(k_1, \dots, k_n) = \sum_{i=1}^n k_i.$$

### 3.2

Ker imamo eksponenten model s funkcijo  $Q = \ln$ , ki je strogo naraščajoča na  $(0, 1)$ , lahko uporabimo Neuman-Pearsonovo lemo za konstrukcijo enakomerno najmočnejših randomiziranih preizkusov. Ker ima model druge momente si lahko pomagamo tudi z normalnimi aproksimacijami (ali asimptotično normalnostjo).

### 3.3

Poskušajmo postopati z Neuman-Pearsonovo lemo. Označimo torej  $\vartheta_0 = 0.7$  in  $\alpha = 0.05$ . Potem imamo enakomerno najmočnejši randomiziran preizkus za  $\vartheta \leq \vartheta_0$  proti alternativni  $\vartheta \geq \vartheta_0$  v obliki

$$\phi(k_1, \dots, k_n) = \begin{cases} 1 & | \quad T(k_1, \dots, k_n) > C \\ \gamma & | \quad T(k_1, \dots, k_n) = C, \\ 0 & | \quad T(k_1, \dots, k_n) < C \end{cases}$$

kjer določimo  $C \in \mathbb{N}$  in  $\gamma \in [0, 1)$  enolično s pogojem

$$\beta(\vartheta_0) = P_{\vartheta_0}[\Phi] = P_{\vartheta_0}(T > C) + \gamma P_{\vartheta_0}(T = C) = \alpha.$$

Idealno bi bilo imeti porazdelitev  $T_n$ , a ta vključuje izračun vsot, ki v splošnem nimajo zaključene oblike ( $\sum_{k=1}^d \vartheta^k/k$ ). Če bi imeli dovolj majhen vzorec, to tudi ne bi bil problem, saj bi izračunali posamezne vrednosti gostote  $T$  z brute-force varianto. Ker pa je vzorec velikosti  $n = 50$  relativno velik, tega ne moremo izvesti. Lahko pa zato  $T$  aproksimiramo z normalno porazdelitvijo (po CLI). Računamo

$$\begin{aligned} \mu_\vartheta &:= P_\vartheta[\text{id}] = -\frac{1}{\ln(1 - \vartheta)} \sum_{k \in \mathbb{N}} \vartheta^k = -\frac{\vartheta}{(1 - \vartheta) \ln(1 - \vartheta)} \\ \sigma_\vartheta^2 &:= V_\vartheta[\text{id}] = P_\vartheta[\text{id}^2] - P_\vartheta[\text{id}]^2 = -\frac{1}{\ln(1 - \vartheta)} \sum_{k \in \mathbb{N}} k \vartheta^k - \mu_\vartheta^2 = -\frac{1}{\ln(1 - \vartheta)} \frac{\vartheta}{(1 - \vartheta)^2} - \mu_\vartheta^2. \end{aligned}$$

Označimo NEP model

$$\tilde{P}_\vartheta := N(\mu_\vartheta, \sigma_\vartheta^2)^{\times n}.$$

Sedaj lahko ocenimo gostoto statistike  $T$  kot

$$P_\vartheta(T = k) \approx \tilde{P}_\vartheta(T \leq k) - \tilde{P}_\vartheta(T \leq k - 1).$$

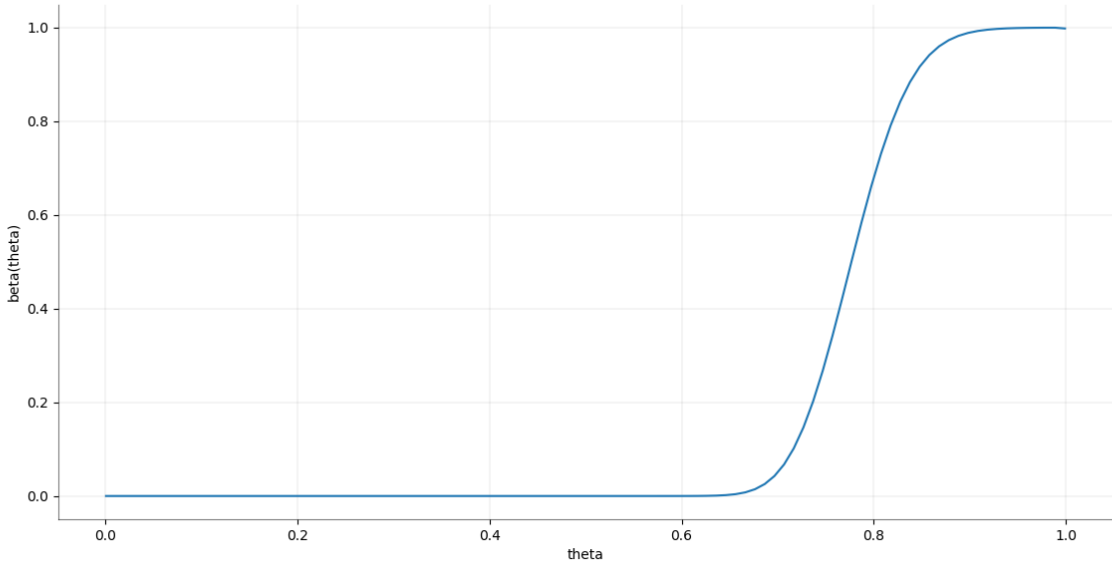
Aproksimirajmo:

$$\begin{aligned} \alpha &= 1 - P_{\vartheta_0}(T \leq C) + \gamma P_{\vartheta_0}(T = C) \\ &\approx 1 - \tilde{P}_{\vartheta_0}(T \leq C) + \gamma \tilde{P}_{\vartheta_0}(T \in [C - 1, C]) \\ &= 1 - \Phi\left(\frac{C/n - \mu_{\vartheta_0}}{\sigma_{\vartheta_0}/\sqrt{n}}\right) + \gamma \left( \Phi\left(\frac{C/n - \mu_{\vartheta_0}}{\sigma_{\vartheta_0}/\sqrt{n}}\right) - \Phi\left(\frac{(C-1)/n - \mu_{\vartheta_0}}{\sigma_{\vartheta_0}/\sqrt{n}}\right) \right). \end{aligned}$$

Vzamemo prvi  $C \in \mathbb{N}$ , za katerega je

$$\gamma = \frac{\alpha + \Phi\left(\frac{C/n - \mu_{\vartheta_0}}{\sigma_{\vartheta_0}/\sqrt{n}}\right) - 1}{\Phi\left(\frac{C/n - \mu_{\vartheta_0}}{\sigma_{\vartheta_0}/\sqrt{n}}\right) - \Phi\left(\frac{(C-1)/n - \mu_{\vartheta_0}}{\sigma_{\vartheta_0}/\sqrt{n}}\right)} \in [0, 1].$$

Dobimo  $C = 117$  in  $\gamma \approx 0.9705$ .



Slika 1: Funkcija moči dobljenega testa, kje so vrednosti aproksimirane z normalno porazdelitvijo.

## 4

### 4.1

Spet fiksirajmo  $n \in \mathbb{N}$  in označimo  $P_{p_0} = \text{Ber}(p_0)^{\times n}$  ter  $T(x_1, \dots, x_n) = \sum_{i=1}^n x_i$ . Neuman-Pearsonov test za hipotezo  $p = p_0$  proti alternativni ima obliko

$$\phi(x) = \begin{cases} 1 & | T(x) \in (C_1(p_0), C_2(p_0)) \\ \gamma_j(p_0) & | T(x) = C_j(p_0) \\ 0 & | T(x) \notin [C_1(p_0), C_2(p_0)] \end{cases},$$

kjer so vrednosti  $C_j(p_0)$ ,  $\gamma_j(p_0)$  določene z zahtevo

$$P_{p_0}[\phi_{p_0}] = \alpha \quad \text{in} \quad \frac{d}{dp_0} P_{p_0}[\phi_{p_0}] = 0.$$

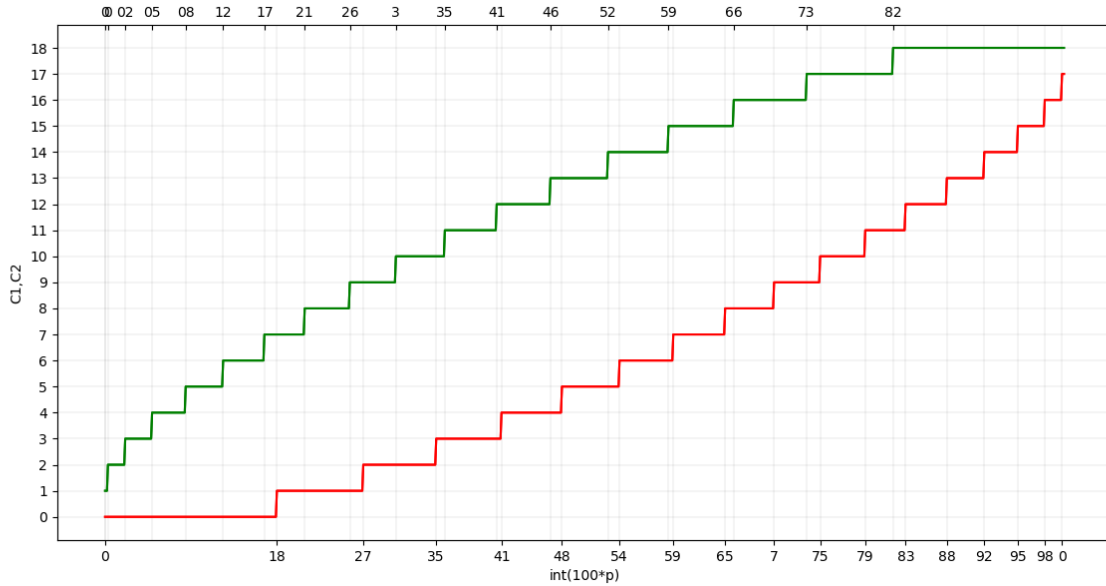
Izračunamo

$$\begin{aligned} P_{p_0}[\phi_{p_0}] &= P_{p_0}(T < C_1(p_0)) + P_{p_0}(T > C_2(p_0)) + \gamma_1(p_0)P_{p_0}(T = C_1(p_0)) + \gamma_2(p_0)P_{p_0}(T = C_2(p_0)) \\ &= 1 - \sum_{i=C_1(p_0)}^{C_2(p_0)} \binom{n}{i} p_0^i (1-p_0)^{n-i} + \sum_{i \in \{1,2\}} \gamma_i(p_0) \binom{n}{C_i(p_0)} p_0^{C_i(p_0)} (1-p_0)^{n-C_i(p_0)} \end{aligned}$$

in

$$\begin{aligned} \frac{d}{dp_0} P_{p_0}[\phi_{p_0}] &= - \sum_{i=C_1(p_0)}^{C_2(p_0)} \binom{n}{i} (i p_0^{i-1} (1-p_0)^{n-i} - (n-i) p_0^i (1-p_0)^{n-i-1}) + \\ &+ \sum_{i \in \{1,2\}} \gamma_i(p_0) \binom{n}{C_i(p_0)} (C_i(p_0) p_0^{C_i(p_0)-1} (1-p_0)^{n-C_i(p_0)} - (n-C_i(p_0)) p_0^{C_i(p_0)} (1-p_0)^{n-C_i(p_0)-1}). \end{aligned}$$

Testiramo vse  $0 \leq C_1(p_0) < n$  in  $C_1(p_0) < C_2(p_0) \leq n$  in na vsaki iteraciji rešimo sistem dveh linearnih enačb za  $\gamma_i(p_0)$ . Vrnemo prve vrednosti, pri katerih velja  $0 \leq \gamma_i(p_0) < 1$  za  $i = 1, 2$ .



Slika 2:  $C_1(p)$  in  $C_2(p)$  na diskretizaciji  $(0, 1)$  na 1000 točkah za  $n = 18$ .

Za interval zaupanja definiramo

$$I(T(x)) := \{p \in (0, 1) \mid \phi_p(x) < 1\} = \{p \in (0, 1) \mid T(x) \in [C_1(p), C_2(p)]\}.$$

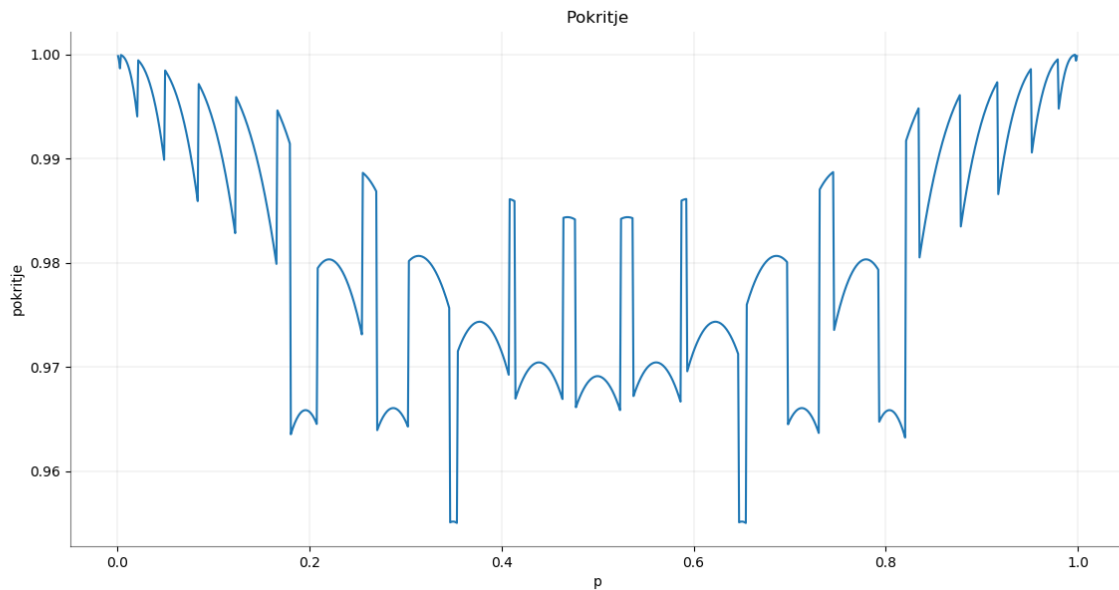
Za vsak  $k = 0, 1, \dots, n$  bomo vzeli vse  $p$  na neki diskretizaciji  $(0, 1)$ , ki so v  $I(k)$ . Spodnja tabela prikazuje aproksimacijo  $I \circ T$  pri delitvi intervala  $(0, 1)$  na 1000 delov, z zaokrožitvijo krajišč intervala na 3 decimalna mesta.

$k$	$I(k)$
0	$(0, 0.18]$
1	$(0, 0.27]$
2	$[0.004, 0.346]$
3	$[0.022, 0.414]$
4	$[0.05, 0.477]$
5	$[0.085, 0.536]$
6	$[0.124, 0.592]$
7	$[0.167, 0.646]$
8	$[0.209, 0.697]$
9	$[0.256, 0.745]$
10	$[0.304, 0.792]$
11	$[0.355, 0.834]$
12	$[0.409, 0.877]$
13	$[0.465, 0.916]$
14	$[0.524, 0.951]$
15	$[0.587, 0.979]$
16	$[0.655, 0.997]$
17	$[0.731, 1)$
18	$[0.821, 1)$

## 4.2

Zaokroženo na 4 decimalna mesta dobimo  $\gamma_1(0.72) \approx 0.1787$  in  $\gamma_2(0.72) \approx 0.6852$ .

## 4.3



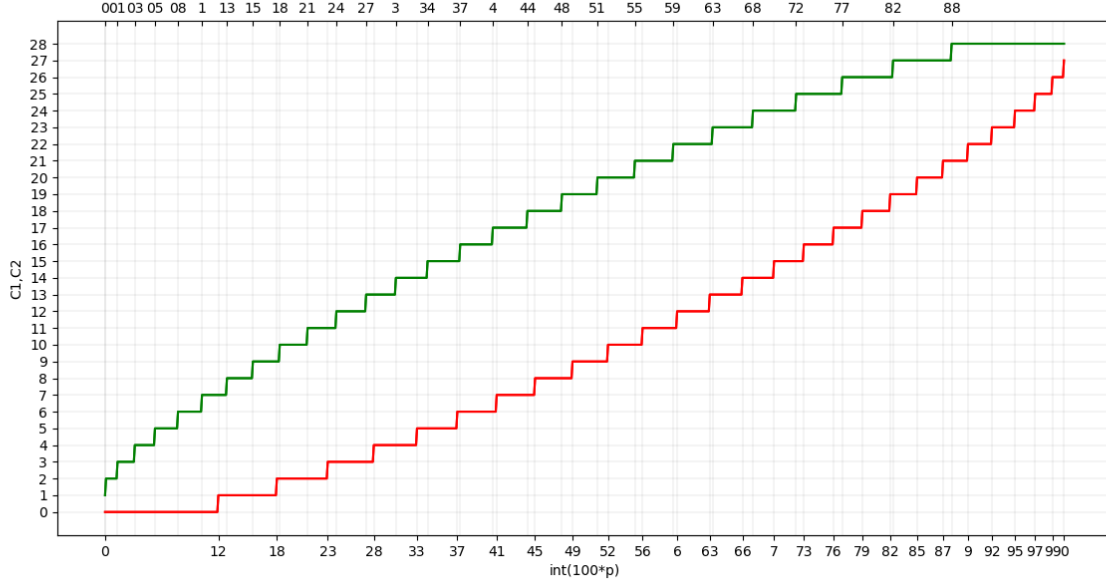
Ocena minimuma vrednosti na zgornjem grafu je 0.95502, kar je naša ocena za koeficient zaupanja.

## 4.4

Program nam vrne

$$C_1 = 15 \quad C_2 = 24 \quad \gamma_1 \approx 0.6922 \quad \gamma_2 \approx 0.0029$$

## 4.5



Slika 3:  $C_1(p)$  in  $C_2(p)$  na diskretizaciji  $(0, 1)$  na 1000 točkah za  $n = 28$

## 5

Označimo torej  $p = (p_0, \dots, p_3)$  in

$$P_p = \begin{pmatrix} \xi_0 & \xi_1 & \xi_2 & \xi_3 \\ p_0 & p_1 & p_2 & p_3 \end{pmatrix}^{\times n}.$$

Vemo, da velja

$$T_i(x_1, \dots, x_n) = \sum_{j=1}^n \mathbb{1}_{\{\xi_i\}}(x_j) \implies P_p \circ (T_0, T_1, T_2, T_3)^{-1} = \text{Multi}(p, n).$$

### 5.1

Računamo moč preizkusa hipoteze  $p = \pi = (\pi_0, \pi_1, \pi_2, \pi_3) = (3/32, 7/32, 9/32, 11/32)$ .

#### 5.1.1

Test je oblike

$$\phi_n^1(x) = \begin{cases} 1 & \text{if } -2 \ln(\lambda_n(Tx)) > \chi_{3;\alpha}^2 \\ 0 & \text{if } -2 \ln(\lambda_n(Tx)) \leq \chi_{3;\alpha}^2 \end{cases}, \text{ kjer je } \lambda_n(t) = \prod_{j=0}^3 \left( \frac{n\pi_j}{t_j} \right)^{t_j}.$$

Dobimo

$$\beta_n^1 = P_\pi[\phi_n^1] = P_\pi(-2 \ln(\lambda_n \circ T) > \chi_{3;\alpha}^2) = \text{Multi}(\pi, n) \left( \{t \mid t_0 + t_1 + t_2 + t_3 = n \wedge -2 \ln(\lambda_n(t)) > \chi_{3;\alpha}^2\} \right).$$

To množico računalniško ni zelo zahtevno parametrizirati. Ko to naredimo, dobimo

$n$	$\beta_n^1$
35	0.000019330779246711946
65	0.00016696632438189583
95	0.0011870347621621064
125	0.007041048805908646

### 5.1.2

Test je oblike

$$\phi_n^2(x) = \begin{cases} 1 & \text{if } \tau_n(Tx) > \chi_{3;\alpha}^2 \\ 0 & \text{if } \tau_n(Tx) \leq \chi_{3;\alpha}^2 \end{cases}, \text{ kjer je } \tau_n(t) = \sum_{j=0}^3 \frac{(t_j - n\pi_j)^2}{n\pi_j}.$$

Spet izrazimo

$$\beta_n^2 = P_\pi[\phi_n^2] = P_\pi(\tau_n \circ T > \chi_{3;\alpha}^2) = \text{Multi}(\pi, n) \left( \{t \mid t_0 + t_1 + t_2 + t_3 = n \wedge \tau_n(t) > \chi_{3;\alpha}^2\} \right)$$

in dobimo

$n$	$\beta_n^2$
35	0.06836781744193415
65	0.09193396478820819
95	0.11436093029694488
125	0.13857104432656328

## 5.2

### 5.2.1

Naj bo

$$H = \{(p_0, p_1, p_2, p_3) \mid p_0 + p_1 + p_2 + p_3 = n \wedge p_1 + 2p_2 = 2/3\}.$$

Za dani  $x$  iščemo

$$\sup_{p \in H} P_p(\{x\}) = \sup_{p \in H} \prod_{i=0}^3 p_i^{T_i(x)},$$

torej maksimiziramo funkcijo

$$l(p) = t_0 \ln(p_0) + t_1 \ln(p_1) + t_2 \ln(p_2) + t_3 \ln(p_3) = t_0 \ln(1/3 + p_2 - p_3) + t_1 \ln(2/3 - 2p_2) + t_2 \ln(p_2) + t_3 \ln(p_3).$$

Po enačenju parcialnih odvodov z 0 dobimo ekstrem

$$\hat{p}_2(t) = \frac{(t_0 - t_1 + t_3) + \sqrt{(t_0 - t_1 + t_3)^2 + 4nt_2}}{6n} \quad \text{in} \quad \hat{p}_3(t) = \frac{t_3(p_2 + 1/3)}{t_0 + t_2}.$$

Podajmo test

$$\phi_n^3(x) = \begin{cases} 1 & \text{if } -2 \ln(\zeta_n(Tx)) > \chi_{3;\alpha}^2 \\ 0 & \text{if } -2 \ln(\zeta_n(Tx)) \leq \chi_{3;\alpha}^2 \end{cases}, \text{ kjer je } \zeta_n(t) = \prod_{i=0}^3 \left( \frac{n\hat{p}_i(t)}{t_i} \right)^{t_i}.$$

Asimptotični izrek za hipotezo  $H$  velja, saj je  $p \mapsto p_1 + 2p_2 - 2/3$  gladka submerzija, drugi pogoji pa so zadoščeni od prej.

### 5.2.2

Izbrali smo  $\pi = (1, 0, 1, 1)/3$  in dobili 34 primerov, ko test zavrne hipotezo, torej delež je  $34/9500 \approx 0.003579$ .

## 6

Fiksirajmo  $(m, n) = (3, 300)$  (dimenzije podatkih) in  $\alpha = 0.05$ .



## 6.1

### 6.1.1

Z reševanjem predoločenega sistema  $X\hat{\beta} = y$ , kjer je  $X = [1 \text{ TCH HDL TRI}]$  in  $y = \text{LDL}$ , dobimo

$$\hat{\beta}_0 = 0.05993031 \quad \hat{\beta}_1 = 0.83909446 \quad \hat{\beta}_2 = -0.64999293 \quad \hat{\beta}_3 = -0.2240437.$$

### 6.1.2

Označimo  $\hat{y} = X \cdot \hat{\beta}$  in  $\bar{y}$  povprečje vektorja  $y$ . Hipotezo  $\beta_1 = \beta_2 = \beta_3 = 0$  testiramo s Fisherjevo statistiko

$$F = \frac{\sum (\hat{y}_i - \bar{y})^2 / m}{\sum (y_i - \hat{y}_i)^2 / (n - m - 1)} \sim F_{m, n-m-1}.$$

Hipotezo zavrnemo, če dobimo  $F > F_{m, n-m-1, \alpha}$ . Na našem vzorcu hipotezo zavrnemo, saj je

$$F \approx 298.0651 \quad \text{in} \quad F_{m, n-m-1, \alpha} \approx 2.6351.$$

### 6.1.3

Hipotezo  $\beta_0 = 0$  testiramo s Studentovo statistiko

$$T = \frac{\hat{\beta}_0 - \beta_0}{\|y - \hat{y}\|_2 \sqrt{((X^T X)^{-1})_{(1,1)} / (n - m - 1)}} \sim t_{n-m-1}.$$

Hipotezo zavrnemo, če je  $|T| > t_{n-m-1, \alpha/2}$ . Na našem vzorcu hipoteze ne moremo zavrniti, saj je

$$T \approx 0.3079 \quad \text{in} \quad t_{n-m-1, \alpha/2} \approx 1.9680.$$

## 6.2

### 6.2.1

Z reševanjem predoločenega sistema  $X\hat{\beta} = y$ , kjer je  $X = [\text{TCH HDL TRI}]$  in  $y = \text{LDL}$ , dobimo

$$\hat{\beta}_1 = 0.84615208 \quad \hat{\beta}_2 = -0.6445336 \quad \hat{\beta}_3 = -0.2225561.$$

### 6.2.2

Hipotezo  $\beta = (1, -1, -0.54)$  testiramo s pomočjo območja zaupanja, konstruiranega v naslednji podpodnalogi. Zavrnemo jo s stopnjo značilnosti  $\alpha$ , če je

$$\left\| \sqrt{\Lambda} Q^T (\hat{\beta} - \beta) \right\|_2^2 > \frac{m \|y - \hat{y}\|_2^2 F_{m, n-m; \alpha}}{n - m}.$$

Na levi strani dobimo 33.667, na desni pa 3.0564 in zato zavrnemo. Hipotezo lahko testiramo tudi s kvadrom zaupanja iz naloge 6.2.4. in spet dobimo zavrnitev.

### 6.2.3

Imamo torej pivotno funkcijo

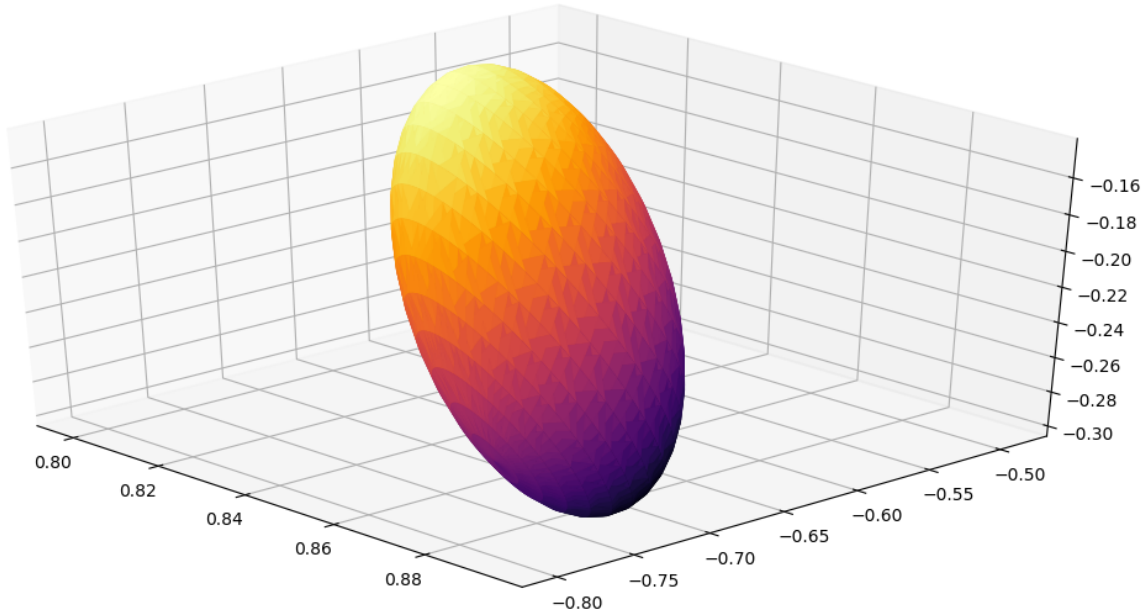
$$F(x, \beta) = \frac{\langle X^T X (\hat{\beta} - \beta), \hat{\beta} - \beta \rangle / m}{\|y - \hat{y}\|_2^2 / (n - m)} \sim F_{m, n-m}$$

s katero konstruiramo elipsoid zaupanja

$$\left\{ \hat{\beta} + Q \Lambda^{-1/2} u \mid \|u\|_2^2 \leq \frac{m \|y - \hat{y}\|_2^2 F_{m, n-m; \alpha}}{n - m} \right\},$$

kjer je  $X^T X = Q \Lambda Q^T$  Schurov razcep. Če s  $S_1(0) \subset \mathbb{R}^3$  označimo enotsko kroglo s središčem v izhodišču, dobimo iz podatkov elipsoid

$$E \approx \begin{bmatrix} 0.84615 \\ -0.64453 \\ -0.22255 \end{bmatrix} + \begin{bmatrix} -0.01289 & -0.02088 & 0.0411 \\ -0.00261 & -0.01985 & -0.1541 \\ -0.00431 & 0.07447 & -0.02955 \end{bmatrix} \cdot S_1(0).$$



Slika 4: Elipsoid zaupanja za  $\beta$ , realiziran na naših podatkih.

#### 6.2.4

Za  $\beta_j$  imamo sedaj intervale zaupanja stopnje zaupanja  $1 - \alpha/m$  oblike

$$I_j := \left[ \hat{\beta}_j \mp t_{n-m; \alpha/(2m)} \|y - \hat{y}\|_2 \sqrt{((X^T X)^{-1})_{(j,j)} / (n-m)} \right].$$

Z Bonferonijevim popravkom dobimo, da je

$$I := \bigtimes_{j=1}^m I_j$$

kvader zaupanja s stopnjo  $\alpha$ . Na našem vzorcu imamo

$$I = [0.805, 0.887] \times [-0.778, -0.511] \times [-0.291, -0.154].$$

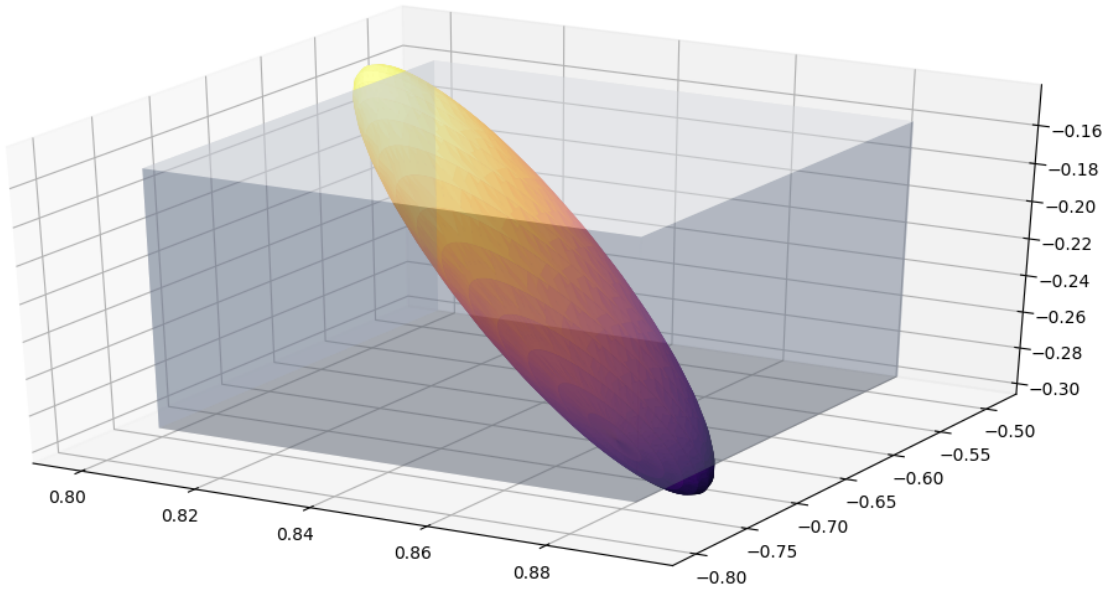
Na sliki vidimo, da je volumen kvadra zaupanja veliko večji od elipsoida in da prvi skoraj vsebuje drugega.

## 7

### 7.1

Vemo, da ima slučajna spremenljivka  $Y$  Lebesgueovo gostoto natanko tedaj, ko obstaja Borelova funkcija  $f_Y : \mathbb{R} \rightarrow \mathbb{R}$ , za katero velja

$$P(Y \leq y) = \int_{(-\infty, y]} f_Y d\mathcal{L} = \int_{-\infty}^y f_Y(x) dx.$$



Slika 5: Kvader zaupanja in elipsoid zaupanja za  $\beta$ , realizirana na naših podatkih.

Upoštevamo izrek o popolni verjetnosti in Tonellijev izrek za zamenjavo vrstnega reda integracije, da dobimo

$$\begin{aligned} P(Y + Z \leq w) &= \sum_{z_k} P(Y \leq w - z_k) P(Z = z_k) = \sum_{z_k} \int_{-\infty}^{w - z_k} f_Y(y) P(Z = z_k) dy \\ &= \sum_{z_k} \int_{-\infty}^w f_Y(x - z_k) P(Z = z_k) dx = \int_{-\infty}^w \sum_{z_k} f_Y(x - z_k) P(Z = z_k) dx = \int_{-\infty}^w f_X(x) dx. \end{aligned}$$

Ker je  $(x, z_k) \mapsto f_Y(x - z_k) P(Z = z_k)$  merljiva in vrsta konvergentna za vsak  $x \in \mathbb{R}$  (je dominirana z 1), je  $f_X \in B_{\mathbb{R}}/B_{\mathbb{R}}$  (po enem iz izrekov pri teoriji mere). Zato je  $f_X$  Lebesgueva gostota slučajne spremenljivke  $X = Y + Z$ .

## 7.2

Dobimo

$$\begin{aligned} f_X(x) &= \sum_{k=0}^{\infty} \frac{c_k \lambda^k}{c(\lambda)} \alpha (x - k)^{\alpha-1} \mathbb{1}_{(0,1)}(x - k) = \frac{c_{[x]} \lambda^{[x]}}{c(\lambda)} \alpha (x - [x])^{\alpha-1} \\ \ln(f_X(x)) &= \ln(\alpha/c(\lambda)) + \ln(c_{[x]}) + [x] \ln(\lambda) + (\alpha - 1) \ln(x - [x]) \end{aligned}$$

## 7.3

Izrazimo

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \left( \frac{\alpha}{c(\lambda)} \right)^n \prod_{i=1}^n c_{[x_i]} \lambda^{\prod_{i=1}^n [x_i]} \left( \prod_{i=1}^n (x_i - [x_i]) \right)^{\alpha-1}$$

Po Fisher-Neymanovem izreku dobimo zadostnost statistike  $T(x) = \sum_{i=1}^n \ln(x_i)$ . Čeprav je model eksponenten pa nima polnega ranga, torej ne moremo sklepati, da je  $T$  kompletna.