Tentamen Kansrekening en Statistiek
MST
14 januari 2016, 14.00–17.00 uur

Het tentamen bestaat uit **15** meerkeuzevragen **2** open vragen. Een formuleblad wordt uitgedeeld.
Normering: 0.4 punt per MC antwoord - 2 punt per open vraag

1. Van de gebeurtenissen A en B zijn de volgende kansen bekend: $P(A) = 0.25$, $P(B) = 0.45$ en $P(A \cap B) = 0.1$. Bereken $P(A^c \cap B)$.

a. 0.1 **b.** 0.15 **c.** 0.2 **d.** 0.25 **e.** 0.3 **f.** 0.35

2. Ik heb twee munten, de ene is vals en gooit altijd kop. De andere is zuiver en gooit kop met kans $\frac{1}{2}$. Ik doe mijn ogen dicht en neem één van deze munten, volstrekt willekeurig. De kans dat ik de valse munt heb gepakt is een half, de kans dat ik de zuivere munt heb gepakt is ook een half. Vervolgens gooi ik de munt. Het resultaat is kop. Hoe groot is nu de kans dat ik de valse munt heb gepakt?

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

3. De stochast (random variable) X heeft uitkomsten tussen 0 en 2. De kansdichtheidsfunctie is $f(x) = kx^2$ voor $0 \leq x \leq 2$. Bepaal k.

a. $\frac{1}{4}$ **b.** $\frac{1}{3}$ **c.** $\frac{3}{8}$ **d.** $\frac{1}{2}$ **e.** $\frac{2}{3}$ **f.** $\frac{3}{4}$

4. Laat X een uniforme $U(0,1)$ variabele zijn. De stochast Y is het kwadraat van X, dus $Y = X^2$. Bepaal de verdelingsfunctie (probability distribution function) $F(y)$ van Y voor $0 \leq y \leq 1$.

a. $F(y) = y^2$ **b.** $F(y) = y$ **c.** $F(y) = \sqrt{y}$
d. $F(y) = 2y$ **e.** $F(y) = y^2/2$ **f.** $F(y) = 2y^2$

5. Bevolkingsstatistieken leren dat van 100 000 vrouwen, 89 835 minstens 60 jaar oud worden, en 57 062 minstens 80 jaar oud. Wat is de kans dat een vrouw van 60 minstens 80 jaar oud wordt?

a. 0.5706 **b.** 0.6352 **c.** 0.7854 **d.** 0.8103 **e.** 0.8547 **f.** 0.8934

6. Volgens de ongelijkheid van Chebyshev geldt dat

$$P\left(\left|\frac{X-\mu}{\sigma}\right| \geq k\right) \leq \frac{1}{k^2}$$

Stel dat $E[X] = 3$ en $\mathrm{Var}(X) = 4$. Hoe groot is de kans dat X in het interval $(-1, 7)$ ligt volgens de ongelijkheid van Chebyshev?

a. ≤ 0.1 **b.** ≤ 0.25 **c.** ≤ 0.5 **d.** ≥ 0.5 **e.** ≥ 0.75 **f.** ≥ 0.9

7. De stochast X heeft een binomiale verdeling $Bin(n, p)$ met $n = 100$ en $p = 0.4$. We benaderen X via de normale verdeling $N(\mu, \sigma^2)$. Wat zijn dan de geschikte waarden van μ en σ^2?

a. $\mu = 100, \sigma^2 = 0.24$ **b.** $\mu = 100, \sigma^2 = 24$ **c.** $\mu = 100, \sigma^2 = 576$
d. $\mu = 40, \sigma^2 = 0.24$ **e.** $\mu = 40, \sigma^2 = 24$ **f.** $\mu = 40, \sigma^2 = 576$

8. De stochasten X, Y nemen alleen waarden aan tussen 0 en 1. De simultane dichtheidsfunctie is $f(x, y) = x + y$ voor $0 \leq x, y \leq 1$. Bereken $E[XY]$ tot op twee decimalen.

a. 0.11 **b.** 0.20 **c.** 0.33 **d.** 0.25 **e.** 0.16 **f.** 0.47

9. Op basis van een dataset van 20 elementen heb ik het volgende 90-procent betrouwbaarheidsinterval berekend $(0.982, 1.018)$ uit $\bar{x}_{20}$ en s_{20} via de t-verdeling. Wat is dan het 95-procent betrouwbaarheidsinterval?

a. $(0.970, 1.030)$ **b.** $(0.973, 1.027)$ **c.** $(0.967, 1.033)$
d. $(0.985, 1.015)$ **e.** $(0.978, 1.022)$ **f.** $(0.980, 1.020)$

10. De stochasten X en Y hebben een gezamenlijke kansmassa zoals in de onderstaande tabel.

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

Bereken de covariantie van X en Y

a. 0.25 **b.** -0.25 **c.** 0 **d.** -0.5 **e.** 0.5 **f.** 1

11. Efking super krachtlijm kan volgens de producent een belasting van tot 10 MPa weerstaan. De consumentenbond twijfelt aan de kracht van deze lijm en onderzoekt de bewering via een statistische toets. We noteren de kritische spanning van Efking super krachtlijm met P. Bepaal de juiste nulhypothese en alternatieve hypothese.

a. $H_0: P = 10;\ H_1: P > 10$ **b.** $H_0: P < 10;\ H_1: P > 10$
c. $H_0: P = 10;\ H_1: P < 10$ **d.** $H_0: P < 10;\ H_1: P = 10$
e. $H_0: P = 10;\ H_1: P \neq 10$ **f.** $H_0: P > 10;\ H_1: P < 10$

12. Welke verdeling past het best bij de volgende dataset:

3	4	4	8
8	3	3	2
2	5	2	8
9	5	1	2
14	10	9	1

a. Bin(20,0.2) **b.** Bin(11,0.2) **c.** Geo(0.3)
d. Bin(14,0.3) **e.** Geo(0.2) **f.** Geo(0.4)

13. Laat X een $Geo(1/3)$ stochast zijn, bereken de kans dat X oneven is. Met andere woorden, bereken de kans dat je een oneven aantal keren moet gooien met een onzuivere munt voordat je kop hebt, afgerond op 1 decimaal.

a. 0.3 **b.** 0.4 **c.** 0.5 **d.** 0.6 **e.** 0.7 **f.** 0.8

14. De levensduur van een ledlamp is exponentieel verdeeld met een verwachting van 10 jaar. Hoe groot is de kans dat de lamp meer dan 15 jaar meegaat, afgerond op 1 decimaal?

a. 0.1 **b.** 0.2 **c.** 0.3 **d.** 0.4 **e.** 0.5 **f.** 0.6

15. Bij de kleinste kwadraten methode heb je te maken met metingen y_i die afhangen van een parameter x_i. Je zoekt het beste lineaire verband tussen deze metingen en parameters via de lijn $y_i = \alpha x_i + \beta$ voor constanten α, β. Daarbij minimaliseer je de som $\sum_{i=1}^{n}(y_i - \alpha x_i - \beta)^2$. Wat minimaliseer je dan precies?

a. De verwachting van de meetfout
b. De variantie van de meetfout
c. De verwachting van de data y_i
d. De variantie van de data y_i
e. De verwachting van het quotiënt y_i/x_i
f. De variantie van het quotiënt y_i/x_i

OPEN VRAGEN

Opgave 1 In Bangladesh stromen drie grote rivieren uit: de Ganges, de Brahmaputra en de Meghna. Tachtig procent van het land is op zeeniveau en vanwege de enorme rivierafvoer zijn er jaarlijks overstromingen. Een overstroming is catastrofaal als er meer dan driekwart van het land overstroomt. In de afgelopen 66 jaar, van 1950 tot en met 2015, waren er catastrofale overstromingen in: 1954, 1955, 1970, 1987, 1988, 1998, 2004, 2007 and 2012.

A. Een catastrofale overstroming is op te vatten als een $Ber(p)$ stochast. Geef een schatting van p; geef ook een korte motivatie van de berekening.

B. Het aantal jaren tussen opeenvolgende overstromingen is op te vatten als een $Geo(p)$ stochast. De dataset is dan, als we beginnen te tellen vanaf 1950, gegeven door 4, 1, 15, 17, 1, 10, 6, 3, 5. Bereken p via het gemiddelde van deze dataset. Licht de berekening toe.

C. De schatting in B kan ook worden uitgerekend via de kans dat het aantal jaar tussen opeenvolgende overstromingen gelijk is aan 1. Welke schatting van p levert dit op?

D. Welke van de schatters in A, B en C zijn zuiver?

Opgave 2

De stochasten $X_1, \ldots, X_{100}$ zijn onafhankelijk en identiek verdeeld. De kansdichtheidsfunctie is

$$f(x) = \begin{cases} \dfrac{x}{24} & \text{als } 4 \leq x \leq 8 \\ 0 & \text{overig} \end{cases}$$

A. Bereken de verwachting en variantie van X_1.

B. Gebruik de centrale limietstelling om de kans $P(X_1 + X_2 + \cdots + X_{100} \geq 600)$ te bepalen.

Antwoorden MC:

1. F,	6. E,	11. C,
2. B,	7. E,	12. E,
3. C,	8. C,	13. D,
4. C,	9. E,	14. B,
5. B,	10. A,	15. B.

Korte beschrijving: opgave 1 komt letterlijk uit een tentamen van MIT, opgave 5, 6, 7, 11, komen uit tentamens van vorig jaar (januari of april), opgave 12 ook, maar daar had ik de data aangepast, opgave 15 komt uit een bonustoets van vorig jaar. Allemaal te vinden op deze pagina! Opgave 2 en 14 komen uit WB tentamens. Opgave 10 komt uit een TB tentamen. Allemaal te vinden op studeersnel. De rest had ik zelf verzonnen, door opgaven uit het boek een beetje aan te passen. Opgave 2 van het open deel staat er omdat heel veel studenten zo'n type opgave vroegen tijdens de tentamenbespreking van gisteren. Bedoeld om jullie te helpen. Opgave 1 heeft een echte dataset (gegevens van Verenigde Naties) uit een onderzoek naar klimaatsverandering en kans op overstromen.

Open deel

1A) E[X]=p bij Bernoulli, schat via het gemiddelde van 66 jaar (als je 2015 meetelt) 9 keer een 1 en 57 keer een nul, gemiddeld 9/66. Je kunt ook zeggen, schat via P(X=1) want dit is gelijk aan E[X] bij Bernoulli.

1B) E[X]=1/p bij geometrisch, schat via het gemiddelde van de wachttijden (4+1+15+17+1+10+6+3+5)/9=52/9 dus 1/p=62/9 oftewel p=9/62

1C) P(X=1)=p bij geometrisch (ook bij Bernoulli, overigens). De dataset heeft 2 keer een 1 onder 9 gegevens, dus kans is 2/9 volgens de dataset.

1D) Een schatter T is zuiver als deze verwachting E[T]=parameter die je wilt schatten. In 1A is E[T]=p. In 1B is E[1/T]=1/p maar dat betekent niet dat E[T]=p. In 1C Schat je via het event (X=1): zet 1 als het gebeurt en 0 als het niet gebeurt. Dit event heeft verwachting = som uitkomst maal kans = 1*P(X=1)+0*P(Xniet 1)=p. Dus dit is ook een zuivere schatter.

Vraag 1 was analoog aan de vragen over rokende vrouwen en zwangerschap, een bekend thema uit het boek.

2A) Bereken E[X] = integraal van 4 tot 8 over x keer x/24. Daar komt 6,22 uit. Bereken E[X^2]= integraal van 4 tot 8 over x^2 keer x/24. Daar komt 40 uit. Variantie 40 - (6,22)^2 en dat is ongeveer 1,3.

2B) Gevraagd kans van som van 100 boven de 600. De som van 100 heeft verwachting 622 en variantie ongeveer 130. De standaardafwijking is dus ongeveer 11. Nu ligt 622 ongeveer 2 standaardafwijkingen boven 600. Voor een standaardnormale geldt P(Z>2)=0.025 dus de kans moet orde 0.98 zijn.