Technische Universiteit Eindhoven
Faculteit Wiskunde en Informatica

Uitwerking tentamen **Mathematische Statistiek (2WS05)**,
maandag 5 januari 2009, van 14.00–17.00 uur.

---

1. **a** Een kansdichtheid is

$$f(x_1,\ldots,x_n;\theta) = (2\theta)^{-n}\prod_{i=1}^n 1\{x_i \in (-\theta,\theta)\} = (2\theta)^{-n}1\{-\theta < x_{(1)} < x_{(n)} < \theta\}$$
$$= (2\theta)^{-n}1\{\max\{-x_{(1)}, x_{(n)}\} < \theta\}.$$

De drager hangt af van θ, dus de familie verdelingen vormt geen exponentiële familie.

b $\max\{-X_{(1)}, X_{(n)}\} = \max_i |X_i|$ is een afdoende statistische grootheid voor $(X_1,\ldots,X_n)$ met betrekking tot θ.

c Merk op dat de verdelingsfunctie van $T := \max\{-X_{(1)}, X_{(n)}\} = \max_i |X_i|$ gelijk is aan

$$F_T(t;\theta) = (F(t) - F(-t))^n = \theta^{-n}\, t^n, \quad t \in (0,\theta),$$

zodat $f_T(t;\theta) = n\,\theta^{-n}\,t^{n-1}$ voor $t \in (0,\theta)$. Als, voor een $\mathbb{R}$-waardige functie u,

$$\mathbb{E}_\theta u(T) = \int_0^\theta u(t)\,n\,\theta^{-n}\,t^{n-1}\,dt = 0$$

voor elke $\theta > 0$, volgt dat $\int_0^\theta u(t)\,t^{n-1}\,dt = 0$ voor elke $\theta > 0$, en dus $\int_a^b u(t)\,t^{n-1}\,dt = 0$ voor alle $0 < a < b < \infty$. Derhalve is $u(t)t^{n-1}$ bijna overal 0, en dus $u(t)$. Derhalve is T volledig.

d Merk op dat

$$\mathbb{E}_\theta T = \int_0^\theta t\,n\,\theta^{-n}\,t^{n-1}\,dt = \frac{n}{n+1}\,\theta.$$

Derhalve is $W = (n+1)\,T/n$ een zuivere schatter gebaseerd op de afdoende statistische grootheid T. Volgens de stelling van Lehmann–Scheffé is er ten hoogste één zuivere schatter gebaseerd op T. W is dus een UMVZ-schatter.

e De Fisher-informatie is

$$I(\theta) = \mathbb{E}_\theta\left(\left[\frac{\partial}{\partial\theta}\log f(X_1;\theta)\right]^2\right) = \int_{-\theta}^{\theta}\frac{1}{2\theta}\left(\frac{-1}{\theta}\right)^2 dx = \frac{1}{\theta^2}.$$

Het tweede moment van T is

$$\int_0^\theta t^2\,n\,\theta^{-n}\,t^{n-1}\,dt = \frac{n}{n+2}\,\theta^2.$$

Dus is $\operatorname{Var} W = \theta^2/(n(n+2))$. Dit is kleiner dan θ^2/n.

2. **a** Omdat $\mathbb{E}X_1 = 1/p$ is de momentenschatter gelijk aan $1/\bar{X}_n$. De aannemelijkheidsvergelijking is

$$\frac{\partial}{\partial p}\left[n \log p + \sum_{i=1}^{n}(X_i - 1)\log(1-p)\right] = 0$$

dus $\hat{p} = 1/\bar{X}_n$. Als alle $X_i = 1$ is de afgeleide strikt dalend en bestaat $\hat{p}$ niet in $(0,1)$. Anders leert analyse van het tekenschema van de afgeleide dat er een uniek optimum is. De schatters zijn gelijk en asymptotisch raak.

b Omdat $p = \mathbb{P}(X_1 = 1)$ is $1\{X_1 = 1\}$ een zuivere schatter. Een afdoende grootheid T voor $(X_1, \ldots, X_n)$ met betrekking tot p is de som $\sum_{i=1}^n X_i$; deze is negatief binomiaal verdeeld. Om een UMVZ-schatter te vinden passen we het Rao–Blackwellprocédé toe: voor $t \in \{n, n+1, \ldots\}$ is

$$\mathbb{P}(X_1 = 1 \mid \sum_{i=1}^{n} X_i = t) = \frac{\mathbb{P}(X_1 = 1; \sum_{i=2}^n X_i = t-1)}{\mathbb{P}(\sum_{i=1}^n X_i = t)} = \frac{p\binom{t-2}{n-2}p^{n-1}(1-p)^{t-n}}{\binom{t-1}{n-1}p^n(1-p)^{t-n}} = \frac{n-1}{t-1}.$$

Een zuivere schatter gebaseerd op de afdoende grootheid is dus $(n-1)/(T-1) = (n-1)/(n\bar{X}_n - 1)$. Aangezien we met een 1-parameter exponentiële familie te maken hebben met $Q(p) = \log(1-p)$ en $\{Q(p) : p \in (0,1)\} = \mathbb{R}^-$ (bevat interval), is er ten hoogste één zuivere schatter van p gebaseerd op T en deze is UMVZ.

3. **a** Met de substituties $z = \lambda x$ en $y = z^2$ zien we dat voor $x \in \mathbb{R}^+$

$$F(x;\lambda) = \int_0^{\lambda x} 2\,z\,e^{-z^2}\,dz = \int_0^{(\lambda x)^2} e^{-w}\,dw = 1 - \exp\left[-(\lambda x)^2\right].$$

Ter bepaling van een betrouwbaarheidsinterval merken we op dat

$$-2\sum_{i=1}^{n}\log\left(1 - F(X_i;\lambda)\right) = 2\lambda^2\sum_{i=1}^{n}X_i^2 \sim \chi^2_{2n}.$$

Een 90% b.i. voor λ is

$$\left(\sqrt{\frac{\chi^2_{2n;0{,}05}}{2\sum_{i=1}^n X_i^2}}, \sqrt{\frac{\chi^2_{2n;0{,}95}}{2\sum_{i=1}^n X_i^2}}\right).$$

b Gebruik als toetsingsgrootheid $T = \sum_{i=1}^n X_i^2$ en verwerp de nulhypothese voor kleine waarden (immers $\mathbb{E}X_1^2 = \lambda^{-2}$ is dalend in λ). Kritiek gebied $(0, c := \chi^2_{2n;0{,}10}/50)$.

c Schrijf F_{2n} voor de verdelingsfunctie van een χ^2-verdeling met $2n$ vrijheidsgraden. Dan is

$$\beta(\lambda) = \mathbb{P}_\lambda(T < c) = P_\lambda(2\lambda^2 T < 2\lambda^2 c) = F_{2n}(\lambda^2\chi^2_{2n;0{,}10}/25).$$

4. **a** Gebruik de t-toets voor gepaarde waarnemingen. Veronderstel aselecte steekproeven uit normale verdelingen. Dus neem $Z_i = X_i - Y_i$ waar Y_i de lengte bij concurrerend en X_i bij eigen product is. De nulhypothese is dat verwachtingen van Z_i ten hoogste nul zijn. Hier is $\bar{Z} \approx 0{,}2$, $S_Z^2 \approx 1{,}55$ zodat de toetsingsgrootheid $T = \sqrt{10}\bar{Z}/S_Z \approx 0{,}51$. Er wordt verworpen als $T \geq t_{9;0{,}95} \approx 1{,}833$. Er is onvoldoende bewijs dat het eigen product beter is.

b Gebruik de twee steekproeven t-toets. Veronderstel aselecte steekproeven uit normale verdelingen met gelijke variantie. De gepoolde steekproefvariantie is $S_{XY}^2 \approx 6,33$. Nu is $T = (\bar{X} - \bar{Y})/(S_{XY}\sqrt{2/10}) \approx 0,18$. We verwerpen als $T \geq t_{18;0,95} \approx 1,734$. Dezelfde conclusie als eerder.

c Kijk bijvoorbeeld naar histogram, Q-Q plot, boxplot, Shapiro-Wilkstoets, ...