TMA4245 Statistikk
Vår 2023

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Anbefalte oppgaver 9
Løsningsskisse

Oppgave 1 Det er oppgitt i oppgaveteksten at estimatoren er forventningsrett, så vi vet allerede at $E(\hat{\mu}) = \mu$. Variansen til $\hat{\mu}$ er

$$\begin{aligned}\mathrm{Var}(\hat{\mu}) &= \left(\frac{\tau_0^2}{\tau_0^2+\sigma_0^2}\right)^2 \mathrm{Var}(X) + \left(\frac{\sigma_0^2}{\tau_0^2+\sigma_0^2}\right)^2 \mathrm{Var}(Y)\\ &= \frac{\tau_0^4\sigma_0^2+\sigma_0^4\tau_0^2}{(\tau_0^2+\sigma_0^2)^2} = \frac{(\tau_0^2+\sigma_0^2)\tau_0^2\sigma_0^2}{(\tau_0^2+\sigma_0^2)^2}\\ &= \frac{\tau_0^2\sigma_0^2}{\tau_0^2+\sigma_0^2},\end{aligned}$$

og fordi $\hat{\mu}$ er en lineærkombinasjon av normalfordelte variable, er den normalfordelt. Altså har vi

$$Z = \frac{\hat{\mu}-\mu}{\sqrt{\frac{\sigma_0^2\tau_0^2}{\tau_0^2+\sigma_0^2}}} \sim N(0,1),$$

slik at vi får et $(1-\alpha)\cdot 100\%$ konfidensintervall fra

$$\begin{aligned}P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) &= 1-\alpha\\ P\left(\hat{\mu} - z_{\alpha/2}\sqrt{\frac{\sigma_0^2\tau_0^2}{\tau_0^2+\sigma_0^2}} \le \mu \le \hat{\mu} + z_{\alpha/2}\sqrt{\frac{\sigma_0^2\tau_0^2}{\tau_0^2+\sigma_0^2}}\right) &= 1-\alpha.\end{aligned}$$

Intervallet er altså

$$\underline{\underline{\left[\hat{\mu} - z_{\alpha/2}\sqrt{\frac{\sigma_0^2\tau_0^2}{\tau_0^2+\sigma_0^2}}, \hat{\mu} + z_{\alpha/2}\sqrt{\frac{\sigma_0^2\tau_0^2}{\tau_0^2+\sigma_0^2}}\right]}}.$$

Oppgave 2

a) Kumulativ fordeling:

Bruk substitusjonen $u = s^2$ med $du = 2sds$:

$$\begin{aligned} P(X \leq x) &= \int_0^x f(s)ds &&= \int_0^x \frac{2s}{\theta} e^{-\frac{s^2}{\theta}} ds \\ &= \int_0^{x^2} \frac{1}{\theta} e^{-\frac{u}{\theta}} du &&= \left[-e^{-\frac{u}{\theta}}\right]_0^{x^2} \\ &= \underline{\underline{1 - e^{-\frac{x^2}{\theta}}}} \end{aligned}$$

Betinget fordeling:
Vi har at $P(X > x) = 1 - P(X \leq x) = e^{-\frac{x^2}{\theta}}$.

$$\begin{aligned} P(X > 15 | X > 10) &= \frac{P(X > 15 \cap X > 10)}{P(X > 10)} &&= \frac{P(X > 15)}{P(X > 10)} \\ &= \frac{e^{-\frac{15^2}{25}}}{e^{-\frac{10^2}{25}}} \\ &= \frac{0.000123}{0.0183} &&\approx \underline{\underline{0.0067}}. \end{aligned}$$

b) Rimelighetsfunksjonen er gitt ved:

$$\begin{aligned} L(\theta; x_1, ..., x_n) &= f(x_1, ..., x_n; \theta) \\ &= f(x_1; \theta) \cdots f(x_n; \theta) \\ &= \frac{2x_1}{\theta} e^{-\frac{x_1^2}{\theta}} \cdots \frac{2x_n}{\theta} e^{-\frac{x_n^2}{\theta}} \\ &= \left(\frac{2}{\theta}\right)^n \cdot (x_1 \cdots x_n) \cdot e^{-\frac{1}{\theta}\sum_{i=1}^n x_i^2}. \end{aligned}$$

Tar logaritmen:

$$\begin{aligned} l(\theta; x_1, ..., x_n) &= \ln\left[L(\theta)\right] \\ &= n \ln\left(\frac{2}{\theta}\right) + \ln(x_1 \cdots x_n) - \frac{1}{\theta}\sum_{i=1}^n x_i^2 \\ &= n\ln(2) - n\ln(\theta) + \ln(x_1 \cdots x_n) - \frac{1}{\theta}\sum_{i=1}^n x_i^2. \end{aligned}$$

Finn maksimumspunkt ved å derivere ln-rimelighetsfunksjonen og sette lik 0.

$$\frac{\partial l}{\partial \theta}\Big|_{\theta=\widehat{\theta}} = 0 - n\frac{1}{\theta} + 0 + \frac{1}{\theta^2}\sum_{i=1}^n x_i^2 = 0$$

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^n x_i^2.$$

SME for θ blir da $\underline{\underline{\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i^2}}$.

Viser at $\hat{\theta}$ er forventningsrett:

$$\mathrm{E}[\hat{\theta}] = \mathrm{E}[\frac{1}{n}\sum_{i=1}^{n} X_i^2] = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[X_i^2] = \frac{1}{n}\sum_{i=1}^{n} \theta = \underline{\underline{\theta}}.$$

For å regne ut variansen til $\hat{\theta}$ trenger vi først variansen til X^2. La $Y = X^2$. Bruker at $\mathrm{Var}[Y] = \mathrm{E}[Y^2] - \mathrm{E}[Y]^2$. Dermed er

$$\begin{aligned}
\mathrm{Var}[X^2] &= \mathrm{Var}[Y] \\
&= \mathrm{E}[Y^2] - \mathrm{E}[Y]^2 \\
&= \mathrm{E}[X^4] - \mathrm{E}[X^2]^2 \\
&= 2\theta^2 - \theta^2 \\
&= \underline{\underline{\theta^2}}.
\end{aligned}$$

Da får vi:

$$\mathrm{Var}[\hat{\theta}] = \mathrm{Var}[\frac{1}{n}\sum_{i=1}^{n} X_i^2] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}[X_i^2] = \frac{\theta^2}{n}.$$

c) Sentralgrenseteoremet sier at hvis $Y_1, ..., Y_n$ er uavhengige og identisk fordelte, så er gjennomsnittet tilnærmet normalfordelt. Mer presist er

$$Z = \frac{\bar{Y} - \mathrm{E}[\bar{Y}]}{\sqrt{\mathrm{Var}[\bar{Y}]}}$$

tilnærmet standard normalfordelt når n er stor (>30). I vårt tilfelle er $Y_i = X_i^2$ uavhengige og $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \hat{\theta}$. Videre er som funnet i forrige punkt (3b), $E(\bar{Y}) = \theta$ og $\mathrm{Var}(\bar{Y}) = \frac{\theta^2}{n}$. Dermed blir

$$Z = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \theta}{\sqrt{\frac{\theta^2}{n}}}$$

tilnærmet normalfordelt når n er stor.

Vi finner så et 95% konfidensintervall med $\alpha = 0.05$. Sett $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$. Bruker at

$$
\begin{aligned}
1-\alpha &\approx \Pr\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}}\right) \\
&= \Pr\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{Y}-\theta}{\sqrt{\frac{\theta^2}{n}}} \leq z_{\frac{\alpha}{2}}\right) \\
&= \Pr\left(-\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}} \leq \frac{1}{\theta}\bar{Y} - 1 \leq \frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}\right) \\
&= \Pr\left(1-\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}} \leq \frac{1}{\theta}\bar{Y} \leq 1+\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}\right) \\
&= \Pr\left(\frac{1}{1+\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}} \leq \frac{\theta}{\bar{Y}} \leq \frac{1}{1-\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}}\right) \\
&= \Pr\left(\frac{\bar{Y}}{1+\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}} \leq \theta \leq \frac{\bar{Y}}{1-\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}}\right)
\end{aligned}
$$

Et tilnærmet 95% konfidensintervall for θ blir da

$$
\underline{\underline{\left[\frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{1+\frac{z_{0.025}}{\sqrt{n}}}, \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{1-\frac{z_{0.025}}{\sqrt{n}}}\right]}}
$$

Siden $p = e^{-\frac{100}{\theta}}$ er en monotont stigende funksjon av θ, har vi at

$$
\begin{aligned}
\Pr\left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{1+\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}} \leq \theta \leq \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{1-\frac{z_{\frac{\alpha}{2}}}{\sqrt{n}}}\right) &\approx 1-\alpha \\
&\Updownarrow \\
\Pr\left(e^{-\frac{100}{\hat{\theta}_L}} \leq e^{-\frac{100}{\theta}} \leq e^{-\frac{100}{\hat{\theta}_U}}\right) &\approx 1-\alpha
\end{aligned}
$$

Et tilnærmet 95% konfidensintervall for $e^{-\frac{100}{\theta}}$ blir da $\underline{\underline{\left[e^{-\frac{100}{\hat{\theta}_L}}, e^{-\frac{100}{\hat{\theta}_U}}\right]}}$

Oppgave 3

a) Den kumulative fordelingsfunksjonen $F(x) = P(X \leq x)$ beregner vi ved å integrere sannsynlighetstettheten $f(x)$. Dvs.

$$
F(x) = \int_{-\infty}^{x} f(t)dt = \int_{1}^{x} \beta t^{-\beta-1}dt = \beta\,\frac{1}{-\beta}\left[t^{-\beta}\right]_1^x = (-1)\left[x^{-\beta}-1\right] = 1 - x^{-\beta}.
$$

Sannsynligheten for at det går mer enn 2 uker mellom to påfølgende feil, når $\beta = 3$, er

$$P(X > 2) = 1 - P(X \leq 2) = 1 - F(2) = 1 - (1 - 2^{-\beta}) = 2^{-3} = 0.125$$

Sannsynligheten for at nettet svikter før det er gått 3.5 uker, gitt at det har gått minst 2 uker siden siste feil, er (med $\beta = 3$)

$$\begin{aligned}
P(X \leq 3.5 \mid X > 2) &= \frac{P(X \leq 3.5 \cap X > 2)}{P(X > 2)} = \frac{P(X \leq 3.5) - P(X \leq 2))}{P(X > 2)} \\
&= \frac{F(3.5) - F(2)}{1 - F(2)} = \frac{(1 - 3.5^{-3}) - (1 - 2^{-3})}{0.125} = 0.813
\end{aligned}$$

b) Sannsynlighetsmaksimeringsestimatoren, SME for β:

Simultantettheten for $X_1, \ldots, X_n$ er $f(x_1, \ldots, x_n; \beta) \overset{\text{uavh.}}{=} \prod_{i=1}^n f(x_i; \beta) = \prod_{i=1}^n \beta x_i^{-\beta-1}$. Rimelighetsfunksjonen er simultanfordelingen sett på som funksjon av β, og kan skrives som

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

SME er den verdien for β som maksimerer $L(x_1, \ldots, x_n; \beta)$. Denne verdien finner vi ved først å ta logaritmen, så derivere og sette lik 0:

$$\begin{aligned}
l(x_1, \ldots, x_n; \beta) &= \ln\left(L(x_1, \ldots, x_n; \beta)\right) = \ln\left(\beta^n \prod_{i=1}^n x_i^{-\beta-1}\right) \\
&= \ln(\beta^n) + \ln\left(\prod_{i=1}^n x_i^{-\beta-1}\right) \\
&= n\ln(\beta) + \sum_{i=1}^n (-(\beta+1)\ln(x_i)) = n\ln(\beta) - (\beta+1)\sum_{i=1}^n \ln(x_i) \\
\frac{dl(x_1, \ldots, x_n; \beta)}{d\beta} &= \frac{n}{\beta} - \sum_{i=1}^n \ln(x_i) = 0 \\
\beta &= \frac{n}{\sum_{i=1}^n \ln(x_i)}
\end{aligned}$$

Dette gir at SME for β er

$$\hat{\beta} = \hat{\beta}_1 = \frac{n}{\sum_{i=1}^n \ln X_i}.$$

Når vi setter inn de observerte verdiene får vi følgende estimat for β:

$$\hat{\beta} = \hat{\beta}_1 = \frac{n}{\sum_{i=1}^n \ln x_i} = \frac{10}{3.39} = 2.95.$$

c) Vi skal først vise at $2\beta\ln(X_i)$ er kjikvadratfordelt med 2 frihetsgrader (som er det samme som en eksponensialfordeling).

La $Y_i = 2\beta \ln(X_i)$. Vi kan finne sannsynlighetsfordelingen til Y_i ved å bruke transformasjonsformelen (vi ser her bort fra indeksen i i utledningen). La

$$\begin{aligned} y &= u(x) = 2\beta \ln(x), \quad \text{slik at} \\ x &= w(y) = \exp(\frac{y}{2\beta}). \end{aligned}$$

La $f_Y(y)$ være sannsynlighetstettheten til Y. Transformasjonsformelen sier da at

$$f_Y(y) = f_X(w(y))|w'(y)|.$$

Vi deriverer $w(y)$ og får $w'(y) = \frac{1}{2\beta}\exp(\frac{y}{2\beta})$. Sannsynlighetstettheten til Y blir da

$$\begin{aligned} f_Y(y) &= f_X(w(y))|w'(y)| = \beta(\exp(\frac{y}{2\beta}))^{-\beta-1}\frac{1}{2\beta}\exp(\frac{y}{2\beta}) \\ &= \frac{1}{2}\exp((-\beta-1)\frac{y}{2\beta})\exp(\frac{y}{2\beta}) = \frac{1}{2}\exp(-\beta\frac{y}{2\beta} - \frac{y}{2\beta} + \frac{y}{2\beta}) \\ &= \frac{1}{2}\exp(-\frac{y}{2}). \end{aligned}$$

Uttrykket for $f_Y(y)$ kan skrives

$$f_Y(y) = \frac{1}{2^{2/2}\Gamma(2/2)} y^{2/2-1}\exp(-\frac{y}{2}),$$

siden $\Gamma(2/2) = \Gamma(1) = 1$. Dette er sannsynlighetstettheten for en kjikvadratfordelt stokastisk variabel med 2 frihetsgrader. Dermed har vi vist at $Y_i = 2\beta \ln(X_i)$ er kjikvadratfordelt med 2 frihetsgrader, dvs. $Y_i \sim \chi^2_2$.

La $Z = 2\beta \sum_{i=1}^n \ln(X_i)$. Med $Y_i = 2\beta \ln(X_i)$ har vi at

$$Z = 2\beta \sum_{i=1}^n \ln(X_i) = \sum_{i=1}^n 2\beta \ln(X_i) = \sum_{i=1}^n Y_i.$$

Vi har vist at $Y_i \sim \chi^2_2$, og siden en sum av uavhengige kjikvadratfordelte stokastiske variabler er kjikvadratfordelt, med summen av frihetsgradene, er $Z = 2\beta \sum_{i=1}^n \ln(X_i)$ kjikvadratfordelt med $\sum_{i=1}^n 2 = 2n$ frihetsgrader.

Konfidensintervall for β:

Vi bruker at $Z = 2\beta \sum_{i=1}^n \ln(X_i) \sim \chi^2_{2n}$. La $\alpha = 0.05$. Vi får da at

$$\begin{aligned} P(\chi^2_{1-\alpha/2,2n} < Z < \chi^2_{\alpha/2,2n}) &= 1-\alpha \\ P(\chi^2_{1-\alpha/2,2n} < 2\beta \sum_{i=1}^n \ln(X_i) < \chi^2_{\alpha/2,2n}) &= 1-\alpha \\ P(\frac{\chi^2_{1-\alpha/2,2n}}{2\sum_{i=1}^n \ln(X_i)} < \beta < \frac{\chi^2_{\alpha/2,2n}}{2\sum_{i=1}^n \ln(X_i)}) &= 1-\alpha \end{aligned}$$

Et 95% konfidensintervall for β blir da

$$\left[\frac{\chi^2_{1-0.025,2n}}{2\sum_{i=1}^{n}\ln(x_i)} < \beta < \frac{\chi^2_{0.025,2n}}{2\sum_{i=1}^{n}\ln(x_i)}\right]$$

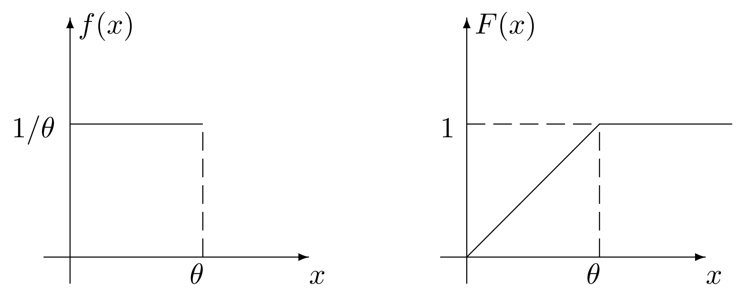
Insatt observerte verdier får vi

$$\left[\frac{9.591}{2\cdot 3.39} < \beta < \frac{34.170}{2\cdot 3.39}\right] = [1.41, 5.04].$$

Oppgave 4

a)

$$F(x) = \int_{-\infty}^{x} f(x)\,dx \;=\; \begin{cases} 0 & \text{hvis } x < 0, \\ \frac{x}{\theta} & \text{hvis } 0 \le x \le \theta, \\ 1 & \text{hvis } x > \theta. \end{cases}$$



$$P(X \le 0,4|\theta = 2) = F(0,4) = \frac{0,4}{2,0} = 0,2$$

b)

$$L(\theta) = \prod_{i=1}^{n} f(x_i) \;=\; \begin{cases} \frac{1}{\theta^n} & \text{hvis } 0 \le x_i \le \theta \;\forall\; i = 1, 2, \ldots, n, \\ 0 & \text{ellers} \end{cases}$$

Ved å skissere opp $L(\theta)$ ser vi at den maksimeres for den minste verdien av θ der $\theta \ge \max(x_1, x_2, \ldots x_n)$. D.v.s $\widehat{\theta} = \max(X_1, X_2, \ldots X_n)$.

c)

$$\begin{aligned} G(y) &= P(Y \le y) = P(\max(x_1, x_2, \ldots, x_n) \le y) \\ &= P(x_1 \le y \cap x_2 \le y \cap \ldots \cap x_n \le y) \\ &= P(x_1 \le y) \cdot P(x_2 \le y) \cdot \ldots \cdot P(x_n \le y) \\ &= F(y) \cdot F(y) \cdot \ldots \cdot F(y) = (F(y))^n \\ &= \begin{cases} 0 & \text{hvis } y < 0, \\ \left(\frac{y}{\theta}\right)^n & \text{hvis } 0 \le y \le \theta, \\ 1 & \text{hvis } y > \theta. \end{cases} \end{aligned}$$

Vi finner sannsynlighetstettheten til Y ved å derivere den kumulative fordelingsfunksjonen:

$$g(y) = G'(y) = \begin{cases} 0 & \text{hvis } y < 0, \\ n\left(\frac{y}{\theta}\right)^{n-1} \cdot \theta^{-1} & \text{hvis } 0 \le y \le \theta, \\ 0 & \text{hvis } y > \theta. \end{cases}$$
$$= \begin{cases} \frac{y^{n-1}}{n \cdot \theta^n} & \text{hvis } 0 \le y \le \theta, \\ 0 & \text{ellers.} \end{cases}$$

d) Skal vise at estimatoren vår ikke er forventningsrett.

$$\begin{aligned} E(\widehat{\theta}) = E(Y) &= \int_{-\infty}^{\infty} y \cdot g(y)\, dy \\ &= \int_0^\theta y \cdot n\frac{y^{n-1}}{\theta^n}\, dy = \int_0^\theta n\frac{y^n}{\theta^n}\, dy \\ &= \left[\frac{n}{n+1} \cdot \frac{y^{n+1}}{\theta^n}\right]_0^\theta = \frac{n}{n+1} \cdot \frac{\theta^{n+1}}{\theta^n} \\ &= \frac{n}{n+1} \cdot \theta \neq \theta \end{aligned}$$

Dvs. $\widehat{\theta}$ er ikke forventingsrett. Skal bestemme k slik at $\widetilde{\theta} = kY$ er forventningsrett.

$$E(\widetilde{\theta}) = \theta$$
$$\Downarrow$$
$$E(kY) = k \cdot E(Y) = k \cdot \frac{n}{n+1} \cdot \theta = \theta$$
$$\Downarrow$$
$$k = \frac{n+1}{n}$$

e) Finner først fordelinga til $Z = \widetilde{\theta}/(k\theta)$ ved hjelp av transformasjonsformelen

$$f_Z(z) = f_Y(y(z)) \cdot |y'(z)|$$

der

$$z(y) = \frac{\widetilde{\theta}}{k\theta} = \frac{y}{\theta} \qquad \Leftrightarrow \qquad \begin{cases} y(z) = z\theta \\ y'(z) = \theta \end{cases}$$

Fordelinga til Z blir da

$$f_Z(z) = \begin{cases} \frac{n(z\theta)^{n-1}}{\theta^n} \cdot \theta = nz^{n-1} & \text{for } 0 \le z \le 1 \\ 0 & \text{ellers} \end{cases}$$

Da har vi at

$$P(L_z \le Z \le U_z) = 0,95$$

der L_z og U_z henholdsvis er 2,5% og 97,5% kvantilene til Z-fordelinga. Dermed gjenstår det bare å bestemme kvantilene.

$$\int_0^{L_z} f_Z(z)\,dz = [z^n]_0^{L_z} = L_z^n = 0,025$$
$$\Downarrow$$
$$L_z = \sqrt[n]{0,025}$$

$$\int_0^{U_z} f_Z(z)\,dz = [z^n]_0^{U_z} = U_z^n = 1 - 0,025$$
$$\Downarrow$$
$$U_z = \sqrt[n]{0.975}$$

Dermed kan vi sette

$$P\left(\sqrt[n]{0,025} \leq \frac{\widetilde{\theta}}{k\theta} \leq \sqrt[n]{0.975}\right) = 0,95$$
$$\Downarrow$$
$$P\left(\frac{\widetilde{\theta}}{k\sqrt[n]{0.975}} \leq \theta \leq \frac{\widetilde{\theta}}{k\sqrt[n]{0.025}}\right) = 0,95$$

Innsatt for k blir da konfidensintervallet

$$\left[\frac{n\widetilde{\theta}}{(n+1)\sqrt[n]{0.975}}, \frac{n\widetilde{\theta}}{(n+1)\sqrt[n]{0.025}}\right].$$

Med $n = 10$ og $y = 0,46$ får vi

$$\widetilde{\theta} = \frac{10+1}{10} \cdot 0,46 = 0,506,$$

og konfidensintervallet blir

$$\left[\frac{10 \cdot 0,506}{11 \cdot \sqrt[10]{0.975}}\,,\,\frac{10 \cdot 0,506}{11 \cdot \sqrt[10]{0.025}}\right]$$
$$= [0,461\ ,\ 0,665].$$