Tentamen Kansrekening en Statistiek 2

woensdag 30 maart 2005, 10.00-13.00 uur

Er zijn 5 opgaven.
Formuleblad: geometrische verdeling, Poissonverdeling en exponentiële verdeling.

Opgave 1.
Laat X een discrete stochastische grootheid zijn. We willen toesten

$H_0 : X + 1$ bezit de geometrische verdeling met parameter $p = 0.05$,

tegen het alternatief

$H_1 : X$ bezit de Poissonverdeling met parameter $\mu = 0.95$.

De formules voor de geometrische verdeling en Poissonverdeling vind je op het formuleblad.
a) Beschrijf de Neyman-Pearson toets voor dit probleem.
b) Bepaal het kritieke gebied (critical region). Kies onbetrouwbaarheidsdrempel (significance level) $\alpha = 0.0975$.
c) Bereken het onderscheidend vermogen (power) van deze toets in H_1.

Opgave 2.
Laat $Y = (Y_1, \ldots, Y_n)^T$ een n-vector van waarnemingen zijn, en $z = (z_1, \ldots, z_n)^T$ een n-vector van gegeven co-variabelen. We bekijken het lineaire model

$$Y_i = \beta_0 + \beta_1 z_i + \beta_2 z_i^2 + U_i,\ i = 1, \ldots, n.$$

Hier is $\beta = (\beta_0, \beta_1, \beta_2)^T$ een vector van onbekende parameters en $U_1, \ldots, U_n$ zijn onderling onafhankelijke meetfouten met verwachting 0.
a) Laat zien dat het model equivalent is met

$$Y_i = \gamma_0 + \gamma_1 x_{i,1} + \gamma_2 x_{i,2} + U_i,\ i = 1, \ldots, n,$$

waarbij $\gamma = (\gamma_0, \gamma_1, \gamma_2)^T$ een vector van onbekende parameters is, en, voor $i = 1, \ldots, n$,

$$x_{i,1} = z_i - \sum_{i=1}^{n} z_i/n,$$

en

$$x_{i,2} = \frac{5}{3}\left(z_i^2 - \sum_{i=1}^{n} z_i^2/n\right).$$

b) Veronderstel nu dat $n = 5$ en dat de waarnemingen zijn $Y = (3, 0, 2, 1, 4)^T$ en $z = (2, 1, 1, 2, -1)^T$. Laat zien dat de kleinste-kwadratenschatter (least squares estimator) $\hat{\gamma}$ van γ gelijk is aan $\hat{\gamma} = (2, -3/2, 1/2)^T$.
c) Bereken nu de kleinste-kwadratenschatter van β en schets de geschatte regressie-functie en de gegevens in één figuur.

Opgave 3.
We beschouwen onderling onafhankelijke en identiek verdeelde waarnemingen $X_1, \ldots, X_n$ met waarden in $\mathbf{R}$. De verdelingsfunctie van een waarneming noemen we

$$F(x) = P(X_1 \leq x),\ x \in \mathbf{R}.$$

Voor alle $x \in R$ definiëren we

$$F_n(x) = \frac{\text{aantal } X_i \leq x}{n}$$

(de empirische verdelingsfunctie). Stel dat F dichtheid f heeft. We nemen als schatter van f in een gegeven punt x:

$$\hat{f}(x) = \frac{F_n(x+h) - F_n(x-h)}{2h},$$

waarbij $h > 0$ een gekozen bandbreedte is. In deze opgave onderzoeken we de verwachte kwadratische fout (mean square error, MSE) van $\hat{f}(x)$.
a) Noem $\hat{p} = F_n(x+h) - F_n(x-h)$ en $p = F(x+h) - F(x-h)$. Ga na dat

$$E\hat{p} = p,$$

en

$$\mathrm{var}(\hat{p}) = p(1-p)/n.$$

b) Stel dat

$$f(x) = \frac{1}{x^2},\ x > 1.$$

Bepaal F.
c) Bepaal, voor gegeven $x > h+1$, de onzuiverheid (bias) en de variantie van $\hat{f}(x)$.
d) Ga na dat voor de verwachte kwadratische fout de volgende bovengrens geldt:

$$\mathrm{MSE}(\hat{f}(x)) \leq \frac{h^4}{x^4} + \frac{1}{2nh}.$$

e) Bepaal de bandbreedte h die de bovengrens, gegeven in d), voor $\mathrm{MSE}(\hat{f}(x))$ minimaliseert.

Opgave 4.
Stel dat de simultane dichtheid van het paar (X, Y) wordt gegeven door

$$f(x,y) = x\mathrm{e}^{-(y+1)x},\ (x,y) \in [0,\infty)^2.$$

a) Bepaal de marginale verdeling van X en van Y.
b) Bepaal de voorwaardelijke verwachtingen $g(Y) = E(X|Y)$ en $h(X) = E(Y|X)$. Hint: gebruik de formules voor verwachting en tweede moment van de exponentiële verdeling (zie formuleblad).
c) Laat zien dat $E(X - g(Y)) = E(Y - h(X)) = 0$.

Opgave 5.
We bekijken een Markovketen $\{X_n\}_{n=0}^{\infty}$ met toestandsruimte $S = \{1,2,3\}$ en overgangswaarschijnlijkheden

$$(p_{i,j}) = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

a) Is de Markovketen reducibel? Motiveer je antwoord.
b) Bereken de kans

$$P(X_6 = 1 | X_0 = 1).$$