

1^e deeltentamen Statistiek

18 april 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 - 20, 2 - 15, 3 - 40, 4 - 25.
- VII Veel succes!

Opgave 1 Zij X_1, X_2, \dots, X_n onafhankelijke, identiek verdeelde stochasten met kansdichtheidsfunctie: $f(x|\theta) = \frac{\Gamma(2\theta)}{(\Gamma(\theta))^2} x^{\theta-1}(1-x)^{\theta-1}$ voor $0 < x < 1$ met $\theta > 0$.

U mag gebruiken dat voor $\alpha > -1$ en $\beta > -1$ geldt dat $\int_0^1 x^\alpha(1-x)^\beta dx = \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}$ en dat $\Gamma(x+1) = x\Gamma(x)$.

- a 7pt) Bepaal de momentschatter T_M van θ .
- b 7pt) Bepaal een voldoende statistiek voor θ .
- c 6pt) Bepaal een minimaal voldoende statistiek voor θ (mag dezelfde zijn als in onderdeel b)) en bewijs dat deze minimaal is.

Opgave 2 Stel dat een bestaand geneesmiddel (middel A) een ziekte in 80% van de gevallen geneest. Een bedrijf heeft een nieuw geneesmiddel (middel B) voor dezelfde ziekte ontwikkeld. Het bedrijf beweert dat het middel significant meer patiënten geneest dan middel A. Om dit te testen wordt een onderzoek gedaan waarbij 200 patiënten die de ziekte hebben middel B toegediend krijgen. Het resultaat van het onderzoek is dat 167 van deze patiënten genezen. De onbetrouwbaarheidsdrempel (α) is 0.05.

- a 6pt) Formuleer een kansmodel en beschrijf het toetsingsprobleem.
- b 9pt) Bepaal door een geschikte toets of de bewering van het bedrijf waar is.

Zie ommezijde!

Opgave 3 Zij X_1, X_2, \dots, X_n onafhankelijke, identiek verdeelde stochasten met kansdicht-

$$\text{heidsfunctie: } f(k|\theta) = \begin{cases} \theta & \text{als } k = 0 \\ \theta(1 - \theta) & \text{als } k = 1 \\ (1 - \theta)^2 & \text{als } k = 2 \end{cases} \text{ met } 0 < \theta < 1$$

Zij n_k het aantal observaties met waarde k ($k \in \{0, 1, 2\}$).

We beschouwen twee schatters, T_1 en T_2 voor θ die als volgt zijn gedefinieerd:

$$T_1 = \frac{n_0}{n}. \quad T_2 \text{ is de meest waarschijnlijke schatter (MLE).}$$

a 7pt) Bereken de meest waarschijnlijke schatter T_2 van θ .

b 8pt) Bepaal of de volgende vier uitspraken waar zijn:

- 1) T_1 is zuiver voor elke steekproefgrootte.
- 2) T_1 is consistent.
- 3) T_2 is zuiver voor elke steekproefgrootte.
- 4) T_2 is consistent.

c 8pt) Bereken de Fisher informatie voor de parameter θ en de Cramér-Rao ondergrens voor de variantie in een zuivere schatter van θ . Is T_1 efficiënt, m.a.w., voldoet T_1 aan de Cramér-Rao ondergrens?

d 3pt) Beargumenteer of u beter T_1 of T_2 kunt gebruiken als de steekproef groot is.

e 7pt) Bepaal een benaderd 99% betrouwbaarheidsinterval voor θ als de steekproef als resultaat gaf: $n_0 = 25$, $n_1 = 24$, $n_2 = 151$.

f 7pt) Bepaal de "Goodness of fit" van het model met een significantieniveau van 0.05 op basis van de steekproef uit onderdeel e).

Opgave 4 Stel dat een bloedbank bloeddonoren test of ze HIV-positief zijn. De test bepaalt de concentratie (X) van een bepaald eiwit in het bloed. Neem aan dat de gemeten concentratie (in mg/l) van het eiwit bij HIV-negatieve individuen normaal verdeeld is met gemiddelde 11 mg/l en standaarddeviatie 1 mg/l en dat de gemeten concentratie bij HIV-positieve individuen normaal verdeeld is met gemiddelde 18 mg/l en standaarddeviatie 2 mg/l. De test gebruikt een concentratie van 13 mg/l als drempelwaarde, d.w.z., als de gemeten concentratie groter is dan 13 mg/l, dan is het testresultaat positief, als de gemeten concentratie kleiner is dan 13 mg/l, dan is de test negatief. De nulhypothese is dat een bloeddonor HIV-negatief is. De alternatieve hypothese is dat een bloeddonor HIV-positief is.

a 5pt) Wat is het significantieniveau van de test?

b 5pt) Wat is het onderscheidend vermogen (de power) van de test?

c 7pt) Neem als prior informatie aan dat 0.15% van de individuen die bloed aanbieden bij de bloedbank HIV positief is. Wat is de posterior kans dat een bloeddonor HIV-positief is gegeven dat de test een positief resultaat gaf.

d 8pt) Is er een test met hetzelfde significantieniveau die een groter onderscheidend vermogen heeft. Zo ja, construeer de test met het grootste onderscheidend vermogen, zo nee, bewijs dat de gegeven test het grootste onderscheidend vermogen heeft.