

WI1102CT – Kansrekening en Statistiek

Oefententamen
juni 2010
oktober 2009

januari 2009
oktober 2008
januari 2008



**Tentamenbundel Civiele Techniek
Het Gezelschap "Practische Studie"**

EEN REPRODUCERENDE
LEERSTIJL
IS SCHADELIJK VOOR
DE ACADEMISCHE
VORMING



63 pagina's

**Oefententamen Statistiek
wi1102CT**

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

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. IKEA heeft 1000 Billy boekenkasten in het magazijn. De kans dat er te weinig schroeven zitten in het bouw pakket van zo'n boekenkast is 0.1%. Hoe groot is de kans dat er in het magazijn precies twee boekenkasten staan met te weinig schroeven?

a. 0.184 b. 0.368 c. 0.512 d. 0.772 e. 0.801 f. 0.920

2. We werpen twee keer met een zuivere dobbelsteen, waarbij we aannemen dat de uitkomst van de eerste worp geen invloed heeft op de tweede worp. Definieer de stochastische variabele M als het maximum van de twee worpen. De kansverdeling van M wordt gegeven door

a.	<table><tr><td>a</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>$P(M=a)$</td><td>$\frac{6}{36}$</td><td>$\frac{6}{36}$</td><td>$\frac{6}{36}$</td><td>$\frac{6}{36}$</td><td>$\frac{6}{36}$</td><td>$\frac{6}{36}$</td></tr></table>	a	1	2	3	4	5	6	$P(M=a)$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	b.	<table><tr><td>a</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>$P(M=a)$</td><td>$\frac{1}{36}$</td><td>$\frac{3}{36}$</td><td>$\frac{5}{36}$</td><td>$\frac{7}{36}$</td><td>$\frac{9}{36}$</td><td>$\frac{11}{36}$</td></tr></table>	a	1	2	3	4	5	6	$P(M=a)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$
	a	1	2	3	4	5	6																								
$P(M=a)$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$																									
a	1	2	3	4	5	6																									
$P(M=a)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$																									
c.	<table><tr><td>a</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>$P(M=a)$</td><td>$\frac{12}{36}$</td><td>$\frac{10}{36}$</td><td>$\frac{8}{36}$</td><td>$\frac{3}{36}$</td><td>$\frac{2}{36}$</td><td>$\frac{1}{36}$</td></tr></table>	a	1	2	3	4	5	6	$P(M=a)$	$\frac{12}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	d.	<table><tr><td>a</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>$P(M=a)$</td><td>$\frac{11}{36}$</td><td>$\frac{9}{36}$</td><td>$\frac{7}{36}$</td><td>$\frac{5}{36}$</td><td>$\frac{3}{36}$</td><td>$\frac{1}{36}$</td></tr></table>	a	1	2	3	4	5	6	$P(M=a)$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$
	a	1	2	3	4	5	6																								
$P(M=a)$	$\frac{12}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$																									
a	1	2	3	4	5	6																									
$P(M=a)$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$																									
e.	<table><tr><td>a</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>$P(M=a)$</td><td>$\frac{11}{36}$</td><td>$\frac{6}{36}$</td><td>$\frac{1}{36}$</td><td>$\frac{1}{36}$</td><td>$\frac{6}{36}$</td><td>$\frac{11}{36}$</td></tr></table>	a	1	2	3	4	5	6	$P(M=a)$	$\frac{11}{36}$	$\frac{6}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$	$\frac{11}{36}$	f.	<table><tr><td>a</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td></tr><tr><td>$P(M=a)$</td><td>$\frac{1}{36}$</td><td>$\frac{6}{36}$</td><td>$\frac{11}{36}$</td><td>$\frac{11}{36}$</td><td>$\frac{6}{36}$</td><td>$\frac{1}{36}$</td></tr></table>	a	1	2	3	4	5	6	$P(M=a)$	$\frac{1}{36}$	$\frac{6}{36}$	$\frac{11}{36}$	$\frac{11}{36}$	$\frac{6}{36}$	$\frac{1}{36}$
	a	1	2	3	4	5	6																								
$P(M=a)$	$\frac{11}{36}$	$\frac{6}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{6}{36}$	$\frac{11}{36}$																									
a	1	2	3	4	5	6																									
$P(M=a)$	$\frac{1}{36}$	$\frac{6}{36}$	$\frac{11}{36}$	$\frac{11}{36}$	$\frac{6}{36}$	$\frac{1}{36}$																									

3. Bereken de verwachting van een stochast X met een kansverdeling gegeven door

a	0	3	4	6
$P(X = a)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$

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

4. Gegeven is een stochastische variabele met kansdichtheid

$$f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ 2 - x & 1 \leq x \leq 2 \\ 0 & \text{anders.} \end{cases}$$

en verdelingsfunctie F . Dan is $F(0.5)$ gelijk aan

a. 0.125 b. 0.150 c. 0.167 d. 0.250 e. 0.333 f. 0.500

5. Drie vrienden gaan een avondje naar het casino. Ze spelen daar een spel met kans $1/4$ om te winnen. Nadat ze alle drie dit spel drie keer hebben gespeeld, wat is de kans dat precies twee van de drie vrienden geen enkele keer het spel hebben gewonnen?

a. 0.052 b. 0.103 c. 0.206 d. 0.309 e. 0.412 f. 0.515

6. En kansspel heeft kans 0.25 op succes. Geert speelt het spel net zo lang tot hij een keer succes heeft gehad. Wat is de verwachting van het aantal spellen dat hij moet spelen?
- a. 2 b. 4 c. 5 d. 8 e. 10 f. 12
7. Een hotelbedrijf heeft bij aardewerk dat in gebruik is geconstateerd dat de kans, dat een bord langer dan 2 jaar heel blijft, gelijk is aan $\frac{1}{2}$. We definiëren X als de levensduur van deze borden, en we veronderstellen dat X een exponentiële verdeling heeft. Bereken $E[X]$ in jaren.
- a. 0.35 b. 0.48 c. 2.00 d. 2.73 e. 2.89 f. 3.00
8. Zij $X_1, X_2, \dots, X_n$ een steekproef van positieve getallen uit een verdeling met verwachting λ en variantie λ , waarbij $\lambda > 0$ een onbekende parameter is. Een schatter voor λ wordt gegeven door $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. De onzuiverheid (bias), en MSE (mean square error) van deze schatter worden gegeven door:
- a. bias = 0, MSE = λ^2/n^2 b. bias = $1/(n\lambda)$, MSE = $2/(n\lambda^2)$
c. bias = λ , MSE = $2/(n\lambda) + \lambda^2$ d. bias = 0, MSE = λ/n
e. bias = $1/(n\lambda)$, MSE = λ/n f. bias = λ , MSE = λ^2/n^2
9. De mediaan van een $Par(1)$ verdeling is gelijk aan
- a. -3 b. 0 c. $\frac{1}{2}$ d. 1 e. 2 f. 4
10. De stochast B is de belasting op een betonnen ligger van een brug. Stel B heeft een normale verdeling met verwachting $\mu_B = 50$ kN en een standaardafwijking $\sigma_B = 8$ kN. De stochast S is de sterkte van de betonnen ligger. Stel S heeft een normale verdeling met verwachting $\mu_S = 60$ kN en een standaardafwijking $\sigma_S = 6$ kN. Gegeven is dat B en S ongecorreleerd zijn. Bereken de kans dat S kleiner is dan B , dus $P(S < B)$. U mag gebruiken dat $S - B$ een normale verdeling heeft.
- a. 0.08 b. 0.92 c. 0.24 d. 0.76 e. 0.84 f. 0.16
11. Gegeven is een stochastische variabele X met $E[X] = 2$, $\text{Var}(X) = 4$. Dan worden de verwachting en variantie van $Y = 3 - 2X$ gegeven door
- a. $E[Y] = -1$ en $\text{Var}(Y) = 5$ b. $E[Y] = -1$ en $\text{Var}(Y) = 8$
c. $E[Y] = -1$ en $\text{Var}(Y) = 16$ d. $E[Y] = -1$ en $\text{Var}(Y) = 19$
e. $E[Y] = 8$ en $\text{Var}(Y) = 5$ f. $E[Y] = 8$ en $\text{Var}(Y) = 8$
12. Stel we hebben een steekproef X_1, X_2 uit een verdeling met onbekende verwachting μ en onbekende variantie σ^2 . Vervolgens doen we een andere meting Y , met verwachting 2μ (let op de factor 2!) en variantie σ^2 , onafhankelijk van de eerste steekproef. We beschouwen een zuivere schatter T voor μ van de volgende vorm:

$$T = aX_1 + aX_2 + \left(\frac{1}{2} - a\right)Y,$$

met $a \in \mathbb{R}$. Voor welke waarde van a is de variantie van T minimaal?

- a. $a = \frac{1}{4}$ b. $a = 0$ c. $a = \frac{1}{6}$ d. $a = \frac{1}{3}$ e. $a = \frac{1}{2}$ f. $a = \frac{3}{8}$

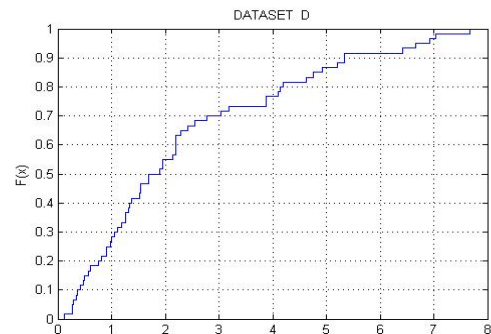
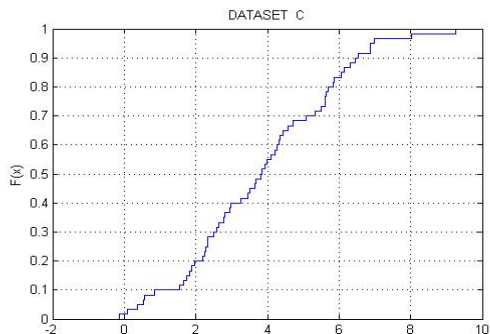
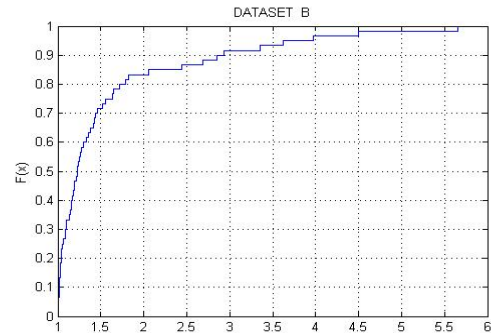
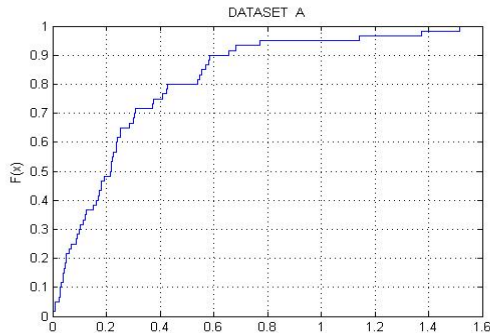
13. Gegeven is de volgende geordende dataset:

0	0	0	2	4	6	8	9	10	10
10	12	15	15	16	21	22	24	26	30
30	31	33	36	44	50	55	58	65	68
75	77	79	81	88	91	97	100	108	108
112	113	114	115	120	122	129	134	138	143
148	160	176	180	193	193	197	227	232	233
236	242	245	255	261	263	281	290	296	300
300	325	330	357	365	369	371	379	386	422
445	446	447	452	457	482	529	529	543	600
648	670	700	707	724	729	748	790	810	816
828	843	860	865	868	875	943	948	983	990
1011	1045	1064	1071	1082	1146	1160	1222	1247	1351
1435	1461	1755	1783	1800	1864	1897	2323	2930	3110
3321	4116	5485	5509	6150					

De hoogte van het histogram van deze data bij de cel $(500, 1000]$ is gelijk aan

- a. 0.00036 b. 0.00163 c. 0.04800 d. 0.17778 e. 0.27000 f. 0.81481

14. Van vier datasets zijn de empirische verdelingfuncties weergegeven. De datasets zijn afkomstig uit vier verschillende verdelingen: een $Exp(3)$, een $Exp(1/3)$, een $Par(3)$ en een $N(4, 4)$ verdeling. Welke dataset komt uit de $Exp(3)$ en welke dataset komt uit de $Exp(1/3)$ verdeling?



- a. $Exp(3)$: A; $Exp(1/3)$: C b. $Exp(3)$: A; $Exp(1/3)$: D
c. $Exp(3)$: B; $Exp(1/3)$: A d. $Exp(3)$: B; $Exp(1/3)$: D
e. $Exp(3)$: D; $Exp(1/3)$: A f. $Exp(3)$: D; $Exp(1/3)$: C

15. We construeren een histogram met cellen $[0,1]$, $(1,3]$, $(3,5]$, $(5,8]$, $(8,11]$, $(11,14]$, en $(14,18]$. Gegeven zijn de waarden van de empirische verdelingsfunctie in de randen van de cellen:

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

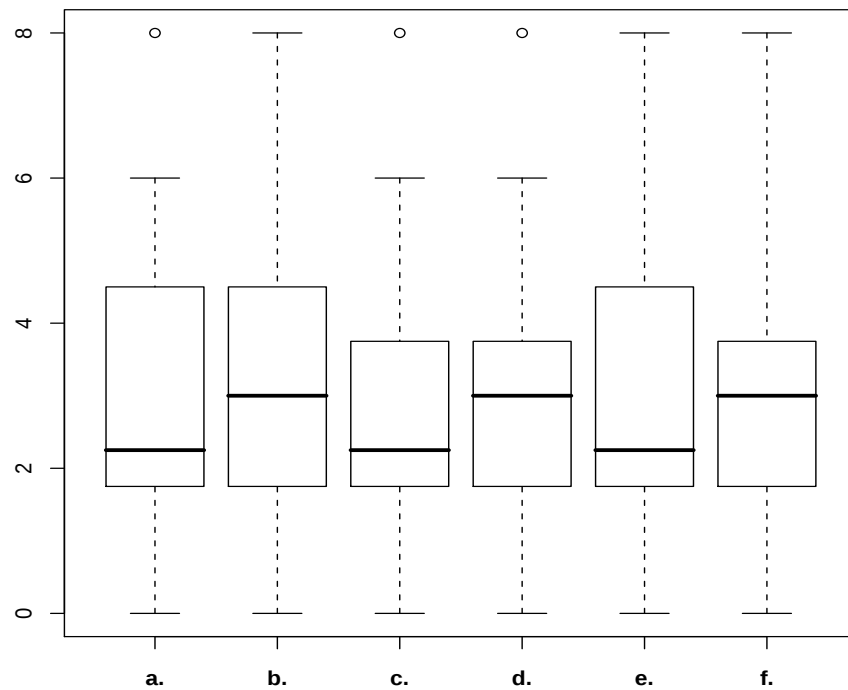
Dan is de hoogte van het histogram op de cel $(8, 11]$ gelijk aan

- a. 0.0233 b. 0.0035 c. 0.0700 d. 0.2100 e. 0.2450 f. 0.2683

16. Gegeven is de dataset

0.00 1.75 1.75 1.75 2.00 2.50 4.50 4.50 6.00 8.00

Welke van de zes onderstaande boxplots hoort bij deze dataset?



17. Aan een tentamen met 20 meerkeuze vragen, elk met 6 mogelijkheden, doen 30 studenten mee. Alle deelnemende studenten beantwoorden de vragen onafhankelijk van elkaar en loten bij elke vraag uit de 6 mogelijkheden met gelijke kans. Zij X het aantal studenten dat minstens 1 vraag goed beantwoord heeft. We modelleren X als een stochastische variabele met een binomiale verdeling met n en p gegeven door

- a. $n = 20$, $p = 1 - (\frac{5}{6})^{30}$ b. $n = 20$, $p = (\frac{1}{6})^{30}$
c. $n = 20$, $p = (\frac{5}{6})^{30}$ d. $n = 30$, $p = 1 - (\frac{5}{6})^{20}$
e. $n = 30$, $p = (\frac{1}{6})^{20}$ f. $n = 30$, $p = (\frac{5}{6})^{20}$

18. Gegeven is een bivariate dataset $(x_1, y_1), (x_2, y_2), \dots, (x_{50}, y_{50})$, waarvan bekend is dat

$$\sum_{i=1}^{50} x_i = 150 \quad \text{en} \quad \sum_{i=1}^{50} y_i = 400.$$

De kleinste kwadraten schatting voor de helling van de regressielijn is $\hat{\beta} = 2$. De kleinste kwadraten schatting $\hat{\alpha}$ voor het intercept is

- a. 0 b. 2 c. 14 d. 100 e. 700 f. niet te berekenen

19. Voor het aantal klanten X_1 dat een winkel binnen loopt tussen 14:00 en 15:00 geldt een Poisson verdeling met parameter $\lambda > 0$. Dat wil zeggen:

$$P(X_1 = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \text{voor } k = 0, 1, 2, \dots$$

Het aantal klanten dat binnen loopt tussen 15:00 en 16:00, noem dit X_2 , heeft dezelfde verdeling, en is onafhankelijk van X_1 . De winkelbediende weet dat er tussen 14:00 en 16:00 uur 1 klant is binnengekomen. Wat is de maximum likelihood schatter voor λ , gebaseerd op dit gegeven?

- a. $\hat{\lambda} = 0.25$ b. $\hat{\lambda} = 0.50$ c. $\hat{\lambda} = 1.0$ d. $\hat{\lambda} = 1.5$ e. $\hat{\lambda} = 2.0$ f. $\hat{\lambda} = 4.0$

20. Piet en Marijke spelen een spel op de computer. Zij hebben beiden kans p om het spel te winnen van de computer. Piet speelt drie keer, en wint 1 keer. Marijke speelt net zolang totdat ze een keer wint; dit gebeurt de vierde keer dat ze speelt. Geef de maximum likelihoodschatter voor p , gebaseerd op deze gegevens.

- a. $\hat{p} = \frac{2}{7}$ b. $\hat{p} = \frac{3}{7}$ c. $\hat{p} = \frac{1}{3}$ d. $\hat{p} = \frac{2}{3}$ e. $\hat{p} = \frac{1}{5}$ f. $\hat{p} = \frac{2}{5}$

21. Een fabrikant van verf wenst de gemiddelde droogtijd van een nieuwe type muurverf te bepalen. Voor 12 testmuren van dezelfde oppervlakte vond hij een gemiddelde droogtijd van 66.3 minuten en een steekproefvariantie van 8.4. De 12 gemeten droogtijden vat men op als een realisatie van een steekproef uit een $N(\mu, \sigma^2)$ verdeling. Het 90% betrouwbaarheidsinterval voor de verwachte droogtijd μ wordt gegeven door

- a. (61.944 , 70.655) b. (61.978 , 70.621) c. (64.730 , 67.869)
d. (64.742 , 67.857) e. (64.797 , 67.803) f. (64.809 , 67.790)

22. Serrano ham wordt in een winkel verkocht in verpakkingen van 100 gram. De manager vreest echter dat het personeel systematisch iets te veel ham in de verpakkingen stopt. Hij meet de hoeveelheid Serrano ham in $n = 12$ verpakkingen. Noem de uitkomsten hiervan $x_1, \dots, x_{12}$. Gegeven is dat

$$\bar{x}_n = 101.03 \text{ g} \quad \text{en} \quad s_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = 4.00 \text{ g}^2.$$

De manager gaat er vanuit dat het gewicht van de ham in een verpakking verdeeld is volgens $N(\mu, \sigma^2)$. We toetsen m.b.v. de t-toets de nul-hypothese $H_0 : \mu = 100$ tegen het alternatief $H_1 : \mu > 100$, bij een significantieniveau van 5%. Noem p de p-waarde van deze toets. Welke van de volgende zes conclusies is de juiste?

- a. $0.025 < p < 0.05$, verwerp H_0 . b. $0.025 < p < 0.05$, verwerp H_0 niet.
c. $0.05 < p < 0.10$, verwerp H_0 . d. $0.05 < p < 0.10$, verwerp H_0 niet.
e. $p > 0.10$, verwerp H_0 . f. $p > 0.10$, verwerp H_0 niet.

23. Iemand voert een toets uit, gebruikmakend van een toetsingsgrootheid T , die waarden heeft tussen 0 en 10. Stel dat hij besluit H_0 te verwerpen ten gunste van H_1 als $T \leq c$. Verder weten we dat onder de nulhypothese geldt: $P(T \leq c) = (c/10)^2$. Hoe groot (tot op drie decimalen) moet hij c nemen als voor het significantieniveau de waarde $\alpha = 0.10$ gekozen wordt?

- a. 1.414 b. 2.173 c. 3.162 d. 4.225 e. 6.666 f. 8.100

24. Vijf maanden achter elkaar wordt de gemiddelde verkoopprijs van huizen bepaald in Delft. Een makelaar vat deze data op als een bivariate dataset $(x_1, y_1), \dots, (x_5, y_5)$, waarbij $x_1, \dots, x_5$ de vijf tijdstippen zijn van meten (hij neemt simpelweg $x_1 = 1, x_2 = 2, \dots, x_5 = 5$), en $y_1, \dots, y_5$ de gemiddelde verkoopprijs van huizen in die repectievelijke maanden. Hij gaat uit van een lineair regressie model voor deze data:

$$Y_i = \alpha + \beta x_i + U_i, \quad \text{voor } i = 1, 2, \dots, 5,$$

waarbij $U_1, U_2, \dots, U_5$ een steekproef is uit een $N(0, \sigma^2)$ verdeling. Hij vindt als kleinste kwadratenschatter

$$\hat{\beta} = -1500 \quad \text{en} \quad S_b^2 = 4.65 \cdot 10^5.$$

Hij wil laten zien dat de huizenprijs daalt door de nul-hypothese $H_0 : \beta = 0$ te toetsen tegen het alternatief $H_1 : \beta < 0$ met behulp van een t-toets, bij een significantieniveau van 5%. Geef de toetsingsgrootheid T en uw conclusie aangaande de nul-hypothese.

- | | |
|-------------------------------------|-------------------------------------|
| a. $T = -3.5$, verwerp H_0 . | b. $T = -4.9$, verwerp H_0 . |
| c. $T = -2.2$, verwerp H_0 . | d. $T = -3.5$, verwerp H_0 niet. |
| e. $T = -4.9$, verwerp H_0 niet. | f. $T = -2.2$, verwerp H_0 niet. |

25. Van een mannenkoor in New York wordt elk lid gevraagd zijn lengte in inches op te geven. Het koor bestaat uit $n = 39$ bassen en $m = 20$ tenoren. De lengtes $X_1, X_2, \dots, X_{39}$ van de bassen worden gemodelleerd met een normale verdeling $N(\mu_1, \sigma_X^2)$. De lengtes $Y_1, Y_2, \dots, Y_{20}$ van de tenoren worden gemodelleerd met een normale verdeling $N(\mu_2, \sigma_Y^2)$. We willen aantonen dat bassen in verwachting langer zijn dan tenoren.

Als $x_1, \dots, x_{39}$ de opgegeven lengtes van de bassen zijn, en $y_1, \dots, y_{20}$ de opgegeven lengtes van de tenoren, dan geldt

$$\bar{x}_n = 70.72 \text{ inch}, S_X^2 = 5.58 \text{ inch}^2, \quad \bar{y}_m = 69.15 \text{ inch}, S_Y^2 = 10.3 \text{ inch}^2.$$

Als nul-hypothese nemen we $H_0 : \mu_1 = \mu_2$. We baseren onze toets op de t-toets toetsingsgrootheid T met niet-gepoolde variantie:

$$T = \frac{\bar{x}_n - \bar{y}_m}{S_d}, \quad \text{met } S_d^2 = \frac{1}{n} S_X^2 + \frac{1}{m} S_Y^2.$$

Geef de alternatieve hypothese H_1 die van toepassing is op dit probleem, en geef uw conclusie aangaande de nulhypothese, als we een significantieniveau van 5% aanhouden.

- | | |
|--|---|
| a. $H_1 : \mu_1 \neq \mu_2$, verwerp H_0 . | b. $H_1 : \mu_1 > \mu_2$, verwerp H_0 . |
| c. $H_1 : \mu_1 < \mu_2$, verwerp H_0 . | d. $H_1 : \mu_1 \neq \mu_2$, verwerp H_0 niet. |
| e. $H_1 : \mu_1 > \mu_2$, verwerp H_0 niet. | f. $H_1 : \mu_1 < \mu_2$, verwerp H_0 niet. |

Antwoorden multiple choice:

1 a.

2 b.

3 e.

4 a.

5 d.

6 b.

7 e.

8 d.

9 e.

10 f.

11 c.

12 c.

13 a.

14 b.

15 a.

16 e.

17 d.

18 b.

19 b.

20 a.

21 e.

22 d.

23 c.

24 f.

25 b.

Uitwerkingen multiple choice:

1 a. Het aantal boekenkasten met te weinig schroeven kan gemodelleerd worden door een stochast X met een $Bin(n, p)$ met $n = 1000$ en $p = 0.001$. Zodoende is de gevraagde kans gelijk aan

$$P(X = 2) = \binom{1000}{2} \times (0.001)^2 \times (0.999)^{998} = 0.184.$$

2 b. Noteer X_1 en X_2 als de uitkomsten van dobbelsteen 1 en 2. Dan is

$$P(M = 6) = P(\text{minstens één } X_i = 6) = 1 - P(X_1 \neq 6, X_2 \neq 6) = 1 - \frac{5}{6} \cdot \frac{5}{6} = \frac{11}{36}$$

Dan zijn alleen nog **b** en **e** over als mogelijkheden. Veder is

$$P(M = 1) = P(X_1 = 1, X_2 = 1) = \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36},$$

zodat **b** het juiste antwoord is. Voor de volledigheid:

$$P(M = 5) = P((5, 5), (4, 5), (3, 5), (2, 5), (1, 5), (5, 4), (5, 3), (5, 2), (5, 1)) = \frac{9}{36}$$

$$P(M = 4) = P((4, 4), (3, 4), (2, 4), (1, 4), (4, 3), (4, 2), (4, 1)) = \frac{7}{36}$$

$$P(M = 3) = P((3, 3), (2, 3), (3, 2), (1, 3), (3, 1)) = \frac{5}{36}$$

$$P(M = 2) = P((1, 2), (2, 1), (2, 2)) = \frac{3}{36}.$$

3 e.

$$E[X] = 0 \cdot \frac{1}{4} + 3 \cdot \frac{1}{2} + 4 \cdot \frac{1}{8} + 6 \cdot \frac{1}{8} = 2\frac{3}{4}.$$

4 a. $F(0.5)$ is de oppervlakte onder f op het interval $(-\infty, 0.5]$. Omdat de onderhavige dichtheid een driehoek vormt, is eenvoudig in te zien dat deze oppervlakte gelijk is aan $\frac{1}{2} \times 0.5 \times 0.5 = 0.125$. Een formele berekening levert uiteraard hetzelfde

$$F(0.5) = \int_{-\infty}^{0.5} f(x) dx = \int_0^{0.5} x dx = \left[\frac{1}{2} x^2 \right]_0^{0.5} = 0.125.$$

5 d. De kans dat vriend nummer 1 geen een keer wint, is gelijk aan $p = (1 - \frac{1}{4})^3 = 0.422$. De kans dat precies twee van de drie vrienden geen een keer wint, is dus de binomiale kans

$$P(Bin(3, p) = 2) = \binom{3}{2} p^2 (1 - p)^1 = 0.309.$$

6 b. Het aantal rondes heeft een $Geo(0.25)$ -verdeling met verwachting $1/0.25 = 4$.

7 e. Zie opgave 5.11. De mediaan $q_{0.50}$ van een $Exp(\lambda)$ verdeling is de oplossing van de vergelijking

$$F(q) = 0.5 \Leftrightarrow 1 - e^{-\lambda q} = 0.5 \Leftrightarrow e^{-\lambda q} = 0.5 \Leftrightarrow q = \frac{-\ln(0.5)}{\lambda} = \frac{\ln(2)}{\lambda}.$$

Hier geldt dat $q = 2$, dus $\lambda = \ln(2)/q = 0.3466$. De verwachting van een $Exp(\lambda)$ verdeling is gelijk aan $1/\lambda = 2.89$.

8 d. Bekend is dat $E[\bar{X}_n] = E[X_1] = \lambda$, en $\text{Var}(\bar{X}_n) = \frac{\text{Var}(X_1)}{n} = \lambda/n$. De schatter is zuiver zodat de bias gelijk is aan 0. Omdat de schatter zuiver is, is de MSE gelijk aan de variantie van de schatter. Het goede antwoord is dus

$$\text{bias} = 0, \quad \text{MSE} = \lambda/n.$$

9 e. Zie huiswerk opgave 5.12. De verdelingsfunctie van de $\text{Par}(1)$ is $F(x) = 1 - x^{-1}$. Om de mediaan te vinden moeten we oplossen

$$F(x) = \frac{1}{2} \Leftrightarrow 1 - \frac{1}{x} = \frac{1}{2} \Leftrightarrow x = 2.$$

10 f. Definieer $W = S - B$. Dan geldt dat $E[W] = E[S] - E[B] = 10 \text{ kN}$, en $\text{Var}(W) = \text{Var}(S) + \text{Var}(B) = 6^2 + 8^2 = 100 \text{ (kN)}^2$, dus W heeft een $N(10, 10^2)$ verdeling. Er geldt dus

$$P(S < B) = P(W < 0) = P\left(\frac{W - 10}{10} < \frac{0 - 10}{10}\right) = P(Z < -1) = P(Z > 1) = 0.1587.$$

Hier is Z een standaard normaal verdeelde stochast.

11 c. Zie huiswerk opgave 7.4. Volgens de lineariteit van verwachting is

$$E[3 - 2X] = 3 - 2 \times E[X] = 3 - 2 \times 2 = -1.$$

Volgens de regel voor de variantie

$$\text{Var}(3 - 2X) = (-2)^2 \times \text{Var}(X) = 4 \times 4 = 16.$$

12 c. Een schatter van de vorm $T = aX_1 + aX_2 + (\frac{1}{2} - a)Y_1$ is zuiver, d.w.z.

$$E[T] = aE[X_1] + aE[X_2] + (\frac{1}{2} - a)E[Y] = (2a + 2(\frac{1}{2} - a))\mu = \mu \text{ voor elke } \mu \in \mathbb{R}.$$

Verder geldt:

$$\text{Var}(T) = \text{Var}(aX_1) + \text{Var}(aX_2) + \text{Var}\left(\left(\frac{1}{2} - a\right)Y_1\right) = (2a^2 + (\frac{1}{2} - a)^2)\sigma^2.$$

Dit is minimaal als

$$4a - 2(\frac{1}{2} - a) = 0 \Leftrightarrow a = \frac{1}{6}.$$

Dit laat zien dat het antwoord **c.** moet zijn. U kunt natuurlijk ook alle zes de varianties uitrekenen:

$$\text{a. } \frac{3}{16}\sigma^2 \quad \text{b. } \frac{1}{4}\sigma^2 \quad \text{c. } \frac{3}{18}\sigma^2 \quad \text{d. } \frac{9}{36}\sigma^2 \quad \text{e. } \frac{1}{2}\sigma^2 \quad \text{f. } \frac{19}{64}\sigma^2$$

13 a. Zie huiswerk opgave 15.4. De hoogte van het histogram op cel $(500, 1000]$ is

$$\frac{\text{aantal elementen in } (500, 1000]}{135 \cdot 500} = 0.00036$$

14 b. De Paretovariabele neemt waarden ≥ 1 aan; dat hoort bij plaatje B.

Plaatje C valt af door de negatieve waarde die daar optreedt.

De $\text{Exp}(1/3)$ verdeling neemt in het algemeen grotere waarden aan dan de $\text{Exp}(3)$ -verdeling (de verwachting van een $\text{Exp}(\lambda)$ -verdeling is $1/\lambda$), dus het eerste plaatje zal bij de $\text{Exp}(3)$ dataset horen. Een andere mogelijkheid is om te kijken naar de empirische mediaan van de verdelingen: deze is te vinden door bij $F(x) = 0.5$ op de verticale as naar rechts te gaan tot de grafiek, en dan naar beneden tot aan de horizontale as.

15 a. Zie huiswerk opgave 15.5. De hoogte op de cel $(8, 11]$ is gelijk aan

$$\begin{aligned} \frac{\text{aantal } x_i \text{ in } (8, 11]}{(11 - 8)n} &= \frac{\text{aantal } x_i \leq 11 - \text{aantal } x_i \leq 8}{3n} \\ &= \frac{\text{aantal } x_i \leq 11}{3n} - \frac{\text{aantal } x_i \leq 8}{3n} \\ &= \frac{1}{3} \left(\frac{\text{aantal } x_i \leq 11}{n} - \frac{\text{aantal } x_i \leq 8}{n} \right) \\ &= \frac{1}{3} (F_n(11) - F_n(8)) = \frac{1}{3} (0.805 - 0.735) = 0.0233. \end{aligned}$$

16 e. We bereken eerst de mediaan: het 5de en 6de getal (van de geordende dataset met 10 punten) is 2 en 2.5, dus de mediaan is 2.25. Dan zijn alleen **a**, **c** en **e** nog over. Vervolgens berekenen we het derde kwartiel: $\frac{3}{4} \cdot 10 = 7.5$ (volgens het boek: $\frac{3}{4} \cdot (10 - 1) + 1 = 7.75$), en het 7de en 8ste getal zijn beide 4.5, dus het derde kwartiel is 4.5. Dan blijven alleen **a** en **e** over. Tot slot berekenen we de Inter Quartile Range: het onderste kwartiel is 1.75 (de waarde van het tweede, derde en vierde getal), dus de IQR = 2.75. De omhooggaande "whisker" heeft maximale lengte $1.5 \cdot \text{IQR} = 4.125$, en $4.5 + 4.125 > 8$. Aangezien 8 het grootste getal is in de dataset, gaat de whisker tot aan 8. Dit leidt tot antwoord **e**.

17 d. X is het aantal studenten met 'succes', dus $n = 30$. Succes betekent minstens 1 vraag goed, dus is $p = 1 - P(\text{alle 20 vragen fout}) = 1 - (\frac{5}{6})^{20}$.

18 b. Zie ook *Quick Exercise 22.2*. Er geldt dat $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$. We hebben $\bar{x} = 150/50 = 3$ en $\bar{y} = 400/50 = 8$, dus $\hat{\alpha} = 8 - 2 \cdot 3 = 2$.

19 b. We weten dat X_1 en X_2 een Poisson verdeling hebben met parameter λ . Gegeven is dat $X_1 + X_2 = 1$, dus ofwel $X_1 = 0$ en $X_2 = 1$, ofwel $X_1 = 1$ en $X_2 = 0$. De kans $L(\lambda)$ op de data, gegeven de parameter λ , is dus:

$$L(\lambda) = P(X_1 = 0) \cdot P(X_2 = 1) + P(X_1 = 1) \cdot P(X_2 = 0) = e^{-\lambda} \cdot \lambda e^{-\lambda} + \lambda e^{-\lambda} \cdot e^{-\lambda} = 2\lambda e^{-2\lambda}.$$

We zoeken de lokatie van het maximum van $L(\lambda)$, en kunnen dit doen door de lokatie van het maximum van

$$l(\lambda) = \log(L(\lambda)) = \log(2) + \log(\lambda) - 2\lambda$$

te vinden. Aangezien

$$l'(\lambda) = 0 \Leftrightarrow \frac{1}{\lambda} - 2 = 0 \Leftrightarrow \lambda = 0.5,$$

vinden we als maximum likelihoodschatter $\lambda = 0.5$.

20 a. De kans $L(p)$ op de data is gegeven door

$$L(p) = P(\text{Piet wint 1 van de 3 keer}) \cdot P(\text{Marijke wint de 4de keer}) = 3p(1-p)^2(1-p)^3p = 3p^2(1-p)^5.$$

Definieer $l(p) = \log(L(p)) = \log(3) + 2\log(p) + 5\log(1-p)$. Vind het maximum van l :

$$l'(p) = 0 \Leftrightarrow \frac{2}{p} - \frac{5}{1-p} = 0 \Leftrightarrow 2(1-p) = 5p \Leftrightarrow p = \frac{2}{7}.$$

Dus $\hat{p} = 2/7$.

21 e. Het 90% betrouwbaarheidsinterval voor de verwachte droogtijd μ wordt gegeven door

$$\left(\bar{x}_n - t_{n-1, \alpha/2} \frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{n-1, \alpha/2} \frac{s_n}{\sqrt{n}} \right).$$

Hierbij is $t_{n-1, \alpha/2} = t_{11, 0.05} = 1.796$ en $s_n = \sqrt{8.4} = 2.898$, zodat het 90% betrouwbaarheidsinterval wordt gegeven door

$$\left(66.3 - 1.796 \frac{2.898}{\sqrt{12}}, 66.3 + 1.796 \frac{2.898}{\sqrt{12}} \right) = (64.797, 67.803).$$

22 d. De toetsingsgrootte wordt gegeven door

$$T = \frac{\bar{x}_n - 100}{s_n / \sqrt{n}} = 1.784.$$

Er geldt dat $t_{11, 0.10} < T < t_{11, 0.05}$ (merk op dat T een t-verdeling heeft met $n - 1 = 11$ vrijheidsgraden!). Aangezien alleen grote waarden van T wijzen op de alternatieve hypothese H_1 , geldt dus dat $0.05 < p < 0.10$. Aangezien p groter is dan het significantieniveau, verwerpen we H_0 niet.

23 c. Onder H_0 moet gelden: $P(T \leq c) \leq 0.1$. De grootste c waarvoor dit het geval is voldoet aan $\frac{c}{10} = \sqrt{0.1}$, en dat is $c = 10\sqrt{0.1} \approx 3.162$.

24 f. De toetsingsgrootte T wordt gegeven door

$$T = \frac{\hat{\beta} - 0}{S_b} = -2.2.$$

Onder de nul-hypothese geldt dat T een t-verdeling heeft met $n - 2 = 3$ vrijheidsgraden! Aangezien alleen sterk negatieve waarden van T wijzen op H_1 , en aangezien $-t_{3, 0.05} < -2.2$, geldt dat de p-waarde groter is dan 0.05, en dus verwerpen we H_0 niet.

25 b. We merken op dat de twee varianties behoorlijk verschillen, dus we werken niet met de pooled variance. De toetsingsgrootte wordt

$$T = \frac{\bar{x}_n - \bar{y}_m}{S_d}, \quad \text{met} \quad S_d^2 = \frac{S_x^2}{n} + \frac{S_y^2}{m}.$$

Uitrekenen geeft $T = 1.93$. Merk op dat alleen grote waarden van T wijzen in de richting van de alternatieve hypothese, dus we verwerpen als $T > t_{n+m-2, 0.05}$. Er geldt dat

$$t_{57, 0.05} \leq t_{50, 0.05} = 1.676,$$

dus $1.93 = T > t_{57, 0.05}$, en we verwerpen H_0 .

Tentamen Statistiek
wi1102CT, 28 augustus 2012

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

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. Aad, Bas en Co willen een zuivere munt gebruiken om op een eerlijke manier een van hun drieën aan te wijzen om de vaat te gaan doen. Zij bedenken drie mogelijke manieren:

- I Ze gooien twee keer met de munt. Als er eerst kop wordt gegooid, en dan munt, dan kiezen ze Aad. Als er eerst munt wordt gegooid, en dan kop, dan kiezen ze Bas. Als er twee keer kop wordt gegooid, dan kiezen ze Co. Als er twee keer munt wordt gegooid, dan spelen ze het spel opnieuw.
- II Ze gooien drie keer met de munt. Bij twee keer kop kiezen ze Aad, bij twee keer munt kiezen ze Bas, en anders kiezen ze Co.
- III Ieder gooit net zo lang met de munt totdat hij kop gooit. Degene die er het langst over doet, moet de afwas doen. Als twee of meer personen er het langst over doen, dan spelen zij het spel opnieuw, totdat er iemand gekozen is.

Welke van deze drie manieren zijn eerlijk, dwz geeft kans $1/3$ aan zowel Aad, Bas als Co?

- a. Manier I en II zijn eerlijk, manier III niet
 - b. Manier I en III zijn eerlijk, manier II niet
 - c. Manier II en III zijn eerlijk, manier I niet
 - d. Manier I is eerlijk, manier II en III niet
 - e. Manier II is eerlijk, manier I en III niet
 - f. Manier III is eerlijk, manier I en II niet
2. Wat is de kans dat de som van het aantal ogen gegooid met drie zuivere zeszijdige dobbelstenen gelijk is aan 6?
- a. $3/216$ b. $7/216$ c. $10/216$ d. $13/216$ e. $18/216$ f. $24/36$
3. Iemand gooit met een dobbelsteen net zolang totdat hij zes gooit; het aantal worpen noemen we X_1 . Daarna gooit hij weer net zolang totdat hij een zes gooit; dit aantal worpen noemen we X_2 . Het totale aantal worpen noemen we X_3 , dus $X_3 = X_1 + X_2$. Welke van de drie stochasten heeft een geometrische verdeling?
- a. alle drie
 - b. alleen X_1
 - c. alleen X_1 en X_3
 - d. alleen X_1 en X_2
 - e. alleen X_3
 - f. geen van de drie
4. Vervolg vorige vraag: Welke van de volgende uitspraken is waar?
- a. X_1 en X_3 zijn onafhankelijk
 - b. X_1 en X_2 zijn negatief gecorreleerd
 - c. X_1 en X_2 zijn positief gecorreleerd
 - d. X_1 en X_3 zijn negatief gecorreleerd
 - e. X_2 en X_3 zijn onafhankelijk
 - f. X_2 en X_3 zijn positief gecorreleerd

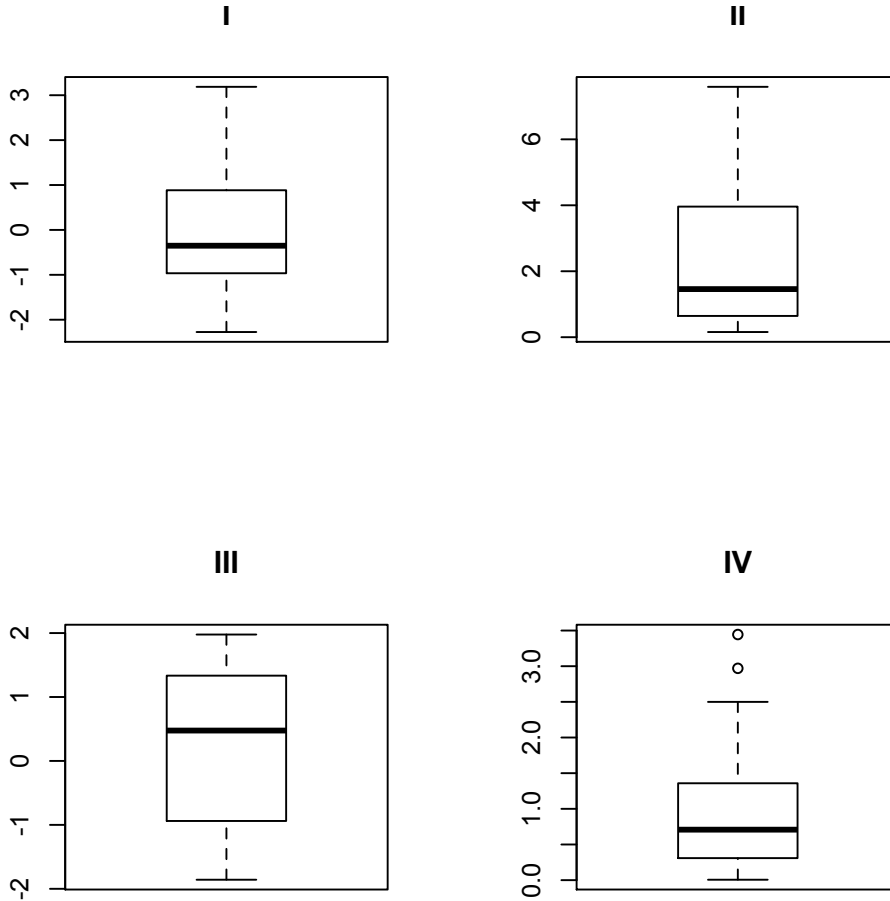
5. In een hockeytoernooi worden 8 partijen gespeeld. Elke partij heeft 15% kans om in gelijk spel te eindigen, onafhankelijk van de andere uitkomsten. Wat is de kans dat er 6 of meer gelijke spelen zijn?
- a. $2.4 \cdot 10^{-4}$ b. $3.0 \cdot 10^{-4}$ c. $5.4 \cdot 10^{-3}$ d. $8.2 \cdot 10^{-3}$ e. $1.2 \cdot 10^{-2}$ f. $9.9 \cdot 10^{-2}$
6. Stel een random variable X neemt alleen de waarden 0 en 2 aan, en $P(X = 0) = 0.5$. Dan
- a. $\text{Var}(X) = 0$ b. $\text{Var}(X) = 0.5$ c. $\text{Var}(X) = 1$
d. $\text{Var}(X) = \sqrt{2}$ e. $\text{Var}(X) = 2$ f. er is niet genoeg informatie.
7. Stel X heeft de dichtheid $f(x) = \begin{cases} 0, & \text{als } x \notin (-1, 1) \\ \frac{1}{2}x + \frac{1}{2}, & \text{als } -1 \leq x \leq 1 \end{cases}$
De verdelingsfunctie van X is voor $-1 \leq x \leq 1$ gelijk aan
- a. 1 b. $\frac{1}{2}$ c. $\frac{1}{2}x + \frac{1}{2}$ d. $\frac{1}{2}x + \frac{1}{4}x^2$
e. $\frac{1}{2}x + \frac{1}{4}x^2 - \frac{1}{4}$ f. $\frac{1}{2}x + \frac{1}{4}x^2 + \frac{1}{4}$
8. De verwachting van de stochast X uit de vorige opgave is gelijk aan
- a. $-\frac{1}{3}$ b. 0 c. $\frac{1}{6}$ d. $\frac{1}{4}$ e. $\frac{1}{3}$ f. $\frac{1}{2}$
9. Een personenlift stopt als hij meer dan 750 kg moet vervoeren. Stel de personen die gebruik maken van de lift hebben een normaal verdeeld gewicht, met verwachting 72 kg en standaarddeviatie 6 kg, onafhankelijk van elkaar. Wat is de kans dat de lift stopt als er 10 personen in staan?
- a. 0.943 b. 0.057 c. 0.78 d. 0.22 e. 0.66 f. 0.34
10. De verdeling van een discrete stochast X wordt gegeven door

a	0	1	2
$P(X = a)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{5}$

Bepaal de verwachting en de variantie van X .

- a. $E[X] = 1$ en $\text{Var}(X) = \frac{14}{25}$ b. $E[X] = 1$ en $\text{Var}(X) = \frac{16}{25}$
c. $E[X] = 1$ en $\text{Var}(X) = \frac{6}{5}$ d. $E[X] = \frac{6}{5}$ en $\text{Var}(X) = \frac{14}{25}$
e. $E[X] = \frac{6}{5}$ en $\text{Var}(X) = \frac{16}{25}$ f. $E[X] = \frac{6}{5}$ en $\text{Var}(X) = \frac{6}{5}$
11. We modelleren de stochastische variabelen $X_1, \dots, X_n$ als onafhankelijk en gelijk verdeeld met dichtheid
- $$f_c(x) = \begin{cases} 0 & x \notin [0, 1] \\ cx + 1 - \frac{1}{2}c & x \in [0, 1], \end{cases}$$
- waarbij $c \in [-2, 2]$ een onbekende parameter is. Welke van de volgende schatters is een zuivere schatter voor c ?
- a. $T = 12\bar{X}_n + 6$ b. $T = 12\bar{X}_n - 6$
c. $T = \bar{X}_n$ d. $T = 4\bar{X}_n - 2$
e. $T = 4\bar{X}_n + 2$ f. $T = 2\bar{X}_n$
12. Vervolg vorige vraag: stel we weten dat $X_1 = 0.63$. Wat is de Maximum Likelihood Schatter $\hat{c}$ voor de parameter $c \in [-2, 2]$?
- a. $\hat{c} = 0.63$ b. $\hat{c} = 1.26$ c. $\hat{c} = -2$ d. $\hat{c} = 2$ e. $\hat{c} = 0.52$ f. $\hat{c} = 1.56$

13. Van vier datasets ter grootte 50 is de boxplot weergegeven. De datasets zijn steekproeven uit vier verschillende verdelingen: een $Exp(1)$, een $Exp(1/3)$, een $N(0, 1)$ en een $U(-2, 2)$ verdeling. Welke dataset komt uit de $Exp(1/3)$ en welke dataset komt uit de $U(-2, 2)$ verdeling?

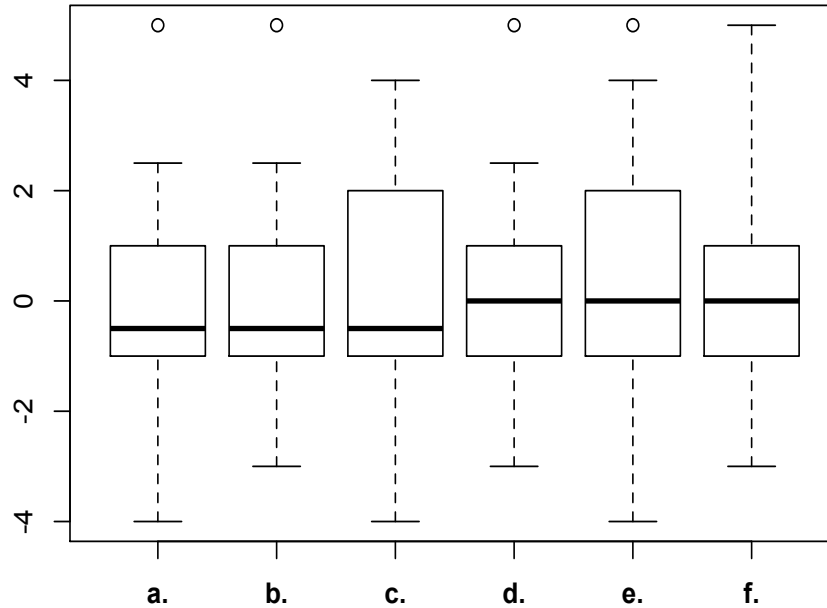


- a. $Exp(1/3)$: I; $U(-2, 2)$: III
b. $Exp(1/3)$: II; $U(-2, 2)$: III
c. $Exp(1/3)$: III; $U(-2, 2)$: I
d. $Exp(1/3)$: IV; $U(-2, 2)$: III
e. $Exp(1/3)$: IV; $U(-2, 2)$: II
f. $Exp(1/3)$: II; $U(-2, 2)$: IV
14. De correlatie tussen de twee stochasten X en Y is gelijk aan -0.5 . Verder geldt dat $Var(X) = 1$ en $Var(Y) = 4$. Wat is de variantie van $X + Y$?
- a. $Var(X + Y) = 2.5$ b. $Var(X + Y) = 5$ c. $Var(X + Y) = 4$
d. $Var(X + Y) = 6$ e. $Var(X + Y) = 3.5$ f. $Var(X + Y) = 3$

15. Gegeven is een dataset met de volgende 11 punten:

−3.0 −2.5 −1.0 −0.8 −0.8 −0.5 0.5 0.7 1.0 2.5 5.0

Welke van de zes onderstaande boxplots hoort bij deze dataset?



16. Biologen beweren dat lieveheersbeestjes uit gebied A genetisch verschillend zijn dan “gewone” lieveheersbeestjes, en dat dit tot uiting komt in een verschillend verwacht aantal stippen μ op hun schild. Om dit kracht bij te zetten, tellen ze van 80 lieveheersbeestjes uit gebied A het aantal stippen. “Gewone” lieveheersbeestjes hebben gemiddeld 4.3 stippen op hun schild. De 80 getelde lieveheersbeestjes hadden gemiddeld 4.1 stippen, met een steekproef standaarddeviatie van 2. Geef de p -waarde behorende bij de hypothese $H_0 : \mu = 4.3$ tegen het alternatief $H_1 : \mu \neq 4.3$. Geef ook aan of de biologen tevreden zullen zijn met deze uitkomst.

- a. p -waarde is 0.032, biologen zijn tevreden b. p -waarde is 0.032, biologen zijn niet tevreden
c. p -waarde is 0.185, biologen zijn tevreden d. p -waarde is 0.185, biologen zijn niet tevreden
e. p -waarde is 0.371, biologen zijn tevreden f. p -waarde is 0.371, biologen zijn niet tevreden

17. Vervolg vorige vraag: op basis van de gegevens over gebied A, hoe schat u de kans dat 40 lieveheersbeestjes uit gebied A samen meer dan 180 stippen hebben?

- a. 38 % b. 31 % c. 21 % d. 16 % e. 10 % f. 5 %

18. We weten dat $X_1, \dots, X_n$ een steekproef vormen uit $U(0, a)$, voor een $a > 0$. Als schatter voor a beschouwen we $T = r\bar{X}_n$, voor een $r > 0$. Bij welke keuze van r is de MSE van T minimaal?

- a. $r = 1$ b. $r = \frac{9n}{4n+1}$ c. $r = \frac{6n}{3n+1}$ d. $r = 2$ e. $r = \frac{3}{2}n$ f. $r = 2n$

19. Voor een bivariate dataverzameling $(x_1, y_1), (x_2, y_2), \dots, (x_{20}, y_{20})$ beschouwen we een lineair regressiemodel $Y_i = \alpha + \beta x_i + U_i$, waarbij $U_1, U_2, \dots, U_{20}$ onafhankelijk zijn en $N(0, \sigma^2)$ verdeeld. We vinden voor onze data

$$\hat{\alpha} = 5, \quad \hat{\beta} = 2.5, \quad s_a = 0.5, \quad s_b = 0.25.$$

We toetsen $H_0 : \alpha = 4.5$ tegen $H_1 : \alpha > 4.5$ door middel van een t -toets bij significantieniveau 0.05. Geef het kritieke gebied van deze toets, en uw conclusie aangaande H_0 .

- a. krit. geb.: $(1.734, \infty)$, verwerp H_0 .
 - b. krit. geb.: $(1.734, \infty)$, verwerp H_0 niet.
 - c. krit. geb.: $(2.101, \infty)$, verwerp H_0 .
 - d. krit. geb.: $(2.101, \infty)$, verwerp H_0 niet.
 - e. krit. geb.: $(1.729, \infty)$, verwerp H_0 .
 - f. krit. geb.: $(1.729, \infty)$, verwerp H_0 niet.
20. Bij een verkiezingsenquête geven 80 van de 400 stemmers aan dat ze op de SP zullen gaan stemmen. Noem p het percentage van de bevolking dat nu denkt op de SP te zullen gaan stemmen. Definieer $\hat{p} = 80/400 = 0.2$. Geef een 95% betrouwbaarheidsinterval van p .
- | | | |
|------------------------|------------------------|------------------------|
| a. $\hat{p} \pm 0.098$ | b. $\hat{p} \pm 0.082$ | c. $\hat{p} \pm 0.047$ |
| d. $\hat{p} \pm 0.039$ | e. $\hat{p} \pm 0.023$ | f. $\hat{p} \pm 0.020$ |
21. Op grond van een enquête met steekproefgrootte n wordt als 90 % betrouwbaarheidsinterval voor een proportie het interval $(0.56, 0.70)$ gevonden. Als op vergelijkbare wijze een nieuwe steekproef wordt genomen, maar dan met grootte $2n$, wat kan er dan met enige zekerheid voorspeld worden over het betrouwbaarheidsinterval dat dit op zal leveren?
- a. het midden ligt tussen 0.62 en 0.64, en de breedte blijft ongeveer 0.14
 - b. het midden ligt tussen 0.62 en 0.64, en de breedte wordt ongeveer 0.1
 - c. het midden ligt tussen 0.56 en 0.70, en de breedte blijft ongeveer 0.14
 - d. het midden ligt tussen 0.56 en 0.70, en de breedte wordt ongeveer 0.1
 - e. het midden ligt tussen 0.62 en 0.64, en de breedte hangt sterk af van de nieuwe $\hat{p}$
 - f. het midden ligt tussen 0.56 en 0.70, en de breedte hangt sterk af van de nieuwe $\hat{p}$
22. Een machine maakt metalen ringen met een instelbare verwachte diameter en een gegeven standaarddeviatie van 0.1 mm. De diameter van de ringen is normaal verdeeld. We meten 25 diameters op, en vinden een gemiddelde diameter van 8.50 mm. Geef een 90% betrouwbaarheidsinterval voor μ , de ingestelde verwachte diameter.
- | | |
|--|---|
| a. $8.50 \pm z_{0.05} \cdot 0.1/\sqrt{24}$ | b. $8.50 \pm t_{0.10,24} \cdot 0.1/\sqrt{24}$ |
| c. $8.50 \pm t_{0.10,24} \cdot 0.02$ | d. $8.50 \pm t_{0.05,24} \cdot 0.02$ |
| e. $8.50 \pm z_{0.10} \cdot 0.02$ | f. $8.50 \pm z_{0.05} \cdot 0.02$ |

- 23.** De eigenaar van een call center beweert dat de kans p dat iemand langer dan 2 minuten moet wachten, kleiner is dan 1%. Om dit te bewijzen meet hij de wachttijd van 150 klanten, en het blijkt dat niemand langer dan 2 minuten hoeft te wachten. We nemen X , het aantal klanten van de 150 dat langer dan 2 minuten moet wachten, als toetsingsgrootheid. Geef de alternatieve hypothese die de eigenaar moet opstellen en de exacte p -waarde van de toets.
- a. $H_1 : p < 0.01$, p -waarde = 0.0337. b. $H_1 : p < 0.01$, p -waarde = 0.910.
c. $H_1 : p < 0.01$, p -waarde = 0.221. d. $H_1 : p > 0.01$, p -waarde = 0.0337.
e. $H_1 : p > 0.01$, p -waarde = 0.910. f. $H_1 : p > 0.01$, p -waarde = 0.221.
- 24.** Vervolg vorige vraag: Stel de kans op een wachttijd langer dan 2 minuten is in werkelijkheid 0.03. Een consumentenorganisatie, die de claim van de eigenaar niet klakkeloos wil accepteren, besluit 50 wachttijden te meten, en verwierpt $H_0 : p = 0.01$ ten gunste van $H_1 : p > 0.01$ als zij 2 of meer keer een wachttijd langer dan 2 minuten vindt. Wat is de kans op een fout van type II?
- a. 0.8009 b. 0.1991 c. 0.9596 d. 0.0404 e. 0.5553 f. 0.4447
- 25.** We willen weten of het percentage VVD stemmers in twee wijken van Delft verschillend is. In wijk 1 blijkt van 200 stemmers er 45 op de VVD te willen stemmen. In wijk 2 blijkt van 150 stemmers er 27 op de VVD te willen stemmen. Noem de echte percentages in de twee wijken p_1 en p_2 . Toets de hypothese $H_0 : p_1 = p_2$ tegen $H_1 : p_1 \neq p_2$, bij een significantieniveau van 5%. Geef de waarde van de juiste toetsingsgrootheid t en uw conclusie. Werk precies! Hint: u kunt $X_1, \dots, X_{200}$ opvatten als de uitkomst van de 200 ondervraagden in wijk 1, waarbij $X_i = 0$ betekent dat persoon i niet op de VVD stemt, en $X_i = 1$ betekent dat persoon i wel op de VVD stemt. Zo kunt u ook $Y_1, \dots, Y_{150}$ voor wijk 2 definiëren.
- a. $t = 1.032$, verwerp H_0 niet b. $t = 1.032$, verwerp H_0 wel
c. $t = 1.045$, verwerp H_0 niet d. $t = 1.045$, verwerp H_0 wel
e. $t = 1.023$, verwerp H_0 niet f. $t = 1.023$, verwerp H_0 wel

Antwoorden multiple choice:

1 b.

2 c.

3 d.

4 f.

5 a.

6 c.

7 f.

8 e.

9 b.

10 d.

11 b.

12 d.

13 b.

14 f.

15 b.

16 f.

17 e.

18 c.

19 b.

20 d.

21 d.

22 f.

23 c.

24 e.

25 a.

Uitwerkingen multiple choice:

1 b. Bij manier I hebben alledrie de jongens $1/4$ kans om gekozen te worden, en er is $1/4$ kans dat het spel opnieuw wordt gespeeld. Aangezien ze allemaal dezelfde kans hebben, is de uiteindelijke kans dat bijvoorbeeld Aad gekozen wordt gelijk aan $1/3$: manier I is dus eerlijk. Bij manier twee wordt Co alleen gekozen als er drie keer kop (kans $1/8$) of drie keer munt (ook kans $1/8$) gegooid wordt; in totaal heeft Co dus kans $2/8$, niet gelijk aan $1/3$, en dus is manier II niet eerlijk.

Bij manier drie worden alle spelers gelijk behandeld, en dus hebben ze ook allemaal gelijke kans om gekozen te worden. Manier III is dus eerlijk.

2 c. De mogelijkheden voor de drie worpen zijn: $\{1, 1, 4\}$ (hier zijn drie mogelijke permutaties van), $\{1, 2, 3\}$ (hier zijn zes mogelijke permutaties van) en $\{2, 2, 2\}$. In totaal zijn er dus 10 mogelijke uitkomsten voor dobbelsteen 1, 2 en 3, van de in totaal $6^3 = 216$ mogelijke uitkomsten.

3 d. Het is duidelijk dat X_1 en X_2 allebei dezelfde geometrische verdeling, en wel $\text{Geo}(1/6)$. Echter, X_3 heeft geen geometrische verdeling (bijvoorbeeld, $P(X_3 = 1) = 0$).

4 f. Hoe groter X_2 , hoe groter X_3 . We hebben: $\text{corr}(X_1, X_2) = 0$, $\text{corr}(X_1, X_3) > 0$ en $\text{corr}(X_2, X_3) > 0$.

5 a. Zij X het aantal gelijke spelen bij 8 gespeelde partijen. Dan heeft X een $\text{Bin}(8, 0.15)$ verdeling. Gevraagd is

$$\begin{aligned} P(X \geq 6) &= P(X = 6) + P(X = 7) + P(X = 8) \\ &= \binom{8}{6} (0.15)^6 (0.85)^2 + \binom{8}{7} (0.15)^7 (0.85) + (0.15)^8 \\ &= 2.4 \cdot 10^{-4}. \end{aligned}$$

6 c. Een manier om dit te zien is de volgende: als Y een Bernoulli variabele is met $p = 1/2$, dan geldt dat $X \sim 2Y$, en dus geldt dat $\text{Var}(X) = 4\text{Var}(Y) = 1$.

7 f. Er geldt voor $x \in [-1, 1]$:

$$P(X \leq x) = \int_{-1}^x \frac{1}{2}t + \frac{1}{2} dt = \left[\frac{1}{4}t^2 + \frac{1}{2}t \right]_{-1}^x = \frac{1}{2}x + \frac{1}{4}x^2 + \frac{1}{4}.$$

8 e. Er geldt

$$E[X] = \int_{-1}^1 x \cdot \left(\frac{1}{2} + \frac{1}{2}x \right) dx = \left[\frac{1}{4}x^2 + \frac{1}{6}x^3 \right]_{-1}^1 = \frac{1}{3}.$$

9 b. Noem G het totale gewicht van de 10 personen. Dan geldt dat $E[G] = 10 \times 72 = 720$ kg, en $\text{Var}(G) = 10 \times 36 = 360 \text{kg}^2$. Dus

$$P(G > 750) = P\left(\frac{G - 720}{\sqrt{360}} > \frac{750 - 720}{\sqrt{360}}\right) = P(Z > 1.58) = 0.057.$$

10 d. Er geldt

$$E[X] = \frac{1}{5} \cdot 0 + \frac{2}{5} \cdot 1 + \frac{2}{5} \cdot 2 = \frac{6}{5}.$$

Net zo geldt

$$E[X^2] = \frac{1}{5} \cdot 0^2 + \frac{2}{5} \cdot 1^2 + \frac{2}{5} \cdot 2^2 = \frac{10}{5}.$$

Dus $\text{Var}(X) = E[X^2] - E[X]^2 = \frac{14}{25}$.

11 b. Er geldt

$$E[X_1] = \int_0^1 cx^2 + x - \frac{1}{2}cx \, dx = \left[\frac{1}{3}cx^3 + \frac{1}{2}x^2 - \frac{1}{4}cx^2 \right]_0^1 = \frac{1}{12}c + \frac{1}{2}.$$

Aangezien $E[\bar{X}_n] = E[X_1]$, volgt nu dus dat

$$E[12\bar{X}_n - 6] = 12\left(\frac{1}{12}c + \frac{1}{2}\right) - 6 = c.$$

12 d. Er geldt voor de likelihood

$$L(c) = cX_1 + 1 - \frac{1}{2}c = 0.13c + 1.$$

Dit is lineair in c , maar we moeten onthouden dat $c \in [-2, 2]$. Dit betekent dat $L(c)$ maximaal is als $c = 2$.

13 b. Uit de vorm van de boxplots is het duidelijk dat II en IV de twee exponentiele verdelingen zijn. Aangezien $Exp(1)$ een kleinere verwachting heeft dan de $Exp(1/3)$, moet II bij $Exp(1/3)$ horen. Verder kan I niet bij $U(-2, 2)$ horen, aangezien er waarnemingen zijn kleiner dan -2 . Dus III moet bij $U(-2, 2)$ horen.

14 f. Er geldt

$$\text{corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}},$$

dus $\text{Cov}(X, Y) = -1$. Dit betekent dat

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) = 3.$$

15 b. We berekenen eerst de mediaan: het 6de getal (van de geordende dataset met 11 punten) is -0.5 , dus de mediaan is -0.5 . Dan zijn alleen **a**, **b** en **c** nog over. Vervolgens berekenen we het derde kwartiel: $\frac{3}{4} \cdot (11 + 1) = 9$, en het 9de getal is 1, dus het derde kwartiel is 1. Dan blijven alleen **a** en **b** over. Tot slot berekenen we de Inter Quartile Range: het onderste kwartiel is -1 (de waarde van het derde $(= (11 + 1)/4)$ getal), dus de IQR = 2. De omlaaggaande "whisker" heeft maximale lengte $1.5 \cdot \text{IQR} = 3$, en $-3 > -1 - 3$. Dit betekent dat de whisker tot aan -3 gaat. Dit leidt tot antwoord **b**. Deze heeft ook een 'outlier' bij 5.

16 f. Door het grote aantal datapunten kunnen we de verdeling van de toetsingsgrootheid

$$T = \frac{\bar{x}_n - 4.3}{s_n / \sqrt{n}}$$

benaderen met een $N(0, 1)$ verdeling, als de nulhypothese waar is. We vinden $t = -0.894$. Alle waarden groter dan 0.894 en kleiner dan -0.894 wijzen op H_1 , dus de p-waarde wordt gegeven door

$$2 \cdot P(Z > 0.894) = 0.371.$$

De biologen zullen met deze hoge p -waarde niet tevreden zijn, aangezien ze geen significant verschil hebben aangetoond.

17 e. Uit de centrale limietstelling volgt dat wanneer we S definiëren als het totaal aantal stippen van de 40 lieveheersbeestjes, dan heeft S bij benadering een normale verdeling, en wel met verwachting $40 \cdot 4.1 = 164$ en variantie $40 \cdot 4 = 160$ (dit zijn schattingen gebaseerd op de vorige vraag). Hieruit volgt dat

$$P(S > 180) = P\left(\frac{S - 164}{\sqrt{160}} > \frac{180 - 164}{\sqrt{160}}\right) = P(Z > 1.26) = 0.104.$$

18 c. We weten dat

$$\begin{aligned}\text{MSE}(T) &= \text{Var}(T) + (\text{E}[T] - a)^2 \\ &= r^2 \frac{a^2}{12n} + \left(\frac{1}{2}ra - a\right)^2 \\ &= a^2\left(\frac{r^2}{12n} + \left(\frac{1}{2}r - 1\right)^2\right).\end{aligned}$$

Om het minimum te vinden stellen we de afgeleide gelijk aan 0:

$$\begin{aligned}0 &= \frac{\partial \text{MSE}(T)}{\partial r} \iff \\ 0 &= a^2\left(\frac{r}{6n} + \frac{1}{2}r - 1\right) \iff \\ r + 3nr &= 6n \iff \\ r &= \frac{6n}{3n+1}.\end{aligned}$$

19 b. De waarde van de toetsingsgrootte is

$$t = \frac{\hat{\alpha} - 4.5}{s_a} = 1.$$

We verwerpen voor waarden die groter zijn dan 0, want sterk negatieve waarden wijzen *niet* op H_1 . Dit betekent dat het kritieke gebied van de vorm (c, ∞) is. Onder H_0 heeft T een $t(n-2) = t(18)$ verdeling. De kans op verwerpen moet 5% zijn als H_0 waar is, dus we moeten hebben dat

$$P(T > c) = 0.05.$$

Dit betekent dat $c = t_{18,0.05} = 1.734$. Aangezien $t = 1 \notin [1.734, \infty)$, valt t niet in het kritieke gebied, en verwerpen we H_0 dus niet.

20 d. De SE van $\hat{p}$ is gelijk aan $\sqrt{\hat{p}(1-\hat{p})/n} = 0.02$. Een 95% betrouwbaarheidsinterval wordt dus gegeven door

$$\hat{p} \pm z_{0.025}\text{SE}(\hat{p}) = \hat{p} \pm 0.039.$$

21 d. De onzekerheid in $\hat{p}$ is groter dan 0.005, dus bij meer metingen kan de schatting ook meer dan 0.005 afwijken. De breedte van het betrouwbaarheidsinterval is min of meer omgekeerd evenredig met $\sqrt{n}$; aangezien we 2.5 keer meer meetpunten hebben, wordt de breedte ongeveer $\sqrt{2.5} \approx 1.6$ keer zo klein.

22 f. we weten dat voor de echte onbekende verwachting μ geldt:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \text{ heeft een } N(0, 1) \text{ verdeling.}$$

Hieruit volgt dat

$$P(\bar{X}_n - z_{0.05} \cdot 0.1/\sqrt{n} \leq \mu \leq \bar{X}_n + z_{0.05} \cdot 0.1/\sqrt{n}) = 0.90.$$

Dit is antwoord **d**.

23 c. De eigenaar wil graag bewijzen dat $p < 0.01$, en dus kiest hij zijn alternatieve hypothese $H_1 : p < 0.01$. Als X het aantal lang wachtende klanten is, dan wijzen lage waarden van X naar het alternatief. De p-waarde wordt dus gegeven door $P(X \leq 0)$. Onder H_0 geldt dat $X \sim \text{Bin}(150, 0.01)$. Dus

$$P(X \leq 0) = (1 - 0.01)^{150} = 22\%.$$

24 e. De consumentenorganisatie maakt een fout van type II als zij niet zou verwerpen. Dit gebeurt als $X = 0$ of $X = 1$, en dit heeft kans

$$P(X \leq 1) = (1 - 0.03)^{50} + 50 \cdot 0.03 \cdot (1 - 0.03)^{49} = 0.5553$$

25 a. Aangezien onder de nul-hypothese de percentages hetzelfde zijn, en dus ook de varianties, werken we met de gepoolde variantie. Er geldt

$$s_1^2 = \hat{p}_1(1 - \hat{p}_1) = 0.174 \quad \text{en} \quad s_2^2 = \hat{p}_2(1 - \hat{p}_2) = 0.148.$$

Dus:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right) = 1.901 \cdot 10^{-3}.$$

De toetsingsgrootte wordt dan:

$$t = \frac{\hat{p}_1 - \hat{p}_2}{s_p} = 1.032.$$

Dit levert een p -waarde van $2 \cdot P(Z > 1.03) = 0.30$, geen enkele reden dus om H_0 te verwerpen.

Tentamen Statistiek
wi1102CT, 21 juni 2011

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

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. De hoofdprijs van een populaire televisiequiz wordt bepaald door blind te trekken uit een bak met tien genummerde ballen: 1,2,2,3,3,3,4,4,4,4. Elke bal is zijn nummer in duizenden euro's waard. De winnaar mag (zonder teruglegging) zoveel ballen trekken als hij/zij wil. Er is een *maar*: als twee ballen met hetzelfde nummer getrokken zijn, dan vervallen de inkomsten en is het spel uit. Jan speelt de volgende strategie: hij trekt twee ballen en stopt dan. Noem de opbrengst voor Jan (in euro's) bij deze strategie X . Dus als Jan eerst een 1 trekt en dan een 3, dan geldt $X = 4000$. Wat is de kans dat $X = 0$?
a. 0.111 b. 0.125 c. 0.167 d. 0.2 e. 0.222 f. 0.25
2. Noem de verdelingsfunctie van de bovengenoemde stochast X maar even F . Bereken $F(6000)$.
a. 0.6 b. 0.625 c. 0.666 d. 0.733 e. 0.8 f. 0.888
3. LET OP: DIT IS EEN VIERKEUZE VRAAG! U kunt uw antwoord gewoon op het antwoordformulier invullen.
Zilverland kent vier politieke partijen, twee 'rechtse', de PvdA, PvdB, en twee 'linkse', de PvdC en de PvdD. Een reporter staat bij een station en vraagt tien willekeurige passanten naar hun politieke voorkeur. Laat X het aantal mensen van die 10 zijn dat op de PvdB stemt, en Y het aantal mensen dat 'links' stemt. Dan geldt
a. X en Y zijn positief gecorreleerd b. X en Y zijn ongecorreleerd en onafhankelijk
c. X en Y zijn negatief gecorreleerd d. X en Y zijn ongecorreleerd en afhankelijk
4. Stel $X_1, X_2, \dots, X_{100}$ zijn onafhankelijke Poisson-variabelen met parameter $\mu = 2$, en Z is een standaard normale variabele.
De centrale limietstelling geeft: $P(X_1 + X_2 + \dots + X_{100} \leq 210) \approx$
a. $P(Z \leq -1/2)$ b. $P(Z \leq -1/20)$ c. $P(Z \leq 1/200)$ d. $P(Z \leq 1/20)$
e. $P(Z \leq 1/2)$ f. $P(Z \leq 1/\sqrt{2})$
5. Een tennisspeler heeft een kans van 7% dat hij een dubbele fout slaat, en zo het punt verliest, onafhankelijk van de daarvoor gespeelde punten. Wat is de kans dat hij bij 15 gespeelde punten, precies 5 dubbele fouten slaat?
a. 0.244% b. 1.05% c. 3.33% d. 5.21% e. 18.1% f. 33.3%
6. Uit ervaring weet een metselaar dat een bepaalde soort baksteen gemiddeld 20 cm lang is met een standaard deviatie van 0.2 cm. Ook weet hij dat de voeg tussen twee bakstenen gemiddeld 1 cm dik is, met een standaard deviatie van 0.1 cm. Hij moet een randje (1 baksteen hoog) metselen van ten minste 200 meter lang. Wat is de kans dat hij genoeg heeft aan 952 stenen?
a. 0.002 b. 0.043 c. 0.097 d. 0.50 e. 0.957 f. 0.998

7. Stel de stochast X heeft de verdelingsfunctie

$$F(x) = \begin{cases} 0 & \text{als } x < 0 \\ \frac{1}{4}x^2 & \text{als } 0 \leq x \leq 2 \\ 1 & \text{als } x > 2. \end{cases}$$

Dan is $E[X]$ gelijk aan

- a. $\frac{3}{32}$ b. $\frac{1}{2}$ c. 1 d. $\frac{4}{3}$ e. $\frac{30}{16}$ f. 2

8. De dichtheid van een stochast X wordt gegeven door

$$f(x) = \begin{cases} \frac{1}{9}x^2, & \text{als } 0 \leq x \leq 3 \\ 0, & \text{anders.} \end{cases}$$

De mediaan van X is:

- a. $\frac{1}{2}$ b. 1 c. $\frac{3}{2}$ d. $3 \cdot 2^{-1/2}$ e. $3 \cdot 2^{-1/3}$ f. 3

9. Een bedrijf organiseert een spel voor een groep van 32 werknemers. Iedere werknemer heeft een kans van $\frac{1}{4}$ om een bedrag van 10 € te winnen. De voorbereidingskosten van het spel zijn uniform verdeeld tussen de 40 en de 70 €, en onafhankelijk van het aantal winnaars. Noem K de totale kosten voor het bedrijf (de voorbereidingskosten plus het uitbetaalde prijzengeld). De verwachting en de variantie van K zijn gelijk aan

- a. $E[K] = 135$ en $\text{Var}(K) = 135$ b. $E[K] = 135$ en $\text{Var}(K) = 675$
c. $E[K] = 135$ en $\text{Var}(K) = 975$ d. $E[K] = 205$ en $\text{Var}(K) = 135$
e. $E[K] = 205$ en $\text{Var}(K) = 675$ f. $E[K] = 205$ en $\text{Var}(K) = 975$

10. De verdeling van een discrete stochast X wordt gegeven door

a	-1	1	4
$P(X = a)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

Bepaal de verwachting en de variantie van X .

- a. $E[X] = \frac{4}{3}$ en $\text{Var}(X) = \frac{3}{2}$ b. $E[X] = \frac{4}{3}$ en $\text{Var}(X) = \frac{9}{2}$
c. $E[X] = \frac{4}{3}$ en $\text{Var}(X) = 4$ d. $E[X] = 2$ en $\text{Var}(X) = \frac{3}{2}$
e. $E[X] = 2$ en $\text{Var}(X) = \frac{9}{2}$ f. $E[X] = 2$ en $\text{Var}(X) = 4$

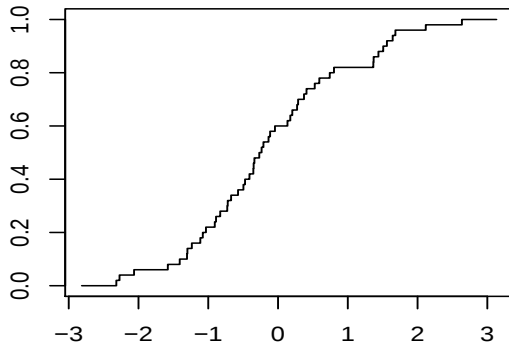
11. X_1 , X_2 en X_3 zijn onafhankelijk met een $Pois(\mu)$ verdeling, waarbij $\mu > 0$ onbekend is. We observeren X_1 en het product X_2X_3 . Definieer de volgende schatter voor μ :

$$T = aX_1 + bX_2X_3 + cX_1^2.$$

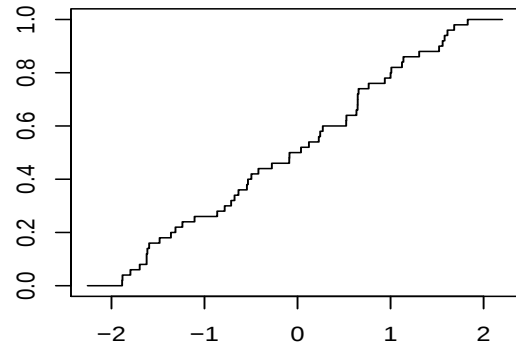
Voor welke keuze van a , b en c is T een zuivere schatter voor μ ? Hint: relateer $E[X_1]$ aan $\text{Var}(X_1)$.

- a. $a = 3$, $b = 2$ en $c = -2$ b. $a = 0$, $b = 1/\mu$ en $c = 0$
c. $a = \frac{1}{3}$, $b = \frac{1}{3}$ en $c = \frac{1}{3}$ d. $a = 0$, $b = 1$ en $c = -1$
e. $a = \frac{1}{2}$, $b = 0$ en $c = \frac{1}{2}$ f. $a = -2$, $b = -1$ en $c = 1$

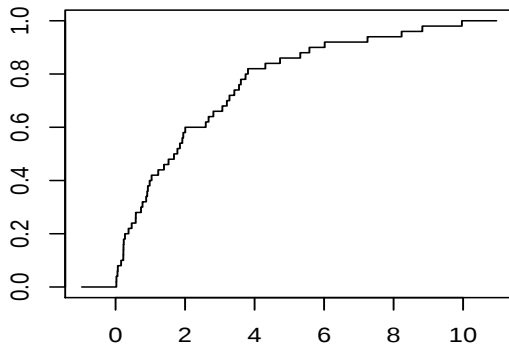
12. Van vier datasets ter grootte 50 is de empirische verdelingsfunctie weergegeven. De datasets zijn steekproeven uit vier verschillende verdelingen: een $Exp(1)$, een $Exp(1/3)$, een $N(0,1)$ en een $U(-2,2)$ verdeling. Welke dataset komt uit de $Exp(1/3)$ en welke dataset komt uit de $N(0,1)$ verdeling?



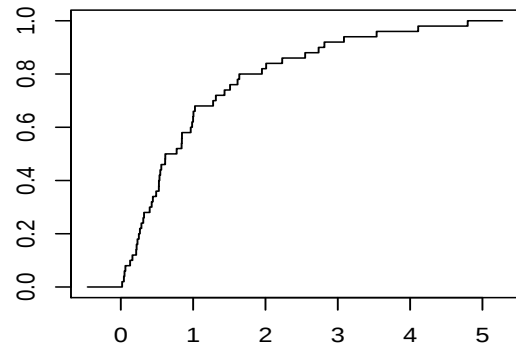
I



II



III



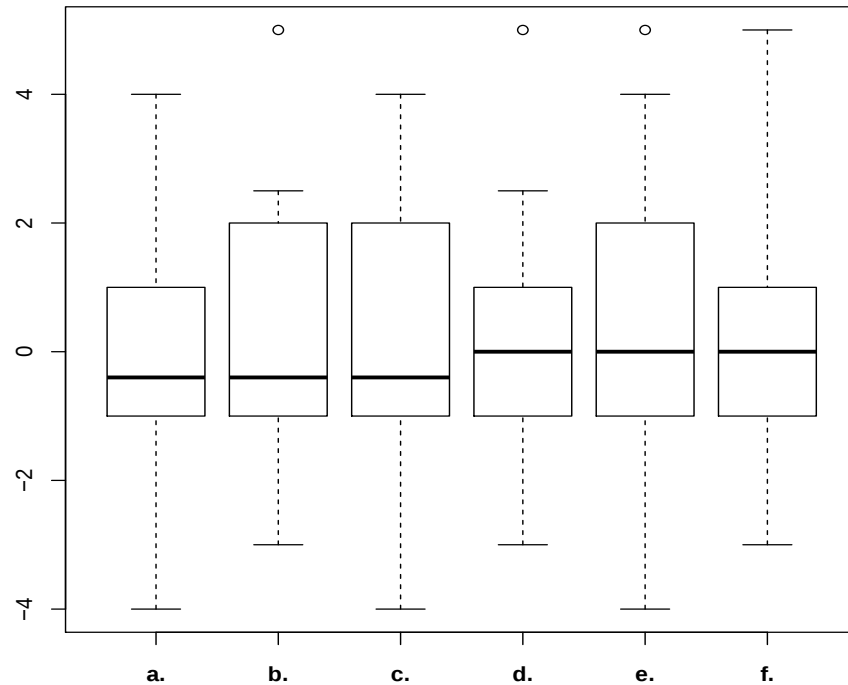
IV

- a. $Exp(1/3)$: I; $N(0,1)$: III
b. $Exp(1/3)$: III; $N(0,1)$: II
c. $Exp(1/3)$: III; $N(0,1)$: I
d. $Exp(1/3)$: IV; $N(0,1)$: II
e. $Exp(1/3)$: IV; $N(0,1)$: I
f. $Exp(1/3)$: II; $N(0,1)$: I
13. Zij X een stochastische variabele met een $Exp(\lambda)$ verdeling. We willen op basis van X een schatting geven voor $1/\lambda$. Beschouw de schatter $T = rX$, voor $r > 0$. Voor welke waarde van r is de MSE van T het kleinst?
- a. $r = \frac{1}{2}$
b. $r = 1$
c. $r = \frac{5}{4}$
d. $r = \frac{3}{2}$
e. $r = \frac{12}{7}$
f. $r = 2$

14. Gegeven is een dataset met de volgende 11 punten:

−3.0 −2.5 −1.0 −0.8 −0.8 0.0 0.5 0.7 1.0 2.5 5.0

Welke van de zes onderstaande boxplots hoort bij deze dataset?



15. Marije kijkt Kees diep in zijn ogen, en het valt haar op dat Kees met zijn ogen knippert. Zij telt 10 perioden dat Kees zijn ogen openhoudt die korter duren dan een seconde, en 2 perioden die langer duren dan een seconde. Als we deze perioden modelleren als onafhankelijke realisaties van een exponentieel verdeelde variabele met parameter λ , dan is de Maximum Likelihoodschatter voor λ op basis van deze gegevens gelijk aan

a. $\lambda = \frac{1}{6}$ b. $\lambda = \frac{5}{6}$ c. $\lambda = \ln(3)$ d. $\lambda = \frac{6}{5}$ e. $\lambda = \ln(6)$ f. $\lambda = 6$

16. DEZE VRAAG IS HIER VERBETERD. OP HET OORSPRONKELIJKE TENTAMEN TELDE HIJ NIET MEE. Het aantal klanten dat een winkel binnenloopt tussen 9:00 en 10:00 noemen we X_1 , en het aantal klanten tussen 10:00 en 11:00 noemen we X_2 . We veronderstellen dat X_1 en X_2 onafhankelijk zijn en $Pois(\mu)$ verdeeld. De bediende herinnert zich dat er tussen 9:00 en 10:00 2 klanten zijn binnengekomen, en tussen 10:00 en 11:00 hoogstens 1. Dit betekent dus dat $(X_1 = 2, X_2 = 0)$ of $(X_1 = 2, X_2 = 1)$. Geef de maximum likelihoodschatter voor μ op basis van deze gegevens.

a. $\mu = 0.333$ b. $\mu = 1.281$ c. $\mu = 1.500$ d. $\mu = 1.667$ e. $\mu = 2.00$ f. $\mu = 2.142$

17. De inhoud van het laadruim van een olietanker wordt gemeten door te meten hoeveel olie erin gespoten wordt bij het volladen. Er zijn 9 van dit soort metingen gedaan, die we modelleren als realisaties van een normale verdeling $N(\mu, \sigma^2)$, waarbij μ de echte inhoud voorstelt. Het gemiddelde van de 9 metingen is 1985 m^3 , en de steekproef standaarddeviatie is 150 m^3 . Geef een 90% betrouwbaarheidsinterval voor μ .

- | | | |
|-----------------------|-----------------------|-----------------------|
| a. (1915.15, 2054.85) | b. (1892.00, 2078.00) | c. (1915.85, 2054.15) |
| d. (1893.35, 2076.65) | e. (1775.45, 2194.55) | f. (1706.00, 2264) |

18. Voor een bivariate dataverzameling $(x_1, y_1), (x_2, y_2), \dots, (x_{20}, y_{20})$ beschouwen we een lineair regressiemodel $Y_i = \alpha + \beta x_i + U_i$, waarbij $U_1, U_2, \dots, U_{20}$ onafhankelijk zijn en $N(0, \sigma^2)$ verdeeld. We vinden voor onze data

$$\hat{\alpha} = 5, \quad \hat{\beta} = 2.5, \quad s_a = 0.5, \quad s_b = 0.25.$$

We toetsten $H_0 : \beta = 2$ tegen $H_1 : \beta \neq 2$ door middel van een t -toets bij significantieniveau 0.05. Geef het kritieke gebied van deze toets, en uw conclusie aangaande H_0 .

- | |
|---|
| a. krit. geb.: $(-\infty, -2.086) \cup (2.086, \infty)$, verwerp H_0 niet. |
| b. krit. geb.: $(-\infty, -2.086) \cup (2.086, \infty)$, verwerp H_0 . |
| c. krit. geb.: $(-\infty, -1.734) \cup (1.734, \infty)$, verwerp H_0 niet. |
| d. krit. geb.: $(-\infty, -1.734) \cup (1.734, \infty)$, verwerp H_0 . |
| e. krit. geb.: $(-\infty, -2.101) \cup (2.101, \infty)$, verwerp H_0 niet. |
| f. krit. geb.: $(-\infty, -2.101) \cup (2.101, \infty)$, verwerp H_0 . |

19. Het zuurstofgehalte van de lucht na afloop van een groot college wordt 16 keer gemeten. De fabrikant van de klimaatinstallatie garandeert dat de standaardafwijking van het zuurstofgehalte gelijk is aan 1%, en dat het zuurstofgehalte goed gemodelleerd wordt door een normale verdeling. Het gemiddelde van de 16 metingen is gelijk aan 19.5%. Geef een 95% betrouwbaarheidsinterval voor het verwachte zuurstofgehalte aan het eind van een groot college.

- | | |
|---|---|
| a. $19.5 \pm z_{0.025} \cdot 1/\sqrt{15}$ | b. $19.5 \pm t_{0.05,15} \cdot 1/\sqrt{15}$ |
| c. $19.5 \pm t_{0.05,15} \cdot 0.25$ | d. $19.5 \pm t_{0.025,15} \cdot 0.25$ |
| e. $19.5 \pm z_{0.05} \cdot 0.25$ | f. $19.5 \pm z_{0.025} \cdot 0.25$ |

20. Bij het oversturen van gegevens komen sommige bits verkeerd door. Bij een test van een bepaald systeem worden $8 \cdot 10^6$ bits doorgestuurd. Er blijken er 110 verkeerd doorgestuurd te zijn. Geef een 95% betrouwbaarheidsinterval van de kans p dat een bit verkeerd wordt doorgestuurd. Gebruik dat als $\hat{p}$ de fractie fout doorgestuurde bits is, dan geldt

$$(p - \hat{p}) \frac{\sqrt{n}}{\sqrt{\hat{p}(1 - \hat{p})}} \sim N(0, 1).$$

- | | |
|---|---|
| a. $[0.875 \cdot 10^{-5}, 1.875 \cdot 10^{-5}]$ | b. $[1.12 \cdot 10^{-5}, 1.63 \cdot 10^{-5}]$ |
| c. $[1.26 \cdot 10^{-5}, 1.49 \cdot 10^{-5}]$ | d. $[0.25 \cdot 10^{-5}, 2.50 \cdot 10^{-5}]$ |
| e. $[0.455 \cdot 10^{-5}, 2.295 \cdot 10^{-5}]$ | f. $[1.35 \cdot 10^{-5}, 1.40 \cdot 10^{-5}]$ |

21. Vervolg op vraag 20. De fabrikant van het systeem denkt dat er tijdens de test uit vraag 20 iets is misgegaan. Hij beweert dat de kans dat een bit fout wordt doorgestuurd zeker kleiner is dan $5 \cdot 10^{-7}$. Hij voert een nieuwe test uit waarbij 10^7 bits worden doorgestuurd. Daarvan blijkt slechts 1 bit verkeerd doorgestuurd te zijn. Neem $H_0 : p = 5 \cdot 10^{-7}$. Stel de juiste alternatieve hypothese op die de fabrikant moet gebruiken en bereken de exacte p -waarde van deze toets. U kunt gebruiken dat $\binom{n}{0} = 1$ en $\binom{n}{1} = n$.

- | |
|--|
| a. $H_1 : p < 5 \cdot 10^{-7}$, p -waarde = 0.0337. |
| b. $H_1 : p < 5 \cdot 10^{-7}$, p -waarde = 0.96. |
| c. $H_1 : p < 5 \cdot 10^{-7}$, p -waarde = 0.0404. |

- d. $H_1 : p > 5 \cdot 10^{-7}$, p -waarde = 0.0337.
- e. $H_1 : p > 5 \cdot 10^{-7}$, p -waarde = 0.96.
- f. $H_1 : p > 5 \cdot 10^{-7}$, p -waarde = 0.0404.

22. Jan koopt hooi van een bepaald merk voor zijn konijnen. Hij denkt echter dat er gemiddeld te weinig hooi in de verpakking zit. Hij meet het gewicht van het hooi in 9 zakken precies op, en hij vindt $\bar{x}_n = 0.977$ kg en een steekproefstandaarddeviatie $s_n = 0.030$ kg. Op de verpakking staat dat er gemiddeld 1 kg in een zak zit. We modelleren het gewicht van het hooi als een $N(\mu, \sigma^2)$ verdeling. Als nul-hypothese nemen we $H_0 : \mu = 1$, en als alternatief $H_1 : \mu < 1$. Verder kiezen we een significantieniveau van 5%. Welke van de volgende zes conclusies is de juiste?

- | | |
|---|---|
| a. $0.01 < p < 0.025$, verwerp H_0 . | b. $0.01 < p < 0.025$, verwerp H_0 niet. |
| c. $0.025 < p < 0.05$, verwerp H_0 . | d. $0.025 < p < 0.05$, verwerp H_0 niet. |
| e. $0.05 < p < 0.10$, verwerp H_0 . | f. $0.05 < p < 0.10$, verwerp H_0 niet. |

23. Stel een meting X van de snelheid van een passerende auto op de snelweg wordt gemodelleerd als een normale variable met parameters $\mu = v$ (in km/u) en $\sigma = 5$ (in km/u), waarbij v de exacte snelheid is van de auto. Om te toetsen of een automobilist harder rijdt dan de toegestane 120 km/u wordt getoetst $H_0: v = 120$ tegen $H_1: v > 120$. Stel verder dat de nulhypothese wordt verworpen (en dus: de automobilist krijgt een boete) als de gemeten snelheid 128 km/u of hoger is. Wat geldt dan voor P_{II} , de kans op een fout van type II, voor een automobilist die *niet harder* rijdt dan 133 km/u, maar wel harder dan 120 km/u?

- | | | |
|----------------------------|-------------------------|-------------------------|
| a. $P_{II} \approx 0.0794$ | b. $P_{II} \geq 0.0794$ | c. $P_{II} \leq 0.0794$ |
| d. $P_{II} \approx 0.159$ | e. $P_{II} \geq 0.159$ | f. $P_{II} \leq 0.159$ |

24. Een laborante bekijkt 100 kweekjes, en telt het aantal besmettingshaarden per kweekje. Zij vindt gemiddeld 3.64 besmettingshaarden per kweekje, met een steekproefstandaarddeviatie van 2. Noem μ het verwachte aantal besmettingshaarden per kweekje. De laborante voert een t -toets uit voor de nulhypothese $H_0 : \mu = 4$, tegen de alternatieve hypothese $H_1 : \mu \neq 4$, bij significantieniveau 0.05. Geef de waarde van de toetsingsgrootte en vermeldt uw conclusie.

- | | |
|----------------------------------|---------------------------------------|
| a. $t = -0.18$; H_0 verwerpen | b. $t = -0.18$; H_0 niet verwerpen |
| c. $t = -1.22$; H_0 verwerpen | d. $t = -1.22$; H_0 niet verwerpen |
| e. $t = -1.80$; H_0 verwerpen | f. $t = -1.80$; H_0 niet verwerpen |

25. Vervolg op vraag **24**. De laborante uit de vorige vraag vindt in de literatuur een vergelijkbare studie met een ander soort kweekje. Bij deze studie, we noemen hem studie 2, werd van $m = 400$ kweekjes het aantal besmettingshaarden geteld. Er werd een gemiddeld aantal besmettingshaarden $\bar{y}_m = 4.2$ gevonden, met een steekproefstandaarddeviatie $s_Y = 2.2$. We willen onderzoeken of voor de kweekjes uit studie 1, namelijk de studie van vraag **24**, het verwachte aantal besmettingshaarden μ_1 lager is dan het verwachte aantal besmettingshaarden μ_2 voor studie 2. We toetsen de nulhypothese $H_0 : \mu_1 = \mu_2$ tegen het alternatief $H_1 : \mu_1 < \mu_2$. Hiervoor gebruiken we een twee-steekproeven t -toets met ongepoolde variantie, vanwege het grote aantal kweekjes. Geef de p -waarde van deze toets.

- | | | | | | |
|-----------|-----------|-----------|-----------|-----------|-----------|
| a. 0.0018 | b. 0.0035 | c. 0.0071 | d. 0.0142 | e. 0.0284 | f. 0.0568 |
|-----------|-----------|-----------|-----------|-----------|-----------|

Antwoorden multiple choice:

1 e.

2 d.

3 c.

4 f.

5 a.

6 c.

7 d.

8 e.

9 b.

10 e.

11 a.

12 c.

13 a.

14 d.

15 e.

16 b.

17 b.

18 e.

19 f.

20 b.

21 c.

22 c.

23 e.

24 f.

25 c.

Uitwerkingen multiple choice:

1 e. Jan verdient niks als hij twee 2'en, twee 3'en of twee 4'en trekt. Twee 2'en heeft kans $(2/10) * (1/9) = 2/90$. Twee 3'en heeft kans $(3/10) * (2/9) = 6/90$. Twee 4'en heeft kans $(4/10) * (3/9) = 12/90$. Samen is dit een kans van $20/90 = 2/9 = 0.222$.

2 d. $F(6000) = P(X \leq 6000) = 1 - P(X > 6000) = 1 - P(X = 7000)$. Er geldt dat $X = 7000$ alleen kan voorkomen als Jan eerst een 3 trekt en dan een 4 (kans $(3/10) * (4/9)$), of als hij eerst een 4 trekt en dan een 3 (kans $(4/10) * (3/9)$). Dus

$$F(6) = 1 - \frac{24}{90} = \frac{11}{15} = 0.733.$$

3 c. Aangezien PvdB een 'rechtse' partij is, impliceert een hoge waarde voor X *minder* 'linkse' stemmers, en dus een lage waarde voor Y . Ze zijn dus negatief gecorreleerd.

4 f. We weten dat $E[X_i] = 2$ en $\text{Var}(X_i) = 2$ (zie formuleblad bij de Poisson verdeling). Dit betekent dat $E[X_1 + \dots + X_{100}] = 200$ en $\text{Var}(X_1 + \dots + X_{100}) = 200$. Dus als we $S = X_1 + \dots + X_{100}$ definiëren, dan geldt volgens de centrale limietstelling:

$$P(S \leq 210) = P\left(\frac{S - 200}{\sqrt{200}} \leq \frac{210 - 200}{\sqrt{200}}\right) \approx P\left(Z \leq 1/\sqrt{2}\right).$$

5 a. Zij X het aantal dubbele fouten bij de 15 gespeelde punten. Dan heeft X een $\text{Bin}(15, 0.07)$ verdeling. Gevraagd is

$$P(X = 5) = \binom{15}{5} (0.07)^5 (0.93)^{10} = 0.244\%.$$

6 c. Als de metselaar 952 stenen achter elkaar metselt, dan heeft hij ook 951 voegen ertussen zitten. Noem X de lengte van dit randje. Er geldt: $E[X] = 952 * 0.2 + 951 * 0.01 = 199.91$ m. Verder geldt

$$\text{Var}(X) = 952 * (0.002)^2 + 951 * (0.001)^2 = 0.004759\text{m}^2.$$

Dus

$$P(X \geq 200) = P\left(\frac{X - 199.91}{\sqrt{0.004759}} \geq \frac{200 - 199.91}{\sqrt{0.004759}}\right) = P(Z \geq 1.30) = 0.097.$$

7 d. De dichtheid van X wordt gegeven door $f(x) = F'(x) = \frac{1}{2}x$. Dus

$$E[X] = \int_0^2 \frac{1}{2}x \cdot x \, dx = \left[\frac{1}{6}x^3\right]_0^2 = \frac{4}{3}.$$

8 e. De mediaan m van X is de oplossing van:

$$\int_{-\infty}^m f(x) \, dx = \frac{1}{2} \Leftrightarrow \int_0^m \frac{1}{9}x^2 \, dx = \frac{1}{2} \Leftrightarrow \left[\frac{1}{27}x^3\right]_0^m = \frac{1}{2} \Leftrightarrow m = 3 \cdot 2^{-1/3}.$$

9 b. Noem X het aantal werknemers dat wint. Dan geldt $X \sim \text{Bin}(32, \frac{1}{4})$. Zij Y de voorbereidingskosten, dus $Y \sim U(40, 70)$. We weten dat

$$E[X] = 8, \text{Var}(X) = 6, E[Y] = 55 \text{ and } \text{Var}(Y) = 75.$$

Er geldt ook dat $K = 10X + Y$, en X en Y zijn onafhankelijk. Dus

$$E[K] = 10E[X] + E[Y] = 135$$

en

$$\text{Var}(K) = \text{Var}(10X) + \text{Var}(Y) = 100\text{Var}(X) + \text{Var}(Y) = 675.$$

10 e. Er geldt

$$E[X] = \frac{1}{4} \cdot -1 + \frac{1}{4} \cdot 1 + \frac{1}{2} \cdot 4 = 2.$$

Net zo geldt

$$E[X^2] = \frac{1}{4} \cdot (-1)^2 + \frac{1}{4} \cdot 1^2 + \frac{1}{2} \cdot 4^2 = 8\frac{1}{2}.$$

$$\text{Dus } \text{Var}(X) = E[X^2] - E[X]^2 = \frac{9}{2}.$$

11 a. Er geldt

$$E[T] = a\mu + b\mu^2 + c(\mu^2 + \mu).$$

We hebben gebruikt dat $E[X_2X_3] = E[X_2]E[X_3]$ en dat $\mu = \text{Var}(X_1) = E[X_1^2] - E[X_1]^2$. Aangezien $E[T] = \mu$ moet gelden voor elke $\mu > 0$, moet dus gelden dat $b + c = 0$ en $a + c = 1$. Aangezien de parameter μ onbekend is (we willen hem tenslotte schatten met T), mogen a, b en c niet van μ afhangen! Het enige antwoord dat aan deze eisen voldoet, is antwoord **a**.

12 c. Uit de vorm van de grafieken is het duidelijk dat III en IV de twee exponentiele verdelingen zijn. Aangezien $\text{Exp}(1)$ een kleinere verwachting heeft dan de $\text{Exp}(1/3)$, moet III bij $\text{Exp}(1/3)$ horen. Verder kan I niet bij $U(-2, 2)$ horen, aangezien er waarnemingen zijn kleiner dan -2 (en ook groter dan 2). Dus I moet bij $N(0, 1)$ horen.

13 a. Er geldt

$$\text{MSE}[T] = \text{Var}(T) + \left(E[T] - \frac{1}{\lambda}\right)^2 = r^2 \text{Var}(X) + \left(\frac{r}{\lambda} - \frac{1}{\lambda}\right)^2 = (r^2 + (r-1)^2) \lambda^{-2}.$$

We moeten nu r zo kiezen dat $\text{MSE}[T]$ wordt geminimaliseerd. Dit doen we door de afgeleide 0 te stellen:

$$\frac{d\text{MSE}[T]}{dr} = (2r + 2(r-1))\lambda^{-2} = 0 \Leftrightarrow r = \frac{1}{2}.$$

14 d. We bereken eerst de mediaan: het 6de getal (van de geordende dataset met 11 punten) is 0, dus de mediaan is 0. Dan zijn alleen **d**, **e** en **f** nog over. Vervolgens berekenen we het derde kwartiel: $\frac{3}{4} \cdot (11 + 1) = 9$, en het 9de getal is 1, dus het derde kwartiel is 1. Dan blijven alleen **d** en **f** over. Tot slot berekenen we de Inter Quartile Range: het onderste kwartiel is -1 (de waarde van het derde $(= (11 + 1)/4)$ getal), dus de IQR = 2. De omhooggaande "whisker" heeft maximale lengte $1.5 \cdot \text{IQR} = 3$, en $2.5 < 1 + 3$. Dit betekent dat de whisker tot aan 2.5 gaat. Dit leidt tot antwoord **d**. Deze heeft ook een 'outlier' bij 5.

15 e. De kans dat een periode korter duurt dan 1 seconde is gelijk aan $1 - e^{-\lambda}$, en de kans dat een periode langer duurt dan 1 seconde is dus gelijk aan $e^{-\lambda}$. De likelihood van de data is dus gegeven door

$$L(\lambda) = \binom{12}{2} (1 - e^{-\lambda})^{10} (e^{-\lambda})^2.$$

De log-likelihood is dan

$$l(\lambda) = \ln \left(\binom{12}{2} \right) + 10 \ln(1 - e^{-\lambda}) - 2\lambda.$$

Dit maximaliseren naar λ geeft de MLE:

$$l'(\lambda) = \frac{10e^{-\lambda}}{1 - e^{-\lambda}} - 2 = 0 \Leftrightarrow 10e^{-\lambda} = 2 - 2e^{-\lambda} \Leftrightarrow \lambda = \ln(6).$$

16 b. De kans op $(X_1 = 2, X_2 = 0)$ of $(X_1 = 2, X_2 = 1)$ wordt gegeven door

$$L(\mu) = P(X_1 = 2)P(X_2 = 0) + P(X_1 = 2)P(X_2 = 1) = \left(\frac{1}{2}\mu^2 + \frac{1}{2}\mu^3\right)e^{-2\mu}.$$

Dit moeten we maximalizeren om de MLE te vinden:

$$L'(\mu) = 0 \Leftrightarrow (-\mu^3 + \frac{1}{2}\mu^2 + \mu)e^{-3\mu} = 0 \Leftrightarrow 2\mu^2 - \mu - 2 = 0 \text{ of } \mu = 0.$$

Aangezien $\mu = 0$ duidelijk een minimum is, krijgen we twee mogelijke oplossingen:

$$\mu = \frac{1 \pm \sqrt{17}}{4} = 1.281 \text{ of } -0.781.$$

De MLE is dan de positieve waarde, dus $\hat{\mu} = 1.281$.

17 b. Omdat de data een realisatie vormen van een normale verdeling wordt het 90% betrouwbaarheidsinterval voor μ gegeven door $\bar{x}_n \pm t_{n-1, 0.05} \cdot s_n / \sqrt{n}$. Dit levert $1985 \pm 1.860 \cdot 150 / 3 = (1892.00, 2078.00)$.

18 e. De toetsingsgrootte is

$$t = \frac{\hat{\beta} - 2}{s_b} = 2.$$

We verwerpen voor waarden die ver van 0 liggen, want zowel sterk positieve als sterk negatieve waarden wijzen op H_1 . Dit betekent dat het kritieke gebied van de vorm $(-\infty, c) \cup (c, \infty)$ is. Onder H_0 heeft T een $t(n-2) = t(18)$ verdeling. De kans op verwerpen moet 5% zijn als H_0 waar is, dus we moeten hebben dat

$$P(T > c) = 0.025.$$

Dit betekent dat $c = t_{18, 0.025} = 2.101$. Aangezien $t = 2 \in [-2.101, 2.101]$, valt t buiten het kritieke gebied, en verwerpen we H_0 dus niet.

19 f. De standaarddeviatie $\sigma = 1\%$ is bekend, en we weten dat voor de echte onbekende verwachting μ geldt:

$$\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} \text{ heeft een } N(0, 1) \text{ verdeling.}$$

Hieruit volgt dat

$$P(\bar{X}_n - z_{0.025}\sigma / \sqrt{n} \leq \mu \leq \bar{X}_n + z_{0.025}\sigma / \sqrt{n}) = P(\bar{X}_n - z_{0.025} \cdot 0.25 \leq \mu \leq \bar{X}_n + z_{0.025} \cdot 0.25) = 0.95.$$

20 b. We weten dat

$$P\left(\left((p - \hat{p}) \frac{n}{\sqrt{\hat{p}(1 - \hat{p})}}\right)^2 < z_{0.025}^2\right) = 95\%.$$

Verder geldt $\hat{p} = 110 / (8 \cdot 10^6) = 1.375 \cdot 10^{-5}$. Dus het betrouwbaarheidsinterval wordt gegeven door

$$\hat{p} \pm z_{0.025} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} = [1.12 \cdot 10^{-5}, 1.63 \cdot 10^{-5}].$$

21 c. De fabrikant wil graag bewijzen dat $p < 5 \cdot 10^{-7}$, en dus kiest hij zijn alternatieve hypothese $H_1 : p < 5 \cdot 10^{-7}$. Als X het aantal foute bits is, dan wijzen lage waarden van X naar het alternatief. De p-waarde wordt dus gegeven door $P(X \leq 1)$. Onder H_0 geldt dat $X \sim \text{Bin}(10^7, 5 \cdot 10^{-7})$. Dus

$$P(X \leq 1) = P(X = 0) + P(X = 1) = (1 - 5 \cdot 10^{-7})^{10^7} + 10^7 \cdot (1 - 5 \cdot 10^{-7})^{10^7 - 1} \cdot 5 \cdot 10^{-7} = 0.0404.$$

22 c. De toetsingsgrootheid wordt gegeven door

$$T = \frac{\bar{x}_n - 1}{s_n/\sqrt{n}} = -2.3.$$

Er geldt dat $t_{8,0.05} < |T| < t_{8,0.025}$ (merk op dat T een t-verdeling heeft met $n - 1 = 8$ vrijheidsgraden!). Aangezien alleen nog negatievere waarden van T wijzen op de alternatieve hypothese H_1 , geldt dus dat $0.025 < p < 0.05$. Aangezien p kleiner is dan het significantieniveau, verwerpen we H_0 .

23 e. Een fout van type II betekent dat $v > 120$, en toch verwerpen we de nul hypothese niet. Er is gegeven dat we niet verwerpen als $X < 128$. Dus

$$P(X < 128) = P\left(\frac{X - v}{5} < \frac{128 - v}{5}\right) = P\left(Z < \frac{128 - v}{5}\right) \geq P\left(Z < \frac{128 - 133}{5}\right).$$

De laatste stap volgt omdat $v \leq 133$ (de automobilist rijdt langzamer dan 133 km/u). Dus

$$P_{II} \geq P(Z < -1) = P(Z > 1) \approx 0.159.$$

Een andere manier om dit te zien is door eerst uit te rekenen wat P_{II} is als je precies 133 km/u rijdt, en je dan te realiseren dat als je zachter rijdt, heb je *meer* kans om geen boete te krijgen.

24 f. De waarde van de toetsingsgrootheid is

$$t = \frac{\bar{x}_n - 4}{s_n/\sqrt{n}} = \frac{3.64 - 4}{2/\sqrt{100}} = -1.80.$$

Aangezien zowel sterk positieve als sterk negatieve waarden wijzen op het alternatief, is de p-waarde gelijk aan $2 \cdot P(Z > 1.80) = 0.0718$. Dit is hoger dan 5%, dus we verwerpen H_0 niet.

25 c. We hebben voor studie 1 dat $n = 100$, $\bar{x}_n = 3.64$ en $s_X = 2$. Dus $s_d^2 = s_X^2/n + s_Y^2/m = 0.0521$, dus $s_d = 0.228$. De toetsingsgrootheid is

$$t = \frac{\bar{x}_n - \bar{y}_m}{s_d} = -2.45.$$

Door de grote waarde van n en m , weten we dat t onder de nulhypothese in goede benadering een $N(0, 1)$ verdeling heeft. Verder verwerpen we H_0 alleen voor sterk negatieve waarden van t , dus de p-waarde is gelijk aan

$$P(Z < -2.45) = P(Z > 2.45) = 0.0071.$$

Tentamen Statistiek
wi1102CT, 22 juni 2010

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

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. Uit een studie blijkt dat 57% van de studenten civiele techniek werk zal vinden dat in het verlengde ligt van hun afstudeerrichting. Dan is de kans, dat onder 8 willekeurig gekozen studenten civiele techniek er *tenminste* 1 student werk zal vinden dat in het verlengde ligt van zijn of haar afstudeerrichting, gelijk aan
a. 0.011 b. 0.125 c. 0.570 d. 0.911 e. 0.989 f. 0.999
2. Piet speelt elke maand mee in twee loterijen, die onafhankelijk van elkaar worden gehouden en zodanig dat de maandelijkse uitslagen onderling onafhankelijk zijn. In de ene loterij heeft hij kans $1/2000$ op een prijs van meer dan 1000 euro, en in de andere loterij is de kans op een dergelijke prijs gelijk aan $1/3000$. Noem X het aantal maanden totdat hij een keer zo'n grote prijs wint. De verwachting van X is ongeveer gelijk aan
a. 600 b. 750 c. 900 d. 1200 e. 2000 f. 3000
3. Een docent ziet dat een bepaalde moeilijke vraag tijdens een multiple choice tentamen met 6 keuze mogelijkheden, door relatief weinig studenten goed is beantwoord. De docent wil graag weten hoeveel studenten het antwoord echt wisten. Hij gebruikt het volgende model: een student heeft kans p om het goede antwoord te *weten*. Als de student het antwoord niet weet, gokt hij of zij volledig willekeurig een van de zes antwoorden. Noem q de kans dat een student het juiste antwoord *geeft*. Welke van de volgende uitspraken is juist volgens het model van de docent?
a. $p = \frac{5}{6}q$ b. $p = \frac{1}{6}q + \frac{1}{6}$ c. $p = q - \frac{1}{6}$
d. $p = q$ e. $p = \frac{6}{5}(q - \frac{1}{6})$ f. $p = \frac{6}{5}q - \frac{1}{6}$
4. Het gewicht van pindakaas in een pot is normaal verdeeld, met verwachting 120g en standaarddeviatie 3g. De pot zelf heeft een gewicht dat ook normaal verdeeld is, met verwachting 60g en standaarddeviatie 2g. Het gewicht van de pot zelf is onafhankelijk van het gewicht van de pindakaas in de pot. Wat is de kans dat 10 potten pindakaas in totaal meer wegen dan 1820g? U mag gebruiken dat de som van onafhankelijke normale verdelingen, weer normaal verdeeld is.
a. 0.0020 b. 0.0397 c. 0.1038 d. 0.3446 e. 0.9000 f. 0.9980
5. Stel de stochast X heeft de dichtheid

$$f(x) = \begin{cases} \frac{3}{8}x^2 & \text{als } 0 \leq x \leq 2 \\ 0 & \text{anders.} \end{cases}$$

Dan is $P(\frac{2}{3} \leq X \leq \frac{4}{3})$ gelijk aan

- a. $\frac{1}{3}$ b. $\frac{1}{2}$ c. $\frac{2}{9}$ d. $\frac{2}{3}$ e. $\frac{7}{27}$ f. $\frac{4}{9}$

6. De verdelingsfunctie van een stochast X wordt gegeven door

$$F(x) = \begin{cases} 0, & \text{als } x \leq 0 \\ \frac{1}{3}\sqrt{x}, & \text{als } 0 \leq x \leq 9 \\ 1, & \text{als } x \geq 9. \end{cases}$$

De mediaan van X is:

- a. $\frac{1}{2}$ b. 1 c. $\frac{9}{4}$ d. 3 e. $\frac{9}{2}$ f. 5

7. Stel U_1 en U_2 zijn onafhankelijke stochasten met een $U(0, 1)$ -verdeling. Dan is de variantie van $2 + 3U_1 - U_2$ is gelijk aan

- a. $\frac{2}{12}$ b. $\frac{4}{12}$ c. $\frac{8}{12}$ d. $\frac{10}{12}$ e. $\frac{26}{12}$ f. 6

8. De verdeling van een discrete stochast X wordt gegeven door

a	0	1	2
$P(X = a)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Bepaal de verwachting en de variantie van X .

- a. $E[X] = \frac{3}{2}$ en $\text{Var}(X) = \frac{1}{2}$ b. $E[X] = 1$ en $\text{Var}(X) = \frac{1}{2}$
c. $E[X] = \frac{3}{2}$ en $\text{Var}(X) = 2$ d. $E[X] = 1$ en $\text{Var}(X) = 2$
e. $E[X] = \frac{3}{2}$ en $\text{Var}(X) = 1$ f. $E[X] = 1$ en $\text{Var}(X) = 1$

9. Een goede benadering voor de kansverdeling van het gemiddelde $\bar{X}_{400}$ van 400 onafhankelijke $Ber(0.25)$ verdeelde stochasten is

- a. $N(0.25, 4.69 \cdot 10^{-4})$ b. $N(0.25, 2.5 \cdot 10^{-3})$ c. $N(0.50, 2.08 \cdot 10^{-4})$
d. $N(0.50, 6.25 \cdot 10^{-4})$ e. $N(1.0, 4.69 \cdot 10^{-4})$ f. $N(1.0, 6.25 \cdot 10^{-4})$

10. Van de stochastische variabele X is gegeven dat $E[X] = 1$ en $\text{Var}(X) = 4$. Van Y is gegeven dat $E[Y^2] = 4$. Dan worden $E[X^2]$ en $E[-2X + 3Y^2]$ gegeven door

- a. $E[X^2] = 4$ en $E[-2X + 3Y^2] = 6$
b. $E[X^2] = 3$ en $E[-2X + 3Y^2] = 6$
c. $E[X^2] = 5$ en $E[-2X + 3Y^2] = 10$
d. $E[X^2] = 4$ en $E[-2X + 3Y^2] = 10$
e. $E[X^2] = 3$ en $E[-2X + 3Y^2] = 10$
f. $E[X^2] = 5$ en $E[-2X + 3Y^2] = 6$

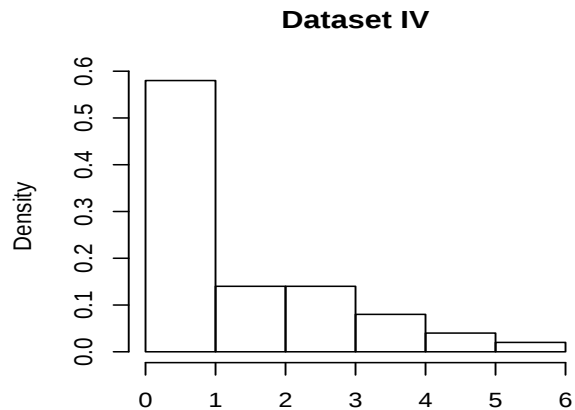
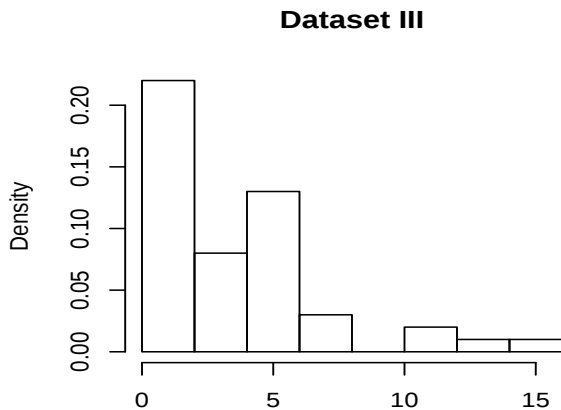
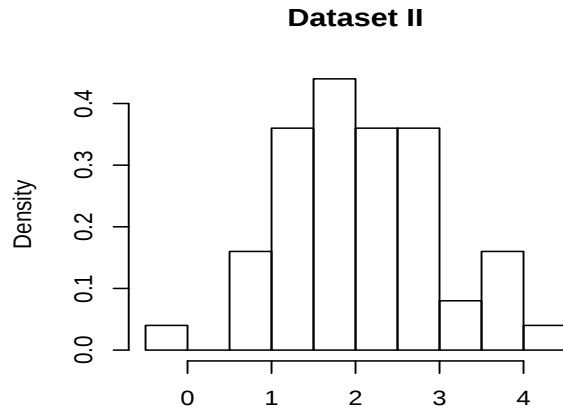
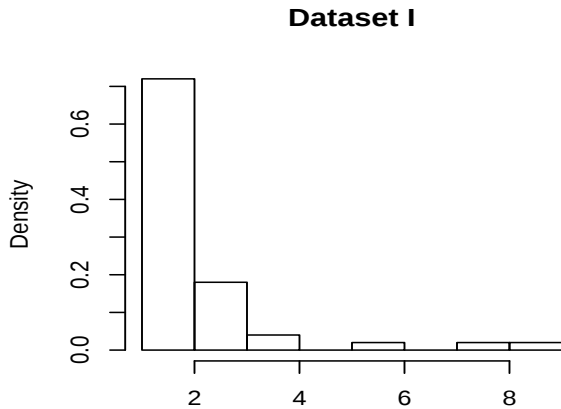
11. Voor de stochastische variabelen $X_1, X_2, \dots, X_n$ geldt $E[X_i] = \mu$ voor alle $i = 1, 2, \dots, n$, waarbij $\mu \in \mathbb{R}$ onbekend is. Voor welke $a, b \in \mathbb{R}$ is

$$T = a(X_1 + X_2 + \dots + X_n) + b$$

een zuivere schatter voor μ ?

- a. $a = 0, b = 0$ b. $a = 0, b = \mu$ c. $a = 1/n, b = \mu$
d. $a = 1/n, b = 0$ e. $a = 1, b = 0$ f. $a = 1, b = \mu$

12. Van vier datasets ter grootte 50 is het histogram weergegeven. De datasets zijn steekproeven uit vier verschillende verdelingen: een $Exp(0.8)$, een $Exp(1/3)$, een $Par(2)$ en een $N(2, 1)$ verdeling. Welke dataset komt uit de $Par(2)$ en welke dataset komt uit de $Exp(1/3)$ verdeling?



- | | |
|--|--|
| <p>a. $Par(2)$: IV; $Exp(1/3)$: I</p> <p>c. $Par(2)$: II; $Exp(1/3)$: IV</p> <p>e. $Par(2)$: I; $Exp(1/3)$: IV</p> | <p>b. $Par(2)$: III; $Exp(1/3)$: IV</p> <p>d. $Par(2)$: II; $Exp(1/3)$: III</p> <p>f. $Par(2)$: I; $Exp(1/3)$: III</p> |
|--|--|
13. Zij $X_1, X_2, \dots, X_n$ onafhankelijke stochastische variabelen, allemaal met dezelfde verwachting μ en variantie σ^2 . Een zuivere schatter voor μ is $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Zij Y een stochastische variabele onafhankelijk van $X_1, X_2, \dots, X_n$ wederom met verwachting μ en variantie σ^2 . Dan is voor elke $r \in [0, 1]$, de combinatie

$$T = r\bar{X}_n + (1 - r)Y$$

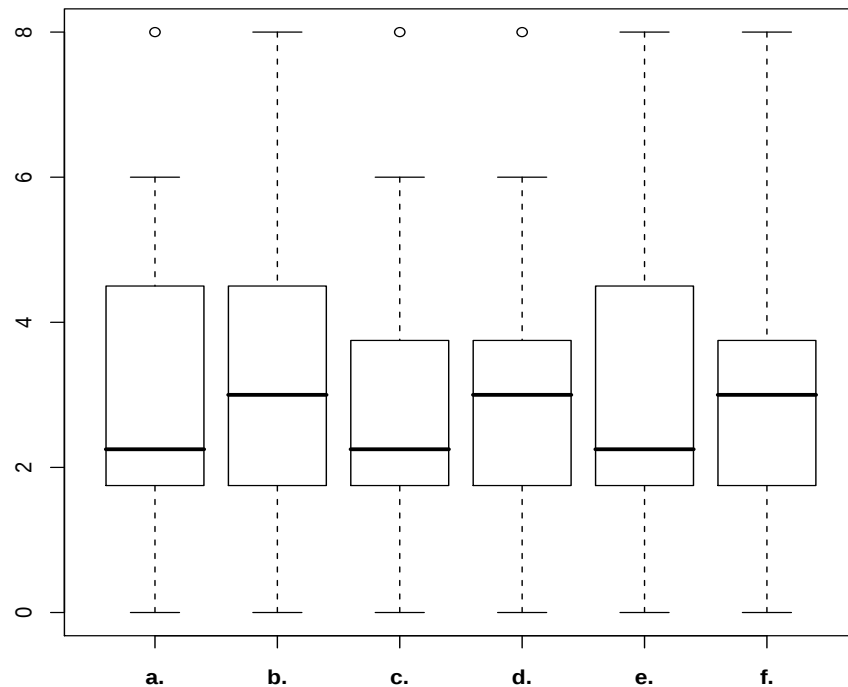
een zuivere schatter voor μ . Welke waarde voor $r \in [0, 1]$ leidt tot de beste (in de zin van MSE) schatter voor μ ?

- | | | |
|--------------------|--------------------|--------------|
| a. $r = 0$ | b. $r = 1/n$ | c. $r = 1/2$ |
| d. $r = n/(n + 1)$ | e. $r = (n - 1)/n$ | f. $r = 1$ |

14. Gegeven is de dataset

0.00 1.5 1.75 1.75 2.00 3.00 3.25 3.75 3.75 6.00 8.00

Welke van de zes onderstaande boxplots hoort bij deze dataset?



15. Een geneticus heeft bevonden dat hypochondrilla's bruine, blauwe of groene ogen hebben, welke voorkomen met kansen respectievelijk 2θ , θ en $(1 - 3\theta)$, voor $\theta \in [0, \frac{1}{3}]$. Een bioloog vindt in een steekproef van 30 hypochondrilla's, 9 exemplaren met bruine ogen, 6 met blauwe en 15 met groene ogen. Noteer C als het aantal combinaties met 9 keer bruin, 6 keer blauw en 15 keer groen. Dan is de likelihood voor de data

- a. $L(\theta) = C \cdot 1620 \cdot \theta^2(1 - 3\theta)$ b. $L(\theta) = C \cdot 2^9 \cdot (\theta - 3\theta^2)^{15}$
c. $L(\theta) = C \cdot (15 - 21\theta)$ d. $L(\theta) = C \cdot ((2\theta)^9 + \theta^6 + (1 - 3\theta)^{15})$
e. $L(\theta) = C \cdot (\frac{1}{9}(2\theta) + \frac{1}{6}\theta + \frac{1}{15}(1 - 3\theta))$ f. niet te berekenen

16. Men schat een onbekende parameter $0 \leq \theta \leq 1$ met de methode van maximum likelihood. Gegeven is dat de likelihood voor de data gelijk is aan

$$L(\theta) = 32\theta^3 - 54\theta^2 + 21\theta + 1$$

Dan is de maximum likelihood schatting voor θ gelijk aan

- a. 0 b. $\frac{1}{8}$ c. $\frac{1}{4}$ d. $\frac{1}{2}$ e. $\frac{7}{8}$ f. 1

17. Voor een bivariate dataverzameling $(x_1, y_1), (x_2, y_2), \dots, (x_{20}, y_{20})$ beschouwen we een lineair regressiemodel $Y_i = \alpha + \beta x_i + U_i$, waarbij $U_1, U_2, \dots, U_{20}$ onafhankelijk zijn en $N(0, \sigma^2)$ verdeeld. We vinden voor onze data

$$\hat{\alpha} = 3, \quad \hat{\beta} = 1.5, \quad s_a = 0.5, \quad s_b = 0.25.$$

We toetsten $H_0 : \beta = 1$ tegen $H_1 : \beta > 1$ door middel van een t -toets bij significantieniveau 0.05. Het kritieke gebied voor deze toets is van de vorm $[c, \infty)$, voor een zekere $c \in \mathbb{R}$. Geef de waarde van de toetsingsgrootheid t en de kritieke waarde c .

- | | | |
|---------------------------|---------------------------|---------------------------|
| a. $t = 2$ en $c = 1.734$ | b. $t = 2$ en $c = 2.101$ | c. $t = 4$ en $c = 1.645$ |
| d. $t = 4$ en $c = 2.101$ | e. $t = 6$ en $c = 1.645$ | f. $t = 6$ en $c = 1.734$ |

18. Een afvulmachine van een frisdrankfabriek vult flessen met een hoeveelheid drank die normaal verdeeld is met verwachtingswaarde μ en standaardafwijking $\sigma = 5$ ml. Van 25 flessen is de gemiddelde inhoud $\bar{x}_n = 1502$ ml. Het 95% betrouwbaarheidsinterval voor de verwachte inhoud van een fles wordt gegeven door

- | | |
|---|---|
| a. $1502 \pm z_{0.025} \cdot 4/\sqrt{24}$ | b. $1502 \pm t_{0.05,24} \cdot 4/\sqrt{24}$ |
| c. $1502 \pm t_{0.05,24}$ | d. $1502 \pm t_{0.025,24}$ |
| e. $1502 \pm z_{0.05}$ | f. $1502 \pm z_{0.025}$ |

19. De inspectie voor de volksgezondheid doet 25 metingen aan de concentratie van een giftige stof in grondwater. De metingen leiden tot een gemiddelde concentratie van $\bar{x}_{25} = 2.25$ ppm en een steekproefvariantie van $s_{25}^2 = 0.25$. Men berekent een 95% betrouwbaarheidsinterval voor de verwachte concentratie μ in het grondwater onder de aanname dat de 25 metingen een realisatie vormen van een steekproef uit een normale verdeling met verwachting μ en (tevens onbekende) variantie σ^2 . Het interval is

- | | | |
|-----------------|-----------------|-----------------|
| a. [2.21, 2.29] | b. [2.15, 2.35] | c. [2.08, 2.42] |
| d. [2.04, 2.46] | e. [1.39, 3.11] | f. [1.22, 3.28] |

20. Om een idee te krijgen hoe vaak rood-groen kleurenblindheid voorkomt bij mannelijke studenten, worden 80 mannelijke studenten getest. Er blijken 10 studenten dit type kleurenblindheid te hebben. Iemand bepaalt met de Wilson-methode een 95% betrouwbaarheidsinterval voor p , de kans dat een student rood-groen kleurenblind is. Deze persoon moet als betrouwbaarheidsinterval vinden

- | | | |
|-------------------|-------------------|-------------------|
| a. [0.053, 0.197] | b. [0.069, 0.215] | c. [0.076, 0.197] |
| d. [0.105, 0.145] | e. [0, 0.076] | f. [0.101, 0.149] |

21. Een studentenblad beweert dat meer dan 25% van de studenten ten minste eenmaal per week pizza eten. Hanna en Kees vragen zich af of hier echt voldoende bewijs voor is. Zij vragen 100 studenten of zij tenminste eenmaal per week pizza eten; het aantal dat “ja” antwoordt noemen we X . Het blijkt dat $X = 31$. Noem de fractie studenten die tenminste één maal per week pizza eten p . Hanna en Kees toetsten de nul-hypothese $H_0 : p = 0.25$ op basis van de toetsingsgrootheid X . Kies de toepasselijke alternatieve hypothese en de juiste p -waarde:

- | |
|---|
| a. $H_1 : p > 0.25$, p -waarde = $P(X \geq 31)$, waarbij $X \sim \text{Bin}(100, 0.25)$. |
| b. $H_1 : p \neq 0.25$, p -waarde = $P(X \geq 31) + P(X \leq 19)$, waarbij $X \sim \text{Bin}(100, 0.25)$. |
| c. $H_1 : p > 0.25$, p -waarde = $P(X \geq 31)$, waarbij $X \sim \text{Bin}(100, 0.31)$. |
| d. $H_1 : p \neq 0.25$, p -waarde = $P(X \geq 31) + P(X \leq 19)$, waarbij $X \sim \text{Bin}(100, 0.31)$. |
| e. $H_1 : p > 0.25$, p -waarde = $P(X \geq 25)$, waarbij $X \sim \text{Bin}(100, 0.31)$. |
| f. $H_1 : p \neq 0.25$, p -waarde = $P(X \neq 25)$, waarbij $X \sim \text{Bin}(100, 0.31)$. |

22. We toetsen een hypothese H_0 op significantieniveau 0.05. Dit betekent dat:
- als H_0 onjuist is, de kans op verwerpen van H_0 minstens 0.95 is.
 - als H_1 onjuist is, de kans op verwerpen van H_1 minstens 0.95 is.
 - als H_0 juist is, de kans op niet verwerpen van H_0 minstens 0.95 is.
 - als H_1 juist is, de kans op niet verwerpen van H_1 minstens 0.95 is.
 - als H_0 onjuist is, de kans op verwerpen van H_0 minstens 0.05 is.
 - als H_0 juist is, de kans op niet verwerpen van H_0 minstens 0.05 is.
23. Men trekt een getal X uit een uniforme verdeling op het interval $[0, \theta]$. Men toetst $H_0 : \theta = 2$ tegen $H_1 : \theta \neq 2$ en verwerpt H_0 ten gunste van H_1 als $X \leq 0.1$ of als $X \geq 1.9$. Als $\theta = 4$, dan is de kans op een type II fout gelijk aan
- 0.2
 - 0.45
 - 0.475
 - 0.525
 - 0.55
 - 0.8
24. Men beschikt over 22 metingen van de trekkracht van een bepaald soort materiaal. De metingen worden opgevat als trekkingen uit een normale verdeling met verwachting μ en onbekende variantie. Het steekproefgemiddelde is 10.71 en de steekproefvariantie is 3.55. Men wil weten of de trekkracht van het materiaal groter is dan 10, en voert een t -toets uit voor de nulhypothese $H_0 : \mu = 10$ bij significantieniveau 0.05. Geef de waarde van de toetsingsgrootte en vermeldt uw conclusie.
- $t = 0.938$; H_0 verwerpen
 - $t = 0.938$; H_0 niet verwerpen
 - $t = 1.767$; H_0 verwerpen
 - $t = 1.767$; H_0 niet verwerpen
 - $t = 4.4$; H_0 verwerpen
 - $t = 4.4$; H_0 niet verwerpen
25. We willen onderzoeken of de kans op orgaandonatie in Nederland verschilt van die in Spanje. Gegeven is dat van $n = 3048$ potentiële donors in Nederland, 212 uiteindelijk een orgaan hebben gedoneerd. In Spanje hebben van de $m = 13040$ potentiële donors er 1152 uiteindelijk gedoneerd. We vatten de Nederlandse dataset op als realisaties van stochasten $X_1, X_2, \dots, X_{3048}$, waarbij

$$X_i = \begin{cases} 1 & \text{als potentiële Nederlandse donor } i \text{ inderdaad doneert} \\ 0 & \text{als potentiële Nederlandse donor } i \text{ niet doneert.} \end{cases}$$

Noem p de kans dat een potentiële donor in Nederland inderdaad doneert. Dan geldt dus dat $E[X_i] = p$. Analooft definiëren we $Y_1, Y_2, \dots, Y_{13040}$ voor de Spaanse dataset, en we definiëren $q = E[Y_j]$ als de kans dat een Spaanse potentiële donor inderdaad doneert. We toetsen de nulhypothese $H_0 : p = q$ tegen de alternatieve hypothese $H_1 : p \neq q$ met behulp van een twee-steekproeven t -toets, gebruik makend van de pooled variance $s_p^2 = 3.54 \cdot 10^{-5}$. Geef de p -waarde.

- 0.0004
- 0.0008
- 0.013
- 0.026
- 0.1032
- 0.2064

Antwoorden multiple choice:

1 f.

2 d.

3 e.

4 b.

5 e.

6 c.

7 d.

8 b.

9 a.

10 c.

11 d.

12 f.

13 d.

14 d.

15 b.

16 c.

17 a.

18 f.

19 d.

20 b.

21 a.

22 c.

23 b.

24 c.

25 b.

Uitwerkingen multiple choice:

1 f. Zij X het aantal studenten van de acht gekozen studenten die inderdaad werk vinden in de richting van hun afstuderen. Dan heeft X een $Bin(8, 0.57)$ verdeling. Gevraagd is

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (1 - 0.57)^8 = 0.999.$$

2 d. Vanwege de maandelijkse onafhankelijkheid heeft X heeft een geometrische kansverdeling met als parameter p de kans dat Piet in een maand zo'n grote prijs wint. Omdat de loterijen onafhankelijk van elkaar opereren is deze kans

$$p = 1 - P(\text{verlies}) = 1 - \left(1 - \frac{1}{2000}\right) \left(1 - \frac{1}{3000}\right) = 0.00083.$$

Dit betekent dat X een $Geo(p)$ verdeling heeft met $p = 0.00083$, dus $E[X] = 1/p \approx 1200$.

3 e. Er geldt dat een student een vraag goed beantwoordt als hij het weet, of als hij het niet weet en goed gokt. Dus

$$q = p + (1 - p) \cdot \frac{1}{6} \quad \Leftrightarrow \quad p = \frac{6}{5} \left(q - \frac{1}{6}\right).$$

4 b. Het totale gewicht van een pot is normaal verdeeld met verwachting $120 + 60 = 180$. De variantie van het totale gewicht is de som van de varianties: $3^2 + 2^2 = 13$. Het totale gewicht W van tien potten is weer normaal verdeeld en heeft verwachting $10 \cdot 180 = 1800$, en variantie $10 \cdot 13 = 130$. Er geldt dus

$$P(W > 1820) = P\left(\frac{W - 1800}{\sqrt{130}} > \frac{1820 - 1800}{\sqrt{130}}\right) = P(Z > 1.75) = 0.0397.$$

5 e.

$$P\left(\frac{2}{3} \leq X \leq \frac{4}{3}\right) = \int_{\frac{2}{3}}^{\frac{4}{3}} \frac{3}{8} x^2 dx = \left[\frac{1}{8} x^3\right]_{\frac{2}{3}}^{\frac{4}{3}} = \frac{1}{8} \left(\frac{64}{27} - \frac{8}{27}\right) = \frac{7}{27}.$$

6 c. De mediaan m van X is de oplossing van:

$$F(m) = \frac{1}{2} \quad \Leftrightarrow \quad \frac{1}{3} \sqrt{m} = \frac{1}{2} \quad \Leftrightarrow \quad m = \frac{9}{4}.$$

7 d. Omdat U_1 en U_2 onafhankelijk zijn, zijn $3U_1$ en $-U_2$ dat ook. Dit betekent dat de variantie van hun som gelijk is aan de som van de varianties:

$$\text{Var}(2 + 3U_1 - U_2) = \text{Var}(3U_1 - U_2) = \text{Var}(3U_1) + \text{Var}(-U_2) = 9\text{Var}(U_1) + (-1)^2\text{Var}(U_2) = \frac{10}{12}.$$

8 b. Er geldt

$$E[X] = \frac{1}{4} \cdot 0 + \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 = 1.$$

Net zo geldt

$$E[X^2] = \frac{1}{4} \cdot 0^2 + \frac{1}{2} \cdot 1^2 + \frac{1}{4} \cdot 2^2 = \frac{3}{2}.$$

$$\text{Dus } \text{Var}(X) = E[X^2] - E[X]^2 = \frac{1}{2}.$$

9 a. Volgens de centrale limietstelling geldt

$$\frac{\bar{X}_{400} - \mu}{\sigma/\sqrt{400}} \approx N(0, 1),$$

hetgeen hetzelfde is als

$$\bar{X}_{400} \approx N\left(\mu, \frac{\sigma^2}{400}\right).$$

Een $Ber(0.25)$ stochast heeft verwachting $\mu = 0.25$ en variantie $\sigma^2 = 0.1625 (= \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16})$. Dus worden 0.25 en $\frac{1}{400} \times \frac{3}{16} = 4.69 \cdot 10^{-4}$ de parameters van de benaderende normale verdeling.

10 c. We weten dat

$$\text{Var}(X) = E[X^2] - (E[X])^2,$$

waaruit volgt dat $E[X^2] = \text{Var}(X) + (E[X])^2 = 4 + 1^2 = 5$. Verder volgt uit de lineariteit van de verwachting dat

$$E[-2X + 3Y^2] = -2E[X] + 3E[Y^2] = 10.$$

11 d. Schatter T is zuiver als $E[T] = \mu$ voor alle $\mu \in \mathbb{R}$. Er geldt vanwege de lineariteit van de verwachting dat

$$E[T] = E[a(X_1 + X_2 + \dots + X_n) + b] = a(E[X_1] + E[X_2] + \dots + E[X_n]) + b = an\mu + b.$$

We moeten nu a en b kiezen zodat voor alle $\mu \in \mathbb{R}$ geldt dat

$$an\mu + b = \mu.$$

Dit kan alleen als $a = 1/n$ en $b = 0$ (vul maar $\mu = 0$ en $\mu = 1/n$ in).

12 f. De Paretovariabele neemt waarden ≥ 1 aan; dat hoort bij I. Een exponentiële dichtheid $f(x) = \lambda e^{-\lambda x}$ is gelijk aan λ in $x = 0$. Omdat $\lambda = 1/3$ en het histogram moet overeenkomen met de dichtheid, komt plaatje III het meest in aanmerking.

Alternatief: de $Exp(1/3)$ verdeling heeft een bijhorende kans om groter dan 6 te zijn: als $X \sim Exp(1/3)$ verdeeld is, dan geldt

$$P(X > 6) = \int_6^\infty \frac{1}{3} e^{-\frac{1}{3}x} dx = \left[-e^{-\frac{1}{3}x}\right]_6^\infty = 0.135.$$

Dit is alleen maar te zien in plaatje III.

13 d. Aangezien T een zuivere schatter is, geldt

$$\text{MSE}[T] = \text{Var}(T) = \text{Var}(r\bar{X}_n + (1-r)Y) = r^2\text{Var}(\bar{X}_n) + (1-r)^2\text{Var}(Y) = \frac{1}{n}r^2\sigma^2 + (1-r)^2\sigma^2.$$

We moeten nu r zo kiezen dat $\text{MSE}[T]$ wordt geminimaliseerd. Dit doen we door de afgeleide 0 te stellen:

$$\frac{d\text{MSE}[T]}{dr} = 2\frac{1}{n}\sigma^2r - 2\sigma^2(1-r) = 0 \Leftrightarrow r = \frac{n}{n+1}.$$

14 d. We bereken eerst de mediaan: het 6de getal (van de geordende dataset met 11 punten) is 3, dus de mediaan is 3. Dan zijn alleen **b**, **d** en **f** nog over. Vervolgens berekenen we het derde kwartiel: $\frac{3}{4} \cdot 11 = 8.25$ (volgens het boek: $\frac{3}{4} \cdot (11 + 1) = 9$), en het 8ste en 9de getal zijn beide 3.75, dus het derde kwartiel is 3.75. Dan blijven alleen **d** en **f** over. Tot slot berekenen we de Inter Quartile Range: het onderste kwartiel is 1.75 (de waarde van het derde en vierde getal), dus de $\text{IQR} = 2$. De omhooggaande "whisker" heeft maximale lengte $1.5 \cdot \text{IQR} = 3$, en $6 < 3.75 + 3 < 8$. Dit betekent dat de whisker tot aan 6 gaat. Dit leidt tot antwoord **d**.

15 b. De likelihood voor de data is gelijk aan de kans op de data. De kans om eerst 9 bruine ogen, dan 6 blauwe ogen en dan 15 groene ogen is gelijk aan

$$(2\theta)^9 \theta^6 (1 - 3\theta)^{15}.$$

De echte kans op de data is de factor C keer deze kans:

$$L(\theta) = C(2\theta)^9 \theta^6 (1 - 3\theta)^{15} = C \cdot 2^9 \cdot (\theta - 3\theta^2)^{15}.$$

16 c. De maximum likelihood schatting wordt gevonden door $L(\theta)$ te maximaliseren. De afgeleide is

$$L'(\theta) = 96\theta^2 - 108\theta + 21 = 96 \left(\theta - \frac{1}{4} \right) \left(\theta - \frac{7}{8} \right).$$

Deze is gelijk aan nul voor $\theta = \frac{1}{4}$ en $\theta = \frac{7}{8}$. Aangezien $L'(\theta)$ een dalparabool is, leidt het tekenverloop van L' er toe dat het maximum ligt bij $\theta = \frac{1}{4}$. Eventuele andere maxima zouden rand-maxima kunnen zijn, maar $L(1) = 0 < L(0) = 1 < L(\frac{1}{4})$.

17 a. De toetsingsgrootheid is

$$t = \frac{\hat{\beta} - 1}{s_b} = 2.$$

We verwerpen als $t \in (c, \infty)$. De kans op een fout van type I moet gelijk zijn aan 5%, dus $P(t > c) = 0.05$. t heeft een $t(18)$ verdeling onder de nulhypothese, dus $c = t_{18,0.05} = 1.734$.

18 f. De variantie $\sigma^2 = 25$ is bekend, en we weten dat voor de echte onbekende verwachting μ geldt:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \text{ heeft een } N(0, 1) \text{ verdeling.}$$

Hieruit volgt dat

$$P(\bar{X}_n - z_{0.025}\sigma/\sqrt{n} \leq \mu \leq \bar{X}_n + z_{0.025}\sigma/\sqrt{n}) = P(\bar{X}_n - z_{0.025} \leq \mu \leq \bar{X}_n + z_{0.025}) = 0.95.$$

19 d. We weten dat voor de echte onbekende verwachting μ geldt:

$$\frac{\bar{X}_n - \mu}{s_n/\sqrt{n}} \text{ heeft een } t(n-1) \text{ verdeling.}$$

Hieruit volgt dat

$$P(\bar{X}_n - t_{0.025,24}s_n/\sqrt{n} \leq \mu \leq \bar{X}_n + t_{0.025,24}s_n/\sqrt{n}) = 0.95.$$

Dit leidt tot $[2.25 - 2.064 \cdot 0.5/5, 2.25 + 2.064 \cdot 0.5/5] = [2.04, 2.46]$.

20 b. We gebruiken als schatter $\hat{p}$ de fractie studenten die rood-groen kleurenblind blijken te zijn. We weten dat dan in goede benadering geldt:

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \text{ heeft een } N(0, 1) \text{ verdeling,}$$

met $n = 80$. Hieruit volgt dat

$$P\left((\hat{p} - p)^2 \leq \frac{z_{0.025}^2}{n} p(1-p)\right) = 0.95.$$

De data geeft $\hat{p} = 0.125$. We zoeken dus alle $p \in [0, 1]$ zodat

$$(0.125 - p)^2 \leq \frac{1.96^2}{80} p(1-p) \Leftrightarrow 1.048 p^2 - 0.298 p + 0.0156 \leq 0 \Leftrightarrow p \in [0.069, 0.215].$$

21 a. De bewering die Hanna en Kees proberen te bewijzen is dat $p > 0.25$, dus $H_1 : p > 0.25$. Dit betekent dat we H_0 willen verwerpen voor grote waarden van X , niet voor kleine waarden. De p-waarde is dus $P(X \geq 31)$, waarbij X de verdeling heeft onder de nul-hypothese, dus X is $Bin(100, 0.25)$ verdeeld.

22 c. Het significantieniveau is de maximaal toelaatbare kans op een type I fout, waarbij een type I fout overeenkomt met verwerpen van H_0 , terwijl deze in werkelijkheid juist is. Als het significantieniveau 0.05 is, dan betekent dit dus dat in het geval dat H_0 juist is, de kans op verwerpen hoogstens 0.05 en dus is de kans op niet verwerpen minstens 0.95.

23 b. De kans op een type II fout bij $\theta = 4$ is

$$P(\text{niet verwerpen} \mid \theta = 4) = P(0.1 < X < 1.9 \mid \theta = 4) = \frac{1.9 - 0.1}{4} = 0.45.$$

24 c. De waarde van de toetsingsgrootte is

$$t = \frac{\bar{x}_n - 10}{s_n/\sqrt{n}} = \frac{10.71 - 10}{\sqrt{3.55}/\sqrt{22}} = 1.767.$$

Omdat men wil weten of de trekkracht van het materiaal groter is dan 10 Newton, toetst men tegen het alternatief $H_1 : \mu > 10$ en hebben we alleen een rechter kritieke waarde $t_{n-1,0.05} = t_{21,0.05} = 1.721$. Aangezien de waarde van de toetsingsgrootte groter is, zal men H_0 verwerpen.

25 b. We hebben $s_p = 0.00595$. De toetsingsgrootte is

$$t = \frac{\bar{x}_n - \bar{y}_m}{s_p} = \frac{\frac{212}{3048} - \frac{1152}{13040}}{0.00595} = -3.36.$$

Door de grote waarde van n en m , weten we dat t onder de nulhypothese een $N(0, 1)$ verdeling heeft. Verder verwerpen we H_0 voor zowel sterk positieve als sterk negatieve waarden van t , dus de p-waarde is gelijk aan

$$P(Z < -3.36) + P(Z > 3.36) = 2P(Z > 3.36) = 0.0008.$$

Tentamen Kansrekening en statistiek wi3102CT

23 januari 2009, 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.

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. De gebeurtenis $(A^c \cup B^c)^c \cup (A^c \cup B)^c$ is gelijk aan
a. A b. B c. $A \cup B$ d. $A \cap B$ e. $\emptyset$ f. Ω
2. De gebeurtenissen A and B zijn onafhankelijk en bovendien is bekend dat $P(A | B) = 1/2$ and $P(B | A) = 1/3$. Dan is $P(A \cup B)$ gelijk aan
a. $1/6$ b. $1/3$ c. $1/2$ d. $2/3$ e. $5/6$ f. 1
3. Een verzekeringsmaatschappij heeft voor brandverzekeringen hoog-risico en laag-risico klanten, die kansen 0.02 respectievelijk 0.0025 hebben op het indienen van een claim in een gegeven jaar. Er geldt dat 30% van de klanten een hoog risico heeft en 70% een laag risico. Wat is de kans dat gegeven dat een klant een claim indient, deze een hoog-risico klant is?
a. 12% b. 54% c. 77% d. 30% e. 63% f. 42%
4. IKEA heeft 1000 Billy boekenkasten in het magazijn. De kans dat er te weinig schroeven zitten in het bouw pakket van zo'n boekenkast is 0.1%. Hoe groot is de kans dat er in het magazijn precies twee boekenkasten staan met te weinig schroeven?
a. 0.184 b. 0.368 c. 0.512 d. 0.772 e. 0.801 f. 0.920
5. Gegeven is een continue stochastische variabele X , die waarden aanneemt op $[0, 1]$, met kansdichtheid $f(x) = c(x - x^3)$, voor $0 \leq x \leq 1$. Dan is c gelijk aan
a. $\frac{1}{2}$ b. $\frac{1}{4}$ c. -1 d. 1 e. 2 f. 4
6. Zij X een stochast met $E[X] = 2$ en $\text{Var}(X) = 4$. Dan is de verwachtingswaarde van $(3 - 2X)^2$ gelijk aan
a. 12 b. 13 c. 14 d. 15 e. 16 f. 17
7. Gegeven is een stochastische variabele X met een $\text{Exp}(2)$ verdeling. De kansdichtheid van de stochast $Y = e^X$ wordt gegeven door $f(y) = 0$ voor $y < 1$ en voor $y \geq 1$
a. e^{1-y} b. $\frac{1}{2}e^{(1-y)/2}$ c. $2(y - 1)$ d. $1/y^2$ e. $2/y^2$ f. $2/y^3$

8. Gegeven is dat de gezamenlijke kansdichtheid van (X, Y) gelijk is aan

$$f(x, y) = \frac{12}{5}xy(1 + y)$$

op het gebied $0 \leq x \leq 1$ en $0 \leq y \leq 1$, en $f(x, y) = 0$ anders. Dan is voor $0 \leq y \leq 1$, de marginale kansdichtheid van Y gelijk aan

- a. $\frac{6}{5}y + \frac{6}{5}y^2$ b. $\frac{12}{5}y^2 + \frac{12}{5}y^3$ c. $\frac{12}{5}y + \frac{12}{5}y^2$
d. $\frac{6}{5}y^2 + \frac{4}{5}y^3$ e. $\frac{12}{5}(1 + 2y)$ f. $2y$

9. Klanten arriveren op tijdstippen $T_1, T_1 + T_2$, etc. Gegeven is dat de tussenaankomsttijden T_1 en T_2 onafhankelijk zijn en een $Exp(\lambda)$ verdeling hebben. Dan wordt de kansdichtheid van de aankomsttijd van de tweede klant, $Z = T_1 + T_2$, gegeven door $f(z) = 0$ voor $z < 0$ en voor $z \geq 0$:

- a. $f(z) = \lambda^2 e^{-\lambda z}$ b. $f(z) = \lambda^2 e^{-2\lambda z}$ c. $f(z) = \lambda^2 z e^{-\lambda z}$
d. $f(z) = 2\lambda e^{-\lambda z}$ e. $f(z) = \lambda e^{-\lambda z^2}$ f. $f(z) = \lambda e^{-\lambda z}$

10. Het aantal bezoekers per dag aan een bank is Poisson verdeeld. De kans op 0 bezoekers per dag is gelijk aan 0.01. Dan is de verwachting van het aantal bezoekers per dag gelijk aan

- a. 3.8 b. 4 c. 4.6 d. 5 e. 5.2 f. 100

11. Gegeven is een rij $X_1, X_2, \dots, X_{100}$ van onafhankelijke gelijkverdeelde stochasten met verwachting 1 en variantie 4. Dan is de normale benadering voor $P(X_1 + X_2 + \dots + X_{100} < 110)$ gelijk aan

- a. 0.0228 b. 0.1587 c. 0.3095 d. 0.6915 e. 0.8413 f. 0.9872

12. De geordende software data zijn gegeven in de onderstaande lijst

0	0	0	2	4	6	8	9	10	10
10	12	15	15	16	21	22	24	26	30
30	31	33	36	44	50	55	58	65	68
75	77	79	81	88	91	97	100	108	108
112	113	114	115	120	122	129	134	138	143
148	160	176	180	193	193	197	227	232	233
236	242	245	255	261	263	281	290	296	300
300	325	330	357	365	369	371	379	386	422
445	446	447	452	457	482	529	529	543	600
648	670	700	707	724	729	748	790	810	816
828	843	860	865	868	875	943	948	983	990
1011	1045	1064	1071	1082	1146	1160	1222	1247	1351
1435	1461	1755	1783	1800	1864	1897	2323	2930	3110
3321	4116	5485	5509	6150					

De hoogte van het histogram op de cel $(500, 1000]$ is gelijk aan

- a. 0.00036 b. 0.00163 c. 0.04800 d. 0.17778 e. 0.27000 f. 0.81481

13. Men beschikt over de volgende bivariate dataset

i	1	2	3	4
x_i	10	20	30	40
y_i	60	100	140	190

Men fit op deze gegevens een lineair regressiemodel $Y_i = \alpha + \beta x_i + U_i$, waarbij $E[U_i] = 0$ en $\text{Var}(U_i) = \sigma^2$. Als kleinste kwadraten schatters vindt men $\hat{\alpha} = 15$ en $\hat{\beta} = 4.3$. De schatting voor σ^2 is dan gelijk aan

- a. $\sqrt{10}$ b. $\sqrt{15}$ c. 10 d. 15 e. 100 f. 225

14. Een fabrikant van verf wenst de gemiddelde droogtijd van een nieuwe type muurverf te bepalen. Voor 12 testmuren van dezelfde oppervlakte vond hij een gemiddelde droogtijd van 66.3 minuten en een steekproefvariantie van 8.4. De 12 gemeten droogtijden vat men op als een realisatie van een steekproef uit een $N(\mu, \sigma^2)$ verdeling. Het 90% betrouwbaarheidsinterval voor de verwachte droogtijd μ wordt gegeven door

- a. (61.944 , 70.655) b. (61.978 , 70.621) c. (64.730 , 67.869)
d. (64.742 , 67.857) e. (64.797 , 67.803) f. (64.809 , 67.790)

15. Een machine vult wijnflessen met een hoeveelheid die $N(\mu, \sigma^2)$ verdeeld is met $\sigma = 2$ (ml). Om na te gaan of de machine juist is afgesteld, wil men een betrouwbaarheidsinterval maken voor μ . Hoeveel metingen moet verrichten om te zorgen dat het 90% betrouwbaarheidsinterval hoogstens 1 ml breed is?

- a. 62 b. 61 c. 44 d. 43 e. 27 f. 26

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. Zij X een stochastische variabele met verdelingsfunctie

$$F(x) = \begin{cases} 0 & x < 1 \\ \frac{1}{7}x^3 - \frac{1}{7} & 1 \leq x \leq 2 \\ 1 & x > 2. \end{cases}$$

- a. Bepaal de median van F .
b. Bereken de verwachting en variantie van $Y = 3 - 2X$.

2. Gegeven zijn twee stochasten X en Y waarvram de gezamenlijke kansen

$$P(X = 0, Y = 0), \quad P(X = 0, Y = 8), \quad P(X = 16, Y = 0), \quad P(X = 16, Y = 16)$$

allemaal gelijk zijn aan $1/4$.

- a. Geef de tabel met de gezamenlijke kansverdeling van X en Y , alsmede de marginale kansverdeling van X en de marginale kansverdeling van Y .
b. Bepaal $\text{Cov}(X, Y)$.

3. Zij $X_1, X_2, \dots, X_n$ een steekproef uit een exponentiële verdeling met onbekende parameter $\lambda > 0$. Definieer de volgende schatter voor λ :

$$T_n = \frac{n}{X_1 + \dots + X_n}.$$

- a. Ga na of T_n een zuivere schatter is voor λ . Zo ja, toon dit aan. Zo nee, geef dan aan of T_n positieve dan wel negatieve bias heeft.
 - b. Gegeven is dat $n = 100$ en $\lambda = 2$. Bepaal voor dit geval $P\{T_n \geq 2.5\}$ met behulp van de normale benadering.
4. 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. Bereken de likelihood voor de data.
- b. Bereken de maximum likelihood schatting voor μ .

Antwoorden multiple choice:

1 a.

2 d.

3 c.

4 a.

5 f.

6 f.

7 f.

8 a.

9 c.

10 c.

11 d.

12 a.

13 d.

14 e.

15 c.

Antwoorden open vragen:

1a 1.651.

1b $E[4X + 3] = 8$ en $\text{Var}(4X + 3) = \frac{51}{5}$.

2a

a	b			$P(X = a)$
	0	8	16	
0	1/4	1/4	0	1/2
16	1/4	0	1/4	1/2
$P(Y = b)$	1/2	1/4	1/4	1

2b $2/81$.

3a T_n is onzuiver voor λ met positieve bias.

3b 0.0288.

4a $L(\mu) = C \cdot \frac{e^{-13\mu} \mu^7 (\mu+1)^{10}}{24}$.

4b $\mu = 0.904$.

Uitwerkingen multiple choice:

1 a. Met behulp van de regels van DeMorgan hebben we $(A^c \cup B^c)^c = A \cap B$ en $(A^c \cup B)^c = A \cap B^c$, zodat

$$(A^c \cup B^c)^c \cup (A^c \cup B)^c = (A \cap B) \cup (A \cap B^c) = A.$$

2 d. Er geldt

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Vanwege de onafhankelijkheid is

$$\begin{aligned} P(A) &= P(A | B) = \frac{1}{2} \\ P(B) &= P(B | A) = \frac{1}{3} \\ P(A \cap B) &= P(A) \times P(B) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}. \end{aligned}$$

Hence

$$P(A \cup B) = \frac{1}{2} + \frac{1}{3} - \frac{1}{6} = \frac{2}{3}.$$

3 c. Noteer H voor de gebeurtenis “iemand is een hoog-risico klant” en C voor ”iemand dient een claim in”. Dan is gegeven dat $P(C | H) = 0.02$, $P(C | H^c) = 0.0025$, $P(H) = 0.3$, en $P(H^c) = 0.7$. Gevraagd wordt om $P(H | C)$. Met de regel van Bayes geldt

$$P(H | C) = \frac{P(C | H)P(H)}{P(C | H)P(H) + P(C | H^c)P(H^c)} = \frac{0.02 \cdot 0.3}{0.02 \cdot 0.3 + 0.0025 \cdot 0.7} = 0.77.$$

4 a. Het aantal boekenkasten met te weinig schroeven kan gemodelleerd worden door een stochast X met een $Bin(n, p)$ met $n = 1000$ en $p = 0.001$. Zodoende is de gevraagde kans gelijk aan

$$P(X = 2) = \binom{1000}{2} \times (0.001)^2 \times (0.999)^{998} = 0.184.$$

5 f. De oppervlakte onder f moet gelijk zijn aan 1. M.a.w.

$$1 = \int_0^1 c(x - x^3) dx = c \left[\frac{1}{2}x^2 - \frac{1}{4}x^4 \right]_0^1 = c/4.$$

Dus $c = 4$.

6 f. Omdat $E[X^2] = \text{Var}(X) + (E[X])^2 = 4 + 2^2 = 8$, vinden we met behulp van de lineariteit van verwachting:

$$E[(3 - 2X)^2] = E[4X^2 - 12X + 9] = 4E[X^2] - 12E[X] + 9 = 4 \cdot 8 - 12 \cdot 2 + 9 = 17.$$

7 f. Omdat X een $Exp(2)$ verdeling heeft is de verdelingsfunctie van $Y = e^X$ is gelijk aan

$$F(a) = P(Y \leq a) = P(e^X \leq a) = P(X \leq \ln(a)) = 1 - e^{-2 \times \ln(a)} = 1 - \frac{1}{a^2}.$$

Hieruit volgt dat de kansdichtheid van Y gegeven wordt door

$$f(y) = \frac{d}{dy}F(y) = \frac{d}{dy} \left(1 - \frac{1}{y^2} \right) = \frac{2}{y^3}$$

8 a. Zie ook huiswerk opgave 9.10. Volgens de regel op p.122 is de marginale kansdichtheid van Y gelijk aan

$$\int_{-\infty}^{\infty} f(x, y) dx = \int_0^1 \frac{12}{5} xy(1+y) dx = \left[\frac{6}{5} x^2 y(1+y) \right]_{x=0}^{x=1} = \frac{6}{5} y(1+y).$$

9 c. Omdat $f_{T_1}(t) = f_{T_2}(t) = \lambda e^{-\lambda t}$ voor $t \geq 0$ is

$$\begin{aligned} f_Z(z) &= f_{T_1+T_2}(z) = \int_{-\infty}^{\infty} f_{T_1}(z-y) f_{T_2}(y) dy \\ &= \int_0^z \lambda e^{-\lambda(z-y)} \cdot \lambda e^{-\lambda y} dy \\ &= \lambda^2 e^{-\lambda z} \int_0^z dy = \lambda^2 z e^{-\lambda z}. \end{aligned}$$

10 c. Het aantal bezoekers per dag kan gemodelleerd worden door een stochast X met een $Pois(\mu)$ verdeling, zodat de kans op 0 bezoekers gelijk is aan

$$0.01 = P(X = 0) = e^{-\mu}$$

De parameter μ is de verwachting van X , dus het verwachte aantal bezoekers per dag is

$$E[X] = \mu = -\ln(0.01) = 4.6.$$

11 d. Omdat $\mu = E[X_i] = 1$ en $\sigma^2 = \text{Var}(X_i) = 4$ hebben we

$$\begin{aligned} P(X_1 + X_2 + \dots + X_{100} < 110) &= P(\bar{X}_n < 1.1) = P\left(\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} < \sqrt{n} \frac{1.1 - \mu}{\sigma}\right) \\ &= P\left(Z_n < 10 \times \frac{1.1 - 1}{2}\right) = P(Z_n < 0.5) \approx P(Z < 0.5) \end{aligned}$$

waarbij Z een $N(0, 1)$ verdeling heeft. Uit de tabel is af te lezen dat

$$P(Z < 0.5) = 1 - P(Z \geq 0.5) = 1 - 0.3085 = 0.6915.$$

12 a. Zie huiswerkopgave 15.4. De hoogte van het histogram op cel $(500, 1000]$ is

$$\frac{\text{aantal elementen in } (500, 1000]}{135 \cdot 500} = 0.00036$$

13 d. De schatting voor σ^2 is gelijk aan

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2.$$

In ons geval is $n = 4$ en kunnen we de residuen $y_i - 15 - 4.3x_i$ als volgt berekenen

i	1	2	3	4
x_i	10	20	30	40
y_i	60	100	140	190
$15 + 4.3x_i$	58	101	144	187
$y_i - 15 - 4.3x_i$	2	-1	-4	3

We vinden $\hat{\sigma}^2 = \frac{1}{2}(4 + 1 + 16 + 9) = 15$.

14 e. Het 90% betrouwbaarheidsinterval voor de verwachte droogtijd μ wordt gegeven door

$$\left(\bar{x}_n - t_{n-1, \alpha/2} \frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{n-1, \alpha/2} \frac{s_n}{\sqrt{n}} \right).$$

Hierbij is $t_{n-1, \alpha/2} = t_{11, 0.05} = 1.796$ en $s_n = \sqrt{8.4} = 2.898$, zodat het 90% betrouwbaarheidsinterval wordt gegeven door

$$\left(66.3 - 1.796 \frac{2.898}{\sqrt{12}}, 66.3 + 1.796 \frac{2.898}{\sqrt{12}} \right) = (64.797, 67.803).$$

15 c. Het interval is van de vorm

$$\left(\bar{x}_n - 1.645 \frac{\sigma}{\sqrt{n}}, \bar{x}_n + 1.645 \frac{\sigma}{\sqrt{n}} \right).$$

Er moet gelden dat de breedte kleiner is dan 1:

$$2 \cdot 1.645 \frac{\sigma}{\sqrt{n}} \leq 1 \Leftrightarrow n \geq (2 \cdot 1.645 \cdot \sigma)^2 = (2 \cdot 1.645 \cdot 2)^2 = 43.3.$$

Dus moet n minstens 44 zijn.

Uitwerkingen open vragen:

1a De mediaan van F is de oplossing van $F(q) = 0.5$. We hebben

$$0.5 = F(q) = \frac{1}{7}q^3 - \frac{1}{7} \Leftrightarrow 4.5 = q^3 \Leftrightarrow q = (4.5)^{1/3} = 1.651.$$

1b Bereken eerst de verwachting en variantie van X . Omdat de kansdichtheid van X wordt gegeven door

$$f(x) = \frac{d}{dx} \left(\frac{1}{7}x^3 - \frac{1}{7} \right) = \frac{3}{7}x^2 \quad \text{voor } 1 \leq x \leq 2,$$

is

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx = \int_1^2 \frac{3}{7} x^3 dx = \left[\frac{3}{28} x^4 \right]_1^2 = \frac{45}{28} \\ E[X^2] &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_1^2 \frac{3}{7} x^4 dx = \left[\frac{3}{35} x^5 \right]_1^2 = \frac{93}{35} \\ \text{Var}(X) &= E[X^2] - (E[X])^2 = \frac{93}{35} - \left(\frac{45}{28} \right)^2 = \frac{291}{3920} \end{aligned}$$

Dit geeft

$$\begin{aligned} E[3 - 2X] &= 3 - 2 \times E[X] = 3 - 2 \times \frac{45}{28} = -\frac{3}{14} \\ \text{Var}(3 - 2X) &= (-2)^2 \text{Var}(X) = 4 \times \frac{291}{3920} = \frac{291}{980}. \end{aligned}$$

2a Zie ook *Exercise 10.1*. Omdat de vier kansen optellen tot 1 volgt uit de gegevens dat X de waarden 0 en 16 aanneemt en Y de waarden 0, 8 en 16, en moeten de kansen op de twee overige combinaties gelijk aan 0 zijn. Hiermee is de gezamenlijke kansverdeling van X en Y bekend. De marginals kunnen worden gevonden door optelling over rijen en kolommen. We vinden de volgende tabel

a	b			$P(X = a)$
	0	8	16	
0	1/4	1/4	0	1/2
16	1/4	0	1/4	1/2
$P(Y = b)$	1/2	1/4	1/4	1

2b De covariantie wordt gegeven door

$$\text{Cov}(X, Y) = E[XY] - E[X] E[Y].$$

We hebben

$$\begin{aligned} E[XY] &= 0 \times 0 \times \frac{1}{4} + 0 \times 8 \times \frac{1}{4} + 0 \times 16 \times 0 \\ &\quad + 16 \times 0 \times \frac{1}{4} + 16 \times 8 \times 0 + 16 \times 16 \times \frac{1}{4} = 64 \\ E[X] &= 0 \times \frac{1}{2} + 16 \times \frac{1}{2} = 8 \\ E[Y] &= 0 \times \frac{1}{2} + 8 \times \frac{1}{4} + 16 \times \frac{1}{4} = 6 \end{aligned}$$

zodat

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 64 - 8 \times 6 = 16.$$

3a Omdat de exponentiële verdeling verwachting $\mu = 1/\lambda$ heeft, is $E[\bar{X}_n] = 1/\lambda$. Verder geldt dat

$$E[T_n] = E\left[\frac{n}{X_1 + \dots + X_n}\right] = E\left[\frac{1}{\bar{X}_n}\right].$$

Omdat de functie $g(x) = 1/x$ strict convex is ($g''(x) = 2/x^3 > 0$) geldt volgens de ongelijkheid van Jensen

$$E[T_n] = E\left[\frac{1}{\bar{X}_n}\right] > \frac{1}{E[\bar{X}_n]} = \frac{1}{1/\lambda} = \lambda.$$

We concluderen dat T_n onzuiver is voor λ met positieve bias.

3b Schrijf eerst

$$P(T_n \geq 2.5) = P\left(\frac{n}{X_1 + \dots + X_n} \geq 2.5\right) = P(\bar{X}_n \leq 0.4).$$

Voor de exponentiële verdeling is $\mu = 1/\lambda = 0.5$ en $\sigma^2 = 1/\lambda^2 = 0.25$ (dus $\sigma = 0.5$), zodat volgens de centrale limietstelling

$$\begin{aligned} P(T_n \geq 2.5) &= P(\bar{X}_n \leq 0.4) = P\left(\frac{\sqrt{100}\bar{X}_n - 0.5}{0.5} \leq \frac{\sqrt{100}(0.4 - 0.5)}{0.5}\right) \\ &= P(Z_{100} \leq -2) = P(Z_{100} \geq 2) \stackrel{\text{Tabel}}{=} 0.0288. \end{aligned}$$

4a De likelihood is gelijk aan de kans op de data:

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

4b 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 $\ell'(0) = +\infty$ en $\ell'(\infty) < 0$). Dit betekent dat $\mu = 0.904$ de maximum likelihood schatting is.

Tentamen Kansrekening en statistiek
wi3102CT
30 oktober 2009, 9.00–12.00 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.

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 of uitgummen, of verwijderen met correctievloeistof of een nieuw formulier invullen. Vergeet niet uw studienummer in te vullen en aan te strepen.

1. We gooien drie keer met een muntstuk. De uitkomstenruimte voor dit experiment is

$$\Omega = \{KKK, MKK, KMK, KKM, MMK, MKM, KMM, MMM\}$$

waarbij K staat voor kop en M voor munt. Beschouw de volgende gebeurtenissen

A : “we gooien precies twee keer munt.”

D : “de eerste worp resulteert in munt.”

Dan wordt de gebeurtenis $A \cap D^c$ gegeven door de verzameling

- a. $\{MKK, MKM, KMM, KKK, KKM, KMK\}$
 - b. $\{MKM, MMK\}$
 - c. $\{KMM, KMK\}$
 - d. $\{KMM\}$
 - e. $\{MMK\}$
 - f. $\emptyset$
2. Gegeven zijn gebeurtenissen A en B , waarvoor $P(A \cup B) = 2/3$ en $P(A^c | B^c) = 1/2$. Dan is $P(B)$ gelijk aan
- a. 1 b. $1/2$ c. $1/3$ d. $1/4$ e. $1/5$ f. $1/6$
3. Iemand test positief op een onbekende ziekte. Het is bekend dat men 1% kans heeft om de ziekte te krijgen en dat de test niet perfect is, d.w.z. 98% van de zieke mensen zal positief testen en 95% van de gezonde mensen zal negatief testen. Dan is de kans dat iemand de ziekte heeft, gegeven dat hij positief test, gelijk aan
- a. 0.0093 b. 0.0098 c. 0.0103 d. 0.0193 e. 0.0573 f. 0.1653
4. Gegeven is de verdelingsfunctie van een discrete stochastische variabels X :

$$F(a) = \begin{cases} 0 & \text{voor } a < 0 \\ \frac{1}{3} & \text{voor } 0 \leq a < \frac{1}{2} \\ \frac{1}{2} & \text{voor } \frac{1}{2} \leq a < \frac{3}{4} \\ 1 & \text{voor } a \geq \frac{3}{4}. \end{cases}$$

De kansverdeling van X wordt gegeven door

a.

a	0	1/2	3/4
$P(X = a)$	1/3	1/6	1/2

b.

a	1/2	1/3	1
$P(X = a)$	1/4	1/2	1/4

c.

a	0	1/2	3/4
$P(X = a)$	1/3	1/2	1/6

d.

a	1/2	1/3	1
$P(X = a)$	1/2	1/4	1/4

e.

a	0	1/2	3/4
$P(X = a)$	1/2	1/6	1/3

f.

a	1/2	1/3	1
$P(X = a)$	1/8	1/8	3/4

5. De mediaan van een $Par(1)$ verdeling is gelijk aan

- a. -3 b. 0 c. $\frac{1}{2}$ d. 1 e. 2 f. 4

6. Gegeven is de verdelingsfunctie $F(x) = 0$ voor $x < 0$, en

$$F(x) = 1 - e^{-5x^2} \quad \text{voor } x \geq 0.$$

Met behulp van een $U(0, 1)$ variabele U kan men een stochast Z met de bovenstaande verdelingsfunctie construeren op de volgende manier:

a. $Z = -\frac{1}{5} \ln(\sqrt{1 - U})$

b. $Z = -\ln(\frac{1}{5}\sqrt{1 - U})$

c. $Z = \sqrt{-\frac{1}{5} \ln(1 - U)}$

d. $Z = 1 - e^{-5U^2}$

e. $Z = 1 - e^{-\sqrt{\frac{1}{5}U}}$

f. $Z = 1 - e^{-5\sqrt{U}}$

7. Gegeven is een stochastische variabele X met $E[X] = 2$, $\text{Var}(X) = 4$. Dan worden de verwachting en variantie van $Y = 3 - 2X$ gegeven door

a. $E[Y] = -1$ en $\text{Var}(Y) = 5$

b. $E[Y] = -1$ en $\text{Var}(Y) = 8$

c. $E[Y] = -1$ en $\text{Var}(Y) = 16$

d. $E[Y] = -1$ en $\text{Var}(Y) = 19$

e. $E[Y] = 8$ en $\text{Var}(Y) = 5$

f. $E[Y] = 8$ en $\text{Var}(Y) = 8$

8. Stel dat X een $Par(\alpha)$ verdeling heeft. Dan heeft $\ln X$ een

a. $Par(\ln \alpha)$ verdeling

b. $Par(e^\alpha)$ verdeling

c. $Exp(1)$ verdeling

d. $Exp(\alpha)$ verdeling

e. $U(0, \alpha)$ verdeling

f. $U(0, 1)$ verdeling

9. Gegeven is de onderstaande gezamenlijke kansverdeling van twee discrete stochastische variabelen X en Y :

b	a			$P(Y = b)$
	0	1	2	
-1	1/6	1/6	1/6	1/2
1	0	1/2	0	1/2
$P(X = a)$	1/6	2/3	1/6	1

Dan geldt

- a. $E[XY] = 0$ en X, Y zijn onafhankelijk
- b. $E[XY] = 1$ en X, Y zijn onafhankelijk
- c. $E[XY] = -1$ en X, Y zijn onafhankelijk
- d. $E[XY] = 0$ en X, Y zijn afhankelijk
- e. $E[XY] = 1$ en X, Y zijn afhankelijk
- f. $E[XY] = -1$ en X, Y zijn afhankelijk

10. Gegeven zijn twee onafhankelijke stochasten X en Y , elk met een $U(0, 1)$ verdeling. De kansverdeling van $Z = X + Y$

- a. is $U(0, 2)$
- b. heeft dichtheid $f(z) = \begin{cases} z & \text{voor } 0 \leq z < 1, \\ 2 - z & \text{voor } 1 \leq z \leq 2, \\ 0 & \text{elders.} \end{cases}$
- c. is $U(1, 1)$
- d. heeft dichtheid $f(z) = \begin{cases} 2 - z & \text{voor } 0 \leq z < 1, \\ z & \text{voor } 1 \leq z \leq 2, \\ 0 & \text{elders.} \end{cases}$
- e. is $U(1, 2)$
- f. heeft dichtheid $f(z) = \begin{cases} \frac{1}{2}z & \text{voor } 0 \leq z < 2, \\ 0 & \text{elders.} \end{cases}$

11. Zij Y een stochastische variabele met een $N(\mu, \sigma^2)$ verdeling. De ondergrenzen volgens de ongelijkheid van Chebyshev voor de kansen $P(|Y - \mu| < k\sigma)$, voor $k = 1, 2, 3$ en de exacte waarden van deze kansen worden gegeven door

a.

k	1	2	3
Cheb.	1	0.2500	0.1111
exact	0.3174	0.0456	0.0026

c.

k	1	2	3
Cheb.	1	0.2500	0.1111
exact	0.1587	0.0228	0.0013

e.

k	1	2	3
Cheb.	1	0.2500	0.1111
exact	niet te berekenen		

b.

k	1	2	3
Cheb.	0	0.7500	0.8889
exact	0.6826	0.9544	0.9974

d.

k	1	2	3
Cheb.	0	0.7500	0.8889
exact	0.8413	0.9772	0.9987

f.

k	1	2	3
Cheb.	0	0.7500	0.8889
exact	niet te berekenen		

12. We construeren een histogram met cellen $[0,1]$, $(1,3]$, $(3,5]$, $(5,8]$, $(8,11]$, $(11,14]$, en $(14,18]$. Gegeven zijn de waarden van de empirische verdelingsfunctie in de randen van de cellen:

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

Dan is de hoogte van het histogram op de cel $(8, 11]$ gelijk aan

- a. 0.0233
- b. 0.0035
- c. 0.0700
- d. 0.2100
- e. 0.2450
- f. 0.2683

13. Gegeven zijn de software data

0	0	0	2	4	6	8	9	10	10
10	12	15	15	16	21	22	24	26	30
30	31	33	36	44	50	55	58	65	68
75	77	79	81	88	91	97	100	108	108
112	113	114	115	120	122	129	134	138	143
148	160	176	180	193	193	197	227	232	233
236	242	245	255	261	263	281	290	296	300
300	325	330	357	365	369	371	379	386	422
445	446	447	452	457	482	529	529	543	600
648	670	700	707	724	729	748	790	810	816
828	843	860	865	868	875	943	948	983	990
1011	1045	1064	1071	1082	1146	1160	1222	1247	1351
1435	1461	1755	1783	1800	1864	1897	2323	2930	3110
3321	4116	5485	5509	6150					

Het 37ste empirische percentiel wordt gegeven door

- a.** 55 **b.** 97 **c.** 144.6 **d.** 810 **e.** 983 **f.** 2275.5

- 14.** Zij $\bar{X}_n$ and $\bar{Y}_m$ de steekproefgemiddelden van twee onafhankelijke steekproeven van omvang n respectievelijk m uit dezelfde kansverdeling met verwachting μ en variantie σ^2 . We combineren de twee schatters tot een nieuwe schatter

$$T = r\bar{X}_n + (1 - r)\bar{Y}_m,$$

waarbij r is een getal is tussen 0 en 1. Dan geldt

- a.** T is zuiver voor μ met de kleinste MSE voor $r = n/(m + n)$
b. T is zuiver voor μ met de kleinste MSE voor $r = m/(m + n)$
c. T is zuiver voor μ met de kleinste MSE voor $r = 1/2$
d. T is onzuiver voor μ met de kleinste MSE voor $r = n/(m + n)$
e. T is onzuiver voor μ met de kleinste MSE voor $r = m/(m + n)$
f. T is onzuiver voor μ met de kleinste MSE voor $r = 1/2$
- 15.** 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 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 μ wordt gegeven door:

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

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. Zij X en Y twee continue stochastische variabelen met gezamenlijke kansdichtheid

$$f(x, y) = \frac{12}{5}xy(1+y) \quad \text{voor } 0 \leq x \leq 1 \text{ en } 0 \leq y \leq 1,$$

en $f(x, y) = 0$ elders.

- Bepaal de uitdrukking voor de gezamenlijke verdelingsfunctie $F(a, b)$ van X en Y voor a en b tussen 0 en 1.
 - Bepaal de marginale kansdichtheid van X .
2. Zij $X_1, X_2, \dots, X_{625}$ onafhankelijke gelijkverdeelde stochastische variabelen, met kansdichtheid

$$f(x) = \begin{cases} 3(1-x)^2 & \text{voor } 0 \leq x \leq 1, \\ 0 & \text{elders.} \end{cases}$$

- Bereken de verwachting en de variantie van X .
 - Geef een benadering voor $P(X_1 + X_2 + \dots + X_{625} < 170)$ met behulp van de centrale limietstelling.
3. Van een bepaalde plantensoort komen de vier typen *starchy-green*, *sugary-white*, *starchy-white*, en *sugary-green* voor met kansen

Type	Kans
1: starchy-green	$p_1 = \frac{1}{4}(2 + \theta)$
2: sugary-white	$p_2 = \frac{1}{4}\theta$
3: starchy-white	$p_3 = \frac{1}{4}(1 - \theta)$
4: sugary-green	$p_4 = \frac{1}{4}(1 - \theta)$
Totaal	1

waarbij $0 < \theta < 1$ een onbekende parameter is. Stel dat men beschikt over n bladeren. Het aantal *starchy-green* bladeren modeleert men met een $\text{Bin}(n, p_1)$ variabele N_1 en aantal *sugary-white* bladeren met een $\text{Bin}(n, p_2)$ variabele N_2 . We beschouwen twee schatters voor θ :

$$T_1 = \frac{4}{n}N_1 - 2 \quad \text{en} \quad T_2 = \frac{4}{n}N_2.$$

- Toon aan data beide schatters zuiver zijn voor θ .
 - Geef aan welke van de twee schatters u zou prefereren en motiveer uw antwoord.
4. We beschikken over een dataset $x_1, x_2, \dots, x_n$, die een realisatie is van een steekproef uit een Rayleigh verdeling met kansdichtheid

$$f_\theta(x) = \frac{x}{\theta^2} e^{-\frac{1}{2}x^2/\theta^2} \quad \text{voor } x \geq 0.$$

- Bepaal de loglikelihood $\ell(\theta)$.
- Bepaal de formule voor de maximum likelihood schatter voor θ .

Antwoorden multiple choice:

1 d.

2 c.

3 f.

4 a.

5 e.

6 c.

7 c.

8 d.

9 d.

10 b.

11 b.

12 a.

13 c.

14 a.

15 e.

Antwoorden open vragen:

1a $F(a, b) = \frac{3}{5}a^2b^2 + \frac{2}{5}a^2b^3.$

1b $f_X(x) = 2x$ voor $0 \leq x \leq 1.$

2a $E[X] = 1/4; \text{Var}(X) = 3/80.$

2b 0.9977.

3a Laat zien dat $E[T_1] = \theta$ en $E[T_2] = \theta.$

3b We prefereren $T_2.$

4a $\sum_{i=1}^n \ln(x_i) - 2n \ln(\theta) - \frac{2}{\theta^2} \sum_{i=1}^n x_i^2$

4b $\sqrt{\frac{1}{2n} \sum_{i=1}^n x_i^2}.$

Uitwerkingen multiple choice:

1 d. Zie huiswerk opgave 2.9. We hebben

$$A = \{MMk, MKM, KMM\}$$
$$D^c = \{KKK, KKM, KMK, KMM\}$$

zodat $A \cap D^c = \{KMM\}$.

2 c. Zie huiswerk opgave 3.7. Met behulp van de voorwaardelijke kans vinden we $P(A^c \cap B^c) = P(A^c | B^c) P(B^c) = \frac{1}{2} (1 - P(B))$. Volgens de wetten van DeMorgan hebben we $P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1/3$. Tesamen geeft dit een vergelijking voor $P(B)$:

$$\frac{1}{2} (1 - P(B)) = 1/3$$

waaruit volgt dat $P(B) = 1/3$.

3 f. Zie huiswerk opgave 3.16. Schrijf D voor de gebeurtenis "men heeft de ziekte", zodat $P(D) = 0.01$ en T voor de gebeurtenis "de test is positief", zodat $P(T | D) = 0.98$ en $P(T^c | D^c) = 0.95$. Volgens de regel van Bayes is

$$P(D | T) = \frac{P(T | D)P(D)}{P(T | D)P(D) + P(T | D^c)P(D^c)} = \frac{0.98 \times 0.01}{0.98 \times 0.01 + (1 - 0.95) \times (1 - 0.01)} = 0.1653.$$

4 a. Zie huiswerk opgave 4.3. Maak eventueel een schets van de grafiek van F . Uit de beschrijving van F is af te lezen dat F sprongen maakt in 0 , $\frac{1}{2}$ en $\frac{3}{4}$, met sprong grootte, respectievelijk $\frac{1}{3}$, $\frac{1}{6}$ en $\frac{1}{2}$. Dit betekent dat X als mogelijke waarden heeft 0 , $\frac{1}{2}$ en $\frac{3}{4}$, met kansen $\frac{1}{3}$, $\frac{1}{6}$ en $\frac{1}{2}$:

a	0	$1/2$	$3/4$
$P(X = a)$	$1/3$	$1/6$	$1/2$

5 e. Zie huiswerk opgave 5.12. De verdelingsfunctie van de $Par(1)$ is $F(x) = 1 - x^{-1}$. Om de mediaan te vinden moeten we oplossen

$$F(x) = \frac{1}{2} \Leftrightarrow 1 - \frac{1}{x} = \frac{1}{2} \Leftrightarrow x = 2.$$

6 c. Zie huiswerk opgave 6.7. We kunnen Z vinden door op te lossen $F(Z) = U$:

$$1 - e^{-5Z^2} = U \Leftrightarrow e^{-5Z^2} = 1 - U \Leftrightarrow -5Z^2 = \ln(1 - U)$$
$$\Leftrightarrow Z^2 = -\frac{1}{5} \ln(1 - U) \Leftrightarrow Z = \sqrt{-\frac{1}{5} \ln(1 - U)}.$$

7 c. Zie huiswerk opgave 7.4. Volgens de lineariteit van verwachting is

$$E[3 - 2X] = 3 - 2 \times E[X] = 3 - 2 \times 2 = -1.$$

Volgens de regel voor de variantie

$$\text{Var}(3 - 2X) = (-2)^2 \times \text{Var}(X) = 4 \times 4 = 16.$$

8 d. Zie huiswerk opgave 8.7. Noteer $Y = \ln X$. We bepalen de verdelingsfunctie van Y :

$$P(Y \leq y) = P(\ln X \leq y) = P(X \leq e^y) = 1 - (e^y)^{-\alpha} = 1 - e^{-\alpha y}.$$

Dit is de verdelingsfunctie van een $Exp(\alpha)$ verdeling.

9 d. Zie huiswerk opgave 10.2. We hebben dat

$$E[XY] = \frac{1}{6} \times 1 \times (-1) + \frac{1}{6} \times 2 \times (-1) + \frac{1}{2} \times 1 \times 1 = 0.$$

Verder zijn X en Y afhankelijk, omdat

$$P(X = 0, Y = 1) = 0 \neq \frac{1}{12} = \frac{1}{6} \times \frac{1}{2} = P(X = 0) \times P(Y = 1).$$

10 b. Zie huiswerk opgave 11.5. We berekenen de kansdichtheid van Z :

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y) dy.$$

Omdat X en Y beide een $U(0, 1)$ verdeling hebben, geldt dat

$$f_X(z-y) = \begin{cases} 1 & 0 \leq z-y \leq 1 \\ 0 & \text{elders.} \end{cases} \quad \text{en} \quad f_Y(y) = \begin{cases} 1 & 0 \leq y \leq 1 \\ 0 & \text{elders.} \end{cases}$$

De integrand is daarom gelijk aan $1 \times 1 = 1$ precies als

$$z-1 \leq y \leq z \quad \text{én} \quad 0 \leq y \leq 1.$$

In het geval dat $0 \leq z < 1$, betekent dat $0 \leq y \leq z$, zodat

$$f_Z(z) = \int_0^z 1 dy = z.$$

In het geval dat $1 \leq z \leq 2$, betekent dat $z-1 \leq y \leq 1$, zodat

$$f_Z(z) = \int_{z-1}^1 1 dy = 2 - z.$$

11 b. Zie huiswerk opgave 13.1. Volgens de ongelijkheid van Chebyshev geldt (zie formule (13.1) in het boek):

$$P(|Y - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

Voor $k = 1, 2, 3$ geeft dit als ondergrenzen 0 , $\frac{3}{4}$ en $\frac{8}{9}$, zodat alleen nog **b**, **d** en **f** in aanmerking komen. De exacte waarde van $P(|Y - \mu| < k\sigma)$ is als volgt te berekenen:

$$P(|Y - \mu| < k\sigma) = P\left(\left|\frac{Y - \mu}{\sigma}\right| < k\right) = P(|Z| < k) = 1 - 2 \cdot P(Z > k)$$

waarbij $Z = (Y - \mu)/\sigma$ een $N(0, 1)$ verdeling heeft. Uit de tabel voor de normale verdeling volgt dat voor $k = 1, 2, 3$, $P(Z > k)$ gelijk is aan respectievelijk 0.1587, 0.0228 en 0.0013. Zodoende is de exacte waarde voor gevraagde kans $P(|Y - \mu| < k\sigma)$ voor $k = 1, 2, 3$ gelijk aan respectievelijk 0.6826, 0.9544 en 0.9974.

12 a. Zie huiswerk opgave 15.5. De hoogte op de cel $(8, 11]$ is gelijk aan

$$\begin{aligned} \frac{\text{aantal } x_i \text{ in } (8, 11]}{(11 - 8)n} &= \frac{\text{aantal } x_i \leq 11 - \text{aantal } x_i \leq 8}{3n} \\ &= \frac{\text{aantal } x_i \leq 11}{3n} - \frac{\text{aantal } x_i \leq 8}{3n} \\ &= \frac{1}{3} \left(\frac{\text{aantal } x_i \leq 11}{n} - \frac{\text{aantal } x_i \leq 8}{n} \right) \\ &= \frac{1}{3} (F_n(11) - F_n(8)) = \frac{1}{3} (0.805 - 0.735) = 0.0233. \end{aligned}$$

13 c. Zie huiswerk opgave 16.1. We hebben $n = 135$ zodat $0.37 \times (n + 1) = 50.32$. Dit betekent dat

$$q_{0.37} = x_{(50)} + 0.32 \times (x_{(51)} - x_{(50)}) = 143 + 0.32 \times (148 - 143) = 144.6.$$

14 a. Zie huiswerk opgave 20.8. Allereerst geldt $E[\bar{X}_n] = \mu$ en $E[\bar{Y}_m] = \mu$. Vanwege de lineariteit van verwachting geldt dan

$$E[T] = E[r\bar{X}_n + (1 - r)\bar{Y}_m] = rE[\bar{X}_n] + (1 - r)E[\bar{Y}_m] = r\mu + (1 - r)\mu = \mu.$$

Dus T is zuiver voor μ . Bovendien is $\text{MSE}(T) = \text{Var}(T)$. Verder is $\text{Var}(\bar{X}_n) = \sigma^2/n$ en $\text{Var}(\bar{Y}_m) = \sigma^2/m$. Omdat $\bar{X}_n$ en $\bar{Y}_m$ onafhankelijk zijn, is

$$\text{MSE}(T) = \text{Var}(T) = r^2 \text{Var}(\bar{X}_n) + (1 - r)^2 \text{Var}(\bar{Y}_m) = r^2 \times \frac{\sigma^2}{n} + (1 - r)^2 \times \frac{\sigma^2}{m}$$

Om de minimale r te vinden, moeten we dit differentiëren naar r en de afgeleide gelijk stellen aan nul:

$$\frac{2r\sigma^2}{n} - \frac{2(1 - r)\sigma^2}{m} = 0 \quad \Leftrightarrow \quad 2rm - 2n(1 - r) = 0 \quad \Leftrightarrow \quad r = \frac{n}{n + m}.$$

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

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

Uitwerkingen open vragen:

1a Omdat $f(x, y) = 0$ voor $x < 0$ of $y < 0$, geldt

$$\begin{aligned} F(a, b) &= P(X \leq a, Y \leq b) = \int_{-\infty}^a \left(\int_{-\infty}^b f(x, y) dy \right) dx \\ &= \int_0^a \left(\int_0^b \frac{12}{5} xy(1+y) dy \right) dx = \frac{3}{5} a^2 b^2 + \frac{2}{5} a^2 b^3. \end{aligned}$$

1b

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^1 \frac{12}{5} xy(1+y) dy = \frac{12}{5} x \int_0^1 y(1+y) dy = 2x.$$

2a We hebben

$$\begin{aligned} E[X] &= \int_0^1 3x(1-x)^2 dx = \int_0^1 (3x^3 - 6x^2 + 3x) dx = \left[\frac{3}{4}x^4 - 2x^3 + \frac{3}{2}x^2 \right]_0^1 = \frac{1}{4} \\ E[X^2] &= \int_0^1 3x^2(1-x)^2 dx = \int_0^1 (3x^4 - 6x^3 + 3x^2) dx = \left[\frac{3}{5}x^5 - \frac{6}{4}x^4 + x^3 \right]_0^1 = \frac{1}{10} \\ \text{Var}(X) &= E[X^2] - (E[X])^2 = \frac{1}{10} - \left(\frac{1}{4}\right)^2 = \frac{3}{80} \end{aligned}$$

2b Since $E[X_i] = 1/4$ and $\text{Var}(X_i) = 3/80$, using the central limit theorem we find

$$\begin{aligned} P(X_1 + X_2 + \dots + X_{625} < 170) &= P\left(\frac{\sqrt{625}\bar{X}_{625} - 1/4}{\sqrt{3/80}} < \frac{\sqrt{625}170/625 - 1/4}{\sqrt{3/80}}\right) \\ &\approx \Phi(2.8402) = 1 - 0.0023 = 0.9977. \end{aligned}$$

3a

$$\begin{aligned} E[T_1] &= \frac{4}{n} E[N_1] - 2 = \frac{4}{n} \times np_1 - 2 = \frac{4}{n} \times n \times \frac{1}{4}(2 + \theta) - 2 = \theta \\ E[T_2] &= \frac{4}{n} E[N_2] = \frac{4}{n} \times np_2 = \frac{4}{n} \times n \times \frac{1}{4}\theta = \theta. \end{aligned}$$

3b We prefereren de schatter met de kleinst variantie. We hebben

$$\begin{aligned} \text{Var}(T_1) &= \left(\frac{4}{n}\right)^2 \text{Var}(N_1) = \left(\frac{4}{n}\right)^2 np_1(1-p_1) \\ &= \left(\frac{4}{n}\right)^2 n \times \frac{1}{4}(2 + \theta) \times \left(1 - \frac{1}{4}(2 + \theta)\right) = \frac{4 - \theta^2}{n} \\ \text{Var}(T_2) &= \left(\frac{4}{n}\right)^2 \text{Var}(N_2) = \left(\frac{4}{n}\right)^2 np_2(1-p_2) \\ &= \left(\frac{4}{n}\right)^2 n \times \frac{1}{4}\theta \times \left(1 - \frac{1}{4}\theta\right) = \frac{\theta(4 - \theta)}{n} \end{aligned}$$

Omdat $4 - \theta^2 > \theta(4 - \theta)$ voor alle $0 < \theta < 1$, is $\text{Var}(T_2) < \text{Var}(T_1)$ zodat we kiezen voor T_2 .

4a We vinden

$$\ln f_{\theta}(x) = \ln(x) - 2 \ln(\theta) - \frac{x^2}{2\theta^2}$$

zodat de loglikelihood wordt gegeven door

$$\begin{aligned} L(\theta) &= \sum_{i=1}^n \ln f_{\theta}(x_i) = \sum_{i=1}^n \left(\ln(x_i) - 2 \ln(\theta) - \frac{x_i^2}{2\theta^2} \right) \\ &= \sum_{i=1}^n \ln(x_i) - 2n \ln(\theta) - \frac{1}{2\theta^2} \sum_{i=1}^n x_i^2 \end{aligned}$$

4b We bepalen eerst de afgeleide

$$L'(\theta) = -\frac{2n}{\theta} + \frac{1}{\theta^3} \sum_{i=1}^n x_i^2.$$

Dan is

$$L'(\theta) = 0 \quad \Leftrightarrow \quad 2n\theta^2 = \sum_{i=1}^n x_i^2 \quad \Leftrightarrow \quad \theta = \pm \sqrt{\frac{1}{2n} \sum_{i=1}^n x_i^2}.$$

Omdat

$$L'(\theta) = -\frac{2n}{\theta} + \frac{1}{\theta^3} \sum_{i=1}^n x_i^2 = \frac{2n}{\theta^3} \left(-\theta^2 + \frac{1}{2n} \sum_{i=1}^n x_i^2 \right)$$

is het tekenverloop van $L'(\theta)$ gelijk aan $- + -$, wat betekent dat het maximum wordt aangenomen bij

$$\theta = \sqrt{\frac{1}{2n} \sum_{i=1}^n x_i^2}.$$