Universiteit Leiden
Wiskunde en Natuurwetenschappen

Vak: INLEIDING STATISTIEK

Naam:

Datum: 31-1-2017

Studierichting:

Docent:

Collegekaartnummer:

1a $\theta \mapsto \prod_{i=1}^{n} f_{Y_i;\theta}(y_i) = \prod_{i=1}^{n} \frac{1}{x_i\sqrt{2\pi}} e^{-\frac{1}{2}(y_i-\theta x_i)^2/x_i^2} = \frac{1}{\prod_{i=1}^{n} x_i}\left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\sum\frac{(y_i-\theta x_i)^2}{x_i^2}}$

1b maximaliseren van de likelihood is equivalent aan minimaliseren van $\theta \mapsto \sum_{i=1}^{n} \frac{(y_i-\theta x_i)^2}{x_i^2}$. De afgeleide van deze functie is $\theta \mapsto \sum_{i=1}^{n} \frac{2(y_i-\theta x_i)}{x_i^2}(-x_i)$.

~~Nu~~ Nulstellen geeft $\sum_{i=1}^{n} \frac{y_i}{x_i} = \theta \sum_{i=1}^{n} \frac{x_i^2}{x_i^2} = \theta n$, oftewel $\theta = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{x_i}$.

Dit is een minimum want de kwadraatsom is een dalparabool.

1c $E_\theta \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} E_\theta Y_i/x_i = \frac{1}{n}\sum_{i=1}^{n} x_i\theta/x_i = \theta$. Dus $\hat{\theta}$ is zuiver voor θ.

1d $\hat{\theta}-\theta$ is normaal verdeeld met $E_\theta(\hat{\theta}-\theta)=0$ en $\mathrm{var}_\theta\hat{\theta} = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{var}_\theta Y_i/x_i^2 = \frac{1}{n}$.

Dus $P_\theta(-1.96 \le \sqrt{n}(\hat{\theta}-\theta) \le 1.96) = 0.95$. Dit geeft betrouwbaarheidsinterval $[\hat{\theta}-\frac{1.96}{\sqrt{n}}, \hat{\theta}+\frac{1.96}{\sqrt{n}}]$.

2a $P_\theta(X_{(1)}-\theta > y) = P_\theta(X_1-\theta>y, \dots, X_n-\theta>y) \overset{o.o.}{=} \left(\int_{y+\theta}^{\infty} e^{-(x-\theta)}\,dx\right)^n = e^{-ny}$, $y>0$.

De dichtheid is $\frac{d}{dy}e^{-ny} = ne^{-ny}$, voor $y>0$.

2b $E_\theta(X_{(1)}+c) = E_\theta(X_{(1)}-\theta)+c+\theta \overset{(2a)}{=} \frac{1}{n}+c+\theta$

$\mathrm{var}_\theta(X_{(1)}+c) = \mathrm{var}(X_{(1)}-\theta) \overset{(2a)}{=} \frac{1}{n^2}$.

Dus MSE is $\frac{1}{n^2} + \left(\frac{1}{n}+c+\theta-\theta\right)^2 = \frac{1}{n^2}+\left(\frac{1}{n}+c\right)^2$. Voor $c=-\frac{1}{n}$ is deze minimaal.

2c $\frac{\int_0^\infty \theta \prod_{i=1}^{n} e^{-(x_i-\theta)} 1_{x_i>\theta}\, e^{-\theta}\,d\theta}{\int_0^\infty \prod_{i=1}^{n} e^{-(x_i-\theta)} 1_{x_i>\theta}\, e^{-\theta}\,d\theta} = \frac{\int_0^{x_{(1)}} \theta e^{n\theta} e^{-\theta}\,d\theta}{\int_0^{x_{(1)}} e^{n\theta} e^{-\theta}\,d\theta} =$

$= \frac{\frac{x_{(1)}e^{(n-1)x_{(1)}}}{n-1} - \frac{1}{(n-1)^2}\left(e^{(n-1)x_{(1)}}-1\right)}{\frac{1}{n-1}\left(e^{(n-1)x_{(1)}}-1\right)} = \frac{x_{(1)} - \frac{1}{n-1}\left(1-e^{-(n-1)x_{(1)}}\right)}{1-e^{-(n-1)x_{(1)}}}$.

(mits $X_{(1)}>0$, wat we kunnen aannemen, want $x_{(1)}>0$ b.z.)

3a $\prod_{i=1}^{n} f_\theta(x_i) = c(\theta)^n \left(\prod_{i=1}^{n} \frac{x_i}{1-x_i}\right)^\theta \prod_{i=1}^{n} x_i \prod_{i=1}^{n}(1-x_i) \; 1_{0<x_{(1)}\leq x_{(n)}<1}$

Vanwege de definitie (factorisatie) is $\prod_{i=1}^{n} \frac{x_i}{1-x_i}$ voldoende.

3b We kunnen de simultane dichtheid in de vorm

$c(\theta)^n e^{\theta \log\left(\prod \frac{x_i}{1-x_i}\right)} \prod x_i \prod (1-x_i)$

schrijven, een exponentiële familie met $Q_1(\theta)=\theta$ en voldoende grootheid $\log \prod_{i=1}^{n}\left(\frac{x_i}{1-x_i}\right)$. Deze is volledig want $\{Q_1(\theta) : \theta \in (0,1)\} = (0,1)$ bevat ~~een~~ inwendig punt.

3c Los op $\overline{X} = \frac{\theta}{2}$, want $E_\theta X_1 = \frac{\theta}{2}$, dwz $\hat{\theta} = 2\overline{X}$

3d De momentenschatter is geen functie van de voldoende grootheid. Volgens de Stelling van Rao-Blackwell is er dan een zuivere schatter wel gebaseerd op de voldoende grootheid met een kleinere variantie. Dus $2\overline{X}$ is niet UMVZ.

4a $\prod_{i=1}^{n} \frac{p_{\theta_1}(x_i)}{p_{\theta_0}(x_i)} = \left(\frac{\theta_1/(1-\theta_1)}{\theta_0/(1-\theta_0)}\right)^{\sum_{i=1}^{n} x_i} \left(\frac{\theta_1}{\theta_0}\right)^{-n} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{2n} \left(\frac{2-\theta_0}{2-\theta_1}\right)^{n}$.

Voor $\theta_1 > \theta_0 = \frac{1}{2}$ is $\frac{\theta_1/(1-\theta_1)}{\theta_0/(1-\theta_0)} = \theta_1/(1-\theta_1) > 1$ en is dit quotiënt

een stijgende functie van $\sum_{i=1}^{n} x_i$. Volgens het Lemma van Neyman-Pearson is de meest onderscheidende lotingstoets voor $H_0: \theta = \theta_0$ tegen $H_1: \theta = \theta_1$,

$\varphi(x) = 1_{\sum x_i > d} + \partial 1_{\sum x_i = d}$, voor d en ∂ bepaald door

$P_{\theta_0}(\sum x_i > d) + \partial P_{\theta_0}(\sum x_i = d) = 0.05$.

4b Uit 4a volgt dat de gegeven toets UMP is voor $H_0: \theta = \theta_0$ tegen $H_1: \theta > \theta_0$, aangezien de toets niet afhangt van θ_1. De toets is ook UMP voor $H_0: \theta \leq \theta_0$ tegen $H_1: \theta > \theta_0$ omdat de gegeven dichtheid een eendimensionale exponentiële familie vormen.

4c $E_{\theta_0} X_1 = \frac{2}{2-\frac{1}{2}} = \frac{4}{3}$ $\mathrm{var}_{\theta_0} X_1 = \frac{2}{9}$. Vanwege de CLT is $\sqrt{n}(\overline{X}_n - 4/3)$ asymptotisch $N(0, 2/9)$ verdeeld. We lossen d uit 4a dus op uit $P_{\theta_0}(\sum x_i > d) = P_{\theta_0}\left(\frac{\sqrt{n}(\overline{X} - 4/3)}{\sqrt{2/9}} > \frac{\sqrt{n}\left(\frac{d}{n} - \frac{4}{3}\right)}{\sqrt{2/9}}\right) = 0.05$.

$(P_{\theta_0}(\sum x_i = d) \to 0)$. Ofwel $\frac{\sqrt{n}\left(\frac{d}{n} - 4/3\right)}{\sqrt{2/9}} = 1.65$. Dit geeft:

verwerpen als $\sqrt{n}(\overline{X} - 4/3) > 1.65\sqrt{2/9}$.

4d $P_{\theta_1}\left(\sqrt{n}(\overline{X} - \frac{4}{3}) > 1.65\frac{\sqrt{2}}{3}\right) = P_{\theta_1}\left(\sqrt{n}(\overline{X} - 3/2) > 1.65\frac{\sqrt{2}}{3} - \left(\frac{3}{2} - \frac{4}{3}\right)\sqrt{n}\right)$

$\approx 1 - \Phi\left(\frac{1.65\sqrt{2}/3 - \left(\frac{3}{2} - \frac{4}{3}\right)\sqrt{25}}{\sqrt{1/4}}\right) \approx 1 - \Phi(-0.11) \approx 0.54$ (want $E_{2/3} X_1 = \frac{3}{2}$, $\mathrm{var}_{2/3} X_1 = \frac{1}{4}$.