

1. (a) Wegens de invariantie-eigenschap van de meest aannemelijke schatter geldt:

$$T_1 = \bar{X}_n (1 - \bar{X}_n).$$

- (b) Er geldt

$$\begin{aligned} S_X^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n-1} \left[\sum_{i=1}^n (X_i^2 - 2X_i\bar{X}_n + \bar{X}_n^2) \right] \\ &= \frac{1}{n-1} \left[\sum_{i=1}^n X_i^2 - 2n\bar{X}_n^2 + n\bar{X}_n^2 \right] \\ &= \frac{n}{n-1} (\bar{X}_n - \bar{X}_n^2) = \frac{n}{n-1} T_1 \end{aligned}$$

- (c) We kunnen dit uitwerken via

$$E_\theta \bar{X}_n (1 - \bar{X}_n) = E_\theta \bar{X}_n - E_\theta \bar{X}_n^2 = \theta - \frac{1}{n^2} E_\theta Y^2 = \theta - \frac{1}{n^2} [\text{var}_\theta(Y) + (E_\theta Y)^2],$$

maar als we gebruik maken van het feit dat S_X^2 een zuivere schatter voor de variantie is, vinden we meteen dat $E_\theta T_1 = \frac{n-1}{n} \theta (1 - \theta)$ geen zuivere schatter voor $\theta(1 - \theta)$ is.

- (d) De Bayes schatting $T(x)$ voor $\theta(1 - \theta)$ wordt verkregen via

$$\begin{aligned} T(x) &= \int_0^1 \theta(1 - \theta) p_{\bar{\Theta}=\theta|Y=x}(\theta) d\theta \\ &= \frac{\int_0^1 \theta(1 - \theta) \theta^{x+\alpha-1} (1 - \theta)^{n-x+\beta-1} d\theta}{B(x + \alpha, n - x + \beta)} \\ &= \frac{B(x + \alpha + 1, n - x + \beta + 1)}{B(x + \alpha, n - x + \beta)} \\ &= \frac{x + \alpha}{n + \alpha + \beta + 1} \frac{B(x + \alpha, n - x + \beta + 1)}{B(x + \alpha, n - x + \beta)} \\ &= \frac{x + \alpha}{n + \alpha + \beta + 1} \frac{n - x + \beta}{n + \alpha + \beta} \end{aligned}$$

Dus de Bayes schatter is

$$T(X) = \frac{(Y + \alpha)(n - Y + \beta)}{(n + \alpha + \beta + 1)(n + \alpha + \beta)}$$

2. (a) De aannemelijkheidsfunctie op basis van de waarnemingen $x_1, \dots, x_n$ is:

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

Dit is een differentieerbare functie in θ . De log-aannemelijkheidsfunctie is:

$$\ell(\theta) = -2n \log \theta + \sum_{i=1}^n \log x_i - \sum_{i=1}^n x_i^2 / (2\theta^2) + \sum_{i=1}^n \log 1_{[0,\infty)}(x_i)$$

Als afgeleide naar θ vinden we de functie $\theta \mapsto \frac{-2n}{\theta} + \frac{\sum_{i=1}^n x_i^2}{\theta^3}$. De afgeleide nul stellen en controleren of extreem een maximum is geeft de meest aannemelijke schatter voor θ en voor θ^2 .

- (b) Laat T_2 de momentenschatter voor θ^2 op basis van het tweede moment zijn. Wegens $E_\theta X_i^2 = 2\theta^2$ krijgen we $2T_2 = \overline{X_n^2}$, en dus

$$T_2 = \frac{1}{2n} \sum_{i=1}^n X_i^2.$$

Voor de momentenschatter op basis van het eerste moment krijgen we, wegens het gegeven $E_\theta X_i = \sqrt{\pi/2} \theta$, dat

$$T_1 = \frac{2}{\pi} \overline{X_n^2} = \frac{2}{\pi} \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2.$$

Er geldt

$$\begin{aligned} E_\theta \frac{2}{\pi} \overline{X_n^2} &= \frac{2}{\pi} [\text{var}_\theta \overline{X_n} + (E_\theta \overline{X_n})^2] \\ &= \frac{2}{\pi n} \text{var}_\theta X_i + \theta^2 \\ &= \frac{2}{\pi n} (2\theta^2 - \frac{\pi}{2} \theta^2) + \theta^2 \end{aligned}$$

Dus $E_\theta T_1 = \frac{4}{\pi n} \theta^2 - \theta^2/n + \theta^2$ en is derhalve geen zuivere schatter voor θ^2 .

- (c) We kunnen de dichtheid van de steekproef herschrijven als

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

Dit heeft de vorm van een 1-parameter exponentiele familie en de verzameling $\{-1/(2\theta^2) \mid \theta > 0\}$ heeft een niet-leeg inwendig. Dus $\sum_{i=1}^n X_i^2$ is een volledige en voldoende statistiek. Omdat $E_\theta \frac{1}{2n} \sum_{i=1}^n X_i^2 = \theta^2$, is $\frac{1}{2n} \sum_{i=1}^n X_i^2$ de UMVZ schatter voor θ^2 .

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

In ons geval is $g'(\theta) = [\theta^2]' = 2\theta$. De Fisher informatie I_θ is gelijk aan $\text{var}_\theta \dot{\ell}_\theta(X_1, \dots, X_n)$.

Er geldt $\dot{\ell}_\theta(X_1, \dots, X_n) = \frac{-2n}{\theta} + \frac{\sum_{i=1}^n X_i^2}{\theta^3}$, dus

$$I_\theta = \text{var}_\theta \left(\frac{-2n}{\theta} + \frac{\sum_{i=1}^n X_i^2}{\theta^3} \right) = \frac{n}{\theta^6} \text{var}_\theta(X_i^2) = \frac{n}{\theta^6} (8\theta^4 - 4\theta^4) = \frac{4n}{\theta^2}.$$

Uiteindelijk krijgen we dus dat de variantie minstens gelijk is aan $\frac{\theta^4}{n}$.
Deze ondergrens wordt gehaald door de UMVZ-schatter:

$$\text{var}_\theta \left(\frac{1}{2n} \sum_{i=1}^n X_i^2 \right) = \frac{1}{4n} \text{var}_\theta(X_i^2) = \frac{\theta^4}{n}.$$

3. (a) Omdat de waarnemingen voor en na de kerstvakantie gecorreleerd zijn (dezelfde personen), voeren we een toets voor paren uit. We nemen aan dat $Z_i \sim \mathcal{N}(\Delta, \sigma^2)$, $i = 1, \dots, n$, en dat de Z_i 's onderling onafhankelijk zijn. We toetsen $H_0 : \Delta = 0$ (of $H_0 : \Delta \leq 0$) tegen $H_1 : \Delta > 0$ en gebruiken de t-Toets voor paren. Als toetsingsgrootte kiezen we

$$\bar{T} = \sqrt{15} \frac{\frac{1}{15} \sum_{i=1}^{15} Z_i}{S_Z},$$

hetgeen onder $\Delta = 0$ een t-verdeling heeft met 14 vrijheidsgraden. We verwerpen de nulhypothese voor grote waarden van gemiddelde gewichtstoename. Als kritiek gebied bij $\alpha_0 = 0.01$ vinden we $[2.62, \infty)$. Het resultaat van de steekproef geeft $t = \sqrt{15} 0.9/1.5 = 2.32$. Omdat dit niet in het kritieke gebied ligt, kunnen we niet concluderen dat het gewicht in de kerstvakantie gemiddeld is toegenomen.

- (b) Het 95% betrouwbaarheidsinterval voor de t-toets voor paren wordt gegeven door

$$\left[\bar{z}_{15} - t_{14,0.975} \frac{s_z}{\sqrt{15}}, \bar{z}_{15} + t_{14,0.975} \frac{s_z}{\sqrt{15}} \right] = \left[0.9 - 2.14 \frac{1.5}{\sqrt{15}}, 0.9 + 2.14 \frac{1.5}{\sqrt{15}} \right] = [0.07, 1.73]$$

Omdat de waarde nul niet in het betrouwbaarheidsinterval ligt, wordt onder tweezijdige toetsing de nulhypothese dat er geen gewichtsverandering is opgetreden verworpen bij $\alpha_0 = 0.05$.

4. (a) Een stochast heeft een chi-kwadraat verdeling met n vrijheidsgraden als het dezelfde verdeling heeft als een som van n gekwadraterde, onderling onafhankelijke, standaard normaal verdeelde stochasten. Op grond van het gegeven, kan $\frac{2}{\lambda} \sum_{i=1}^n T_i$ geschreven worden als de som van $2n$ gekwadraterde, onderling onafhankelijke, standaard normaal verdeelde stochasten, en heeft dus een chi-kwadraat verdeling met $2n$ vrijheidsgraden.
- (b) Als toetsingsgrootte kiezen we $\bar{T}_{15} = \frac{1}{15} \sum_{i=1}^{15} T_i$ (of $\frac{2}{12} \sum_{i=1}^{15} T_i = \frac{1}{6} \sum_{i=1}^{15} T_i$). Onder de nulhypothese heeft $\frac{1}{6} \sum_{i=1}^{15} T_i$ een chi-kwadraat verdeling met 30 vrijheidsgraden. We verwerpen de nulhypothese voor grote waarden van de toetsingsgrootte. Het kritiek gebied vinden we via

$$\begin{aligned} P_{\lambda=12}(\bar{T} \geq k) &= P_{\lambda=12} \left(\frac{1}{15} \sum_{i=1}^{15} T_i \geq k \right) \\ &= P_{\lambda=12} \left(\frac{1}{6} \sum_{i=1}^{15} T_i \geq k \frac{30}{12} \right) = 0.05 \end{aligned}$$

Uit de tabel vinden we dat $k \frac{30}{12} = 43.8$, oftewel $K = [17.52, \infty)$. Wegens $\bar{t}_{15} = 15.4 \notin K$ verwerpen we de nulhypothese niet.

(c) Er geldt voor de kans op een fout van de tweede soort bij $\lambda = 14$

$$\begin{aligned}P_{\lambda=14}(\bar{T}_{100} \leq 13.968) &= P_{\lambda=14}\left(\frac{2}{14} \sum_{i=1}^{100} T_i \leq \frac{200}{14} 13.968\right) \\&= P_{\lambda=14}\left(\frac{\frac{2}{14} \sum_{i=1}^{100} T_i - 200}{20} \leq \frac{\frac{200}{14} 13.968 - 200}{20}\right) \\&= P(U \leq -0.023) \\&= P(U \geq 0.023) = 1 - 0.5092 = 0.4908\end{aligned}$$

NB In het oorspronkelijke tentamen bevond zich een fout, waardoor het antwoord wordt:

$$P_{\lambda=14}(\bar{T}_{100} \leq 13.968/100) = P(U \leq -9.9) = 0$$