Uitwerkingen Tentamen Mathematische Statistiek (2S990) en Mathematische Statistiek 1 (2S210) op vrijdag 25 januari 2002, 9.00-12.00 uur.

---

1. (a) Als Y exponentieel verdeeld is met verwachting 1, dan geldt $P(Y+\theta \le x) = P(Y \le x-\theta) = 1-e^{x-\theta}$ voor $x \ge \theta$, dus via differentiëren vinden we dat $f_{X+\theta} = e^{-(x-\theta)}$ voor $x \ge \theta$.
   Verder geldt dat $P(U \ge t) = \prod_{i=1}^n P(X_i \ge t) = \prod_{i=1}^n P(Y+\theta \ge t) = \prod_{i=1}^n P(Y \ge t-\theta) = e^{-n(t-\theta)}$. Analoog aan het eerste gedeelte volgt nu dat U verdeeld is als $Z+\theta$, waarbij Z exponentieel verdeeld is met verwachting $1/n$.

   (b) We zoeken een interval van de vorm $(U-c_1, U+c_2)$ (vanwege de scheve verdeling van U kunnen we geen symmetrisch interval verwachten):
   $$P(\theta < U+c_2) = P(U-\theta > -c_2) = \frac{\alpha}{2} \Leftrightarrow e^{nc_2} = \frac{\alpha}{2} \Leftrightarrow c_2 = \frac{1}{n}\log\left(\frac{\alpha}{2}\right).$$
   Analoog geldt
   $$P(\theta > U-c_1) = P(U-\theta < c_1) = \frac{\alpha}{2} \Leftrightarrow 1-e^{-nc_1} = \frac{\alpha}{2} \Leftrightarrow -c_1 = \frac{1}{n}\log\left(1-\frac{\alpha}{2}\right).$$
   Hieruit volgt dat een $100(1-\alpha)\%$-betrouwbaarheidsinterval gegeven wordt door
   $$\left(U + \frac{1}{n}\log\left(1-\frac{\alpha}{2}\right), U + \frac{1}{n}\log\left(\frac{\alpha}{2}\right)\right).$$

   (c) Merk op dat de dichtheid alleen positief is als $x \ge \theta$. Dus
   $$\begin{aligned} L(\theta \mid x_1,\ldots,x_n) &= \prod_{i=1}^n e^{-(x_i-\theta)}\, 1_{\{x_i \ge \theta\}} \\ &= e^{n\theta - \sum_{i=1}^n x_i} \prod_{i=1}^n 1_{\{x_i \ge \theta\}} \\ &= e^{n\theta - \sum_{i=1}^n x_i}\, 1_{\{\min(x_1,\ldots,x_n) \ge \theta\}}. \end{aligned}$$

   (d) Volgens het factorisatiecriterium is het voldoende te laten zien dat $L(\theta \mid x_1,\ldots,x_n)$ herschreven kan worden in de vorm $g(U,\theta)h(x_1,\ldots,x_n)$. Dat kunnen we hier bereiken door
   $$g(U,\theta) = 1_{U \ge \theta}\, e^{n\theta} \text{ en } h(x_1,\ldots,x_n) = e^{-\sum_{i=1}^n x_i}$$
   te kiezen.

   (e) Merk op dat de likelihood een stijgende, positieve functie van θ is op het gebied $\theta \le \min(x_1,\ldots,x_n)$. De likelihood onder $H_0 \cup H_1$ (m.a.w. zonder beperking op θ neemt dus zijn maximum aan in $\theta = \min(x_1,\ldots,x_n)$. Vervolgens maximaliseren we de likelihood onder $H_0 : \theta \le \theta_0$. In dit geval neemt de likelihood een maximum aan in $\theta = \min(\theta_0, \min(x_1,\ldots,x_n))$. Hieruit volgt dat $-2\log\Lambda = 2n\,(U - \min(U,\theta_0))$.

Aangezien $2\log\Lambda$ bij benadering χ_1^2-verdeeld is, volgt dat het kritieke gebied gegeven wordt door

$$-2\log\Lambda > \chi^2_{1;\alpha} \Leftrightarrow (U - \min(U,\theta_0)) > \frac{\chi^2_{1;\alpha}}{2n} \Leftrightarrow U - \theta_0 > \frac{\chi^2_{1;\alpha}}{2n} \Leftrightarrow U > \theta_0 + \frac{\chi^2_{1;\alpha}}{2n}\ .$$

2. (a) Het 95%-betrouwbaarheidsinterval voor μ is $\overline{x} \pm z_{0,025}\frac{\sigma}{\sqrt{n}}$, dus de halve breedte (de marge) is $z_{0,025}\frac{\sigma}{\sqrt{n}}$. Hieruit volgt dat

$$z_{0,025}\frac{\sigma}{\sqrt{n}} < 1,2 \Leftrightarrow \sqrt{n} > \frac{z_{0,025}\sigma}{1,2} \Leftrightarrow n > \frac{1,96^2\, 8,6}{1,44} = 22,94 \Rightarrow n \geq 23.$$

   (b) $\overline{X} = 88,775$, $\sigma = \sqrt{8,6} = 2,93$, dus $T = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{88,775-90}{2,93/\sqrt{4}} = -0,835$. Kritiek gebied: $T < -z_\alpha = -z_{0,05} = -1,645$. Conclusie: H_0 mag niet verworpen worden.

   (c) $S^2 = 11,4825$. $T = \frac{(n-1)S^2}{\sigma_0^2} = \frac{3*11,4825}{8,6} = 4,0$. Kritiek gebied: $T > \chi^2_{3,\alpha/2} = 9,35$ en $T < \chi^2_{3,1-\alpha/2} = 0,216$. Conclusie: $H_0 : \sigma^2 = 8,6$ mag niet verworpen worden ten gunste van $H_1 : \sigma^2 \neq 8,6$.

3. (a) We toetsen $H_0 : p = 0,10$ tegen $H_1 : p > 0,10$ (wat we willen aantonen zetten we in de alternatieve hypothese). $T = \frac{X-np_0}{\sqrt{np_0(1-p_0)}} = \frac{31-200*0,1}{\sqrt{200*0,1*0,9}} = 2,59$. Kritiek gebied: $T > z_{0,05} = 1,645$. Conclusie: we verwerpen H_0, dus de oven functioneert niet goed.

   (b) $\widehat{p} \pm z_{0,025}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 0,155 \pm 1,96\sqrt{0,0256} = 0,155 \pm 0,050 = (0,105; 0,205)$.

   (c) $P(X \geq 2 \mid X \sim \text{Bin}(10;0,1)) = 1 - P(X \leq 1 \mid X \sim \text{Bin}(10;0,1)) = 0,2639$.

   (d) $P(X \leq 1 \mid X \sim \text{Bin}(10;0,2)) = 0,3758$.

4. (a) De verwachte frequenties zijn als $\theta = \frac{1}{2}$ gelijk aan $\frac{1}{9}$, $\frac{4}{9}$ en $\frac{4}{9}$. De toetsingsgrootheid is dus gelijk aan $X^2 = \sum_{i=1}^3 \frac{(E_i-O_i)^2}{E_i} = \frac{(50/9-8)^2}{50/9} + \frac{(200/9-17)^2}{200/9} + \frac{(200/9-25)^2}{200/9} = 2,65$. Kritiek gebied: $X^2 > \chi^2_{3-1;0,05} = 5,99$. Conclusie: $H_0 : \theta = \frac{1}{2}$ niet verwerpen ten gunste van $H_1 : \theta \neq \frac{1}{2}$.

   (b) $L(\theta) = 130172\frac{\theta^{33}}{(\theta+1)^{100}}$, dus $\frac{\partial \log L}{\partial\theta} = \frac{33}{\theta} - \frac{100}{\theta+1}$. Dit gelijk aan 0 stellen levert $\widehat{\theta} = \frac{33}{67}$.

   (c) De verwachte frequenties zijn als $\theta = \frac{33}{67}$ gelijk aan $\frac{1089}{10000}$, $\frac{2211}{5000}$ en $\frac{4489}{10000}$. De toetsingsgrootheid is dus gelijk aan $X^2 = \sum_{i=1}^3 \frac{(E_i-O_i)^2}{E_i} = \frac{\left(50*\frac{1089}{10000}-8\right)^2}{50*\frac{1089}{10000}} + \frac{\left(50*\frac{2211}{5000}/9-17\right)^2}{50*\frac{2211}{5000}} + \frac{\left(50*\frac{4489}{10000}-25\right)^2}{50*\frac{4489}{10000}} = 2,67$. Kritiek gebied: $X^2 > \chi^2_{3-1-1;0,05} = 3,84$. Conclusie: $H_0 : P(GG) = \left(\frac{\theta}{\theta+1}\right)^2, P(GR) = \frac{2\theta}{(\theta+1)^2}$ en $P(RR) = \frac{1}{(\theta+1))^2}$ niet verwerpen ten gunste van $H_1 : P(GG) \neq \left(\frac{\theta}{\theta+1}\right)^2$ of $P(GR) \neq \frac{2\theta}{(\theta+1)^2}$ of $P(RR) \neq \frac{1}{(\theta+1)^2}$.