

TEORIA

③ Xi no Bassantall $x = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100$

①

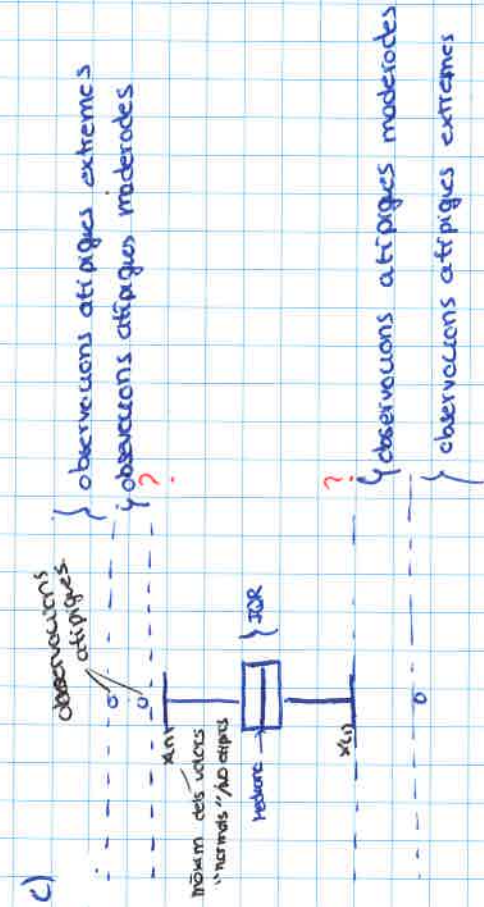
a) Que un estadístic sigui robust en presència d'atípics significa que el seu valor no es veurà gaire influenciat per un valor extrem de la mostra.

És a dir, que l'observació atípic no es veu reflectida en el valor de l'estadística.

b) La mediana i la mitjana retallada són robustes en presència d'atípics.

També ho són el IQR, el quartil Q_1 i el percentil 95.

La



En un box-plot tenim dibuixades línies discontinues que ens mostren els límits inferiors i superiors per als valors considerats atípics moderats i atípics extremes (e). Les observacions atípiques tenen representades per punts (e) aquests límits es calculen a partir dels quartils

② $X \sim N(\mu, \sigma^2)$ $x_1, \dots, x_n$ mos.

a) $k_1 \frac{\bar{X} - \mu}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2}} = k_1 \frac{\bar{X} - \mu}{\sqrt{(n-1)S}}$

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

• Si posem $|k_1| = \sqrt{n(n-1)}$ obtenim $\sqrt{n} \frac{\bar{X} - \mu}{S} = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$

b) $k_2 \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{\sum_{i=1}^n (x_i - \mu)^2} = k_2 \frac{(n-1)S^2}{n\sigma^2}$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

• Sabem que $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ i que $\frac{n\sigma^2}{\sigma^2} \sim \chi_n^2$

Per tant, $\frac{(n-1)S^2}{n\sigma^2} \sim \frac{\chi_{n-1}^2}{\chi_n^2}$

Sabem que $\frac{\chi_{n-1}^2/(n-1)}{\chi_n^2/n} \sim F_{n-1, n}$

Així doncs si $|k_2| = \frac{n}{n-1}$ obtenim $\frac{\frac{\chi_{n-1}^2}{n-1}}{\frac{\chi_n^2}{n}} = \frac{\chi_{n-1}^2}{\chi_n^2} \sim F_{n-1, n}$

c) $k_3 \frac{(\bar{X} - \mu)^2}{\sum_{i=1}^n (x_i - \mu)^2} = k_3 \frac{\left(\frac{\bar{X} - \mu}{\sigma}\right)^2}{\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2} = k_3 n \frac{\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2}{\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2} \sim k_3 n \frac{\chi_1^2}{\chi_n^2}$

$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \sum_{i=1}^n N(0,1)^2 = \chi_n^2$

$N^2(0,1) = \chi_1^2$ Per tant $\frac{\chi_1^2}{\chi_n^2} \sim F_{1, n}$

Així doncs amb $|k_3| = 1$ tenim $k_3 \frac{(\bar{X} - \mu)^2}{\sum_{i=1}^n (x_i - \mu)^2} = \frac{\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2}{\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right)^2} \sim \frac{\chi_1^2}{\chi_n^2} \sim F_{1, n} = F_n$

Assignatura: ESTADÍSTICA

Estudiant/a: _____

Data: 26/03/2021

TEORIA

③ $X \sim \text{Poisson}(\lambda)$ $X = (x_1, \dots, x_n) \text{ i.i.d.}$

a) Calculem primer l'estimador màxim versemblant de λ , $\hat{\lambda}_{ML}$, i després apliquem el principi d'invariança per calcular l'estimador de la probabilitat.

$$f_X(x|\lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$$

Funció de versemblança: $\mathcal{L}(\lambda|X) = \prod_{i=1}^n f(x_i|\lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{1}{\prod_{i=1}^n x_i!} e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}$

$X \text{ mas}$

$$\ell(\lambda|x) = \log \mathcal{L}(\lambda|x) = \log\left(\frac{1}{\prod_{i=1}^n x_i!}\right) - n\lambda + \sum_{i=1}^n x_i \log \lambda$$

$$\frac{\partial \ell}{\partial \lambda} = -n + \frac{\sum_{i=1}^n x_i}{\lambda}$$

Igualem a 0 la funció d'isore per trobar un extrem de la funció:

$$-n + \frac{\sum_{i=1}^n x_i}{\lambda} = 0 \Leftrightarrow \frac{\sum_{i=1}^n x_i}{\lambda} = n \Leftrightarrow \lambda = \frac{\sum_{i=1}^n x_i}{n} = \bar{X}$$

Per tant, $\boxed{\hat{\lambda}_{ML} = \bar{X}}$

• Cal comprovar que l'extrem que hem trobat és un màxim:

$$\frac{\partial^2 \ell}{\partial \lambda^2} = -\frac{\sum_{i=1}^n x_i}{\lambda^2} < 0 \quad \checkmark \quad x_i \geq 0$$

Volem $P(X > 0) = 1 - P(X = 0) = 1 - \frac{e^{-\lambda} \lambda^0}{0!} = 1 - e^{-\lambda}$

El domini de X és $X \in [0, \infty)$

Del principi d'invariança sabem que l'estimador màxim-versemblant de $\tau(\theta)$ és $\tau(\hat{\theta}_{ML})$

Altores, podem estimar la probabilitat per $1 - e^{-\hat{\lambda}_{ML}} = 1 - e^{-\bar{X}}$, on $\bar{X}$ és la mitjana mostral

b) $I_X(\tau(\lambda)) = \frac{I_X(\lambda)}{\tau'(\lambda)^2}$ Per tant calculem primer $I_X(\lambda)$

$$I_X(\lambda) = \mathbb{E} \left[\frac{-\partial^2}{\partial \lambda^2} \log f_X(x|\lambda) \right] = \mathbb{E} \left[\frac{\sum_{i=1}^n x_i}{\lambda^2} \right] = \frac{1}{\lambda^2} \sum_{i=1}^n \mathbb{E}[x_i] = \frac{1}{\lambda^2} n\lambda = \frac{n}{\lambda}$$

$\frac{\partial^2 \ell}{\partial \lambda^2} = -\frac{\sum x_i}{\lambda^2}$ (calculat prèviament) $\frac{x_i}{\mathbb{E}[x_i]} = \lambda$

$$\frac{x_i}{\mathbb{E}[x_i]} = \lambda$$

$$\tau(\lambda) = 1 - e^{-\lambda}$$

$$\tau'(\lambda) = e^{-\lambda}$$

$$\tau'(\lambda)^2 = e^{-2\lambda}$$

$$\Rightarrow I_{\tilde{X}}(\tau(\lambda)) = \frac{\frac{n}{\lambda}}{e^{2\lambda}} = \frac{\frac{n e^{2\lambda}}{\lambda}}{\quad}$$

4)

- a) Significa que la probabilitat de l'interval no contingui el paràmetre es reparteix equitativament a ambos costats de l'interval

- b) Quan es calcula un interval de confiança ~~del 50%~~

Assignatura: ESTADÍSTICA

Estudiant/a: [REDACTED]

Data: 26/03/2021

PROBLEMA 1

$Y_1, \dots, Y_n$ var. i.i.d. $Y_i = \beta x_i + \varepsilon_i$ $x_1, \dots, x_n$ constants
 $\varepsilon_1, \dots, \varepsilon_n$ i.i.d. $\varepsilon_i \sim N(0, \sigma^2)$

(Sabem que $x_i \sim N(\mu, \sigma^2)$ independents)
 $\Rightarrow \sum_{i=1}^n a_i x_i \sim N(\sum_{i=1}^n a_i \mu, \sum_{i=1}^n a_i^2 \sigma^2)$

Per tant, sabem que $\varepsilon_i \sim N(0, \sigma^2)$

$$Y_i = \beta x_i + \varepsilon_i \sim N(\beta x_i, \sigma^2)$$

Aleshores
$$f_Y(y_i | \beta, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - \beta x_i)^2}{2\sigma^2}}$$

1) Un estadístic $T(X)$ és suficient si la funció de distribució conjunta descompon com

$$f_X(x | \beta, \sigma^2) = g(T(X) | \beta, \sigma^2) h(x)$$

Calculem f_X : $f_X(y | \beta, \sigma^2) = \prod_{i=1}^n f_Y(y_i | \beta, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2}$
 $x_1, \dots, x_n$ indep. perquè són

Podem prendre $h(y) = 1$ i $g(T(X)) = (T_1(X), T_2(X) | \beta, \sigma^2) = T_1(X) \cdot e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2}$

on $T_1(X) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n$ i $T_2(X) = \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2$

No son estadístics
-0.8

2) $\hat{\beta}_1$?

Calculem la funció de versemblança (suposem σ^2 fix)

$$\mathcal{L}(\beta | X) = \prod_{i=1}^n f_Y(y_i) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2}$$

$$\ell(\beta | X) = -n \log(\sqrt{2\pi}) - n \log(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2$$

$$\frac{\partial \ell}{\partial \beta} = -\frac{1}{2\sigma^2} \cdot 2 \cdot \sum_{i=1}^n (y_i - \beta x_i)(-x_i) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i x_i - \beta x_i^2) = 0 \Leftrightarrow \sum_{i=1}^n y_i x_i = \beta \sum_{i=1}^n x_i^2 \Leftrightarrow \hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i^2}$$

3) $\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i^2}$, on $y_i \sim N(\beta x_i, \sigma^2)$

$\sum_{i=1}^n y_i x_i \stackrel{(*)}{\sim}_{(y_i=x_i)} N(\sum_{i=1}^n x_i^2 \beta, \sum_{i=1}^n x_i^2 \sigma^2)$

$\left[\hat{\beta}_1 = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i^2} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^n x_i^2}\right) \right]$

(*) $\left[\text{volem que si } X_i \sim N(\mu_i, \sigma_i^2) \text{ indep.} \Rightarrow \sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right) \right]$

• I que les Y_i són indep (perquè ε_i ho són)

En el nostre cas tenim $\mu_i = \beta x_i$
 $\sigma_i^2 = \sigma^2$

4) $\left[\hat{\beta}_2 = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \stackrel{(*)}{\sim}_{(y_i=x_i)} \frac{1}{\sum_{i=1}^n x_i} N\left(\beta \sum_{i=1}^n x_i, n \sigma^2\right) \sim N\left(\beta, \frac{n \sigma^2}{\left(\sum_{i=1}^n x_i\right)^2}\right) \right]$

$\left[\hat{\beta}_3 = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{x_i} \stackrel{(*)}{\sim}_{(y_i=x_i)} \frac{1}{n} N\left(\sum_{i=1}^n \frac{\beta x_i}{x_i}, \sum_{i=1}^n \frac{\sigma^2}{x_i^2}\right) = \frac{1}{n} N\left(n \beta, \sum_{i=1}^n \frac{\sigma^2}{x_i^2}\right) \sim N\left(\beta, \frac{\sigma^2}{n^2} \sum_{i=1}^n \frac{1}{x_i^2}\right) \right]$

5) Definim $\hat{\beta}_4 = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i}$

Volem que y_i té baix:

$E[\hat{\beta}_4] = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i} E[y_i x_i] = \sum_{i=1}^n \beta x_i$
 $y_i \sim N(\beta x_i)$

5) Definim $\hat{\beta}_4 := \frac{1}{\sum_{i=1}^n x_i} \cdot \sum_{i=1}^n \frac{y_i}{x_i}$

Tenim un estimador no esbiaixat ja que

$E[\hat{\beta}_4] = \frac{1}{\sum_{i=1}^n x_i} \cdot \sum_{i=1}^n \frac{E[y_i]}{x_i} = \frac{1}{\sum_{i=1}^n x_i} \cdot \sum_{i=1}^n \frac{\beta \sum_{i=1}^n x_i}{x_i} = \beta \Rightarrow [B_\beta(\hat{\beta}_4) = E[\hat{\beta}_4] - \beta = 0]$

$y_i \sim N(\beta x_i, \sigma^2) \Rightarrow E[y_i] = \beta x_i$

6) Reestimem tot β

Em quedaria amb l'estimador que tingues variances mínima, i.e. ~~$\hat{\beta}_3$~~ ~~NO~~
ja que tots són no esbiaixats.

Assignatura: **ESTADÍSTICA**

Estudiant/a:

Data: 26/03/2021

PROBLEMA 2

$X = \#$ vegades cara superior negre

$n=10 \rightarrow y=3$

$z = 20 \cdot \frac{2}{10} \cdot \frac{1}{6} = \frac{2}{3}$

1) $H = \{Y \sim \text{Bin}(10, p) \mid p \in [0, 1]\}$ model estadístic experiment

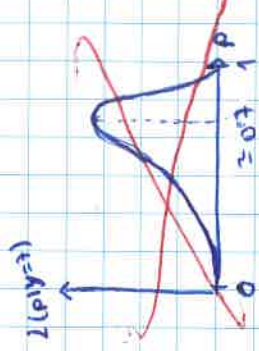
2) $P(Y=y) = \binom{10}{y} p^y (1-p)^{10-y}$

$Y_i \sim \text{Bern}(p) \rightarrow Y = \sum_{i=1}^{10} Y_i \sim \text{Bin}(10, p)$

$L_Y(p) = \mathcal{L}(p|Y) = \binom{10}{y} p^y (1-p)^{10-y} = 120 p^3 (1-p)^7$

$L_Y(0) = 0$
 $L_Y(1) = 0$
 $L_Y(0.5) = \frac{15}{128} \approx 0.117$
 $L_Y(0.25) \approx 0.1003$
 $L_Y(0.3) \approx 0.117$

$\frac{\partial \mathcal{L}}{\partial p} = 3 \cdot 120 p^2 (1-p)^7 - 120 \cdot 3 p^3 (1-p)^2$

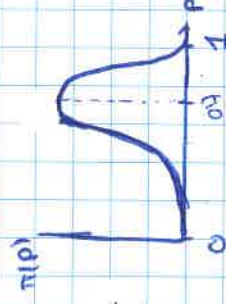


seu punt final
0.25

Busco el ^{extremo} máximo de la función
 $\frac{\partial \mathcal{L}}{\partial p} = 360 p^2 (1-p)^7 - 120 \cdot 3 p^3 (1-p)^2 = 0$ cuando $p=0.25$

3) Podem triar una prior $\pi(p)$ amb una distribució, per exemple $N(0.7, 1)$

No
seu punt final



0.25

Per calcular la distribució a posteriori fem:

$\pi(p|Y) = \frac{\mathcal{L}(p|Y) \pi(p)}{\int_0^1 \mathcal{L}(p|Y) \pi(p) dp}$

Si $y=3$, $\pi(p|Y) = \frac{120 p^3 (1-p)^7 \frac{1}{\sqrt{2\pi}} e^{-\frac{(p-0.7)^2}{2}}}{\int_0^1 p^3 (1-p)^7 \frac{1}{\sqrt{2\pi}} e^{-\frac{(p-0.7)^2}{2}} dp}$

4) En aquest cas, ~~binomial~~ $2^n(p)$ la moneda

la probabilitat de cara o creu d'una moneda segueix una Bernoulli ($\frac{1}{2}$)

Així que # carres negres $\sim$ ~~$\text{Bin}(n, \frac{1}{2}) = \text{Bin}(10, \frac{1}{2})$~~

Teoria, doncs una miora $\left| \frac{1}{6} \text{Bin}(10, \frac{1}{2}) \right|$

Assignatura: ESTADÍSTICA

Estudiant/

Data: 28/03/2021

PROBLEMA 3

$x_1, \dots, x_n$ amb densitat de prob $f(x; \theta) = \frac{3\theta^3}{(x+\theta)^4}$ $x > 0, \theta > 0$.

1) Utilitzem la introducció i comencem calculant $E(x+\theta)$

$$E(x+\theta) = \int_0^{+\infty} (x+\theta) f(x; \theta) dx = \int_0^{+\infty} \frac{3\theta^3}{(x+\theta)^3} dx = 3\theta^3 \int_0^{+\infty} \frac{1}{(x+\theta)^3} dx = ?$$

$$E[(x+\theta)^2] = \int_0^{+\infty} (x+\theta)^2 \frac{3\theta^3}{(x+\theta)^4} dx = \int_0^{+\infty} \frac{3\theta^3}{(x+\theta)^2} dx = 3\theta^3 \int_0^{+\infty} \frac{1}{(x+\theta)^2} dx = ?$$

$$\Rightarrow E[x+\theta] = E[x] + \theta \Rightarrow E[x] = E[x+\theta] - \theta$$

Aleshores pel mètode dels moments $\Rightarrow \bar{X} = E[x] = \hat{\theta}$ $\Rightarrow$ d'aquí obtenim $\hat{\theta}_{MH}$

2) $V(\hat{\theta}_{MH}) = E(\hat{\theta}_{MH}^2) - E(\hat{\theta}_{MH})^2 = ?$

Un estimador és eficient si la té baix $(E[\hat{\theta}_{MH}] = \theta)$ i desdax la cota de Cramer-Rao, i.e

$$V(\hat{\theta}_{MH}) = \frac{1}{I_X(\theta)}$$

• Calculem doncs la informació de Fisher.

$$I_X(\theta) = n I_X(\theta) \quad I_X(\theta) = E \left[-\frac{\partial^2}{\partial \theta^2} \log f_X(x; \theta) \right]$$

$$\log f(x; \theta) = \log(3\theta^3) - 4 \log(x+\theta) = \log(3) + 3 \log(\theta) - 4 \log(x+\theta)$$

$$\frac{\partial \log f(x; \theta)}{\partial \theta} = \frac{3}{\theta} - \frac{4}{x+\theta} \quad , \quad \frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} = -\frac{3}{\theta^2} + \frac{4}{(x+\theta)^2}$$

$$\Rightarrow I_X(\theta) = E \left[\frac{3}{\theta^2} - \frac{4}{(x+\theta)^2} \right] = \frac{3}{\theta^2} - \frac{4}{E[(x+\theta)^2]}$$

No $\Rightarrow$ ~~$\frac{3}{\theta^2} - \frac{4}{E[(x+\theta)^2]}$~~

0.3

0.8

Així doncs si la funció de versemblança amb $\frac{1}{I_X(\theta)}$ i veiem que $\hat{\theta}_{ML}$ no té bias $\Rightarrow \hat{\theta}_{ML}$ eficient.

3) Sabem que un estimador mèdian veient de la següent distribució asimptòtica.

$$" \hat{\theta}_{ML} \rightarrow N\left(\theta, \frac{1}{I_X(\theta)}\right) \rightarrow \sqrt{n I_X(\theta)} (\hat{\theta}_{ML} - \theta) \rightarrow N(0, 1)$$

-0.3

$$4) P(X \leq \text{med}) = 0.5 \quad \text{c} \quad P(X \geq \text{med}) = 0.5$$

$$P(X \leq \text{med}) = \int_0^{\text{med}} \frac{3e^3}{(x+\theta)^4} dx = 0.5 \quad P(X \geq \text{med}) = \int_{\text{med}}^{\infty} \frac{3e^3}{(x+\theta)^4} dx = 0.5$$

Estimem aquestes probabilitats per $\hat{\theta}_{ML}$ en lloc de θ i trobem el valor de la mediana (med)
~~de veritat~~ tal que compleixi les igualtats anteriors

$$\hat{\theta}_{ML} \quad \mathcal{L}(\theta | X) = \prod_{i=1}^n \frac{3e^3}{(x_i + \theta)^4} = (3e^3)^n \frac{1}{\prod_{i=1}^n (x_i + \theta)^4} \quad x_i > 0$$

$$l(\theta | X) = \log \mathcal{L}(\theta | X) = n \log(3e^3) - \sum_{i=1}^n 4 \log(x_i + \theta) = n \log(3) + 3n \log(e) - 4 \sum_{i=1}^n \log(x_i + \theta)$$

$$\frac{\partial l}{\partial \theta} = \frac{3n}{e} - 4 \sum_{i=1}^n \frac{1}{x_i + \theta} = 0 \Leftrightarrow \frac{3n}{e} = 4 \sum_{i=1}^n \frac{1}{x_i + \theta} \Leftrightarrow \frac{3}{4} n = \sum_{i=1}^n \frac{e}{x_i + \theta}$$

?