Hertentamen Statistiek
22 augustus 2012

I *Schrijf je naam en studentnummer op elk vel dat je inlevert.*
II *Het gebruik van het boek van J.A. Rice, aantekeningen, handouts en zakrekenmachines is toegestaan.*
III *U mag in ieder onderdeel de conclusies van voorgaande onderdelen gebruiken, ook als u die (nog) niet bewezen hebt.*
IV *Motiveer steeds uw antwoord door duidelijk aan te geven welke argumenten en welke resultaten u gebruikt om een bepaalde conclusie te trekken.*
V *U heeft 3 uur de tijd voor het tentamen.*
VI *Achter elke vraag staat het aantal punten dat met de vraag te behalen is. De puntenverdeling is: 1 - 40, 2 - 15, 3 - 15, 4 - 15, 5 -15.*
VII *Veel succes!*

Opgave 1 Zij X_1, X_2, ..., X_n onafhankelijke, identiek verdeelde stochasten met cumulatieve dichtheidsfunctie: $F(x|\theta) = \begin{cases} 0 & \text{als } x < 1 \\ 1 - \left(\frac{1}{x}\right)^\theta & \text{als } x \geq 1 \end{cases}$ met $\theta > 2$.

a 2pt) Bepaal de kansdichtheidsfunctie van X_1.

b 6pt) Bepaal de momentschatter van θ.

c 6pt) Stel dat $n = 50$, $\bar{X} = 1,28$, er 19 punten in het interval (1-1,2) liggen, 17 punten in het interval (1,2-1,4), 8 punten in het interval (1,4-1,6) en er 6 punten in het interval (1,6-∞) liggen. Bepaal de goodness of fit op basis van de momentschatter met een significantieniveau van $\alpha = 0{,}1$. (Gebruik $\hat{\theta} = 4,5$ als u onderdeel b) niet hebt gedaan.)

d 6pt) Bepaal de meest waarschijnlijke schatter (MLE) van θ.

e 6pt) Bepaal een minimaal voldoende statistiek voor θ en bewijs dat deze minimaal is.

f 6pt) Bepaal de Fisher informatie in 1 waarneming.

g 8pt) Bepaal de Cramér-Rao ondergrens. Bewijs of de momentschatter asymptotisch efficiënt is.

Opgave 2 Zij X_1, X_2, ..., X_n onafhankelijke, identiek verdeelde stochasten met kansdichtheidsfunctie $f(x|\theta) = \begin{cases} 0 & \text{als } x < 0 \\ \theta e^{-\theta x} & \text{als } x \geq 0 \end{cases}$, voor $\theta > 0$, en zij x_1, x_2, ..., x_n de waarde die de stochasten hebben aangenomen. Veronderstel als prior kansdichtheid voor θ dat elke waarde van θ groter dan 0 even waarschijnlijk is.

a 3pt) Waarom is dit een oneigenlijke prior?

b 3pt) Stel dat de prior gelijk is aan $p(\theta) = \begin{cases} 0 & \text{als } \theta \leq 0 \\ \lambda e^{-\lambda\theta} & \text{als } \theta > 0 \end{cases}$, met $\lambda = 10^{-10}$.
Wat is de interpretatie van de prior kansdichtheid voor θ (en wat is de invloed van het feit dat λ klein is)?

c 9pt) Gebruik de oneigenlijke prior om de posterior kansdichtheid van θ te berekenen. Schrijf deze kansdichtheid als de dichtheid van een bekende kansverdeling.

Zie ommezijde!

Opgave 3 Een boswachter wil het aantal dassen in een gebied bepalen. Daartoe heeft hij op een dag vallen gezet en de dassen die in de vallen terecht zijn gekomen zijn geoormerkt. In 10 vallen bleek een das te zitten. Korte tijd later zijn er opnieuw vallen geplaatst (op andere plaatsen in het natuurgebied). Dassen in deze vallen werden gevangen gehouden tot er 22 dassen gevangen waren. 6 van de 22 dassen hadden een eerder uitgedeeld oormerk. We negeren geboorte en sterfte van dassen gedurende het experiment en we veronderstellen dat geoormerkte dassen en dassen zonder oormerk dezelfde kans hebben om gevangen te worden, i.e., we veronderstellen dat het aantal geoormerkte dassen in de populatie van 22 dassen uit een hypergeometrische verdeling komt. Zij $n = 22$ de steekproefgrootte, zij $D = 10$ het aantal geoormerkte dassen, zij N het aantal dassen in het gebied en $d = 6$ is het aantal geoormerkte dassen in de steekproef.

a 7pt) Bepaal de meest waarschijnlijke schatter van N.

b 8pt) Veronderstel als prior kansdichtheid voor N: $p(N = i) = 1/100$ als $21 < i \leq 121$ en 0 anders. Bereken de volgende ratio's van de posterior kansen: $\frac{P_{posterior}(N=25)}{P_{posterior}(N=40)}$, $\frac{P_{posterior}(N=140)}{P_{posterior}(N=40)}$ en $\frac{P_{posterior}(N=33)}{P_{posterior}(N=40)}$.

Opgave 4 Een hoogleraar wil onderzoeken of het tentamencijfer afhangt van de leeftijd van de kandidaat. Er zijn 5 kandidaten geweest, ($1 \leq i \leq 5$). De hoogleraar veronderstelt het volgende model voor het tentamencijfer t_i van kandidaat i: $t_i = \beta_0 + \beta_1 a_i + \epsilon_i$ waarbij ϵ_i i.i.d. normaal verdeelde stochasten zijn met gemiddelde 0 en standaarddeviatie σ en a_i de leeftijd van kandidaat i is. De data staan in de tabel hieronder:

individu	i	1	2	3	4	5
leeftijd	a_i	17	21	22	23	25
tentamencijfer	t_i	9	7	5	6	5

De nulhypothese is dat leeftijd geen invloed heeft op het tentamencijfer. De hoogleraar gebruikt als significantieniveau $\alpha = 0,05$.

a 12pt) Toon aan of de nulhypothese verworpen kan worden.

b 3pt) Wat kunt u zeggen over het verwachte tentamencijfer van een kandidaat van 65 jaar?

Opgave 5 Hoe gevoelig een individu is om een mazelen-infectie te krijgen hangt af van de concentratie afweerstoffen (in mIU/ml) in het bloed. In een studiepopulatie van 12 individuen die ooit de mazelen hebben gehad, wil men onderzoeken of 65-plussers een lagere concentratie afweerstoffen hebben dan individuen die jonger zijn dan 65. Als significantieniveau wordt $\alpha = 0,05$ gebruikt. De concentratie afweerstoffen staat in de volgende tabel:

individu	1	2	3	4	5	6	7	8	9	10	11	12
leeftijd	33	82	71	23	67	17	69	66	3	87	71	12
concentratie	181	120	130	140	170	200	142	78	128	143	127	180

a 4pt) Beschrijf het toetsingsprobleem en kies een geschikte toets om de onderzoeksvraag te beantwoorden.

b 11pt) Gebruik deze toets om de onderzoeksvraag te beantwoorden.

Einde.