

1. Stel $X_1, \dots, X_9 \sim \mathcal{N}(\nu_1, \tau_1^2)$ en $Y_1, \dots, Y_9 \sim \mathcal{N}(\nu_2, \tau_2^2)$ het aantal studieuren vóór en na de bijeenkomst. We mogen aannemen dat $Z_i = Y_i - X_i \sim \mathcal{N}(\mu, \sigma^2)$.
Er geldt $\bar{z}_9 = 2$ en $s_9^2 = \frac{1}{8} \sum_{i=1}^9 (z_i - \bar{z}_9)^2 = 54/8 = 6.75$.

- (a) Nulhypothese: $H_0: \mu \leq 0$ tegen alternatieve hypothese $H_1: \mu > 0$.
($H_0: \mu = 0$ tegen $H_1: \mu \neq 0$ mag ook.)

Toetsingsgrootte:

$$T = \sqrt{9} \frac{\bar{Z}_9 - \mu}{S_9}$$

Onder H_0 heeft T een t_8 -verdeling.

Kritiek gebied: $[2.90, \infty)$

Realisatie: $t = 3 \times 2 / \sqrt{6.75} = 2.31 \notin K$, dus we verwerpen de nulhypothese niet.

- (b) Er geldt

$$\mathbb{P}_\mu \{ -t_{8;0.025} \leq \sqrt{9} (\bar{Z}_9 - \mu) / S_9 \leq t_{8;0.025} \} = 0.95$$

Dit levert als 95% betrouwbaarheidsinterval

$$\left[\bar{z}_9 - t_{8;0.025} \times \frac{s_9}{\sqrt{9}}, \bar{z}_9 + t_{8;0.025} \times \frac{s_9}{\sqrt{9}} \right]$$

Invullen levert $[-0.0005, 4.0005]$

2. (a) De log-likelihood heeft de volgende vorm

$$L(p; k) = \log \binom{k-1}{r-1} + (k-r) \log(1-p) + r \log p$$

Als $k = r$, dan heeft de functie een randmaximum in $p = 1$.

Als $k > r$, dan geldt in ieder geval dat het maximum wordt aangenomen in $(0, 1)$. Differentiëren van de likelihood geeft $L'(p; x) = -\frac{k-r}{1-p} + \frac{r}{p}$. Nulstellen van afgeleide, oplossen naar p , en controleren of extreem maximum is levert als meest aannemelijke schatter voor p : $\hat{p}(X) = r/X$. Ook voor $k = r$ geldt deze formule.

Wegens de invariantie van de meest aannemelijke schatter onder monotone transformaties geldt dat de meest aannemelijke schatter voor $1/p$ gelijk is aan X/r .

- (b) We kunnen de kansverdeling herschrijven als een 1-parameter exponentiële familie:

$$P(X = k) = (p/(1-p))^r \binom{k-1}{r-1} e^{k \log(1-p)}$$

Aangezien $\{\log(1-p) \mid p \in (0, 1]\}$ een niet-leeg inwendig heeft, geldt dat X een volledige en voldoende statistiek is.

- (c) Omdat de zuivere schatter voor $1/p$ die een functie is van X de UMVZ schatter is, en omdat $E_p(X) = \frac{r}{p}$, geldt dat X/r de UMVZ schatter is.

- (d) De Cramér-Rao ongelijkheid zegt dat ieder zuivere schatter van $g(p)$ een variantie heeft die minstens gelijk is aan $\frac{g'(p)^2}{I_p}$.

In ons geval is $g'(p) = [1/p]' = -1/p^2$. De Fisher informatie I_p is gelijk aan $\text{var}_p \dot{\ell}_p(X)$. Er geldt $\dot{\ell}_p(X) = -\frac{X-r}{1-p} + \frac{r}{p}$, dus

$$I_\theta = \frac{1}{(1-p)^2} \frac{r(1-p)}{p^2} = \frac{r}{p^2(1-p)}.$$

Uiteindelijk krijgen we dus dat de variantie minstens gelijk is aan: $\frac{1-p}{rp^2}$.

Deze ondergrens wordt gehaald door X/r : $\text{var}_p(X/r) = \frac{1}{r^2} \text{var}_p(X) = \frac{1-p}{rp^2}$.

3. (a) De likelihood ratio statistiek (LRS) wordt gegeven door

$$\begin{aligned} \text{LRS} &= \frac{\sup_{\theta > 0} \prod_{i=1}^n f_\theta(X_i)}{\prod_{i=1}^n f_{\theta=\theta_0}(X_i)} \\ &= \frac{\prod_{i=1}^n f_{\theta=\hat{\theta}}(X_i)}{\prod_{i=1}^n f_{\theta=\theta_0}(X_i)} \\ &= \frac{\theta_0^{2n}}{\left(\frac{1}{2n} \sum_{i=1}^n X_i^2\right)^n} e^{-n + \sum_{i=1}^n X_i^2 / (2\theta_0^2)} \end{aligned}$$

- (b) De score functie op basis van één waarneming is gelijk aan:

$$\dot{\ell}_\theta(x) = \frac{-2}{\theta} + \frac{x^2}{\theta^3}.$$

De Fisher informatie op basis van één waarneming is derhalve

$$\text{var}_\theta \dot{\ell}_\theta(X) = \frac{1}{\theta^6} \text{var}_\theta(X^2) = \frac{4}{\theta^2}$$

Dus de plug-in schatter wordt gegeven door $\frac{4}{\frac{1}{2n} \sum_{i=1}^n X_i^2} = \frac{8}{\frac{1}{n} \sum_{i=1}^n X_i^2}$.

De waargenomen Fisher informatie wordt gegeven door $-\frac{1}{n} \sum_{i=1}^n \ddot{\ell}_{\hat{\theta}}(X_i)$.

Vanwege $\ddot{\ell}_\theta(x) = \frac{2}{\theta^2} - \frac{3x^2}{\theta^4}$, krijgen we

$$-\frac{1}{n} \sum_{i=1}^n \ddot{\ell}_{\hat{\theta}}(X_i) = -\frac{1}{n} \sum_{i=1}^n \left[\frac{2}{\frac{1}{2n} \sum_{i=1}^n X_i^2} - \frac{3X_i^2}{\left(\frac{1}{2n} \sum_{i=1}^n X_i^2\right)^2} \right] = \frac{8}{\frac{1}{n} \sum_{i=1}^n X_i^2}$$

- (c) Een benaderend 95% betrouwbaarheidsinterval voor θ gebaseerd op de meest aannemelijke schatter is

$$\theta = \hat{\theta} \pm \frac{1}{\sqrt{n} \sqrt{\hat{i}_\theta}} \xi_{0.025} = \sqrt{\frac{1}{2n} \sum_{i=1}^n X_i^2} \pm 1.96 \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}}{\sqrt{8n}}$$

Een benaderend 95% betrouwbaarheidsinterval voor θ gebaseerd op de LRS is gebaseerd op de eigenschap dat $2 \log LRS \rightsquigarrow \chi_1^2$ -verdeling als $n \rightarrow \infty$. Met andere woorden, het 95% betrouwbaarheidsinterval wordt gegeven door

$$\begin{aligned} &\{ \theta : 4n \log \theta - 2n \log \frac{1}{2n} \sum_{i=1}^n X_i^2 - 2n + \sum_{i=1}^n X_i^2 / \theta^2 \leq \chi_{1,0.95}^2 \} \\ &= \{ \theta : 4n \log \theta + \sum_{i=1}^n X_i^2 / \theta^2 \leq 2n \log \frac{1}{2n} \sum_{i=1}^n X_i^2 + 2n + 3.84 \} \end{aligned}$$

4. (a) Er geldt

$$\int_0^\theta x \frac{2(\theta - x)}{\theta^2} dx = \frac{1}{3}\theta.$$

Gelijkstellen van de verwachting aan het steekproefgemiddelde $\bar{x}_n$ levert

$$T_1 = 3\bar{X}_n$$

- (b) De likelihood is gelijk aan

$$L(\theta; (x_1, \dots, x_n)) = \prod_{i=1}^n \frac{2(\theta - x_i)}{\theta^2} 1_{[0, \theta]}(x_i) = \frac{2^n}{\theta^{2n}} 1_{[x_{(n)}, \infty]}(\theta) \prod_{i=1}^n (\theta - x_i)$$

Invullen van $\theta = x_{(n)}$ levert $L(x_{(n)}; (x_1, \dots, x_n)) = 0$, terwijl de likelihood voor $\theta < x_{(n)}$ ook 0 is, en voor $\theta > x_{(n)}$ is de likelihood groter dan nul. Dus het maximum van de likelihoodfunctie wordt aangenomen in een waarde groter dan $x_{(n)}$.

NB: De m.a.s. wordt gevonden uit de vergelijking

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

- (c) Alle informatie over θ zit in een voldoende statistiek. We gaan op zoek naar een voldoende statistiek met een zo laag mogelijke dimensie. De simultane kansdichtheid heeft de vorm

$$\prod_{i=1}^n \frac{2(\theta - x_i)}{\theta^2} 1_{[0, \theta]}(x_i)$$

Als T een voldoende statistiek is, dan kunnen we deze dichtheid schrijven in de vorm

$$p_\theta(x_1, \dots, x_n) = g_\theta(T(x_1, \dots, x_n))h(x_1, \dots, x_n).$$

Omdat $\prod_{i=1}^n (\theta - x_i)$ niet te vereenvoudigen is, is iedere voldoende statistiek van dimensie n .

- (d) We kunnen de simultane kansdichtheid herschrijven tot

$$\frac{2^n}{\theta^{2n}} e^{\log \prod_{i=1}^n (\theta - x_i) + \log \prod_{i=1}^n 1_{[0, \theta]}(x_i)}$$

De functie in de exponent kunnen we niet schrijven in de vorm $\sum_{j=1}^k Q_j(\theta) V_j(x_1, \dots, x_n)$.