Mathematatisch Instituut
Universiteit Leiden

Tentamen Inleiding Statistiek

20 januari 2012, 10:00–13:00

Je mag geen boek of aantekeningen gebruiken, zakrekenmachines (ook grafisch) zijn wel toegestaan. Schrijf leesbaar en motiveer je antwoorden.

Opgave 1

De *Rayleigh-verdeling* met parameter $\theta > 0$ heeft kansdichtheid

$$f(x) = \frac{x}{\theta^2}\, \mathrm{e}^{-\frac{x^2}{2\theta^2}}, \qquad x \geq 0.$$

We nemen aan dat $X_1, \ldots, X_n$ onderling onafhankelijk en identiek verdeelde trekkingen zijn uit een *Rayleigh-verdeling* met parameter θ.

De verwachtingswaarde van de Rayeigh-verdeling is $\theta\sqrt{\pi/2}$. Je krijgt bonuspunten, als je dit bewijst.

(a) Gebruik de momenten-methode om een schatter voor θ te berekenen.

(b) Bereken de meest aannemelijke schatter (*maximum likelihood estimator*) voor θ.

(c) Bepaal de asymptotische variantie van de meest aannemelijke schatter.

(d) Laat zien dat $\sum_{i=1}^n X_i^2$ een voldoende (*sufficiënte*) statistiek is voor θ.

Opgave 2

We nemen aan dat $X_1, \ldots, X_n$ onderling onafhankelijk en identiek verdeelde trekkingen zijn uit een *normale verdeling* met gemiddelde μ en variantie σ^2. Beide parameter zijn onbekend.

De steekproefvariantie s^2 is gedefineerd als

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2,$$

We hebben op college bewezen dat $\frac{(n-1)\, s^2}{\sigma^2}$ chi-kwadraat (χ^2) verdeelt is met $n-1$ vrijheidsgraden. De χ^2-verdeling met $n-1$ vrijheidsgraden heeft verwachtingswaarde $n-1$ en variantie $2(n-1)$.

(a) Laat zien dat s^2 een zuivere schatter is voor σ^2, en bereken de variantie van s^2.

(b) Bepaal c (als functie van n) zodat de gemiddelde kwadratische fout (*mean-square error*) van $c \sum_{i=1}^n (X_i - \bar{X})^2$ minimaal is.

Opgave 3

De stochast X is binomiaal verdeeld met parameters n en p. Voor het gemak kunnen we aannemen dat n even is.

(a) Wat is de gegeneralizeerde likelihood quotiënt Λ voor het toetsen van de nulhypothese $H_0\colon p = 0.5$ tegen het alternatief $H_1\colon p \neq 0.5$?

(b) Laat zien dat de likelihood ratio toets de nulhypothese verwerpt als $|X - n/2|$ groot is. In andere woorden, bewijs het volgende: Er bestaat $c > 0$ zodat H_0 wordt geaccepteert als $|X - n/2| < c$ en H_0 wordt verworpen zodra $|X - n/2| > c$.

(c) Stel $n = 10$. Wat is het significantie-niveau van een toets die de nulhypothese verwerpt zodra $|X - n/2| \geq 2$? Wat is het onderscheidingsvermogen (*power*) van deze toets als feitelijk $p = 0,3$?

Opgave 4

We beschouwen het lineaire model in matrixvorm:

$$Y = X \cdot \beta + e$$

met $Y \in \mathbb{R}^n, X \in \mathbb{R}^{n \times p}, \beta \in \mathbb{R}^p$, de matrix X heeft volle rang en e is een n-dimensionale toevalsvector met gemiddelde 0 en covariantiematrix

$$\Sigma_e = \sigma^2 I_n = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}.$$

De p-dimensionale vector $\hat{\beta} = (X^T X)^{-1} X^T Y$ is de kleinste kwadraten schatter voor β. Verder is $\hat{Y} = X\hat{\beta}$ en $\hat{e} = Y - \hat{Y}$ de vector van residuen; de bijbehorende covariantiematrices zijn $\Sigma_{\hat{Y}}$ en $\Sigma_{\hat{e}}$. Bewijs dat

$$\Sigma_{\hat{Y}} + \Sigma_{\hat{e}} = \sigma^2 I_n,$$

en leid vervolgens af dat

$$\sum_{i=1}^{n} \mathrm{Var}(\hat{Y}_i) + \sum_{i=1}^{n} \mathrm{Var}(\hat{e}_i) = n\sigma^2.$$

Hint: Je kunt $\hat{Y}$ schrijven als $\hat{Y} = P \cdot Y$ met $P = X(X^T X)^{-1} X^T$. Merk op dat $P^T = P = P^2$, en hetzelfde geldt voor $I - P$.

Voor een n-dimensionale toevalsvector Y en een matrix $A \in \mathbb{R}^{n \times n}$ hebben we bewezen dat

$$\Sigma_{A \cdot Y} = A \cdot \Sigma_Y \cdot A^T.$$