

**Kansrekening en statistiek**  
**wi2105IN—deel 2**  
**11 april 2011, 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.

**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. Laat gegeven zijn dat  $X_1, X_2, \dots, X_{60}$  onafhankelijke en uniform verdeelde stochasten op het interval  $[0, 3]$  zijn. Er geldt  $\text{Var}(X_1) = \frac{3}{4}$ . De centrale limietstelling geeft dat  $P(X_1 + X_2 + \dots + X_{60} \geq 80)$  bij benadering gelijk is aan
- a. 0.9319      b. 0.7967      c. 0.8159      d. 0.8289      e. 0.9236      f. 0.9871
2. We beschikken over een dataset van omvang 20 die we opvatten als realisatie van onafhankelijke stochasten  $X_1, X_2, \dots, X_{20}$ . Gegeven is dat elke  $X_i$  een continue verdeling heeft met kansdichtheid  $f(x) = \frac{1}{2\theta}$ , als  $-\theta \leq x \leq \theta$ , en  $f(x) = 0$  voor alle andere waarden van  $x$ . Verder is  $\theta > 0$  onbekend. Definieer de volgende schatter voor  $\theta$ :

$$T = a(|X_1| + |X_2| + \dots + |X_{20}|).$$

Voor welke  $a$  is  $T$  een zuivere schatter voor  $\theta$ ?

- a.  $\frac{1}{40}$       b.  $\frac{1}{20}$       c.  $\frac{3}{20}$       d.  $\frac{1}{10}$       e.  $\frac{5}{40}$       f.  $\frac{3}{40}$
3. Gegeven is de volgende dataset

|       |       |       |       |       |
|-------|-------|-------|-------|-------|
| .0018 | .1571 | .1687 | .2006 | .2093 |
| .2691 | .2774 | .3233 | .3279 | .3922 |
| .4417 | .4931 | .5704 | .5998 | .6013 |
| .6221 | .6719 | .9470 | .9531 | .9870 |

Voor welk getal in de dataset is de waarde van de empirische verdelingsfunctie van deze dataset gelijk aan 0.6?

- a. 0.2691      b. 0.5704      c. 0.4931      d. 0.9470      e. 0.4417      f. 0.4991
4. Gegeven zijn twee steekproefgrootheden (Engels: sample statistics)  $T$  en  $S$ . We willen een parameter  $\theta$  schatten, die uit het interval  $(1, \infty)$  komt. Bekend is dat  $E[T] = 2\theta$ , en dat  $E[S] = \sqrt{\theta}$ . Welke van de volgende schatters is een zuivere schatter voor  $\theta$ ?
- a.  $2T$       b.  $T^2$       c.  $T - 2$   
d.  $T + S$       e.  $\sqrt{S}$       f. i.h.a. géén van deze vijf
5. Gegeven is een dataverzameling die een realisatie is van een steekproef van omvang 16 uit een  $N(\mu, \sigma^2)$  verdeling, met  $\mu$  en  $\sigma^2 > 0$  onbekend. Verder is gegeven dat het gemiddelde van de data 15 is, en de steekproefvariantie gelijk is aan 4. Het 90% betrouwbaarheidsinterval voor  $\mu$  wordt dan gegeven door:
- a. (13.934, 16.066)      b. (13.940, 16.006)      c. (14.020, 15.980)  
d. (14.124, 15.877)      e. (14.127, 15.873)      f. (14.177, 15.822)

6. Gegeven data  $x_1, x_2, \dots, x_{15}$  die we opvatten als een steekproef uit een verdeling met kansdichtheid

$$f(x) = \begin{cases} 0 & x < \theta \\ e^{-(x-\theta)} & x \geq \theta \end{cases}.$$

We toetsen de nulhypothese  $\theta = 0$  tegen de alternatieve hypothese  $\theta > 0$ . Als toetsingsgrootte nemen we  $T = \min\{X_1, X_2, \dots, X_{15}\}$ . Grote waarden van  $T$  duiden op de alternatieve hypothese. Stel dat de geobserveerde waarde van  $T$  gelijk is aan  $t = 0.1$ . De  $p$ -waarde bij deze toets (afgerond op 2 decimalen) is gelijk aan:

- a.** 0.14      **b.** 0.90      **c.** 0.22      **d.** 0.06      **e.** 0.12      **f.** 0.43

*Hint: Als  $X_1, X_2, \dots, X_n$  een steekproef uit een  $\text{Exp}(\lambda)$  verdeling is, dan heeft*

$$\min\{X_1, X_2, \dots, X_n\}$$

*een  $\text{Exp}(n\lambda)$  verdeling.*

7. Gegeven is een dataverzameling die een realisatie is van een steekproef van omvang 16 uit een kansverdeling met verwachting  $\mu$ . Het gemiddelde van de data is 10 en de steekproefvariantie 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 90% bootstrap betrouwbaarheidsinterval voor  $\mu$  wordt gegeven door:

- a.** (9.305, 11.470)      **b.** (8.290, 10.825)      **c.** (8.530, 10.695)  
**d.** (8.995, 10.555)      **e.** (9.175, 11.710)      **f.** (9.445, 11.005)

8. Stel dat  $X_1, X_2, \dots, X_n$  een steekproef is uit een verdeling met verwachting  $\mu$  en variantie  $\sigma^2$ . Beschouw de beweringen

- A Het steekproefgemiddelde is een zuivere schatter voor  $\mu$ .  
 B De steekproefvariantie  $S_n^2$  is een zuivere schatter voor  $\sigma^2$ .  
 C De steekproefstandaarddeviatie  $S_n$  is een zuivere schatter voor  $\sigma$ .

Welke van deze beweringen zijn waar?

- a.** geen      **b.** B en C      **c.** A  
**d.** A, B en C      **e.** A en C      **f.** A en B

9. Stel dat  $X_1, X_2, \dots, X_n$  een steekproef is uit een normale verdeling met verwachting  $\mu$  en variantie  $\sigma^2$ . Beschouw de volgende beweringen:

- A Het 95%-betrouwbaarheidsinterval voor  $\mu$  wordt smaller naarmate  $n$  groter wordt.  
 B Het 95%-betrouwbaarheidsinterval voor  $\mu$  is smaller dan het 99%-betrouwbaarheidsinterval voor  $\mu$

Welke van deze beweringen zijn waar?

- a.** geen      **b.** A      **c.** B      **d.** A en B

10. Voor een zekere statistische toets wordt de toetsingsgrootte gegeven door  $T$ . Het significantie-niveau is gelijk aan 0.1. Het kritieke gebied behorende bij deze toets noteren we met  $K$ . De kans dat we de nulhypothese verwerpen als deze niet waar is, is gelijk aan 0.6. De kans op een fout van de tweede soort is gelijk aan:

- a. 0.1                      b. 0.9                      c. 0.4  
d. 0.6                      e. 0.5                      f. 0.7

### 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. Men beschikt over een dataset van 13 elementen die men opvat als een realisatie van een steekproef uit een  $Pois(\mu)$  verdeling. Door omstandigheden zijn de 0-en en 1-en op een hoop terecht gekomen. Zodoende is alleen het volgende bekend:

| waarde | 0 of 1 | 2 | 3 |
|--------|--------|---|---|
| aantal | 10     | 2 | 1 |

- a. Beargumenteer dat de likelihood wordt gegeven door

$$L(\mu) = C \cdot \frac{e^{-13\mu} \mu^7 (\mu + 1)^{10}}{24},$$

waarbij de constante  $C$  het aantal volgorden is om 10 keer een 0 of een 1, 2 keer een 2 en 1 keer een 3 te hebben.

- b. Bepaal de loglikelihood en bereken de maximum likelihood schatting voor  $\mu$ .
2. Stel dat  $X$  een uniforme verdeling heeft op het interval  $[0, \theta]$ , waarbij  $\theta > 0$  een onbekende parameter is. We toetsen  $H_0 : \theta = 3$  versus  $H_1 : \theta > 3$ . Als toetsingsgrootte nemen we  $X$  zelf.
- a. Bereken het kritieke gebied bij significantieniveau 0.1.
- b. Stel dat we  $x = 1$  observeren. Bereken de bijbehorende  $p$ -waarde.

### Normering:

Voor iedere open vraag is 1 punt te behalen.

Voor de open vragen geldt:

1a: 1 punt

1b: 2 punten

2a: 1 punt

2b: 1 punt

In totaal kun je dus 15 punten halen.

Het eindcijfer wordt gegeven door  $1 + \text{aantal behaald punten} * \frac{9}{15}$ .

### Antwoorden multiple choice:

1 a.  $E[X] = 1.5$ ,  $\text{Var}(X) = \frac{9}{12} = \frac{3}{4}$ . De cls geeft

$$\frac{\sum_{i=1}^{60} X_i - 60 \cdot 1.5}{\sqrt{\frac{3}{4} \cdot 60}} \text{ heeft bij benadering een } N(0, 1) - \text{verdeling}$$

Dus

$$\begin{aligned} P(X_1 + X_2 + \dots + X_{60} \geq 100) &= P\left(\frac{\sum_{i=1}^{60} X_i - 60 \cdot 1.5}{\sqrt{\frac{3}{4} \cdot 60}} \geq \frac{80 - 90}{\sqrt{45}}\right) \\ &\approx P(Z \geq -1.4907) = 1 - 0.0681 = 0.9319 \end{aligned}$$

2 d.  $E[X_i] = \theta/2$ , dus  $E[T] = a \times 20 \times \theta/2 = 10a\theta$ . Voor zuiverheid moet dit gelijk zijn aan  $\theta$ . Dus moet  $a = 1/10$ .

3 c.  $0.6 = \frac{12}{20}$ , dus zoeken we  $x_{12}$  (i.e., het twaalfde punt in de dataset in volgorde van grootte). Dit is 0.4931.

4 f. Gegeven is, dat  $\theta$  uit het interval  $(1, \infty)$  komt, dus  $\theta \neq 0$ . Dit is handig om te weten, omdat gegeven is dat  $E[T] = 2\theta$ , en dus is  $E[T] \neq \theta$ , en is **dus**  $T$  geen zuivere schatter voor  $\theta$ . Ook is gegeven dat  $E[S] = \sqrt{\theta}$ . Omdat  $\theta > 1$ , zien we dus direct dat  $E[S] < \theta$ , en dus  $S$  is ook geen zuivere schatter voor  $\theta$ . Je gaat op eenvoudige wijze naa dat **f** het enig juiste antwoord is.

5 d. De beschreven situatie is die uit paragraaf 23.3 in het boek. In het bijzonder hoeft je alleen maar bovenstaande gegevens in te vullen in de formule bovenaan blz. 350 (of deze formule even af te leiden; dit hebben we een aantal malen op het college gedaan), en klaar is Kees!

6 c. Onder de nulhypothese heeft iedere  $X_i$  een  $\text{Exp}(1)$  verdeling. De verdeling van  $T$  onder de nulhypothese is dus  $\text{Exp}(15)$ . De  $p$ -waarde is dus

$$P(T > 0.1 \mid H_0 \text{ waar}) = e^{-15 \cdot 0.1} \approx 0.22.$$

7 a. Voor het 90% bootstrap betrouwbaarheidsinterval moeten we gebruik maken van de 50-ste en 951-ste bootstrap waarden. Het interval is dan

$$\left(10 - 1.39 \cdot \frac{2}{4}, 10 - (-2.94) \cdot \frac{2}{4}\right) = (9.305, 11.470).$$

8 f. Beweringen A en B zijn waar, zie MIPS. Jensen's ongelijkheid toont aan dat C niet waar is.

9 d. Beweringen A, B zijn waar, zie de uitdrukkingen voor de betrouwbaarheidsintervallen in MIPS.

10 c. Kans op fout van tweede soort  $= P(\text{verwerp } H_0 \text{ niet} \mid H_0 \text{ niet waar})$  Dit is echter gelijk aan  $1 - P(\text{verwerp } H_0 \mid H_0 \text{ niet waar}) = 1 - 0.6 = 0.4$ .

### Antwoorden open vragen:

**1a** De likelihood is gelijk aan de kans op de data:

$$\begin{aligned}
 L(\mu) &= C \cdot P(X = 0 \text{ of } 1)^{10} \cdot P(X = 2)^2 \cdot P(X = 3) \\
 &= C \cdot (e^{-\mu} + \mu e^{-\mu})^{10} \cdot \left(\frac{\mu^2 e^{-\mu}}{2}\right)^2 \cdot \frac{\mu^3 e^{-\mu}}{6} \\
 &= C \cdot (e^{-\mu}(1 + \mu))^{10} \cdot \left(\frac{\mu^2 e^{-\mu}}{2}\right)^2 \cdot \frac{\mu^3 e^{-\mu}}{6} \\
 &= C \cdot \frac{e^{-13\mu} \mu^7 (\mu + 1)^{10}}{24}.
 \end{aligned}$$

**1b** De loglikelihood is:

$$\ell(\mu) = \ln(L(\mu)) = \ln(C) - \ln(24) - 13\mu + 7\ln(\mu) + 10\ln(\mu + 1).$$

Om het maximum te bepalen stellen we de afgeleide gelijk aan 0:

$$\begin{aligned}
 \ell'(\mu) = -13 + \frac{7}{\mu} + \frac{10}{\mu + 1} = 0 &\Leftrightarrow -13\mu(\mu + 1) + 7(\mu + 1) + 10\mu = 0 \\
 &\Leftrightarrow -13\mu^2 + 4\mu + 7 = 0 \\
 &\Leftrightarrow \mu = 0.904 \quad \text{of} \quad \mu = -0.596.
 \end{aligned}$$

Uit het tekenverloop van  $\ell'(\mu)$  volgt dat  $\ell(\mu)$  een maximum heeft bij  $\mu = 0.904$ . (bijvoorbeeld  $\ell'(0) = +\infty$ ). Dit betekent dat  $\mu = 0.904$  de maximum likelihood schatting is.

**2a** Het kritieke gebied is van de vorm  $[c, \infty)$  voor zekere  $c > 0$ . De kritieke waarde  $c$  volgt uit

$$P(X \in K \mid \theta = 3) = P(X \geq c \mid \theta = 3) = 0.1.$$

Nu geldt

$$P(X \geq c \mid \theta = 3) = \max(1 - c/3, 0)$$

en dit is dus gelijk aan 0.1 als  $1 - c/3 = 0.1$ . Hieruit volgt dat  $c = 2.7$ .

**2b** De  $p$ -waarde wordt gegeven door

$$P(X \geq 1 \mid \theta = 3) = \frac{2}{3}.$$