

15 januari 2007, 14.00–17.00 uur

Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan. Tevens krijgt u een formuleblad uitgereikt—na afloop inleveren alstublieft. Normering: De meerkeuze vragen en de open vragen tellen elk voor de helft van het cijfer. Bij de open vragen telt elk (vraag)onderdeel even zwaar.

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 èn aan te strepen.

1. X , Y en Z zijn onafhankelijke stochasten. $\text{Cov}(Z + X, Z + Y)$ kan ook geschreven worden als:
 - a. $\text{Var}(Z)$
 - b. $\text{Var}(X + Y)$
 - c. $\text{Cov}(X, Y)$
 - d. $\text{Var}(Z + X, Z + Y)$
 - e. 0
 - f. Geen van de voorgaande manieren
2. Gegeven is een rij onafhankelijke stochasten $X_1, X_2, \dots$ met dezelfde kansverdeling, namelijk kansdichtheid f en verdelingsfunctie F . Definieer

$$T_n = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

Volgens de wet van de grote aantallen geldt

$$\lim_{n \rightarrow \infty} \mathbb{P}(|T_n - a| > \epsilon) = 0$$

voor elke $\epsilon > 0$. Hierbij is a gelijk aan

- a.** $f(X_1^2)$. **b.** $F(X_1^2)$. **c.** $E[X_1^2]$.
d. $(E[X_1])^2$. **e.** $E[X_1^2]/n$. **f.** $\text{Var}(X_1)$

3. Vijf kaarten worden uit een kaartspel van 52 getrokken. De eerste is een koning. Wat is de kans op minstens 1 andere koning?

- a.** $1 - \frac{\binom{49}{5}}{\binom{52}{5}}$ **b.** $1 - \frac{\binom{48}{4}}{\binom{51}{4}}$ **c.** $1 - \frac{\binom{47}{3}}{\binom{51}{4}}$ **d.** $1 - \frac{\binom{49}{4}}{\binom{51}{4}}$ **e.** $\frac{\binom{48}{4}}{\binom{51}{4}}$ **f.** $\frac{\binom{47}{3}}{\binom{51}{4}}$

4. De stochast X heeft verdelingsfunctie

$$F(x) = \begin{cases} 0 & \text{als } x < 0 \\ 1 - e^{-x^2/4} & \text{als } x \geq 0 \end{cases}$$

Het 95-ste percentiel van de verdeling is gelijk aan

- a. 2.507 b. 2.888 c. 3.462 d. 4.217 e. 4.382 f. 5.013**

- 5.** Van twee gebeurtenissen A en B is gegeven

$$P(B) = \frac{2}{3}, \quad P(A|B) = \frac{1}{4} \quad \text{en} \quad P(A|A \cup B) = \frac{2}{5}.$$

Bereken $P(A)$.

- a. $\frac{1}{2}$ b. $\frac{1}{3}$ c. $\frac{1}{4}$ d. $\frac{1}{5}$ e. $\frac{2}{5}$ f. $\frac{1}{6}$

6. De stochastische variabele X heeft kansdichtheid $f(x) = \begin{cases} \frac{1}{2\sqrt{x}}, & \text{als } 1 \leq x \leq 4 \\ 0, & \text{elders.} \end{cases}$

Dan is $\text{Var}(3X + 2)$ gelijk aan

- a. $\frac{26}{15}$ b. $\frac{62}{15}$ c. $\frac{49}{15}$ d. $\frac{34}{5}$ e. $\frac{49}{5}$ f. $\frac{62}{5}$

7. Gegeven de $\text{Exp}(\lambda)$ verdeelde stochast X met momentgenererende functie $M(t) = \lambda/(\lambda - t)$, voor $t < \lambda$. Daaruit kunnen we afleiden dat $E[X^4]$ gelijk is aan:

- a. $4/\lambda$ b. $1/\lambda^4$ c. $4/\lambda^4$ d. $24/\lambda^4$ e. λ^4 f. $\lambda/(\lambda - 4)$

8. Een bottelmachine in een bierbrouwerij vult flesjes met een hoeveelheid bier B die normaal verdeeld is met $\mu = 300$ (ml) en $\sigma^2 = 25$ (ml²). Hoe groot is de kans dat een krat van 24 flesjes in totaal minder dan 7150 ml bier bevat?

- a. 0.0059 b. 0.021 c. 0.0485 d. 0.0668 e. 0.113 f. 0.123

9. Het aantal manieren waarop je met 12 dobbelstenen 30 kunt gooien komt overeen met de coëfficiënt van

- a. x^{12} uit $(1 - x^6)^{30} (1 - x)^{-30}$ b. x^{18} uit $(1 - x^6)^{12} (1 - x)^{-12}$
c. x^{18} uit $(1 - x^7)^{12} (1 - x)^{-12}$ d. x^{30} uit $(1 - x^5)^{12} (1 + x)^{12}$
e. x^{30} uit $(1 - x^6)^{12} (1 - x)^{-12}$ f. x^{30} uit $(1 - x^7)^{12} (1 - x)^{-12}$

10. Uit een verzameling objecten (bijv. tanks) genummerd 1 tot en met K worden *met* teruglegging 20 objecten getrokken. We willen toetsen $H_0 : K = 100000$ tegen $H_1 : K < 100000$, met het hoogste rangnummer M van onze steekproef als toetsingsgrootte. We vinden als realisatie voor M de waarde 85115. De p -waarde van deze uitkomst is:

- a. 0.01 b. 0.04 c. 0.05 d. 0.1 e. 0.15 f. 0.85

11. (Vervolg vorige vraag) Iemand gebruikt als kritiek gebied voor deze toets: $\{1, 2, 3, \dots, 86089\}$. Wat is de kans op een fout van type II als de werkelijke waarde van K gelijk is aan 96000?

- a. 0.83 b. 0.86 c. 0.89 d. 0.90 e. 0.95 f. 0.99

12. Over een histogram is het volgende gegeven:
- | cel | hoogte |
|-----------|--------|
| $[0, 2]$ | 0.05 |
| $(2, 6]$ | 0.10 |
| $(6, 9]$ | 0.10 |
| $(9, 13]$ | 0.05 |

De waarden van de empirische verdelingsfunctie in de punten 2 en 4 zijn dan respectievelijk

- a. 0.05 en 0.10 b. 0.10 en 0.20 c. 0.10 en 0.30
d. 0.05 en niet te bepalen e. 0.10 en niet te bepalen f. $1/6$ en $1/3$

13. Beschouw de volgende dataset: 1 2 3 4 6 10. De MAD van deze dataset is (ongeveer):

- a. 1.5 b. 2.0 c. 2.3 d. 2.4 e. 2.5 f. 3.5

14. Voor een dataset met $n = 135$, steekproefgemiddelde 656.88 en steekproefstandaardafwijking 1037.3 voer ik een gestudentiseerde bootstrap uit. Ik bepaal c_l^* en c_u^* zo dat

$$P\left(c_l^* < \frac{\bar{X}_n^* - \mu^*}{S_n^*/\sqrt{n}} < c_u^*\right) \approx 0.9.$$

Ik vind $c_l^* = -2.107$ en $c_u^* = 1.389$. Het 90% bootstrapbetrouwbaarheidsinterval voor de verwachting μ is dan

- | | | |
|-------------------|-------------------|--------------------|
| a. (468.8, 780.9) | b. (481.9, 831.9) | c. (510.0, 803.7) |
| d. (500.8, 812.9) | e. (532.9, 845.0) | f. niet te bepalen |

15. Stel we hebben een steekproef $X_1, \dots, X_6$ van een verdeling met onbekende verwachting $\mu > 2$. We bekijken de volgende vier schatters:

$$T_1 = X_1, \quad T_2 = \sqrt{X_5 X_6}, \quad T_3 = \frac{X_1 + X_3 + X_5}{9}, \quad T_4 = \frac{3X_2 - X_4 + X_5}{3}.$$

Dan zijn zuivere schatters voor μ :

- | | | |
|-------------------|-------------------|---------------------|
| a. alleen T_1 | b. T_1 en T_2 | c. T_1 en T_4 |
| d. T_2 en T_3 | e. T_2 en T_4 | f. T_3 en T_4 . |

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.

- We noemen twee getallen van n cijfers (inclusief “leading zeros”) *equivalent* als de ene uit de andere kan ontstaan door cijfers van positie te wisselen. Zo zijn bijvoorbeeld 12033, 20331 en 01332 equivalente getallen van vijf cijfers.
 - Van de lijst van getallen van vijf cijfers schrappen we getallen die equivalent zijn aan een getal dat al eerder in de lijst voorkomt. Hoeveel getallen van vijf cijfers houden we dan over?
 - We eisen bovendien dat (elk van) de cijfers 1, 3 en 7 niet meer dan eenmaal voor mogen komen in een getal. Hoeveel zijn het er dan?
- Een darter werpt een pijl op een cirkelvormig bord met een straal van 1 meter. Het is niet zo’n beste speler: hij raakt het op een uniform over het bord verdeelde lokatie. Noem de afstand tot de roos X . Voor $0 \leq t \leq 1$ geldt $P(X \leq t) = t^2$.
 - Bepaal $E[X]$.
 - Hij gooit in totaal drie pijlen. Neem aan dat de worpen onafhankelijk zijn. Beschouw de pijl die het verst van de roos terechtkomt; noem de afstand tot de roos Y . Bepaal de verwachting van Y . Hint: bepaal eerst de verdelingsfunctie.
- Een bepaalde soort komt voor in typen A, B en C, in de verhoudingen θ , 2θ resp. $1 - 3\theta$. Een steekproef van 60 levert 13 exemplaren van type A, 20 van type B en 27 van type C.
 - Bepaal de likelihood en de log-likelihood.
 - Bepaal de maximum likelihood schatting van θ .
- Een zestal metingen van de sterkte θ van een nieuw type plastic levert de waarden

27.57 28.64 26.92 29.54 28.22 26.82 29.91.

Deze getallen hebben gemiddelde 28.23 en standaardafwijking 1.21. De metingen worden verondersteld normaal verdeeld te zijn met verwachting θ .

- Bepaal een 95% betrouwbaarheidsinterval voor θ .
- Hoeveel metingen zijn er nodig om een 95% betrouwbaarheidsinterval van lengte hoogstens 0.4 te krijgen? (Neem aan dat de (verdelings)variantie kleiner is dan 2.)
- Je bent van plan dit plastic te gaan gebruiken als het sterk genoeg is. Bepaal het van toepassing zijnde eenzijdige 95% betrouwbaarheidsinterval voor θ .

Antwoorden multiple choice:

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

2 c. $E[X_1^2]$.

3 b. Na het trekken van de eerste kaart hebben we nog 51 kaarten met daarin 3 koningen. De kans op minstens 1 andere koning is 1 minus de kans op *geen* andere koning. We kunnen op $\binom{51}{4}$ manieren vier kaarten trekken, en als we het trekken van een koning uitsluiten: op $\binom{51-3}{4}$ manieren.

4 c. Gevraagd wordt de q waarvoor $P(X \leq q) = 0.95$.

$$\begin{aligned} 1 - e^{-q^2/4} = 0.95 &\iff e^{-q^2/4} = 0.05 \\ &\iff -q^2/4 = \ln 0.05 = -2.9957... \\ &\iff q = \sqrt{4 \cdot 2.9957} = 3.4616... \end{aligned}$$

5 b. Ten eerste: uit $P(A|B) = \frac{1}{4}$ volgt $P(A \cap B) = \frac{1}{4}P(B)$.
Ten tweede: uit $P(A|A \cup B) = \frac{2}{5}$ volgt $P(A) = \frac{2}{5}P(A \cup B)$, dus

$$\begin{aligned} P(A) &= \frac{2}{5}(P(A) + P(B) - P(A \cap B)) \\ &= \frac{2}{5}(P(A) + P(B) - \frac{1}{4}P(B)) \\ &= \frac{2}{5}(P(A) + \frac{3}{4}P(B)) = \frac{2}{5}P(A) + \frac{1}{5}P(B) \end{aligned}$$

Daaruit volgt via $\frac{3}{5}P(A) = \frac{1}{5}$ dat $P(A) = \frac{1}{3}$.

6 d.

$$\begin{aligned} E[X] &= \int_1^4 x \cdot \frac{1}{2\sqrt{x}} dx = \int_1^4 \frac{1}{2} \sqrt{x} dx = \left[\frac{1}{3} x \sqrt{x} \right]_1^4 = \frac{7}{3}. \\ E[X^2] &= \int_1^4 x^2 \cdot \frac{1}{2\sqrt{x}} dx = \int_1^4 \frac{1}{2} x \sqrt{x} dx = \left[\frac{1}{5} x^2 \sqrt{x} \right]_1^4 = \frac{31}{5}. \end{aligned}$$

Dus $\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{31}{5} - \frac{49}{9}$, en $\text{Var}(3X + 2) = 3^2 \text{Var}(X) = 9 \cdot \frac{31}{5} - 49 = \frac{34}{5}$.

7 d. $E[X^4] = 24/\lambda^4$; zie voorbeeld 11.5 van de reader.

8 b. Voer in: $X = B_1 + B_2 + \dots + B_{24}$, de totale hoeveelheid in 24 flesjes. Dan geldt: $X \sim N(24 \cdot 30, 24 \cdot 25) = N(7200, 600)$, dus $\frac{X-7200}{\sqrt{600}} \sim N(0, 1)$. Dan:

$$P(X \leq 7150) = P\left(\frac{X - 7200}{\sqrt{600}} \leq \frac{7150 - 7200}{\sqrt{600}} = -2.0412\right) \approx 0.021.$$

9 b. Het gevraagde aantal komt overeen met het aantal geheeltallige oplossingen van de vergelijking $x_1 + \dots + x_{12} = 30$, waarbij $1 \leq x_i \leq 6$ het aantal ogen op dobbelsteen i voorstelt. Door de substitutie $y_i = x_i - 1$ krijgen we $y_1 + \dots + y_{12} = 18$, waarbij $0 \leq y_i \leq 5$. Het aantal oplossingen komt overeen met de coëfficiënt van x^{18} uit $(1 + x + x^2 + \dots + x^5)^{12} = [(1 - x^6)/(1 - x)]^{12}$.

10 b. De p -waarde is hier de linkerstaartkans van 85115 onder H_0 , dus: $P(M \leq 85115) = (85115/100000)^{20} = 0.0398$.

11 c. De kans op een fout van de tweede soort is $P(M > 86089)$. Als $K = 96000$, dan is dit: $P(M > 86089) = 1 - P(M \leq 86089) = 1 - (86089/96000)^{20} \approx 0.8869$.

12 e. De waarde van de empirische verdelingsfunctie kan alleen op de celgrenzen bepaald worden. De eerste cel heeft oppervlak $0.1 = F_n(2)$. Verder kunnen we alleen concluderen $0.1 \leq F_n(4) \leq 0.5$, maar de precieze waarde is niet te bepalen.

13 b. De mediaan is $(3+4)/2 = 3.5$. De absolute verschillen met de mediaan zijn dus: 0.5 0.5 1.5 2.5 2.5 6.5. De MAD is de mediaan van deze getallen: $(1.5 + 2.5)/2 = 2$.

14 e. Zie het voorbeeld in paragraaf 23.3 van het boek.

15 c. Natuurlijk is $E[X_1] = \mu$, en via de eigenschap $E[aX + bY] = aE[X] + bE[Y]$ volgt $E\left[\frac{3X_2 - X_4 + X_5}{3}\right] = \frac{1}{3}E[3X_2 - X_4 + X_5] = \frac{3}{3}E[X_2] - \frac{1}{3}E[X_4] + \frac{1}{3}E[X_5] = \left(\frac{3}{3} - \frac{1}{3} + \frac{1}{3}\right)\mu = \mu$.

Antwoorden open vragen:

1 Dit is Exercise 11 uit 1.4 van Grimaldi, een huiswerkopgave.

1a Het is alleen van belang hoe vaak elk cijfer voorkomt, met andere woorden, de vraag is dezelfde als: op hoeveel manieren kunnen we uit tien objecten er vijf kiezen, als herhaling is toegestaan. Dus $\binom{14}{5}$.

1b Dit bekijken we als volgt: het aantal cijfers uit de verzameling $\{1, 3, 7\}$ kan 0, 1, 2 of 3 zijn. De overige cijfers kiezen we uit de resterende 7. Bij nul uit $\{1, 3, 7\}$ kunnen we $\binom{7+5-1}{5}$ niet-equivalente getallen formeren uit de overige cijfers; bij één zijn dat er $\binom{10}{4}$; bij twee $\binom{9}{3}$; en bij drie $\binom{8}{2}$. Deze aantallen moeten we nog vermenigvuldigen met het aantal manieren waarop de cijfers uit $\{1, 3, 7\}$ kunnen kiezen; als het aantal daaruit te kiezen cijfers k is, is $\binom{3}{k}$ het aantal mogelijkheden. In totaal vinden we:

$$\binom{11}{5} + \binom{3}{1}\binom{10}{4} + \binom{3}{2}\binom{9}{3} + \binom{8}{2}.$$

2a De kansdichtheid van X is $f(t) = 2t$ voor $0 \leq t \leq 1$, en $f(t) = 0$ elders. Er volgt $E[X] = \int_0^1 t \cdot 2t dt = \dots = \frac{2}{3}$.

2b Er geldt $Y = \max\{X_1, X_2, X_3\}$ waarbij X_1, X_2, X_3 onafhankelijk zijn, en verdeeld als de gegeven X . Er volgt (zie paragraaf 8.4): $F_Y(t) = t^6$ voor $0 \leq t \leq 1$. De kansdichtheid van Y is $f(t) = 6t^5$ voor $0 \leq t \leq 1$, en $f(t) = 0$ elders. Er volgt $E[Y] = \int_0^1 t \cdot 6t^5 dt = \dots = 6/7$.

3a De likelihood $L(\theta)$ is de kans op 13 exemplaren van type A, 20 van type B en 27 van type C, bij 60 trekkingen, als functie van θ , dus

$$L(\theta) = \binom{60}{13, 20, 27} \theta^{13} (2\theta)^{20} (1 - 3\theta)^{27},$$

waabij de multinomiaalcoëfficiënt staat voor het aantal volgorden waarop de gegeven uitkomst tot stand kan komen. De loglikelihood:

$$\ell(\theta) = \ln L(\theta) = \dots = \ln \binom{60}{13, 20, 27} + 20 \ln 2 + 33 \ln \theta + 27 \ln(1 - 3\theta),$$

na enig vereenvoudigen en hergroeperen van termen.

3b We zoeken het maximum van $\ell(\theta)$ bij een stationair punt. Eerst differentieren:

$$\ell'(\theta) = \frac{33}{\theta} - 3 \frac{27}{1-3\theta} \quad (\text{kettingregel})$$

en vervolgens

$$\ell'(\theta) = 0 \Leftrightarrow 33(1-3\theta) = 81\theta \Leftrightarrow 33 = (81+99)\theta,$$

waaruit volgt $\theta = 33/180 \approx 0.183$. Omdat $\ell'(\theta)$ dichtbij nul positief is en bij $33/180$ van teken wisselt, is dit punt inderdaad een maximum.

4a Het gemiddelde is $\bar{x}_7 = 28.23$ en de (steekproef)standaardafwijking $s_7 = 1.21$. $T = \frac{\bar{X}_7 - \theta}{S_7/\sqrt{7}}$ heeft een $t(6)$ -verdeling, dus halen we de kritieke waarde uit de tabel B.2: $t_{6,0.025} = 2.447$. Het 95% betrouwbaarheidsinterval wordt gegeven door

$$\left(\bar{x}_7 - 2.447 \cdot \frac{s_7}{\sqrt{7}}, \bar{x}_7 + 2.447 \cdot \frac{s_7}{\sqrt{7}} \right) = (27.11, 29.35)$$

4b Als σ bekend zou zijn, dan is het 95% betrouwbaarheidsinterval:

$$\left(\bar{x}_n - z_{0.025} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{0.025} \cdot \frac{\sigma}{\sqrt{n}} \right) = \left(\bar{x}_n - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \bar{x}_n + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right)$$

met breedte $2 \cdot 1.96 \cdot \sigma / \sqrt{n}$, wat vanwege de aanname omtrent σ niet groter is dan $2 \cdot 1.96 \cdot \sqrt{2} / \sqrt{n}$, wat we dus kleiner dan 0.4 proberen te maken. De steekproefgrootte n moet dus minstens $2 \cdot 1.96^2 / 0.2^2 = 192.08$ zijn, dus $n \geq 193$ voldoet.

4c Als in onderdeel **a.**, dit keer gebruiken we de kritieke waarde bij 0.05. Als gebruiker wil je een betrouwbaarheids~~sonder~~grens voor de sterkte, dus we maken een rechtszijdig betrouwbaarheidsinterval voor θ :

$$\left(\bar{x}_7 - 1.943 \cdot \frac{s_7}{\sqrt{7}}, \infty \right) = (27.34, \infty).$$