

**Tentamenuitwerking van het tentamen Mathematische Statistiek van
14 augustus 1997**

1.

$$F(x) = \int_0^x 2(1-\tilde{x}) d\tilde{x} = -(1-\tilde{x})^2 \Big|_0^x = 1 - (1-x)^2,$$

zodat voor $b_n > 0$,

$$P(b_n(X_{(n)} - a_n) \leq x) = P\left(X_1 \leq a_n + \frac{x}{b_n}\right)^n = \left(1 - \left(1 - a_n - \frac{x}{b_n}\right)^2\right)^n$$

als $a_n + \frac{x}{b_n} \in [0, 1]$. Kies $a_n = 1$ (merk ook op dat als $n \rightarrow \infty$, $X_{(n)} \xrightarrow{P} 1$). Dan geldt voor $x \leq 0$ dat

$$P(b_n(X_{(n)} - a_n) \leq x) = \left(1 - \frac{x^2}{b_n^2}\right)^n.$$

Kies $b_n = \sqrt{n}$, dan geldt voor $x \leq 0$ dat

$$P(b_n(X_{(n)} - a_n) \leq x) = \left(1 - \frac{x^2}{n}\right)^n \rightarrow e^{-x^2}$$

als $n \rightarrow \infty$, zodat $\sqrt{n}(X_{(n)} - 1)$ in verdeling convergeert naar een niet-triviale limiet.

2. **a.** Definieer $\alpha_i = EX^i$ en $\beta_i = EY^i$. Uit de Multivariate Centrale Limietstelling volgt dat

$$\sqrt{n} \left(\begin{pmatrix} \frac{\bar{X}}{\bar{X}^2} \\ \frac{\bar{Y}}{\bar{Y}^2} \end{pmatrix} - \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{pmatrix} \right) \rightsquigarrow \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma \right)$$

met

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \alpha_3 & 0 & 0 \\ \alpha_3 & \alpha_4 - \alpha_2^2 & 0 & 0 \\ 0 & 0 & \sigma_Y^2 & \beta_3 \\ 0 & 0 & \beta_3 & \beta_4 - \beta_2^2 \end{pmatrix},$$

aangezien $\alpha_1 = \beta_1 = 0$.

Omdat $S_{n,X}^2 = \bar{X}^2 - \bar{X}^2$ kan de limietverdeling van $S_{n,X}^2/S_{n,Y}^2$ bepaald worden met de Deltamethode met als afbeelding $\phi(x_1, x_2, y_1, y_2) = \frac{x_2 - x_1^2}{y_2 - y_1^2}$. Aangezien

$$\phi'(\alpha_1, \alpha_2, \beta_1, \beta_2) = \begin{pmatrix} \frac{-2\alpha_1}{\beta_2 - \beta_1^2} & \frac{1}{\beta_2 - \beta_1^2} & \frac{2\beta_1(\alpha_2 - \alpha_1^2)}{(\beta_2 - \beta_1^2)^2} & \frac{-(\alpha_2 - \alpha_1^2)}{(\beta_2 - \beta_1^2)^2} \end{pmatrix}$$

volgt dat

$$\begin{aligned} \sqrt{n} \left(\frac{S_{n,X}^2}{S_{n,Y}^2} - \frac{\sigma_X^2}{\sigma_Y^2} \right) &= \sqrt{n} \left(\phi(\bar{X}, \bar{X}^2, \bar{Y}, \bar{Y}^2) - \phi(\alpha_1, \alpha_2, \beta_1, \beta_2) \right) \\ &\rightsquigarrow \mathcal{N}(0, \Sigma'), \end{aligned}$$

met

$$\Sigma' = \phi'(\alpha_1, \alpha_2, \beta_1, \beta_2) \Sigma \phi'(\alpha_1, \alpha_2, \beta_1, \beta_2)^T = \frac{1}{\sigma_Y^4} \left(\alpha_4 - \alpha_2^2 + \frac{\sigma_X^4}{\sigma_Y^4} (\beta_4 - \beta_2^2) \right)$$

aangezien $\alpha_1 = \beta_1 = 0$.

b.

$$\hat{\sigma}^2 = \frac{1}{S_{n,Y}^4} \left(\overline{X^4} - (\overline{X^2})^2 + \frac{S_{n,X}^4}{S_{n,Y}^4} \left(\overline{Y^4} - (\overline{Y^2})^2 \right) \right).$$

c. Een betrouwbaarheidsinterval met asymptotische onbetrouwbaarheid 2α is

$$\left(\frac{S_{n,X}^2}{S_{n,Y}^2} - Z_\alpha \frac{\sqrt{\hat{\sigma}^2}}{\sqrt{n}}, \frac{S_{n,X}^2}{S_{n,Y}^2} + Z_\alpha \frac{\sqrt{\hat{\sigma}^2}}{\sqrt{n}} \right),$$

waar Z_α het bovenste α -kwantiel is van de standaardnormale verdeling.

3. **a.** In de notatie uit het dictaat, paragraaf 2.5: $\hat{a}_1 = \frac{15}{50} = \frac{3}{10}$, $\hat{a}_2 = \frac{35}{50} = \frac{7}{10}$, $\hat{b}_1 = \frac{24}{50} = \frac{12}{25}$, $\hat{b}_2 = \frac{26}{50} = \frac{13}{25}$. Een toetsingsgrootheid is

$$D_n^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(X_{n,ij} - n\hat{a}_i\hat{b}_j)^2}{n\hat{a}_i\hat{b}_j}.$$

In dit geval dus:

$$\begin{aligned} D_{50}^2 &= \frac{(7 - 50 * \frac{3}{10} * \frac{12}{25})^2}{50 * \frac{3}{10} * \frac{12}{25}} + \frac{(8 - 50 * \frac{3}{10} * \frac{13}{25})^2}{50 * \frac{3}{10} * \frac{13}{25}} + \frac{(17 - 50 * \frac{7}{10} * \frac{12}{25})^2}{50 * \frac{7}{10} * \frac{12}{25}} + \frac{(18 - 50 * \frac{7}{10} * \frac{13}{25})^2}{50 * \frac{7}{10} * \frac{13}{25}} \\ &= \frac{(7 - \frac{3*12}{5})^2}{\frac{3*12}{5}} + \frac{(8 - \frac{3*13}{5})^2}{\frac{3*13}{5}} + \frac{(17 - \frac{7*12}{5})^2}{\frac{7*12}{5}} + \frac{(18 - \frac{7*13}{5})^2}{\frac{7*13}{5}}. \end{aligned}$$

b. De limietverdeling van D_n voor $n \rightarrow \infty$ is chikwadraat met $(2-1)(2-1) = 1$ vrijheidsgraad. Een toets met asymptotische onbetrouwbaarheid α is dus: verwerp de nulhypothese als $D_n > \chi_{1,\alpha}^2$ waar $\chi_{1,\alpha}^2$ het bovenste α -kwantiel is van de chikwadraatverdeling met 1 vrijheidsgraad.

4.

$$\begin{pmatrix} X_1 - \overline{X}_n \\ X_2 - \overline{X}_n \\ \vdots \\ X_n - \overline{X}_n \end{pmatrix} = \begin{pmatrix} 1 - \frac{1}{n} & -\frac{1}{n} & \cdot & \cdot & \cdot & -\frac{1}{n} \\ -\frac{1}{n} & 1 - \frac{1}{n} & -\frac{1}{n} & \cdot & \cdot & -\frac{1}{n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -\frac{1}{n} & \cdot & \cdot & \cdot & -\frac{1}{n} & 1 - \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \cdot & \cdot & \cdot & \frac{1}{n} \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}.$$

Dit is een lineaire transformatie zodat met Corollary 2.5 uit het dictaat volgt dat de linkerkant multivariaat normaal verdeeld is. Volgens som II-4 van het werkcollege zijn dan $(X_1 - \overline{X}_n \quad \cdot \quad \cdot \quad X_n - \overline{X}_n)^T$ en $\overline{X}_n$ onderling onafhankelijk dan en slechts dan als $\text{cov}(X_i - \overline{X}_n, \overline{X}_n) = 0$ voor $i = 1, \dots, n$. Dat is inderdaad het geval:

$$\text{cov}(X_i - \overline{X}_n, \overline{X}_n) = \text{cov}\left(\frac{n-1}{n}X_i, \frac{1}{n}X_i\right) + \sum_{j \neq i} \text{cov}\left(-\frac{1}{n}X_j, \frac{1}{n}X_j\right) = 0.$$

Omdat $\sum_{i=1}^n |X_i - \bar{X}_n|/S_n$ een functie is van alleen $(X_1 - \bar{X}_n \quad \dots \quad X_n - \bar{X}_n)^T$ zijn ook $\bar{X}_n$ en $\sum_{i=1}^n |X_i - \bar{X}_n|/S_n$ onderling onafhankelijk.

5. **a.** Voer in:

$$M_n(\theta) = \frac{1}{n} \sum_{i=1}^n (Y_i - f_\theta(X_i))^2.$$

Stel θ_0 is de ware parameter, dus $Y = f_{\theta_0}(X) + e$ met e onafhankelijk van X . Als aan de voorwaarden voor de wet van de grote aantallen voldaan is gaat $M_n(\theta)$ in kans naar $M(\theta) = E_{\theta_0}(M_n(\theta))$. Er geldt:

$$\begin{aligned} M(\theta) &= E_{\theta_0}(f_{\theta_0}(X) + e - f_\theta(X))^2 \\ &= E_{\theta_0}(f_{\theta_0}(X) - f_\theta(X))^2 + Ee^2 + 2E_{\theta_0}(f_{\theta_0}(X) - f_\theta(X))Ee \end{aligned}$$

omdat e en X onderling onafhankelijk zijn. Als $Ee = 0$ dan

$$M(\theta) = E_{\theta_0}(f_{\theta_0}(X) - f_\theta(X))^2 + Ee^2.$$

Dus als $Ee = 0$ dan is dit minimaal voor $\theta = \theta_0$. Dit is de belangrijkste voorwaarde. Het minimum is uniek als voor $\theta \neq \theta_0$, $f_\theta(X)$ en $f_{\theta_0}(X)$ niet bijna zeker gelijk zijn. Dit geeft de tweede voorwaarde.

Eventueel nog checken wanneer aan de voorwaarden voor de wet van de grote aantallen is voldaan zodat $M_n(\theta) \xrightarrow{P} M(\theta)$. Uit de formule voor $M(\theta)$ is te zien dat het voldoende is als $Ee^2 < \infty$ en $E_{\theta_0}(f_{\theta_0}(X) - f_\theta(X))^2 < \infty$ voor alle θ .

Deze voorwaarden zijn niet voldoende (zie dictaat), maar als verder aan redelijke voorwaarden is voldaan zal het minimum $\hat{\theta}_n$ van $M_n(\theta)$ in kans convergeren naar het minimum θ_0 van $M(\theta)$.

b. Onder voorwaarden die hier niet gecheckt worden kan de formule uit het dictaat bij Stelling 4.8 gebruikt worden om de limietverdeling van $\sqrt{n}(\hat{\theta}_n - \theta_0)$ te bepalen. $m_\theta(x, y) = (y - f_\theta(x))^2$. Dus $\dot{m}_\theta(x, y) = -2(y - f_\theta(x))\dot{f}_\theta(x)$ en omdat e en X onafhankelijk zijn

$$E_{\theta_0}(\dot{m}_{\theta_0}(X, Y))^2 = 4E_{\theta_0}((Y - f_{\theta_0}(X))^2 \dot{f}_{\theta_0}^2(X)) = 4Ee^2 E_{\theta_0}(\dot{f}_{\theta_0}^2(X)).$$

Verder:

$$Pm_\theta(X, Y) = M(\theta) = E_{\theta_0}(f_{\theta_0} - f_\theta)^2(X) + Ee^2,$$

zodat als verwachting nemen en afgeleide naar θ verwisseld kunnen worden

$$\frac{d}{d\theta} Pm_\theta = -2E_{\theta_0}((f_{\theta_0} - f_\theta)(X) \dot{f}_\theta(X))$$

en

$$\frac{d^2}{d\theta^2} Pm_\theta = 2E_{\theta_0}(\dot{f}_\theta^2(X)) - 2E_{\theta_0}((f_{\theta_0} - f_\theta)(X) \ddot{f}_\theta(X)).$$

Dus $V = 2E_{\theta_0}(\dot{f}_{\theta_0}^2(X))$. De formule bij Stelling 4.8 uit het dictaat invullen levert dat onder voorwaarden

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow \mathcal{N}\left(0, \frac{Ee^2}{E_{\theta_0}\dot{f}_{\theta_0}^2(X)}\right).$$

5b. kan ook door eerst over te gaan op $\psi_\theta = \dot{m}_\theta$ en de bijbehorende formule uit het dictaat te gebruiken.

6. **a.** Merk eerst op dat

$$f_{Y,X}(y, x) = f_X(x) \Psi(\theta x)^y (1 - \Psi(\theta x))^{1-y}$$

voor $y = 0, 1$. Dus $\log(f_{Y,X}(y, x)) = \log f_X(x) + y \log(\Psi(\theta x)) + (1 - y) \log(1 - \Psi(\theta x))$
en

$$\dot{l}_\theta = \frac{y}{\Psi(\theta x)} \Psi'(\theta x) x - \frac{1-y}{1-\Psi(\theta x)} \Psi'(\theta x) x = \frac{y - \Psi(\theta x)}{\Psi(\theta x)(1 - \Psi(\theta x))} \Psi'(\theta x) x.$$

Neem eerst de voorwaardelijke verwachting gegeven X en gebruik dat Y alternatief($\Psi(\theta X)$) verdeeld is gegeven X om te zien dat

$$\begin{aligned} i(\theta) &= E_\theta \dot{l}_\theta^2 = E \left(E_\theta \left((Y - \Psi(\theta X))^2 | X \right) \frac{\Psi'(\theta X)^2 X^2}{\Psi^2(1 - \Psi)^2(\theta X)} \right) \\ &= E \Psi(1 - \Psi)(\theta X) \frac{\Psi'(\theta X)^2 X^2}{\Psi^2(1 - \Psi)^2(\theta X)} \\ &= E \frac{\Psi'(\theta X)^2 X^2}{\Psi(\theta X)(1 - \Psi(\theta X))}. \end{aligned}$$

b. Een betrouwbaarheidsinterval voor θ met asymptotische onbetrouwbaarheid 2α is

$\left(\hat{\theta}_n - \frac{Z_\alpha}{\sqrt{n} \sqrt{i(\hat{\theta}_n)}}, \hat{\theta}_n + \frac{Z_\alpha}{\sqrt{n} \sqrt{i(\hat{\theta}_n)}} \right)$ met Z_α het bovenste α -kwantiel van de standaard-normale verdeling. Dus $i(\hat{\theta}_n)$ schatten.

$$\begin{aligned} \widehat{i(\hat{\theta}_n)} &= \frac{1}{n} \sum_{i=1}^n \frac{\Psi'(\hat{\theta}_n X_i)^2 X_i^2}{\Psi(\hat{\theta}_n X_i)(1 - \Psi(\hat{\theta}_n X_i))} \\ &= \frac{1}{100} 50 \frac{\Psi'(\log 2)^2 (\log 2)^2}{\Psi(\log 2)(1 - \Psi(\log 2))} \\ &= \frac{1}{2} \frac{\left(\frac{2}{9}\right)^2 (\log 2)^2}{\frac{1}{3} * \frac{2}{3}} \\ &= \frac{1}{9} (\log 2)^2, \end{aligned}$$

want $\Psi(\log 2) = \frac{1}{3}$ en $\Psi'(\log 2) = -\frac{2}{9}$. Een numeriek betrouwbaarheidsinterval met asymptotische onbetrouwbaarheid 2α is dus

$$\left(1 - \frac{1}{\frac{10}{3} \log 2} Z_\alpha, 1 + \frac{1}{\frac{10}{3} \log 2} Z_\alpha \right).$$