

Algemene Statistiek – 2BWI

2E DEELTENTAMEN/VOLLEDIG TENTAMEN

WOENSDAG 1 JUNI 2005

INSTRUCTIES

- Schrijf uw naam en student nummer op iedere pagina
- Geef op de eerste pagina duidelijk aan *deeltentamen* of *volledig tentamen*
- Gebruik van een rekenmachine is niet nodig en niet toegestaan
- Gebruik van geschreven materiaal is niet toegestaan
- U krijgt beschikking over een tabellenboekje

TWEEDE DEELTENTAMEN

- Opgaven: 3, 4 en 5 – Duur: 2 uur
- Normering ($Cijfer = (totaal\ 3, 4\ en\ 5)/70$)

3(a):	5	4(a):	25	5(a):	5
3(b):	5			5(b):	5
3(c):	5			5(c):	5
3(d):	5			5(d):	5
				5(e):	5

VOLLEDIG TENTAMEN

- Opgaven: 1, 2, 4 en 5 – Duur: 3 uur
- Normering ($Cijfer = (totaal\ 1, 2, 4\ en\ 5)/100$)

1(a):	5	2(a):	10	4(a):	25	5(a):	5
1(b):	5	2(b):	10			5(b):	5
1(c):	5	2(c):	5			5(c):	5
1(d):	5					5(d):	5
1(e):	5					5(e):	5

SLA DEZE PAGINA PAS OM ZODRA DAARVOOR DE INSTRUCTIE GEGEVEN IS

PROBLEEM 1 (*alléén voor volledig tentamen*)

De stochastische grootheden $X_1, \dots, X_n$ zijn onafhankelijk en identiek verdeeld. Iedere X_i neemt de waarde 1 of de waarde 2 aan en heeft marginale kansdichtheid

$$p_\theta(x) = \frac{2}{2+\theta} \left(\frac{\theta^{x-1}}{x} \right), \quad (x = 1 \text{ of } x = 2).$$

Hierin is $\theta > 0$ een onbekende parameter. Dit vraagstuk behandelt verschillende methodes voor schatting van θ op basis van de waarneming $(X_1, \dots, X_n)$.

- Bepaal de meest aannemelijke schatter $\hat{\theta}(X_1, \dots, X_n)$ voor θ .
Laat zien dat $\hat{\theta}(X_1, \dots, X_n) > 0$.
- Laat zien dat $E_\theta X_1 = (2 + 2\theta)/(2 + \theta)$.
- Bepaal de momentenschatter $\tilde{\theta}(X_1, \dots, X_n)$ voor θ .

We kiezen op de parameterruimte voor θ een a-priori verdeling met kansdichtheid

$$\pi(\theta) = \begin{cases} c(\beta, \lambda) \theta^\beta (2 + \theta)^n e^{-\lambda\theta}, & \text{als } \theta > 0, \\ 0 & \text{als } \theta \leq 0. \end{cases}$$

voor β en λ positieve constanten en $c(\beta, \lambda)$ een normeringsconstante.

- Laat zien dat de dichtheid van de a-posteriori verdeling voor θ relatief ten opzichte van deze a-priori verdeling een $\Gamma(\alpha, \lambda)$ -dichtheid is.
Geef een uitdrukking voor bijbehorende α in termen van β , n en $(X_1, \dots, X_n)$.
Hint: de dichtheid $f_{\alpha, \lambda}$ van de $\Gamma(\alpha, \lambda)$ -verdeling heeft de vorm:

$$f_{\alpha, \lambda}(\theta) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\lambda\theta}, & \text{als } \theta > 0, \\ 0, & \text{als } \theta \leq 0, \end{cases}$$

$$\text{waarbij } \Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du.$$

- Bepaal de Bayes schatter voor θ behorend bij de gegeven a-priori verdeling. U mag hierbij gebruiken dat $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ voor alle $\alpha > 0$.

PROBLEEM 2 (*alléén voor volledig tentamen*)

De stochastische grootheden $X_1, \dots, X_n$ zijn *i.i.d.*- P_θ verdeeld, waarbij P_θ wordt voorgeschreven door de kansdichtheid

$$p_\theta(x) = \begin{cases} \frac{2x}{\theta^2} & \text{als } 0 \leq x \leq \theta, \\ 0 & \text{als } x < 0 \text{ of } x > \theta. \end{cases}$$

Hierin is $\theta > 0$ een onbekende parameter. We beschouwen de schatters $T_c = c\bar{X}$ voor $c \in \mathbb{R}$.

- Voor welke c is T_c een zuivere schatter voor θ ?
- Voor welke c heeft T_c de kleinste verwachte kwadratische fout (MSE)?
- Welke schatter T_c verdient de voorkeur?

PROBLEEM 3 (*alléén voor deeltentamen*)

De gemiddelde temperatuur gedurende een zomer in Nice, Zuid-Frankrijk kan goed worden gemodelleerd met behulp van een normale verdeling met verwachting 26.6 graden Celcius en standaardafwijking 1.6 graden. Onderzoekers geïnteresseerd in het broeikaseffect willen aantonen dat in de laatste jaren een stijging van de gemiddelde temperatuur in de zomer heeft plaatsgevonden. Daartoe beschouwen ze de gemiddelde temperatuur X_i tijdens de zomers van de laatste negen jaren ($i = 1, 2, \dots, 9$). De gemeten gemiddelde zomertemperaturen zijn:

$$\begin{aligned} x_1 = 27.5, \quad x_2 = 30.3, \quad x_3 = 26.0, \quad x_4 = 26.0, \quad x_5 = 28.1, \\ x_6 = 27.9, \quad x_7 = 27.6, \quad x_8 = 25.6, \quad x_9 = 26.5, \end{aligned}$$

(in graden Celcius). U kunt in het vervolg van dit vraagstuk gebruiken dat $\sum_{i=1}^9 x_i = 245.5$.

- Beschrijf een geschikt statistisch model voor deze situatie.
Geef duidelijk aan welke parameter(s) bekend danwel onbekend geacht worden en welke onafhankelijkheidsveronderstelling(en) u maakt.
- Formuleer een geschikte nulhypothese en alternatieve hypothese.
- Toets de nulhypothese bij onbetrouwbaarheidsdrempel 5%.
- Bepaal de overschrijdingskans corresponderend met de antwoorden op a) en b).

PROBLEEM 4 (voor zowel deeltentamen als volledig tentamen)

Een politieke partij behaalde bij de laatste verkiezingen 20% van de stemmen. Door middel van een opiniepeiling willen we nagaan of dit percentage is gestegen. We toetsen daartoe de nulhypothese $H_0 : p \leq 0.20$ op basis van het percentage aanhangers van de partij in een steekproef ter grootte n met teruglegging.

- a. Hoe groot moeten we n minimaal kiezen (bij benadering) om een fout van de eerste soort van ten hoogste 4% en een onderscheidend vermogen in $p = 0.25$ van ten minste 96% te bewerkstelligen?

PROBLEEM 5 (voor zowel deeltentamen als volledig tentamen)

De stochastische grootheden $X_1, \dots, X_n$ zijn onderling onafhankelijk en identiek verdeeld met kansdichtheid

$$p_\theta(x) = \begin{cases} 4\theta x^3 e^{-\theta x^4}, & \text{als } x > 0, \\ 0, & \text{als } x \leq 0. \end{cases}$$

Hierin is $\theta > 0$ een onbekende parameter. De waarneming is de vector $X = (X_1, \dots, X_n)$.

- a. Bepaal de meest aannemelijke schatter voor θ .
- b. Bepaal de waargenomen informatie voor θ (als functie van X).
- c. Geef een formule voor een benaderend betrouwbaarheidsinterval voor θ met betrouwbaarheid $1 - \alpha$ gebaseerd op de meest aannemelijke schatter.
- d. Bepaal de log-likelihood-ratio statistiek voor het toetsen van de nulhypothese $H_0 : \theta = \theta_0$.
- e. Geef een formule voor een benaderend (voor n groot) betrouwbaarheidsinterval voor θ met betrouwbaarheid $1 - \alpha$ gebaseerd op de likelihood ratio statistiek.