Hertentamen Inleiding Kansrekening en Statistiek (WISB161) 2018-2019
16 april 2019

Opgave 1

Zij $\theta > 0$ een parameter en zij X een continue stochastische variabele op de reële getallen met kansdichtheidsfunctie

$$f_\theta(x) = \begin{cases} 2\frac{\theta^2}{x^3} & \text{voor } x \geq \theta \\ 0 & \text{voor } x < \theta. \end{cases}$$

Stel dat we een willekeurig steekproef $X_1, X_2, \ldots, X_n$ van grootte n hebben waarbij alle X_i onafhankelijk en identiek verdeeld zijn als X en zij $x_1, x_2, \ldots, x_n$ de realisatie van de steekproef.

a 5pt) Laat zien dat f_θ inderdaad een kansdichtheidsfunctie is.
ANTWOORD: Een stuksgewijs continue functie f is een kansdichtheidsfunctie als $\int\limits_{-\infty}^{\infty} f(x)dx = 1$. Hier geldt: $\int\limits_{-\infty}^{\infty} f(x)dx = \int\limits_{\theta}^{\infty} 2\frac{\theta^2}{x^3}dx = -\frac{\theta^2}{x^2}|_\theta^\infty = -0+1 = 1$.

b 10pt) Bepaal de meest-waarschijnlijke schatter (maximum likelihood estimate) van θ.
ANTWOORD: De Likelihoodfunctie wordt gegeven door

$$L(\theta) = \begin{cases} \prod\limits_{i=1}^{n} 2\frac{\theta^2}{x_i^3} & \text{als } \min\{x_1, x_2, \ldots, x_n\} \geq \theta \\ 0 & \text{anders} \end{cases}$$

Deze functie van θ kan geschreven worden als:

$$L(\theta) = \begin{cases} C\theta^{2n} & \text{als } \min\{x_1, x_2, \ldots, x_n\} \geq \theta \\ 0 & \text{anders} \end{cases}$$

met $C = \prod\limits_{i=1}^{n} 2\frac{1}{x_i^3}$

Op het interval $(\min\{x_1, x_2, \ldots, x_n\}, \infty)$ is de likelihood een strikt stijgende functie. De maximum likelihood is dus de grootste waarde van θ die aan de conditie
$\min\{x_1, x_2, \ldots, x_n\} \geq \theta$ voldoet, dus de MLE is $\hat{\theta} = \min\{x_1, x_2, \ldots, x_n\}$.

Opgave 2

Beschouw een erfelijke aandoening. Het is bekend dat 1% van de vaders en 1% van de moeders drager is van de aandoening. Partnerkeuze hangt niet af van dragerschap. Als beide ouders drager zijn van de aandoening, is de kans 1/4 dat hun kind de aandoening heeft, onafhankelijk van of hun andere kinderen de aandoening hebben. Als slechts een van de ouders of geen van de ouders drager is van de aandoening hebben hun kinderen de aandoening niet. Een test kan de aandoening in kinderen aantonen. De sensitiviteit van de test (de kans op een positief testresultaat als het geteste kind de aandoening heeft) is 1, de specificiteit (de kans op een negatief testresultaat als het geteste kind de aandoening niet heeft) is 0.99. U mag er in deze opgave vanuit gaan dat er geen ouders zijn die de aandoening zelf hebben.

a 6pt Wat is de kans dat een kind zonder broers en zussen de aandoening heeft als het testresultaat positief was?

ANTWOORD: $P(\text{aandoening}|\text{test positief}) = \frac{P(\text{aandoening,test positief})}{P(\text{test positief})} =$
$\frac{P(\text{test positief}|\text{aandoening})P(\text{aandoening})}{P(\text{test positief})} =$
$\frac{P(\text{test positief}|\text{aandoening})P(\text{aandoening})}{P(\text{test positief}|\text{aandoening})P(\text{aandoening})+P(\text{test positief}|\text{aandoening niet})P(\text{aandoening niet})}$
$P(\text{aandoening} = 0.01 \times 0.01 \times 0.025.$
Dus $P(\text{aandoening}|\text{test positief}) = \frac{1\times 0.01\times 0.01\times 0.25}{1\times 0.01\times 0.01\times 0.25+0.01\times(1-0.01\times 0.01\times 0.25)}$
$\approx \frac{0.01\times 0.01\times 0.25}{0.01} = 0.25 \times 0.01 = 0.0025.$

b 9pt) Wat is de kans dat een kind de aandoening heeft als zowel hij als zijn broer een positief testresultaat hadden en er verder niets bekend is over dragerschap van de aandoening in de familie? Om een numerieke uitkomst te krijgen mag u in het rekenwerk redelijke benaderingen gebruiken (zo lang u geen onnodige benaderingen doet in de trend van $1 \approx 0.99 \approx 0.99^2 \approx 0.99^3 \ldots \approx 0$).

ANTWOORD: $P(\text{aandoening}|\text{test positief en test broer positief}) = \frac{P(\text{test positief en test broer positief}|\text{aandoening})P(\text{aandoening})}{P(\text{test positief en test broer positief})}$
Er geldt $P(\text{test positief en test broer positief}) =$
$P(\text{test positief en test broer positief}|\text{ouders allebei drager})P(\text{ouders allebei drager})+$
$P(\text{test positief en test broer positief}|\text{ouders niet allebei drager})P(\text{ouders niet allebei drager})$

Er geldt: $P(\text{test positief en test broer positief}|\text{ouders allebei drager}) = (0.25 + 0.75 \times 0.01)^2 = (0.25 + 0.0075)^2 \approx 0.25^2 = \frac{1}{16}$ en $P(\text{ouders allebei drager}) = 0.01^2$.

Tevens geldt:
$P(\text{test positief en test broer positief}|\text{ouders niet allebei drager})P(\text{ouders niet allebei drager}) = 0.01 \times 0.01 \times (1 - 0.01^2) \approx 0.01^2$. Dus geldt:

$P(\text{test positief en test broer positief}) \approx \frac{1}{16} \times 0.01^2 + 0.01^2$

Er geldt $P(\text{aandoening}) = 0.01 \times 0.01 \times 0.25$.

En er geldt:
$P(\text{test positief en test broer positief}|\text{aandoening}) = P(\text{test broer positief}|\text{aandoening}) =$
$P(\text{test broer positief}|\text{aandoening en broer aandoening})P(\text{broer aandoening}|\text{aandoening})+$
$P(\text{test broer positief }|\text{aandoening en broer aandoening niet})P(\text{broer aandoening niet}|\text{aandoening}) =$
$1 \times \frac{1}{4} + 0.01 \times \frac{3}{4} = 0.2575$.
Er geldt dus: $P(\text{aandoening}|\text{test positief en test broer positief}) =$
$\frac{P(\text{test positief en test broer positief}|\text{aandoening})P(\text{aandoening})}{P(\text{test positief en test broer positief})} =$
$\frac{0.2575 \times (0.01^2 \times 0.25)}{17/16 \times 0.01^2} = \frac{0.2575 \times 0.25}{17/16} \approx \frac{1}{17}$.

Opgave 3
Een studente wil de lichtsnelheid bepalen. Daartoe meet zij de snelheid van het licht 10 keer. Zij veronderstelt dat de gemeten waarden onafhankelijk van elkaar zijn, dat er geen systematische fout optreedt en dat de metingen normaal verdeeld zijn. De variantie in de uitkomsten is niet van te voren bekend. Op basis van de 10 metingen construeert ze een symmetrisch 95% betrouwbaarheidsinterval voor de lichtsnelheid. Ze vindt als 95% betrouwbaarheidsinterval: (299792454 - 299792462) meter/seconde.

a 5pt) Wat is het gemiddelde van de 10 metingen geweest?

ANTWOORD: Voor een symmetrisch 95% betrouwbaarheidsinterval (l_n, u_n) geldt $P(\theta < l_n) = 0.025$ en $P(\theta > u_n) = 0.025$. In deze situatie geldt dat het berekende betrouwbaarheidsinterval er als volgt uit zal zien: $(\bar{x}_{10} - t_{9,0.025}\frac{s_{10}}{\sqrt{10}}, \bar{x}_{10} + t_{9,0.025}\frac{s_{10}}{\sqrt{10}})$, i.e., gebaseerd op de student-t verdeling. Het gemiddelde van de 10 metingen $\bar{x}_{10}$ ligt dus in het midden van het betrouwbaarheidsinterval en dus $\bar{x}_{10} = 299792458$ m/s.

b 9pt) Is het op basis van deze data mogelijk een symmetrisch 90% betrouwbaarheidsinterval voor de lichtsnelheid te bepalen? Zo ja, bepaal dit 90% betrouwbaarheidsinterval, zo nee, welke data heeft u nodig om dit wel te kunnen.

ANTWOORD: Een 90% betrouwbaarheidsinterval is van de vorm $(\bar{x}_{10} - t_{9,0.05}\frac{s_{10}}{\sqrt{10}}, \bar{x}_{10} + t_{9,0.05}\frac{s_{10}}{\sqrt{10}})$. We weten dat $t_{9,0.025}\frac{s_{10}}{\sqrt{10}} = 4$, dus $t_{9,0.05}\frac{s_{10}}{\sqrt{10}} = t_{9,0.025}\frac{s_{10}}{\sqrt{10}}\frac{t_{9,0.05}}{t_{9,0.025}} = 4\frac{t_{9,0.05}}{t_{9,0.025}}$
Volgens de tabel van de student-t-verdeling geldt: $t_{9,0.05} = 1.833$ en $t_{9,0.025} = 2.262$.
Dus geldt: $t_{9,0.05}\frac{s_{10}}{\sqrt{10}} = 4\frac{1.833}{2.262} \approx 3.24$.

Een symmetrisch 90% betrouwbaarheidsinterval voor de lichtsnelheid is dus: $(299792458 - 3.24, 299792458 + 3.24) = (299792454.76 - 299792461.24)$

Opgave 4
Veronderstel dat de kans om te winnen met roulette in een casino $\frac{1}{2}$ is en de kans om

te verliezen ook $\frac{1}{2}$. Het casino krijgt een tip dat een bezoeker mogelijk vals speelt met roulette. Om dat te onderzoeken worden 4 opeenvolgende roulette-spellen waar de bezoeker aan deelneemt geobserveerd. Als deze observaties aanleiding geven om vals spel te vermoeden, wordt er een grootschaliger onderzoek gestart. Het casino neemt hiervoor een significantieniveau van $\alpha = 0.2$.

a 4pt) Formuleer de nulhypothese en de alternatieve hypothese.

ANTWOORD: De nulhypothese H_0 is: De kans dat de bezoeker een roulettespel wint is $\frac{1}{2}$.
De alternatieve hypothese H_1 is: De kans dat de bezoeker een roulettespel wint is groter dan $\frac{1}{2}$.

De bezoeker wint het spel bij zijn 1^{e}, 2^{e} en 4^{e} spel, maar verliest tijdens het 3^{e} spel.

b 10pt) Zal het casino de bezoeker verder gaan onderzoeken?

ANTWOORD: We bepalen wat de kans op een soortgelijk resultaat of extremer is onder de nulhypothese. De kans dat de bezoeker 3 van de 4 keer wint of alle vier de keren wint als hij eerlijk speelt is $\binom{4}{3}\frac{1}{2}^4 + \binom{4}{4}\frac{1}{2}^4 = 5\frac{1}{2}^4 = \frac{5}{16} > \alpha = 0.2$. Het casino heeft dus geen reden om de bezoeker verder te onderzoeken.

Opgave 5 *Gebruik een nieuw vel papier!*
Zij Ω de uitkomstenruimte en P een kansfunctie op Ω. Veronderstel dat A, B en C deelverzamelingen van Ω zijn. Toon aan of de volgende stellingen waar zijn voor alle deelverzamelingen A, B en C die aan de condities voldoen: Merk op, een Venn-diagram is geen bewijs.

a 6pt) Als $P(A) = P(B)$, dan geldt $A = B$.

ANTWOORD: De stelling is niet waar. Stel $\Omega = \{0, 1\}$, en $P(\{0\}) = P(\{1\}) = \frac{1}{2}$, i.e., we hebben een Bernoulli-stochast met parameter $1/2$. $P(\{0\}) = P(\{1\})$, maar $\{0\} \neq \{1\}$.

b 8pt) Als $A \subseteq B$ en $P(A) = P(B)$, dan geldt $P(A \cap C) = P(B \cap C)$.

ANTWOORD: De stelling is waar. Er geldt $B = B \cap \Omega = B \cap (A \cup A^c) = (B \cap A) \cup (B \cap A^C)$. De verzamelingen $B \cap A$ en $B \cap A^c$ zijn disjunct (omdat A en A^c disjunct zijn), dus geldt volgens eigenschap 2 van een kansfunctie dat:
$P(B) = P(B \cap A) + P(B \cap A^c)$.
Omdat $A \subseteq B$ geldt: $B \cap A = A$ en dus $P(B \cap A) = P(A)$. Er geldt dus: $P(B) = P(A) + P(B \cap A^c)$. Maar omdat $P(B) = P(A)$ geldt: $P(B \cap A^c) = 0$.

Gebruik nu dat $B = A \cup (B \cap A^c)$.
$P(B \cap C) = P((A \cup (B \cap A^c)) \cap C) = P((A \cap C) \cup (B \cap A^c) \cap C)) = P(A \cap C) + P((B \cap A^c) \cap C)$ omdat $A \cap C$ en $(B \cap A^c) \cap C$ disjunct zijn. Dus geldt: $P(B \cap C) = P(A \cap C) + P((B \cap A^c) \cap C)$, maar $P((B \cap A^c) \cap C) = 0$ omdat $(B \cap A^c) \cap C \subseteq B \cap A^c$ en dus $P((B \cap A^c) \cap C) \leq P(B \cap A^c) = 0$. We concluderen dus dat $P(B \cap C) = P(A \cap C)$.

Opgave 6
Zij $X_1, X_2, X_3 \ldots$ onafhankelijke, identiek verdeelde Bernoulli stochasten met parameter p. Voor $i \geq 4$, definieer de stochast $S_i := \sum_{j=i-3}^{i} X_i$ en de stochast B_i door

$$B_i = \begin{cases} 1 & \text{als } S_i = 2 \\ 0 & \text{als } S_i \neq 2. \end{cases}$$

a 3pt) Bepaal de kansfunctie van S_i voor alle $i \geq 4$.

ANTWOORD: S_i is $Bin(4,p)$-verdeeld, dus $P(S_i = k) = \binom{4}{k} p^k(1-p)^{(4-k)}$ als $0 \leq k \leq 4$ en 0 anders.

b 3pt) Bepaal de verwachtingswaarde van B_i voor alle $i \geq 4$.

ANTWOORD: $E(B_i) = 0P(B_i = 0) + 1P(B_i = 1) = P(B_i = 1) = P(S_i = 2) = \binom{4}{2} p^2(1-p)^2 = 6p^2(1-p)^2$.

c 3pt) Bepaal de variantie van B_i voor alle $i \geq 4$.

ANTWOORD: B_i is een Bernoulli-stochast met parameter $r = 6p^2(1-p)^2$. Dus geldt $\text{Var}(B_i) = r(1-r) = 6p^2(1-p)^2(1-6p^2(1-p)^2)$.

d 10pt) Bepaal $\text{Cov}(B_i, B_j)$ als $i, j \geq 4$ in de situaties dat $|i-j| = 0$, dat $|i-j| = 1$ en dat $|i-j| \geq 4$.

ANTWOORD:

$i = j$ Dan is $\text{Cov}(B_i, B_j) = \text{Var}(B_i) = 6p^2(1-p)^2(1-6p^2(1-p)^2)$.

$|i-j| \geq 4$ Dan zijn S_i en S_j onafhankelijk, en dus zijn B_i en B_j ook onafhankelijk, dus $\text{Cov}(B_i, B_j) = 0$.

$|i-j| = 1$ Neem aan dat $i > j$, draai anders de rol van i en j om.
$\text{Cov}(B_i, B_j) = E(B_iB_j) - E(B_i)E(B_j) = E(B_iB_j) - \left(6p^2(1-p)^2\right)^2$.
$E(B_iB_j) = 0P(B_iB_j = 0) + 1P(B_iB_j = 1) = P(B_iB_j = 1) = P(B_i = 1, B_j = 1) = P(B_i = 1|B_j = 1)P(B_j = 1) = P(B_i = 1|B_j = 1)6p^2(1-p)^2$
Beschouw nu $P(B_i = 1|B_j = 1)$.
$P(B_i = 1|B_j = 1) = P(S_i = 2|S_j = 2) =$
$P(X_{j+1} + X_j + X_{j-1} + X_{j-2} = 2|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2) =$
$P(X_{j+1} + X_j + X_{j-1} + X_{j-2} + X_{j-3} - X_{j-3} = 2|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2) =$
$P(X_{j+1} + 2 - X_{j-3} = 2|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2) =$
$P(X_{j+1} - X_{j-3} = 0|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2)$
Conditioneer nu op de toestand van X_{j-3}.
$P(X_{j+1} - X_{j-3} = 0|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2) =$
$P(X_{j+1} - X_{j-3} = 0|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2, X_{j-3} = 0)P(X_{j-3} = 0|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2) + P(X_{j+1} - X_{j-3} = 0|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2, X_{j-3} = 1)P(X_{j-3} = 1|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2) =$
$P(X_{j+1} - X_{j-3} = 0|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2, X_{j-3} = 0)\frac{1}{2} + P(X_{j+1} - X_{j-3} = 1|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2, X_{j-3} = 0)\frac{1}{2} =$
$P(X_{j+1} = 0|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2, X_{j-3} = 0)\frac{1}{2} + P(X_{j+1} = 1|X_j + X_{j-1} + X_{j-2} + X_{j-3} = 2, X_{j-3} = 1)\frac{1}{2} =$
$(1-p)\frac{1}{2} + p\frac{1}{2} = \frac{1}{2}$.
Hieruit volgt: $\text{Cov}(B_i, B_j) = E(B_iB_j) - \left(6p^2(1-p)^2\right)^2 = \frac{1}{2}6p^2(1-p)^2 - \left(6p^2(1-p)^2\right)^2 = 3p^2(1-p)^2 - \left(6p^2(1-p)^2\right)^2$

Opgave 7
Zij X een uniform verdeelde stochast op $(0,1)$. Definieer de stochast Z door $Z = X + X^2$.

a 9pt) Bepaal zowel de cumulatieve verdeling $F_Z(z)$ als de kansdichtheidsfunctie $f_Z(z)$ van de stochast Z.

ANTWOORD: Stel $0 < z < 2$) (omdat Z geen uitkomsten buiten deze range kan hebben).
$F_Z(z) = P(Z \leq z) = P(X + X^2 \leq z) = P(X^2 + X - z \leq 0)$. Los op: $x^2 + x - z = 0$. Dit geeft $x = \frac{-1 \pm \sqrt{1+4z}}{2}$.
Merk op dat de functie $g(z) = \frac{-1+\sqrt{1+4z}}{2}$ een strikt monotoom stijgende functie in z

is met $g(0) = 0$ en $g(2) = 1$, dus $0 < \frac{-1+\sqrt{1+4z}}{2} < 1$ voor $0 < z < 2$. Hieruit volgt dat voor alle $x \in (0, \frac{-1+\sqrt{1+4z}}{2}]$ geldt dat $x^2 + x - z \leq 0$.

$F_Z(z) = P(X^2 + X - z \leq 0) = \int\limits_0^{\frac{-1+\sqrt{1+4z}}{2}} 1dx = \frac{-1+\sqrt{1+4z}}{2}$. Dus geldt:

$$F_Z(z) = \begin{cases} 0 & \text{als } z \leq 0 \\ \frac{-1+\sqrt{1+4z}}{2} & \text{als } 0 < z \leq 2 \\ 1 & \text{als } z \geq 2 \end{cases}$$

$$f_Z(z) = \frac{d}{dz} F_Z(z) \begin{cases} 0 & \text{als } z \leq 0 \\ \frac{1}{\sqrt{1+4z}} & \text{als } 0 < z \leq 2 \\ 0 & \text{als } z \geq 2 \end{cases}$$

Einde.