

1. (a) $\mathcal{P}_p = \{P_p; P_p \text{ kansverdeling van } \sum_{i=1}^{100} Y_i, \text{ met } P_p(Y_i = 5) = p = 1 - P_p(Y_i = -1)\}$
 Stel X_i is "succes" bij vraag i . Dan geldt dus $W = \sum_{i=1}^{100} Y_i = \sum_{i=1}^{100} (6X_i - 1) = 6 \sum_{i=1}^{100} X_i - 100$ en derhalve

$$P(W = 6k - 100) = \binom{n}{k} p^k (1-p)^{100-k}, \quad k = 0, 1, \dots, 100$$

(NB. Ieder antwoord dat in ieder geval $P_p(Y_i = 5) = p = 1 - P_p(Y_i = -1)$ geeft is goed gerekend)

- (b) $E_p W = E_p \sum_{i=1}^{100} Y_i = \sum_{i=1}^{100} E_p Y_i = 100 \times [5p - (1-p)] = 100 \times (6p - 1)$
 (c) Omwerken van $E_p W = 100(6p - 1)$ levert $E_p \frac{1}{6} \left(\frac{W}{100} + 1 \right) = p$, dus $T = \frac{1}{6} \left(\frac{W}{100} + 1 \right)$ is een zuivere schatter voor p .
 (d) Omdat T een zuivere schatter is, is de verwachte kwadratische fout gelijk aan de variantie. We krijgen dan

$$\text{MSE}_p(T) = \text{var}_p \left(\frac{1}{6} \left[\frac{W}{100} + 1 \right] \right) = \frac{1}{36} \frac{1}{100^2} \text{var}_p(W) = \frac{1}{36} \frac{1}{100^2} \sum_{i=1}^{100} \text{var}_p(Y_i) = \frac{p(1-p)}{100},$$

$$\text{wegens } \text{var}_p(Y_i) = E_p Y_i^2 - (E_p Y_i)^2 = 25p + (1-p) - (6p-1)^2 = 36p(1-p)$$

2. (a) De aannemelijkheidsfunctie op basis van de waarnemingen $(x_1, y_1), \dots, (x_n, y_n)$ is:

$$\theta \mapsto L(\theta; (x_1, y_1), \dots, (x_n, y_n)) = \prod_{i=1}^n 3\theta^2 y_i^2 e^{-\theta(x_i + y_i^3)}$$

Dit is een continue functie in θ . Afgeleide nemen naar θ van de log-aannemelijkheidsfunctie levert

$$\frac{\partial}{\partial \theta} \sum_{i=1}^n [2 \log \theta + \log 3 y_i^2 - \theta(x_i + y_i^3)] = \frac{2n}{\theta} - \sum_{i=1}^n (x_i + y_i^3)$$

Nulstellen levert $\theta = \frac{2n}{\sum_{i=1}^n (x_i + y_i^3)}$, en inspectie van de afgeleide laat zien dat dit een

maximum is. Dus de meest aannemelijke schatter is gelijk aan $\hat{\theta} = \frac{2n}{\sum_{i=1}^n (X_i + Y_i^3)}$

- (b) De Bayes schatter voor θ wordt gegeven door

$$\begin{aligned} E[\Theta | (x_1, y_1), \dots, (x_n, y_n)] &= \int \theta p(\theta | (x_1, y_1), \dots, (x_n, y_n)) d\theta \\ &= \frac{\int \theta [\prod_{i=1}^n p_\theta(x_i, y_i)] \pi(\theta) d\theta}{\int [\prod_{i=1}^n p_\theta(x_i, y_i)] \pi(\theta) d\theta} \\ &= \frac{\int_0^\infty \theta^{2n+1} e^{-\theta[\sum_{i=1}^n (x_i + y_i^3) + 1]} d\theta}{\int_0^\infty \theta^{2n} e^{-\theta[\sum_{i=1}^n (x_i + y_i^3) + 1]} d\theta} \\ &= \frac{[\sum_{i=1}^n (x_i + y_i^3) + 1]^{2n+1} (2n+1)!}{[\sum_{i=1}^n (x_i + y_i^3) + 1]^{2n+2} (2n)!} = \frac{2n+1}{\sum_{i=1}^n (x_i + y_i^3) + 1} \end{aligned}$$

Dus de Bayes schatter ten opzichte van de gegeven dichtheid voor Θ is $\frac{2n+1}{\sum_{i=1}^n (X_i + Y_i^3) + 1}$.

(c) De simultane dichtheid heeft de vorm

$$p_{\theta}((x_1, y_1), \dots, (x_n, y_n)) = 3^n \theta^{2n} \prod_{i=1}^n y_i^2 e^{-\theta \sum_{i=1}^n (x_i + y_i^3)}$$

Dit vormt een exponentiële familie, met $c(\theta) = 3^n \theta^{2n}$, $h((x_1, y_1), \dots, (x_n, y_n)) = \prod_{i=1}^n y_i^2$, $Q_1(\theta) = -\theta$ en $V_1((x_1, y_1), \dots, (x_n, y_n)) = \sum_{i=1}^n (x_i + y_i^3)$ en een niet-leeg inwendig voor $\{-\theta | \theta \in (0, \infty)\}$. Dus $\sum_{i=1}^n (X_i + Y_i^3)$ vormt een volledige en voldoende statistiek voor deze familie van kansverdelingen.

3. (a) Plaatje: De empirische verdelingsfunctie $\hat{F}_5(x)$ heeft sprongen ter grootte 0.2 bij de waarden uit de data set (0.04, 0.71, 0.76, 0.94, 0.96), en loopt van 0 naar 1.
 (b) $T = \sup_{x \in \mathbb{R}} |\hat{F}_5(x) - F_0(x)|$, met $F_0(x) = x$ op $[0, 1]$. Dus $T = |\lim_{x \uparrow 0.72} \hat{F}_5(x) - F_0(0.71)| = |0.2 - 0.71| = 0.51$.
4. (a) Laat Y het aantal kaaltjes in 1000 meter pijp zijn. Er geldt $P_{\lambda}(Y = k) = e^{-\lambda} \frac{\lambda^k}{k!}$.
 Ten onrechte opnieuw afregelen als $\lambda = 8$ en $Y \geq 10$.
 Deze kans is: $1 - P_8(Y \leq 9) = 1 - 0.7166 = 0.2834$.
 $P_{10}(Y \leq 9) = 0.4579$.
 (b) Kritieke gebied: $K = \{k, k+1, \dots\}$.
 Onder nulhypothese moet gelden: $P_8(Y \geq k) \leq 0.05$.
 Wegens $P_8(Y \geq 13) = 1 - 0.9362 = 0.0638$ en $P_8(Y \geq 14) = 1 - 0.9658 = 0.0342$, geldt $k = 14$.
 Het werkelijke aantal kaaltjes is 12, dus we verwerpen de nulhypothese niet.
 (c) Onderscheidingsvermogen: $\omega(\lambda) = P_{\lambda}(Y \in K)$.
 $\omega(10) = P_{10}(Y \geq 14) = 1 - 0.8645 = 0.1355$.
 $\omega(15) = P_{15}(Y \geq 14)$
 Y is bij benadering $\mathcal{N}(15, 15)$ verdeeld. Dus er geldt, met continuïteitscorrectie,
 $P_{15}(Y \geq 14) = P_{15}(Y \geq 13.5) \approx P(U \geq (13.5 - 15)/\sqrt{15}) = P(U \geq -0.387) = P(U \leq 0.387) = 0.6509$. Hierbij is gebruik gemaakt van lineaire interpolatie:
 $P(U \leq 0.39) = 0.6517$ en $P(U \leq 0.38) = 0.6480$.
 $\omega(8) = 0.0342$ (tabellenboekje)
 $\omega(13.5) \approx P(U \geq 0) = 0.5$
 $\omega(17.5) \approx P(U \leq 0.956) = 0.8305$ (ook weer via lineaire interpolatie)
 Op basis van deze waarden kan een grafiek van het onderscheidend vermogen geschetst worden.
5. (a) Stel $X \sim \text{Bin}(n, p)$ Het 95% betrouwbaarheidsinterval op grond van de normale benadering wordt gegeven door $p = \frac{X}{n} \pm \frac{1}{\sqrt{n}} \sqrt{\frac{X}{n} (1 - \frac{X}{n})} \xi_{0.025}$. Invullen levert $p = 0.05 \pm \sqrt{\frac{19}{20^3}} \times 1.96 = 0.05 \pm 0.0955$
 Er zitten waarden kleiner dan nul in het betrouwbaarheidsinterval, hetgeen niet mogelijk is. (Dit komt doordat de normale benadering onnauwkeurig is bij kleine p en n . Er geldt $np \approx 20 \times 1/20 = 1$, terwijl als vuistregel geldt dat voor het kunnen toepassen van de normale benadering moet gelden $np \geq 5$ en $n(1-p) \geq 5$).

- (b) Vanwege de relatie tussen betrouwbaarheidsintervallen en toetsen, kan de waarde 0 alleen in het betrouwbaarheidsinterval liggen als de nulhypothese $H_0 : p = 0$ niet verworpen zou worden op grond van de data. Echter, $p = 0$ is onmogelijk als je minstens één “succes” vindt.
- (c) De linkerkant p_l en rechterkant p_r van het exacte betrouwbaarheidsinterval voldoen respectievelijk aan $P_{p_l}(X \geq 1) = 0.025$ en $P_{p_r}(X \leq 1) = 0.025$.
 $0.025 = P_{p_l}(X \geq 1) = 1 - P_{p_l}(X = 0) = 1 - (1 - p_l)^{20}$ levert $p_l = 1 - \sqrt[20]{0.975} = 0.0013$
 $P_{p_r}(X \leq 1) = 0.025$ lezen we af uit het tabellenboek: $P_{0.25}(X \leq 1) = 0.0243$. Interpolatie met $P_{0.20}(X \leq 1) = 0.0692$ geeft $p_r = 0.249$. Dus het exacte 95% betrouwbaarheidsinterval is $[0.0013; 0.249]$.

Normering:

1(a): 4	2(a): 8	3(a): 5	4(a): 5	5(a): 5
1(b): 3	2(b): 8	3(b): 4	4(b): 8	5(b): 5
1(c): 4	2(c): 6		4(c): 9	5(c): 8
1(d): 8				

Eindcijfer = (totaal + 10)/10