

1. (a) Voor de binomiale verdeling geldt

$$P_\theta(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \dots, n$$

De log-likelihood heeft de volgende vorm

$$L(\theta; x) = \log \binom{n}{x} + x \log \theta + (n - x) \log(1 - \theta)$$

Als $x = 0$ (resp. $x = n$), dan heeft de functie een randmaximum in $\theta = 0$ (resp. $\theta = 1$). Als $1 \leq x \leq n-1$, dan geldt in ieder geval dat het maximum wordt aangenomen in $(0, 1)$. Differentiëren van de likelihood geeft $L'(\theta; x) = \frac{x}{\theta} - \frac{n-x}{1-\theta}$. Nulstellen van afgeleide, oplossen naar θ , en controleren of extreem maximum is levert als meest aannemelijke schatter voor θ : $\hat{\theta}(X) = X/n$. Ook voor $x = 0$ en $x = n$ geldt deze formule. Wegens de invariantie onder monotone transformaties geldt dat

$$(\hat{\theta}(X))^2 = \left(\frac{X}{n} \right)^2$$

de meest aannemelijke schatter voor θ^2 is.

- (b) Er moet gelden $E_\theta T(X) = \theta^2$ voor iedere θ .

Er geldt $\text{var}_\theta(X) = E_\theta X^2 - (E_\theta X)^2 = n\theta(1 - \theta) = n\theta - n\theta^2$.

Omdat $(E_\theta X)^2 = n^2\theta^2$, levert invullen $E_\theta X^2 = n\theta - n\theta^2 + n^2\theta^2$.

Omdat $n\theta = E_\theta X$ krijgen we $E_\theta X^2 - E_\theta X = (n^2 - n)\theta^2$, en dus wordt een zuivere schatter voor θ^2 gegeven door

$$T(X) = \frac{X^2 - X}{n^2 - n}$$

- (c) De Bayes schatting $\widehat{g(\theta)}$ voor $g(\theta)$ t.o.v. een a-priori verdeling $\bar{\Theta} \sim \pi(\theta)$ wordt verkregen via

$$\begin{aligned} T(x) = E(g(\bar{\Theta}) | X = x) &= \int g(\theta) p_{\bar{\Theta}=\theta | X=x}(\theta) d\theta \\ &= \frac{\int g(\theta) p_\theta(x) \pi(\theta) d\theta}{\int p_\theta(x) \pi(\theta) d\theta} \end{aligned}$$

De Bayes schatting kan dus gezien worden als een voorwaardelijke verwachting van een stochast $g(\bar{\Theta})$, gegeven $X = x$. Verwachtingen (en dus ook voorwaardelijke verwachtingen) zijn in het algemeen niet invariant onder transformaties. Dus in het algemeen geldt

$$g(\hat{\theta}(x)) = g(E(\bar{\Theta} | X = x)) \neq E(g(\bar{\Theta}) | X = x) = \widehat{g(\theta)}(x)$$

(We kunnen ook het antwoord uit deel (d) als tegenvoorbeeld gebruiken)

(d) De Bayes schatting $T(x)$ voor θ^2 wordt verkregen via

$$\begin{aligned}
 T(x) &= \int_0^1 \theta^2 p_{\bar{\Theta}=\theta|X=x}(\theta) d\theta \\
 &= \frac{\int_0^1 \theta^2 \theta^{x+\alpha-1} (1-\theta)^{n-x+\beta-1} d\theta}{B(x+\alpha, n-x+\beta)} \\
 &= \frac{B(x+\alpha+2, n-x+\beta)}{B(x+\alpha, n-x+\beta)} \\
 &= \frac{x+\alpha+1}{x+\alpha+1+n-x+\beta} \frac{B(x+\alpha+1, n-x+\beta)}{B(x+\alpha, n-x+\beta)} \\
 &= \frac{x+\alpha+1}{n+\alpha+\beta+1} \frac{x+\alpha}{n+\alpha+\beta}
 \end{aligned}$$

Dus de Bayes schatter is

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

2. (a) We nemen aan dat $X_i \sim \mathcal{N}(\mu_1, \sigma^2)$ en $Y_i \sim \mathcal{N}(\mu_2, \sigma^2)$. Als we van tevoren niet weten welke invloed stress precies heeft, toetsen we

$H_0 : \mu_1 = \mu_2$ tegen

$H_1 : \mu_1 \neq \mu_2$.

(We kunnen ook $H_0 : \mu_1 \leq \mu_2$ tegen $H_1 : \mu_1 > \mu_2$ toetsen.)

Als toetsingsgrootte nemen we

$$T = \frac{\bar{X}_{26} - \bar{Y}_{26}}{S_{XY} \sqrt{\frac{1}{26} + \frac{1}{26}}}$$

Hierbij is $S_{XY}^2 = \frac{1}{50} \left(\sum_{i=1}^{26} (X_i - \bar{X})^2 + \sum_{i=1}^{26} (Y_i - \bar{Y})^2 \right)$.

Onder de nulhypothese geldt $T \sim t_{50}$, en het kritieke gebied voor de toets is derhalve $(-\infty, -2.01] \cup [2.01, \infty)$.

Als realisatie van T vinden we (gebruik hier dat $\sum_{i=1}^{26} (x_i - \bar{x})^2 = 25 \times (11.6)^2$ en $\sum_{i=1}^{26} (y_i - \bar{y})^2 = 25 \times (13.2)^2$):

$$t = \frac{8}{\sqrt{\frac{25}{50} [(11.6)^2 + (13.2)^2]} \sqrt{\frac{2}{26}}} = 2.32$$

Deze waarde ligt in het kritieke gebied, dus we verwwerpen de nulhypothese en komen tot dezelfde conclusie als in het genoemde artikel.

3. (a) Er geldt

$$E\bar{X}_n = \frac{1}{n} \sum_{i=1}^n EX_i = \mu$$

Voor de variantie geldt:

$$E \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} E \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{n}{n-1} \left(\sigma^2 - \frac{\sigma^2}{n} \right) = \sigma^2$$

want

$$\begin{aligned}\sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n (X_i - \bar{X}_n + \bar{X}_n - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \bar{X}_n)^2 + \sum_{i=1}^n (\bar{X}_n - \mu)^2 + 2 \times (\bar{X}_n - \mu) \times \sum_{i=1}^n (X_i - \bar{X}_n)\end{aligned}$$

Wegens $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n) = \bar{X}_n - \bar{X}_n = 0$, valt de laatste term weg. Halen we de tweede term naar links, dan krijgen we

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - \frac{1}{n} \sum_{i=1}^n (\bar{X}_n - \mu)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Verwachting nemen levert

$$E \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \frac{1}{n} \sum_{i=1}^n (\text{var}(X_i) - \text{var}(\bar{X}_n)) = \sigma^2 - \frac{\sigma^2}{n}$$

In de laatste stap gebruiken we onafhankelijkheid als aanname: $\text{var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i)$.

- (b) Laat μ de verwachte bloeddruk van geïnfecteerde vrouwen zijn. We toetsen dan
 $H_0 : \mu = 114.8$ tegen $H_1 : \mu \neq 114.8$
 Toetsingsgrootheid: $T = \sqrt{16} \frac{\bar{X}_{16} - 114.8}{S_{16}}$, onder H_0 t_{15} verdeeld
 Kritieke gebied: $K = (-\infty, -2.13] \cup [2.13, \infty)$
 Realisatie toetsingsgrootheid: $t = 1.67$. Dus we verwerpen H_0 niet.
- (c) Laat σ^2 de variantie van de bloeddrukverdeling van geïnfecteerde vrouwen zijn. We toetsen dan
 $H_0 : \sigma = 13.1$ tegen $H_1 : \sigma \neq 13.1$
 Toetsingsgrootheid: $T = 15 \frac{S_X^2}{(13.1)^2}$, onder H_0 χ_{15}^2 verdeeld
 Kritieke gebied: $K = (0, 6.26] \cup [27.5, \infty)$
 Realisatie toetsingsgrootheid: $t = 44.3$. Dus we verwerpen H_0 .
- (d) Op grond van bovenstaande twee toetsen lijkt de virusinfectie invloed te hebben op de bloeddruk, in die zin dat de variantie van de verdeling verhoogd wordt.
- (e) Het onderscheidend vermogen is een functie van σ , die gedefinieerd wordt door

$$P_\sigma(T \in K) = P_\sigma \left(15 \frac{S_X^2}{(13.1)^2} \leq 6.26 \right) + P_\sigma \left(15 \frac{S_X^2}{(13.1)^2} \geq 27.5 \right)$$

Herschrijven levert:

$$P_\sigma \left(15 \frac{S_X^2}{(13.1)^2} \leq 6.26 \right) = P_\sigma \left(15 \frac{S_X^2}{\sigma^2} \leq \frac{(13.1)^2 \times 6.26}{\sigma^2} \right)$$

en

$$P_\sigma \left(15 \frac{S_X^2}{(13.1)^2} \geq 27.5 \right) = P_\sigma \left(15 \frac{S_X^2}{\sigma^2} \geq \frac{(13.1)^2 \times 27.5}{\sigma^2} \right),$$

waarbij $15 \frac{S_X^2}{\sigma^2}$ onder σ een χ_{15}^2 verdeling heeft.