

Tentamen Kansrekening en Statistiek
TI2116M, 27 januari 2015, 14.00–17.00 uur

Het tentamen bestaat uit 15 meerkeuzevragen en 3 open vragen. Een formuleblad wordt uitgedeeld.
S.v.p. na afloop weer inleveren. Normering: MC deel 2 punten per vraag. Open vragen deel: 8 + 7 + 6.
Eindcijfer = (totaal aantal punten + 4) gedeeld door 5.5.

1. De kans dat precies één van de gebeurtenissen A en B optreedt is
 - a. $1 - P(A) - P(B)$
 - b. $P(A \cup B)$
 - c. $P(A | B^c) + P(B | A^c)$
 - d. $1 - P(A)P(B)$
 - e. $P(A) + P(B) - 1$
 - f. $P(A) + P(B) - 2P(A \cap B)$
2. Van een bevolkingsgroep is bekend dat 2% lijdt aan een bepaalde hartafwijking. Een medisch centrum hanteert een test die deze hartafwijking bij 95% van de patiënten aan het licht brengt, maar ook bij niet-patiënten in 10% van de gevallen ten onrechte de afwijking diagnosticeert. De (conditionale) kans voor iemand om de hartafwijking te hebben gegeven dat de test hierop wijst is dan ongeveer gelijk aan
 - a. 95%
 - b. 89%
 - c. 74%
 - d. 64%
 - e. 47%
 - f. 16%
3. De tijd (in jaren) nodig voor een grootschalig project wordt gemodelleerd via een stochast X met de volgende kansdichtheid:

$$f(x) = \begin{cases} \frac{1}{2} & \text{als } 1 \leq x < 2 \\ \frac{1}{3} & \text{als } 2 \leq x \leq 3 \\ \frac{1}{6} & \text{als } 3 \leq x \leq 4 \\ 0 & \text{elders} \end{cases}$$

De verwachting van X is gelijk aan

- a. $2\frac{1}{12}$
 - b. $2\frac{1}{6}$
 - c. $2\frac{1}{4}$
 - d. $2\frac{1}{3}$
 - e. $2\frac{1}{2}$
 - f. $2\frac{2}{3}$
4. Stochast X heeft een Bernoulli verdeling met parameter p . De verwachting van $Y = 2^X$ is gelijk aan
 - a. 2^p
 - b. $p + 1$
 - c. $2p + 1$
 - d. p
 - e. $(1 - p)2^p$
 - f. $p2^p + (1 - p)2^{1-p}$
 5. Een bepaalde aandelenkoers heeft elke dag kans $\frac{1}{3}$ om 1 punt te stijgen, $\frac{1}{3}$ kans om 1 punt te dalen, en $\frac{1}{3}$ kans om gelijk te blijven (onafhankelijk van de voorgaande dagen). Geef de waardeverandering over zes opeenvolgende dagen aan met X_6 (dus X_6 neemt een van de waarden $-6, -5, \dots, 6$ aan. Er geldt: $P(X_6 = 4) =$
 - a. $\frac{21}{3^6}$
 - b. $\frac{41}{3^6}$
 - c. $\frac{2}{3^6}$
 - d. $\frac{15}{3^6}$
 - e. $\frac{6}{3^6}$
 - f. $\frac{11}{3^6}$
 - g. $\frac{20}{3^6}$
 - h. $\frac{7}{3^6}$
 6. Gegeven zijn twee onafhankelijke stochasten X en Y , met een exponentiële verdeling met parameter $\lambda = 1$. Dan heeft $M = \max\{X, Y\}$ als dichtheidsfunctie $f(x) = 0$ voor $x < 0$ en voor $x \geq 0$:
 - a. $f(x) = 1 - e^{-2x}$
 - b. $f(x) = (1 - e^{-x})^2$
 - c. $f(x) = 1 - (1 - e^{-x})^2$
 - d. $f(x) = 2e^{-2x}$
 - e. $f(x) = 2e^{-x} - e^{-2x}$
 - f. $f(x) = 2e^{-x} - 2e^{-2x}$
 7. Hoe kun je met Matlab een trekking simuleren uit een verdeling met dichtheid
$$f(x) = \begin{cases} 2x - 2, & \text{als } 1 \leq x \leq 2 \\ 0, & \text{anders} \end{cases} \quad ? \quad \text{Stap 1: } U = \text{rand}; \quad \text{Stap 2:}$$
 - a. $X = 1 + U$
 - b. $X = 1 + U^2$
 - c. $X = 1 + \text{sqrt}(U)$
 - d. $X = 2 - U$
 - e. $X = 2 - U^2$
 - f. $X = 2 - \text{sqrt}(U)$

8. $X_1, X_2, \dots, X_{300}$ zijn stochasten met onafhankelijke uniforme verdelingen op $\{1, 2, 3\}$, dus $P(X_i = k) = \frac{1}{3}$, voor $k = 1, 2, 3$.

De som $S_{300} = \sum_{i=1}^{300} X_i$ heeft dan bij benadering de normale verdeling

- a. $\mathcal{N}(200, 300)$ b. $\mathcal{N}(300, 300)$ c. $\mathcal{N}(600, 300)$ d. $\mathcal{N}(200, 200)$
e. $\mathcal{N}(300, 200)$ f. $\mathcal{N}(600, 200)$

9. Bereken voor een normaal verdeelde stochastische variabele X , met verwachting 1 en variantie $\sigma^2 = 1/4$ de kans $P\left(\frac{3}{4} \leq X \leq \frac{3}{2}\right)$

- a. 0.6587 b. 0.5987 c. 0.4013 d. 0.5328 e. 0.3413 f. 0.8413 g. 0.1587 h. 0.4311

10. 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

11. Orden volgende vier rijtjes ('datasets') naar oplopende steekproefstandaardafwijking:

- A: 1000, 1001, 1002, 1003, 1004, 1005 B: 10, 20, 30, 40, 50
C: 0, 1, 2, 3, 4, $\dots$, 10 D: -10, -5, 0, 5, 10

(Het is echt niet nodig om de waarden uit te rekenen.)

- a. $s_D < s_B < s_C < s_A$ b. $s_A < s_C < s_D < s_B$ c. $s_A < s_C < s_B < s_D$
d. $s_D < s_C < s_B < s_A$ e. $s_C < s_A < s_D < s_B$ f. $s_D < s_A < s_C < s_B$

12. We beschikken over een dataset van omvang 20 die we opvatten als een realisatie van steekproef $X_1, X_2, \dots, X_{20}$ uit de kansdichtheid

$$f(x) = \begin{cases} \frac{2x}{\theta^2} & , \text{ als } 0 \leq x \leq \theta, \\ 0 & , \text{ elders,} \end{cases}$$

waarbij $\theta > 0$ een onbekende parameter is. Definieer de volgende schatter voor θ :

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

Voor welke a is T een zuivere schatter voor θ ?

- a. $\frac{1}{40}$ b. $\frac{1}{20}$ c. $\frac{3}{20}$ d. $\frac{1}{10}$ e. $\frac{5}{40}$ f. $\frac{3}{40}$

13. In het kader van een vruchtbaarheidsonderzoek zijn gedurende twee jaar vrouwen gevolgd die zwanger wilden worden. Geregistreerd zijn de perioden (in maanden) tot zwangerschap (als in het voorbeeld in het boek). Onder zekere vereenvoudigende aannamen zal X , de periode tot zwangerschap van een willekeurige vrouw een geometrische verdeling hebben met parameter p . Als $X_1, \dots, X_n$ een steekproef is, welk van de twee schatters is/zijn zuiver voor $1/p$?

$$T_1 = \frac{n}{\text{aantal } X_i \text{ gelijk aan } 1}, \quad T_2 = \bar{X}_n$$

- a. beide zijn zuiver b. T_1 is zuiver, T_2 onzuiver
c. beide onzuiver d. T_1 is onzuiver, T_2 zuiver
14. Beschouw betrouwbaarheidsintervallen voor het gemiddelde van normale data met bekende variantie σ^2 . Welk van de beweringen is/zijn waar: het betrouwbaarheidsinterval wordt groter als
(A) de steekproefomvang groter wordt (bij gelijkblijvende betrouwbaarheid en variantie);
(B) de betrouwbaarheid groter wordt (bij gelijkblijvende steekproefomvang en variantie);
(C) de variantie groter wordt (bij gelijkblijvende steekproefomvang en betrouwbaarheid).
- a. alle drie onwaar b. alleen (A) waar c. alleen (B) waar d. alleen (C) waar
e. (A) en (B) waar f. (A) en (C) waar g. (B) en (C) waar h. alle drie waar
15. Men beschikt over een dataverzameling die een realisatie vormt van een steekproef $X_1, \dots, X_{16}$ uit een $N(\mu, 4)$ verdeling. Men toetst $H_0: \mu = 1$ tegen $H_1: \mu > 1$ met toetsingsgrootheid $T = \bar{X}_{16}$. Men verworpt H_0 ten gunste van H_1 als voor de data $\bar{x}_{16} \geq 2$. In dit geval is de kans op een type I fout
- a. 0.1151 b. 0.2734 c. 0.0556 d. 0.1587 e. 0.3085 f. 0.2302

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. Stel X en Y zijn de uitkomsten van twee (onafhankelijke) worpen met een dobbelsteen. Definieer $M = \max\{X, Y\}$, $Z = \min\{X, Y\}$ en $V = |X - Y|$, de absolute waarde van het verschil.
- a. Bereken $P(M = V)$, $P(M = Z)$ en $P(V = Z)$.
b. Beargumenteer (zonder precieze berekening) of M en Z positief of negatief gecorreleerd zullen zijn. Net zo voor V en Z .
2. Drie spelers mikken op het middelpunt (de roos) van een rond dartbord met straal R . De afstanden van het geraakte punt tot het middelpunt geven we aan met respectievelijk Y_1, Y_2, Y_3 . Stel dat de dichtheden – voor $0 \leq y \leq R$ – gegeven worden door

$$f_1(y) = \frac{2y}{R^2}, \quad f_2(y) = \frac{1}{2\sqrt{R}} \frac{1}{\sqrt{y}}, \quad f_3(y) = \frac{1}{R}$$

- a. Bereken voor speler 1 en voor speler 3 de verwachting (van de afstand tot het midden).
b. Welke van de drie dichtheden correspondeert met ‘uniform’ gooien op het bord. (Anders gezegd: als een speler een volledig random punt op het bord raakt, wat is dan de bijbehorende dichtheid voor ‘de afstand tot het midden’?)

- c. Bereken voor speler 3 het 10 % percentiel (anders gezegd: het 0.1 kwantiel).
3. Volgens de wet van Hardy-Weinberg zijn gen-frekwenties in evenwicht als de genotypen AA , Aa en aa voorkomen met kansen respectievelijk $(1-\theta)^2$, $2\theta(1-\theta)$ en θ^2 . Hierbij is $0 < \theta < 1$ een onbekende parameter. We willen de maximum likelihood schatting berekenen voor θ . Bij een dataverzameling omtrent de Chinese bevolking van Hong Kong in 1937 kwamen de genotypen voor met de volgende aantallen:

Type	Aantal
1: AA	342
2: Aa	500
3: aa	187
Totaal	1029

- a. Leid af dat de likelihoodfunctie wordt gegeven door

$$L(\theta) = C \cdot 2^{500} (1 - \theta)^{1184} \theta^{874},$$

waarbij C een constante is die het aantal volgordes aangeeft met 342 keer genotype 1, 500 keer genotype 2, en 187 keer genotype 3.

Bepaal tevens de maximum likelihood schatting op grond van de bovenstaande dataverzameling.

- b. Bereken de maximum likelihood schatting voor θ voor een willekeurige dataverzameling waarbij de genotypen voorkomen met aantallen n_1 , n_2 en n_3 (alle drie > 0).