# Uitwerkingen Tentamen Mathematische Statistiek (2S990) vrijdag 25 november 2005, 9.00-12.00 uur.

**De notatie voor quantielen van de $\chi^2$-verdeling volgt de conventie $P(\chi^2_n < \chi^2_{n;\alpha}) = \alpha$.**

---

1. (a) Het stochastisch model is $X_1, \ldots, X_{20}$; onderling onafhankelijke stochasten die ieder Bernoulli(1,$p$) verdeeld zijn (d.w.z. binomiaal met parameters 1 en $p$). De toetsingsgrootheid is $T = \sum_{i=1}^{20} X_i$. De hypothesen zijn $H_0 : p = 0,4$ en $H_a : p > 0,4$ (eenzijdig alternatief omdat training niet tot verslechtering kan leiden). Het kritieke gebied is $\left\{\sum_{i=1}^{20} X_i \geq 13\right\}$.
   Alternatief model: $X \sim \text{Bin}(20, p)$ met kritiek gebied $\{X \geq 13\}$.

   (b) De kans op een type I fout is

   $$P\left(T \geq 13 \mid T \sim \text{Bin}(20; 0,4)\right) = 1 - P\left(T \leq 12 \mid T \sim \text{Bin}(20; 0,4)\right) = 1 - 0,9790 = 0,0210.$$

   (c) Het onderscheidingsvermogen bij $p = 0,6$ is gelijk aan

   $$P\left(T \geq 13 \mid T \sim \text{Bin}(20; 0,6)\right) = P\left(V \leq 7 \mid V \sim \text{Bin}(20; 0,4)\right) = 0,4159.$$

2. We gebruiken de parametrisering van de exponentiële verdeling waarbij $\mathrm{E}\, X_i = \theta$. Merk op dat $\sum_{i=1}^{n} X_i$ een Gammaverdeling heeft met parameters $n$ en $\theta$. Aangezien deze verdeling in het algemeen niet uitgebreid getabelleerd is, gebruiken we liever dat $2n\overline{X}/\theta$ een $\xi^2_{2n}$-verdeling bezit. Hieruit volgt dat $P\left(\chi^2_{2n;\alpha/2} \leq 2n\overline{X}/\theta \leq \chi^2_{2n;1-\alpha/2}\right) = 1 - \alpha$. Met $\alpha = 0,05$, concluderen we dus dat $\left(\frac{2n\overline{X}}{\chi^2_{2n;0,975}}, \frac{2n\overline{X}}{\chi^2_{2n;0,025}}\right)$ een 95%-betrouwbaarheidsinterval is voor $\theta$.

3. (a) Merk eerst op dat de verdelingsfunctie van deze verdeling gegeven wordt door $F(u) = \int_0^u 3x^2\theta^{-3}\, dx = (u/\theta)^3$. Dus $P\left(\max_{1\leq i\leq n} X_i/\theta < x\right) = \prod_{i=1}^{n} P\left(X_i < x\theta\right) = x^{3n}$. Hieruit volgt dat $P\left(c < \max_{1\leq i\leq n} X_i/\theta < 1\right) = 1 - P\left(\max_{1\leq i\leq n} X_i/\theta < c\right) = 1 - c^{3n}$.

   (b) Als $c = \sqrt[12]{0,05}$, dan volgt dat $1 - c^{3\times 4} = 0,95$. Uit a) volgt nu dat $\left(2,3; 2,3/\sqrt[12]{0,05}\right) = (2,3; 2,95)$ een 95%-betrouwbaarheidsinterval is voor $\theta$.

4. (a) Door de verdelingsfunctie te differentiëren naar $x$ vinden we dat $2x\exp(-x^2/\theta)/\theta$ de dichtheid van deze verdeling is voor $x \geq 0$. De loglikelihood $\log L(x_1, \ldots, x_n; \theta)$ is dus gelijk aan $n\log 2 + \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} x_i^2/\theta - n\log\theta$. Differentiëren naar $\theta$ en de afgeleide gelijk aan 0 stellen levert dat $\sum_{i=1}^{n} x_i^2/\theta - n/\theta = 0$. We concluderen dat $\widehat{\theta}_{ML} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

   (b) Door de verdelingsfunctie te differentiëren naar $x$ vinden we dat $2x\exp(-x^2/\theta)/\theta$ de dichtheid van deze verdeling is voor $x \geq 0$. De likelihood $L(x_1, \ldots, x_n; \theta) = 2^n\theta^{-n} \prod i = 1^n\, x_i \exp\left(\sum_{i=1}^{n} x_i^2/\theta\right)$ herschrijven we als

   $$\left[\exp\left(\sum_{i=1}^{n} x_i^2/\theta\right)\theta^{-n}\right]\left(2^n\prod_{i=1}^{n} x_i\right) = g\left(\sum_{i=1}^{n} x_i^2, \theta\right) h(x_1, \ldots, x_n).$$

   Uit de factorisatiestelling volgt nu dat $\sum_{i=1}^{n} X_i^2$ een voldoende steekproefgrootheid is voor $\theta$.

(c) De Cramér-Rao ondergrens is gelijk aan

$$-\left(n\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\log\left\{2x\theta^{-1}\exp\left(-X^2/\theta\right)\right\}\right]\right)^{-1} = \left(n\mathrm{E}\left[-\frac{1}{\theta^2}+2\frac{X^2}{\theta^3}\right]\right)^{-1}.$$

Gebruik makend van de substitutie $y = x^2/\theta$ zien we dat

$$\mathrm{E}\left(X^2\right) = \int_0^\infty x^2\left(2x\theta^{-1}\exp(-x^2/\theta)\right)dx = \theta\int_0^\infty ye^{-y}dy = \theta.$$

De laatste integraal kan door partiële integratie uitgerekend worden of door op te merken dat dit de verwachting is van een exponentiële verdeling met parameter 1. We concluderen dat de Cramér-Rao ondergrens voor $\theta$ gelijk is aan $\frac{1}{n}\left(-1/\theta^2+2/\theta^2\right)^{-1} = \theta^2/n$.

(d) Gebruik makend van de substitutie $y = x^2/\theta$ zien we dat

$$\mathrm{E}\left(X^4\right) = \int_0^\infty x^4\left(2x\theta^{-1}\exp(-x^2/\theta)\right)dx = \theta^2\int_0^\infty y^2e^{-y}dy = 2\theta^2.$$

De laatste integraal kan door partiële integratie uitgerekend worden of door op te merken dat het tweede moment is van een exponentiële verdeling met parameter 1. Hieruit volgt dat $\mathrm{Var}\left(X_i^2\right) = \mathrm{E}\left(X_i^4\right) - \left(\mathrm{E}\left(X_i^2\right)\right)^2 = 2\theta^2-\theta^2 = \theta^2$. De Centrale Limietstelling mag dus toegepast worden en levert dat

$$\frac{\sum_{i=1}^n X_i^2 - \sum_{i=1}^n \mathrm{E}\left(X_i^2\right)}{\sqrt{\sum_{i=1}^n \mathrm{Var}\left(X_i^2\right)}} = \frac{n\widehat{\theta}_{ML}-n\theta}{\sqrt{n\theta^2}} = \frac{\sqrt{n}\left(\widehat{\theta}_{ML}-\theta\right)}{\theta} \xrightarrow{d} N(0,1)$$

als $n\to\infty$. We concluderen dus dat $\sqrt{n}\left(\widehat{\theta}_{ML}-\theta\right)$ in verdeling naar een normale verdeling met verwachting 0 en variantie $\theta^2$ convergeert. Delen door $\theta$ levert op dat $\sqrt{n}\left(\widehat{\theta}_{ML}/\theta-1\right)$ asymptotisch standaard normaal verdeeld is. Hieruit volgt dat $P\left(-z_{0,05} < \sqrt{n}\left(\widehat{\theta}_{ML}/\theta-1\right) < z_{0,05}\right) = 0,90$ en dus is $\left(\frac{\widehat{\theta}_{ML}}{1-z_{0,05}/\sqrt{n}}, \frac{\widehat{\theta}_{ML}}{1+z_{0,05}/\sqrt{n}}\right)$ een asymptotisch 90%-betrouwbaarheidsinterval voor $\theta$.

5. De verdelingsfunctie $F$ van $X_i$ in het punt $y$ is gelijk aan $\int_0^y \theta(1-x)^{\theta-1}\,dx = 1-(1-y)^\theta$. We gebruiken het lemma van Neyman-Pearson om eerst een UMP toets af te leiden voor de enkelvoudige alternatieve hypothese $H_a : \theta = \theta_0$ met $\theta_0 > 1$. Het likelihoodquotiënt is gelijk aan $\theta_0^n\prod_{i=1}^n(1-x_i)^{\theta_0-1}$, aangezien we onder $H_0$ een uniforme verdeling hebben. Omdat $\theta_0 > 1$, kunnen we een equivalente UMP toets baseren op $\prod_{i=1}^n(1-X_i)$ met eveneens een rechtseenzijdig kritiek gebied. Om van het product een som van onafhankelijke stochasten te maken en zodoende de verdeling beter te kunnen bepalen, gaan we over op nog een andere equivalente toets, nl. die gebaseerd op $-\sum_{i=1}^n\log(1-X_i)$. Merk op dat $P_{H_0}(-\log(1-X_i) < x) = P_{H_0}\left(1-X_i < e^{-x}\right) = P\left(X_i < e^{-x}\right) = e^{-x}$ (we gebruikten dat $1-X_i$ ook uniform verdeeld is op $[0,1]$). M.a.w., $-\log(1-X_i)$ heeft een exponentiële verdeling met verwachting 1 en dus heeft $T = -2\sum_{i=1}^n\log(1-X_i)$ een $\chi^2_{2n}$-verdeling. Het kritieke gebied is hier linkseenzijdig: $\{T < \chi^2_{2n;\alpha}\}$. Aangezien deze UMP toets niet afhangt van $\theta_0$ als $\theta_0 > 1$, volgt dat deze toets ook een UMP toets is voor alle $H_a : \theta > 1$.

**De notatie voor quantielen van de $\chi^2$-verdeling volgt de conventie $P(\chi^2_n < \chi^2_{n;\alpha}) = \alpha$.**