Technische Universiteit Delft | Mekelweg 4
Fac. der Technische Wiskunde en Informatica | 2628 CD Delft

Tentamen wi1604 Kansrekening en Statistiek I
26 augustus 2004, 9.00–12.00 uur

Toelichting. Er wordt verwacht dat je bij elke vraag een uitwerking geeft voorzien van toelichting en motivering. Alleen het antwoord levert niets op. Alle onderdelen hebben alle hetzelfde gewicht. Bij dit examen is het gebruik van een rekenmachine en het standaard Kanstat formuleblad toegestaan.

1. Gegeven is een stochast X met verdelingsfunctie $F(x) = 0$ voor $x < \sqrt{3}$ en

$$F(x) = 1 - \frac{3}{x^2} \qquad \text{voor } x \geq \sqrt{3}.$$

- **a.** Bepaal $\mathrm{E}[X]$.
- **b.** Laat zien hoe je een stochast X met deze verdelingsfunctie construeert uit een $U(0,1)$-verdeelde stochast U.

2. Gegeven de $Exp(\lambda)$ verdeelde stochast X.

- **a.** Laat zien dat de momentgenererende functie van X gelijk is aan $M(t) = \lambda/(\lambda - t)$, voor $t < \lambda$.

Beantwoord de volgende vragen met behulp van de momentgenererende functie, en leg telkens uit welke algemene regel je hebt gebruikt (bewijs niet gevraagd).

- **b.** Bepaal $\mathrm{E}\left[X^3\right]$.
- **c.** Bepaal de verdeling van $Y = \lambda X$.

3. X, Y en Z zijn onafhankelijke stochasten. Herleid $\mathrm{Cov}(Z + X, Z + Y)$ tot een zo eenvoudig mogelijke uitdrukking.

4. Bij het Commissariaat voor de Media heeft de omroep BNN haar ledenbestand met 224 964 leden ingeleverd. Alleen degenen die aan bepaalde eisen voldoen, mogen daadwerkelijk meetellen als lid. Voor verlenging van de zendmachtiging zijn minimaal 150 000 (echte) leden vereist.[1] Slechts 143 977 leden voldoen aan de wettelijke criteria om mee te tellen als lid (dit is het Commissariaat niet bekend).

- **a.** Bepaal de kans dat bij selectie van 8 leden, tenminste 6 daarvan voldoen aan de eisen.

[1] Deze getallen werden in 2004 in de kranten vermeld. De volgende getallen in deze opgave zijn verzonnen.

Er wordt een steekproef van 2000 leden uit het bestand genomen en geteld hoeveel daarvan voldoen aan de wettelijke criteria. We beschouwen dit als een toetsingsprobleem, waarbij het Commissariaat de aanvraag alléén afwijst als overtuigend is aangetoond dat het aantal leden onvoldoende is.

b. Formuleer de bij deze toets horende hypothesen, kies een toetsingsgrootheid en bepaal met behulp van de centrale limietstelling het kritieke gebied, voor $\alpha = 0.05$.

5. Gegeven is een steekproef $X_1, \ldots, X_n$ uit een $Pois(\mu)$ verdeling.

a. Laat zien dat $\bar{X}_n$ de maximum likelihood schatter voor μ is. Is het een zuivere schatter? (Onderbouw je antwoord.)

b. Bepaal de maximum likelihood schatter voor $\mathrm{P}(X_1 = 0)$ en een (zo scherp mogelijke) uitspraak over de zuiverheid.

6. Gegeven is een dataverzameling die een realisatie is van een steekproef $X_1, X_2, \ldots, X_n$ uit een $Par(\theta)$ verdeling. Men wil $H_0 : \theta = 2$ toetsen tegen $H_1 : \theta > 2$ bij significantieniveau 0.05 met toetsingsgrootheid

$$T = \frac{1}{n} \ln(X_1 X_2 \cdots X_n).$$

Op grond van de centrale limietstelling kan de kansverdeling van $\sqrt{n}(\theta T - 1)$ benaderd worden door een $N(0, 1)$ verdeling. Voor de dataverzameling is $n = 100$ en $t = 0.2$.

a. Gegeven is dat het kritieke gebied van de vorm is $K = (0, c]$. Bereken de kritieke waarde c met behulp van de normale benadering en geef uw conclusie omtrent de nulhypothese.

b. Bereken de kans op een type II fout als in werkelijkheid $\theta = 3$. (Indien je c niet kon bepalen, gebruik dan $c = 0.385$.)

c. Bereken een tweezijdig 95% betrouwbaarheidsinterval voor θ, gebruikmakend van de genoemde normale benadering.

Beknopte uitwerkingen

1a

1b

2b

2c

3 Gebruik $\text{Cov}(Z+X, Z+Y) = \text{E}[(Z+X)(Z+Y)] - \text{E}[Z+X]\,\text{E}[Z+Y]$. Uitwerken geeft: $\text{E}[(Z+X)(Z+Y)] = \text{E}\left[Z^2\right] + \text{E}[ZY] + \text{E}[XZ] + \text{E}[ZY]$. Omdat X, Y en Z onfhankelijk zijn geldt $\text{E}[ZY] = \text{E}[Z]\,\text{E}[Y]$, etc. Wanneer we gebruiken $\text{E}[Z+X] = \text{E}[Z] + \text{E}[X]$ (en hetzelfde voor $\text{E}[Z+Y]$) en alles uitwerken en uitvermenigvuldigen, dan blijken alle termen met X en Y erin tegen elkaar weg te vallen. We houden over: $\text{E}\left[Z^2\right] - (\text{E}[Z])^2$.

4a Het aantal leden X dat voldoet aan de eisen, heeft een $Bin(8, 0.64)$ verdeling. We berekenen

$$\begin{aligned} \text{P}(X \geq 6) &= \text{P}(X=6) + \text{P}(X=7) + \text{P}(X=8) \\ &= \binom{8}{6} 0.64^6 \cdot 0.36^2 + \binom{8}{7} 0.64^7 \cdot 0.36 + 0.64^8 \\ &= 0.2494 + 0.1267 + 0.0281 = 0.4042. \end{aligned}$$

4b Noem het aantal echte leden l. De hypothesen worden dan $H_0 : l = 150\,000$ en $H_1 : l < 150\,000$. Het aantal echte leden X in de steekproef heeft een $Bin(2000, p)$ verdeling. Omdat we moeten verwerpen bij te kleine waarden moeten we oplossen: $\text{P}(X \leq c) = 0.05$, ($X$ verdeeld zoals onder H_0). Onder H_0 geldt: $p = p_0 = 150000/224964$, en dus $\text{E}[X] = 2000p_0 = 1333.5$ en $\sigma_X = \sqrt{2000p_0(1-p_0)} = 21.08$. Wegens

$$\text{P}(X \leq c) = \text{P}\left((\frac{X - 1333.5}{21.08} \leq \frac{c - 1333.5}{21.08} \right) = \text{P}\left(Z \leq \frac{c - 1333.5}{21.08} \right) = 0.05$$

volgt dus $c = 1333.5 - 1.645 \cdot 21.08 = 1298.8 \to 1298$ (we ronden natuurlijk naar beneden af).

6a Voor c moet gelden $\text{P}(T \leq c) = 0.05$. Met behulp van de normale benadering is onder $H_0 : \theta = 2$:

$$\text{P}(T \leq c) = \text{P}\big(\sqrt{n}(2T-1) \leq \sqrt{n}(2c-1)\big) \approx \text{P}(Z \leq 10(2c-1))\,,$$

waarbij Z een $N(0,1)$ verdeling heeft. Dus moet c voldoen aan $\text{P}(Z \leq 10(2c-1)) = 0.05$, hetgeen betekent dat $10(2c-1) = -1.645$, of te wel $c = 0.41775$.
Aangezien $t = 0.2$ kleiner is dan $c = 0.41775$ en dus in het kritieke gebied ligt, moeten we de nulhypothese verwerpen.

6b De kans op een type II fout is gelijk aan $\mathrm{P}(T > c \mid \theta = 3)$. Met behulp van de normale benadering is voor $\theta = 3$:

$$\mathrm{P}(T > c) = \mathrm{P}\big(\sqrt{n}(3T - 1) > \sqrt{n}(3c - 1)\big) \approx \mathrm{P}(Z > 10(3c - 1))\,.$$

waarbij Z een $N(0,1)$ verdeling heeft. Omdat $c = 0.41775$ is de laatste kans gelijk aan $\mathrm{P}(Z > 2.53) = 0.0057$.

6c Op grond van de normale benadering geldt

$$\mathrm{P}\big(-1.96 < \sqrt{n}(\theta T - 1) < 1.96\big) \approx 0.95,$$

hetgeen equivalent is met

$$\mathrm{P}\left(\frac{1 - 1.96/\sqrt{n}}{T} < \theta < \frac{1 + 1.96/\sqrt{n}}{T}\right) \approx 0.95$$

Zodoende is het asymptotisch 95% betrouwbaarheidsinterval voor θ gelijk aan

$$\left(\frac{1 - 1.96/10}{0.2}, \frac{1 + 1.96/10}{0.2}\right) = (4.02,\ 5.98).$$