

Uitwerkingen Tentamen Mathematische Statistiek (2S990/2WS05) vrijdag 23 maart 2007, 14.00-17.00 uur.

De notatie voor quantielen van de χ^2 -verdeling volgt de conventie $P(\chi_n^2 < \chi_{n,\alpha}^2) = \alpha$.

1. (a) Het 95%-betrouwbaarheidsinterval voor het verschil in gemiddelde wordt gegeven door

$$\bar{x}_1 - \bar{x}_2 \pm z_{0,025} \sqrt{\frac{1,5}{15} + \frac{1,2}{20}} = -2,9 \pm 1,96 \cdot 0,4 = -2,9 \pm 0,784 = (-3,684, -2,116).$$

- (b) Aangezien de fabrikant aan wil tonen dat de tweede soort een hoger octaangehalte heeft, moet deze uitspraak als alternatieve hypothese genomen worden. We nemen dus $H_0 : \mu_1 = \mu_2$ en $H_1 : \mu_1 > \mu_2$. De toetsingsgrootheid is $T = (\bar{x}_1 - \bar{x}_2) / \sqrt{\frac{1,5}{15} + \frac{1,2}{20}}$ met als kritiek gebied $T < -z_{0,05} = -1,645$. Invullen van de gegevens leidt tot $T = -2,9/0,4 = -7,25$, dus we moeten H_0 verwerpen. Conclusie: de tweede soort heeft een hoger octaangehalte.

2. (a) Voor de likelihood geldt

$$f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n \frac{1}{\theta} e^{-\frac{1}{\theta} x_i} = \frac{1}{\theta^n} e^{-\frac{1}{\theta} \sum_{i=1}^n x_i} = \frac{1}{\theta^n} e^{-\frac{n}{\theta} \bar{x}} = g(\bar{x}, \theta) h(x_1, \dots, x_n).$$

Uit het factorisatiecriterium volgt nu dat \bar{X} een voldoende steekproefgrootheid is.

- (b) De loglikelihood is gelijk aan $-n \log \theta - \frac{n\bar{x}}{\theta}$, dus de afgeleide van de loglikelihood is gelijk aan $-n/\theta - n\bar{x} \cdot \theta^{-2}$. Afgeleide gelijk aan nul stellen suggereert dat $\hat{\theta} = \bar{X}$. Om aan te tonen dat dit inderdaad een maximum is, is het voldoende om op te merken dat $n/\theta^2 - n\bar{x}^2/\theta^3$, de tweede afgeleide van de loglikelihood, geëvalueerd in $\theta = \bar{x}$ gelijk is aan $-n/\bar{x}^2 < 0$ of om op te merken dat de eerste afgeleide positief is voor $\theta < \bar{x}$ en negatief voor $\theta > \bar{x}$.
- (c) Aangezien θ de verwachting is van X_i ($1 \leq i \leq n$), geldt dat $E(\bar{X}) = \theta$. Uit de voldoendeheid van \bar{X} en de volledigheid van de klasse van exponentiële verdelingen volgt met het lemma van Lehmann-Scheffé dat \bar{X} een zuivere schatter is voor θ met minimale variantie. Alternatief: $\text{Var}(\bar{X}) = \theta^2/n$, wat gelijk is aan de ondergrens van Cramér-Rao.
- (d) Uit het invariantieprincipe (lemma van Zehna) volgt dat $1/\bar{X}$ een Maximum Likelihoodschatter is voor $1/\theta$.
- (e) Merk eerst op dat $n\bar{X}$ een Gammaverdeling heeft met parameters n en θ . Gebruik maken van de formule voor de dichtheid van de Gammaverdeling en de definitie van de Gammafunctie (zie Statistisch Compendium, inleiding hoofdstuk 3) volgt dat

$$\begin{aligned} E\left(\frac{n-1}{n\bar{X}}\right) &= (n-1) \int_0^\infty \frac{1}{x} f_{n,\theta}(x) dx = (n-1) \int_0^\infty \frac{1}{x} \theta^{-n} \frac{x^{n-1} e^{-x/\theta}}{\Gamma(n)} dx \\ &= (n-1) \frac{\Gamma(n-1)}{\Gamma(n)} \frac{1}{\theta} \int_0^\infty f_{n-1,\theta}(x) dx = \frac{n-1}{n-1} \frac{1}{\theta} = \frac{1}{\theta}. \end{aligned}$$

Aangezien \bar{X} ook een voldoende schatter is voor $1/\theta$, volgt net als bij (c) uit het lemma van Lehmann-Scheffé dat $(n-1)/(n\bar{X})$ een zuivere schatter met minimale variantie (UMVU schatter) is voor $1/\theta$.

(f)

$$\text{Var } \bar{X} = \frac{1}{n^2} (n\theta^2) = \frac{\theta^2}{n}.$$

De berekening van $\text{Var } (n-1)/(n\bar{X})$ volgt hetzelfde patroon als bij e). We berekenen eerst het tweede moment (stilzwijgend nemen we aan dat $n \geq 3$):

$$\begin{aligned} \text{E} \left(\frac{n-1}{n\bar{X}} \right)^2 &= (n-1)^2 \int_0^\infty \frac{1}{x^2} \theta^{-n} \frac{x^{n-1} e^{-x/\theta}}{\Gamma(n)} dx \\ &= (n-1)^2 \frac{\Gamma(n-3)}{\Gamma(n)} \frac{1}{\theta^2} \int_0^\infty f_{n-2,\theta}(x) dx = \frac{n-1}{n-2} \frac{1}{\theta^2}. \end{aligned}$$

Hieruit volgt dat

$$\text{Var} \left(\frac{n-1}{n\bar{X}} \right) = \text{E} \left(\frac{n-1}{n\bar{X}} \right)^2 - \left(\text{E} \frac{n-1}{n\bar{X}} \right)^2 = \frac{n-1}{n-2} \frac{1}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{(n-2)\theta^2}.$$

De parametrisering met $\frac{1}{\theta} e^{-x/\theta}$ i.p.v. $\theta e^{-x\theta}$ heeft dus een lichte voorkeur vanwege de iets kleinere variantie.

3. (a) Merk eerst op dat $Y = \sum_{i=1}^{10} X_i$ Poisson verdeeld is met parameter 10λ . Het significantieniveau is de kans dat onder H_0 de toetsingsgrootheid in het kritieke gebied valt, dus

$$P \left(\sum_{i=1}^{10} X_i \geq 3 \mid \lambda = 0,1 \right) = 1 - P(Y \leq 2 \mid Y \sim \text{Poi}(1)) = 1 - 0,9197 = 0,0803.$$

- (b) Het onderscheidingsvermogen is de kans dat onder de alternatieve hypothese, de toetsingsgrootheid in het kritieke gebied valt:

$$P \left(\sum_{i=1}^{10} X_i \geq 3 \mid \lambda = 0,5 \right) = 1 - P(Y \leq 2 \mid Y \sim \text{Poi}(5)) = 1 - 0,1247 = 0,8753.$$

- (c) We maken gebruik van het lemma van Neyman-Pearson.

$$\lambda(x_1, \dots, x_{10}, \lambda_0 = 0,1, \lambda_1 = 0,5) = \frac{\prod_{i=1}^{10} e^{-\lambda_0} \frac{\lambda_0^{x_i}}{x_i!}}{\prod_{i=1}^{10} e^{-\lambda_1} \frac{\lambda_1^{x_i}}{x_i!}} = e^{10(0,5-0,1)} 5^{-\sum_{i=1}^{10} x_i}.$$

De toets met als kritiek gebied $\lambda(x_1, \dots, x_{10}, \lambda_0 = 0,1, \lambda_1 = 0,5) < k$ is equivalent met een toets van de vorm $\sum_{i=1}^{10} x_i \geq k^*$. Hieruit volgt het gevraagde.

- (d) Omdat volgens d) toetsen met kritiek gebied van de vorm $\sum_{i=1}^{10} x_i \geq k^*$ optimaal zijn voor $H_0 : \lambda = 0,1$ tegen $H_1 : \lambda = 0,5$ en in de afleiding alleen maar gebruikt is dat $0,5 > 0,1$, volgt dat deze toetsen optimaal zijn voor $H_0 : \lambda = 0,1$ tegen $H_1 : \lambda > 0,1$.

4. (alleen voor 2S990)

- (a) De gevraagde ondergrens wordt gegeven door de ondergrens van Cramér en Rao. In de laatste integralen gebruiken we de substitutie $y = \tan u$, $dy = 1/\cos^2(u) du$, $1 + y^2 = 1/\cos^2(u)$ en partiële integratie:

$$\begin{aligned}
 \left(nE \left[\frac{\partial}{\partial \theta} \log \frac{1}{\pi(1 + (X - \theta)^2)} \right]^2 \right)^{-1} &= \frac{1}{n} \left(E \left[-\frac{2(X - \theta)}{(1 + (X - \theta)^2)} \right]^2 \right)^{-1} \\
 &= \frac{\pi}{4n} \left(\int_{-\infty}^{\infty} \frac{(x - \theta)^2}{(1 + (x - \theta)^2)^3} dx \right)^{-1} \\
 &= \frac{\pi}{4n} \left(\int_{-\infty}^{\infty} \frac{y^2}{(1 + y^2)^3} dy \right)^{-1} \\
 &= \frac{\pi}{4n} \left(\int_{-\infty}^{\infty} \frac{y^2}{(1 + y^2)^3} dy \right)^{-1} \\
 &= \frac{\pi}{4n} \left(\int_{-\pi/2}^{\pi/2} \sin^2(u) \cos^2(u) du \right)^{-1} \\
 &= \frac{\pi}{4n} \left(\int_{-\pi/2}^{\pi/2} \sin(u) (\sin(u) \cos^2(u)) dy \right)^{-1} \\
 &= \frac{3\pi}{4n} \left([-\sin(u) \cos^3(u)]_{-\pi/2}^{\pi/2} + \int_{-\pi/2}^{\pi/2} \cos^4(u) du \right)^{-1} \\
 &= \frac{3\pi}{4n} \left(\int_{-\pi/2}^{\pi/2} \cos^4(u) du \right)^{-1}
 \end{aligned}$$

De trigonometrische formule $\cos^2(u) = \frac{1}{2}(1 + \cos(2u))$ levert nu dat de laatste integraal overgaat in

$$\begin{aligned}
 \frac{3\pi}{4n} \left(\int_{-\pi/2}^{\pi/2} \frac{(1 + \cos(2u))^2}{4} du \right)^{-1} &= \frac{3\pi}{n} \left(\int_{-\pi/2}^{\pi/2} [1 + 2 \cos(2u) + \cos^2(2u)] du \right)^{-1} \\
 &= \frac{3}{n\pi} \left(\pi + 0 + \int_{-\pi/2}^{\pi/2} \cos^2(2u) du \right)^{-1} \\
 &= \frac{3\pi}{n} \left(\pi + \frac{1}{2} \int_{-\pi/2}^{\pi/2} (1 + \cos(4u)) du \right)^{-1} \\
 &= \frac{3\pi}{n} \left(\pi + \frac{1}{2}(\pi + 0) \right)^{-1} = \frac{2}{n}
 \end{aligned}$$

- (b) De asymptotische verdeling van een Maximum-Likelihoodschatter is een normale verdeling met als verwachting θ en variantie de ondergrens van Cramér-Rao. In dit geval is de asymptotische verdeling dus een $N(\theta, 2/n)$ -verdeling.

5. (alleen voor 2WS05)

- (a) De formule is $\sum_{i=1}^n \frac{(X_i - n/k)^2}{n/k}$.
- (b) Het is duidelijk dat de toetsingsgrootte T niet-negatief is. Verder geldt onder H_0 dat $E(X_i) = n/k$, dus onder H_0 verwachten we kleine waarden. Dit leidt tot een eenzijdig kritiek gebied van de vorm $T \geq c$. Uit het gegeven dat T (bij benadering) χ^2 verdeeld is met $n - k - 1$ vrijheidsgraden, volgt dat het kritieke gebied gegeven wordt door $T > \chi_{n-k+1;1-\alpha}^2$.

De notatie voor quantielen van de χ^2 -verdeling volgt de conventie $P(\chi_n^2 < \chi_{n;\alpha}^2) = \alpha$.