

Kansrekening en statistiek
wi2105IN-deel 2
16 april 2010, 14.00–16.00 uur

Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Tevens krijgt u een formuleblad uitgereikt—na afloop inleveren alstublieft.

Het multiple-choice deel telt voor 2/3 mee voor het eindcijfer; de open-vragen voor 1/3.

Meerkeuzevragen

Toelichting: In het algemeen zijn niet altijd vijf van de zes alternatieven 100% fout, het juiste antwoord is het meest volledige antwoord. Maak op het bijgeleverde antwoordformulier het hokje behorende bij het door u gekozen alternatief zwart of blauw. Doorstrepen van een fout antwoord heeft geen zin: u moet het òf uitgummen, òf verwijderen met correctievloeistof òf een nieuw formulier invullen. Vergeet niet uw studienummer in te vullen en aan te strepen.

1. Beschouw de volgende twee beweringen over de kernschatter:

- (A) Verkleinen van de bandbreedte bij een kernschatter zorgt voor een gladdere kernschatting.
- (B) De keuze van de kernfunctie heeft een naar verhouding kleinere invloed dan de keuze van de bandbreedte op de kernschatting.

Welke beweringen zijn waar?

- a. Alleen A b. Alleen B c. A en B d. Geen van beide

2. We genereren een bootstrap dataset $(x_1^*, x_2^*, x_3^*, x_4^*)$ uit de empirische verdelingsfunctie van de dataset

1 3 4 6.

De kans dat precies 2 elementen in de bootstrap dataset zijn dan 2 is (afgerond op 2 decimalen) gelijk aan

- a. 0.42 b. 0.06 c. 0.04 d. 0.05 e. 0.21 f. 0.25

3. In een rapport staat een 95% betrouwbaarheidsinterval $(1.6; 7.8)$ voor de parameter μ van een $N(\mu, \sigma^2)$ -verdeling. Het interval is gebaseerd op 16 waarnemingen, geconstrueerd met behulp van het gestudentiseerd gemiddelde. Het gemiddelde van de dataset is gelijk aan:

- a. Voor het beantwoorden van de vraag moeten we de waarde van σ kennen.
- b. Voor het beantwoorden van de vraag moeten we de waarde van de s_n kennen.
- c. Voor het beantwoorden van de vraag moeten we de waarde van μ kennen.
- d. 6.2
- e. 5.4
- f. 4.7

4. Gegeven onafhankelijke stochasten $X_1, \dots, X_n$ uit een geometrische verdeling met parameter p . We hebben twee schatters voor p , namelijk

$$S = \frac{1}{\bar{X}_n} \quad \text{en} \quad T = \frac{\#\{i : X_i = 1\}}{n}.$$

Hierbij gebruiken we de notatie $\#A$ voor het aantal elementen in de verzameling A . Dan geldt

- a. S en T zijn beide zuiver voor p
 - b. S is zuiver en T heeft een negatieve bias
 - c. T is zuiver en S heeft een negatieve bias
 - d. S is zuiver en T heeft een positieve bias
 - e. T is zuiver en S heeft een positieve bias
 - f. S en T zijn beide onzuiver
5. Stel dat $X_1, \dots, X_n$ onafhankelijke stochasten zijn met een exponentiële verdeling met parameter λ . Beschouw de schatter $T = nM_n$ voor $1/\lambda$, waarbij $M_n = \min\{X_1, \dots, X_n\}$. We kunnen laten zien dat de verdeling van M_n exponentieel met parameter $n\lambda$ is. De mean squared error van T (als schatter voor $1/\lambda$) is gelijk aan
- a. $\frac{1}{\lambda}$
 - b. $\frac{n}{\lambda}$
 - c. $\frac{1}{n\lambda}$
 - d. $\frac{1}{n^2\lambda^2}$
 - e. $\frac{1}{\lambda^2}$
 - f. $\frac{1}{n\lambda^2}$
6. In een zevendaagse studie naar het effect van ozon is een groep van 14 ratten in een ozon-vrije omgeving, en een groep van 13 ratten in een ozon-rijke omgeving geplaatst. Voor iedere rat is de gewichtstoename (in grammen) gemeten. We vragen ons af of ozon een negatieve invloed heeft op gewichtstoename. We onderzoeken dit door de hypothese $H_0 : \mu_1 = \mu_2$ te toetsen tegen de alternatieve hypothese $H_1 : \mu_1 > \mu_2$. Hierbij zijn μ_1 en μ_2 de gewichtstoename voor een rat in de ozon-vrije en ozon-rijke omgeving respectievelijk. In de ozon-vrije omgeving is de gemiddelde gewichtstoename $\bar{x}_{14} = 22.40$; in de ozon-rijke omgeving $\bar{y}_{13} = 11.01$. De gepoolde standaard-deviatie is 4.58. De niet gepoolde standaard-deviatie is 4.14.
- Veronderstel dat de data normaal verdeeld zijn, met gelijke variantie in beide groepen. De p -waarde van de t -toets is ongeveer gelijk aan: (kies het antwoord dat het beste past)
- a. 0.005
 - b. 0.05
 - c. 0.025
 - d. 0.001
 - e. 0.01
 - f. geen van de antwoorden bij (a) tot en met (e)
7. Gegeven is een dataverzameling die een realisatie is van een steekproef van omvang 16 uit een kansverdeling met verwachting μ . Het gemiddelde van de data is 10 en de steekproef-variantie is 4. Men voert een bootstrapsimulatie uit voor het gestudentiseerde gemiddelde met 1000 herhalingen. Van de 1000 geordende bootstrapwaarden is het volgende bekend:
- | | | | | | |
|--------|--------|---------|---------|---------|---------|
| 25-ste | 50-ste | 100-ste | 901-ste | 951-ste | 976-ste |
| -3.42 | -2.94 | -2.01 | 1.11 | 1.39 | 1.62 |
- Het 95% bootstrap betrouwbaarheidsinterval voor μ wordt gegeven door:
- a. (8.290, 10.825)
 - b. (8.530, 10.695)
 - c. (8.995, 10.555)
 - d. (9.190, 11.710)
 - e. (9.305, 11.470)
 - f. (9.445, 11.005)
8. Voor een dataverzameling construeert men een histogram met cellen $[0, 1]$, $(1, 3]$, $(3, 5]$, $(5, 8]$, $(8, 11]$, $(11, 14]$ and $(14, 18]$. Gegeven zijn de waarden van de empirische verdelingsfunctie op de grenzen van de cellen:

t	$F_n(t)$
0	0
1	0.225
3	0.445
5	0.615
8	0.735
11	0.805
14	0.910
18	1.000

Het 40ste percentiel van de dataverzameling ligt in cel

- a.** $(1, 3]$ **b.** $(3, 5]$ **c.** $(5, 8]$
d. $(8, 11]$ **e.** $(11, 14]$ **f.** $(14, 18]$

9. Gegeven is de dataverzameling:

11 14 18 20 21 23 31 32 41
43 67 77 80 83 92 134 183 201 215

en de empirische quantielen

p	0.25	0.50	0.75
$q_n(p)$	21	43	92

Het aantal elementen van de dataverzameling dat buiten de snorharen ("whiskers") van de boxplot valt is gelijk aan

- a.** 0 **b.** 2 **c.** 3 **d.** 4 **e.** 6 **f.** 8

10. We beschikken over een dataverzameling $x_1, x_2, \dots, x_n$ met gemiddelde $\bar{x}_n$, mediaan m_n en standaardafwijking s_n . De dataverzameling is een realisatie van een steekproef $X_1, X_2, \dots, X_n$ uit een $N(\mu, \sigma^2)$ verdeling. Om het gedrag te onderzoeken van de mediaan M_n van $X_1, X_2, \dots, X_n$ als schatter voor μ wil men de kansverdeling van de stochast $T_n = M_n - \mu$ benaderen door middel van een bootstrapsimulatie. Bij het uitvoeren van de simulatie genereert men in elke iteratie een bootstrap dataverzameling $x_1^*, x_2^*, \dots, x_n^*$ en berekent men t_n^* . Noteer $\bar{x}_n^*$ en m_n^* als het gemiddelde en de mediaan van $x_1^*, x_2^*, \dots, x_n^*$. Dan

- a.** is x_i^* getrokken uit een $N(m_n, s_n^2)$ verdeling, en $t^* = m_n^* - m_n$
b. is x_i^* getrokken uit een $N(m_n, s_n^2)$ verdeling, en $t^* = \bar{x}_n^* - \bar{x}_n$
c. is x_i^* getrokken uit een $N(m_n, s_n^2)$ verdeling, en $t^* = m_n^* - \mu$
d. is x_i^* geloot uit $x_1, x_2, \dots, x_n$ en $t^* = m_n^* - m_n$
e. is x_i^* geloot uit $x_1, x_2, \dots, x_n$ en $t^* = \bar{x}_n^* - \bar{x}_n$
f. is x_i^* geloot uit $x_1, x_2, \dots, x_n$ en $t^* = m_n^* - \mu$

Open vragen

Toelichting: Een antwoord alleen is *niet* voldoende: er dient een berekening, toelichting en/of motivatie aanwezig te zijn. Dit alles goed leesbaar en in goed Nederlands.

1. Gegeven de dataset $x_1, x_2, \dots, x_n$. We vatten deze dataset op als een realisatie van de onafhankelijke stochasten $X_1, X_2, \dots, X_n$ uit een Poisson verdeling met parameter μ . Leid een formule af voor de meest aannemelijke (maximum likelihood) schatter voor μ , gegeven de dataset $x_1, x_2, \dots, x_n$.
2. We genereren een getal x uit een uniforme verdeling op het interval $[0, \theta]$ ($\theta > 0$). We toetsen $H_0 : \theta = 3$ tegen $H_1 : \theta \neq 3$ door H_0 te verwerpen als $x \leq 0.5$ of $x \geq 2.5$.
 - a. Bereken de kans op het begaan van een type I fout.
 - b. Bereken de kans op het begaan van een type II fout als de echte waarde van θ gelijk is aan 3.5.
 - c. Bepaal het kritieke gebied van de toets (met als toetsingsgrootte de waarneming X), als we eisen dat de kans op een type I fout precies 0.1 moet zijn.
3. Stel dat we data $x_1, \dots, x_n$ hebben die we opvatten als een realisatie van een steekproef $X_1, \dots, X_n$ uit een normale verdeling met onbekende verwachting μ en bekende variantie σ^2 . In dit geval wordt een $100(1 - \alpha)\%$ betrouwbaarheidsinterval voor μ gegeven door

$$\left(\bar{x}_n - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right). \quad (*)$$

- a. Stel dat $\sigma = 2$ en $\alpha = 0.05$. Hoe groot moet n zijn opdat de lengte van het betrouwbaarheidsinterval maximaal 1 is?
- b. Stel dat we berekend hebben dat het betrouwbaarheidsinterval voor μ gelijk is aan $(1\frac{1}{2}, 3)$. Geef een betrouwbaarheidsinterval voor $1 - \mu$ (bij dezelfde data en dezelfde betrouwbaarheid).