Inleiding Kansrekening en Statistiek (WISB 161)
Hertentamen

Sjoerd Dirksen

14 april 2020, 09:00-12:00

Dit tentamen bestaat uit 6 vragen en 1 bonusvraag. Schrijf op het eerste blad van jouw uitwerkingen jouw naam en studentnummer en nummer alle pagina's. Voorzie elk antwoord van een zorgvuldige motivatie. Vereenvoudig je antwoord voor zover mogelijk. Je mag gebruik maken van een rekenmachine.

Vraag 1 [6 punten]

Zij (X, Y) een discrete kansvector die waarden aanneemt in $\{-1, 0, 1\} \times \{-1, 0, 1\}$. De waarden van de kansfunctie van (X, Y) worden gegeven in de onderstaande tabel.

	-1	0	1
-1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$
0	$\frac{3}{16}$	0	$\frac{3}{16}$
1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$

Bereken de covariantie tussen X en Y. Zijn X en Y onafhankelijk?

Uitwerking:

Merk op dat X en Y identiek verdeeld zijn,

$$\mathbb{P}(X = -1) = \frac{1}{16} + \frac{3}{16} + \frac{1}{16} = \frac{5}{16}, \qquad \mathbb{P}(X = 0) = \frac{3}{16} + \frac{3}{16} = \frac{6}{16}$$

en

$$\mathbb{P}(X = 1) = \frac{1}{16} + \frac{3}{16} + \frac{1}{16} = \frac{5}{16}.$$

Hieruit volgt dat

$$\mathbb{E}(Y) = \mathbb{E}(X) = -1 \cdot \mathbb{P}(X = -1) + 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1) = 0.$$

Zij $Z = XY$. Uit de tabel zien we dat Z waarden aanneemt in $\{-1, 0, 1\}$ en

$$\mathbb{P}(Z = 1) = \mathbb{P}(Z = -1) = \frac{2}{16}, \qquad \mathbb{P}(Z = 0) = \frac{12}{16},$$

zodat $\mathbb{E}(Z) = 0$. We concluderen dat

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0.$$

Hoewel X en Y ongecorreleerd zijn, zijn ze niet onafhankelijk. Bijvoorbeeld is

$$\mathbb{P}(X = 0, Y = 0) = 0 \neq \frac{6}{16} \cdot \frac{6}{16} = \mathbb{P}(X = 0)\mathbb{P}(Y = 0).$$

Vraag 2 [7 punten]

Zij X een continue kansvariabele, $\alpha \in \mathbb{R}$, en definieer $F : \mathbb{R} \to [0,1]$ door

$$F(z) = \mathbb{P}(X \leq z \mid X \geq \alpha).$$

(a) Laat zien dat voor $z \geq \alpha$ geldt

$$F(z) = \frac{F_X(z) - F_X(\alpha)}{1 - F_X(\alpha)},$$

waarbij F_X de verdelingsfunctie van X is. Toon aan dat F de verdelingsfunctie van een kansvariabele Z is.

(b) Stel dat X de levensduur van een accu in uren is. Neem aan dat de kansdichtheid van X gegeven wordt door

$$f_X(x) = \begin{cases} xe^{-x^2/2} & \text{als } x \geq 0 \\ 0 & \text{elders.} \end{cases}$$

Geef een interpretatie van de verwachting van Z en bereken deze. Druk je antwoord uit in termen van de standaard normale verdelingsfunctie.

Uitwerking:

(a) Uit de definitie van de conditionele verwachting volgt dat

$$F(z) = \mathbb{P}(X \leq z \mid X \geq \alpha) = \frac{\mathbb{P}(\{X \leq z\} \cap \{X \geq \alpha\}}{\mathbb{P}(X \geq \alpha)} = \frac{\mathbb{P}(\alpha \leq X \leq z)}{1 - \mathbb{P}(X < \alpha)}$$

Als $z < \alpha$ dan geldt

$$\mathbb{P}(\alpha \leq X \leq z) = \mathbb{P}(\emptyset) = 0.$$

Stel dat $z \geq \alpha$. Aangezien X continu is (en dus $\mathbb{P}(X = \alpha) = 0$), vinden we

$$\mathbb{P}(\alpha \leq X \leq z) = \mathbb{P}(X \leq z) - \mathbb{P}(X < \alpha) = \mathbb{P}(X \leq z) - \mathbb{P}(X \leq \alpha) = F_X(z) - F_X(\alpha)$$

en

$$\mathbb{P}(X < \alpha) = \mathbb{P}(X \leq \alpha) = F_X(\alpha).$$

Hieruit volgt dat

$$F(z) = \begin{cases} \frac{F_X(z) - F_X(\alpha)}{1 - F_X(\alpha)} & \text{als } z \geq \alpha \\ 0 & \text{als } z < \alpha. \end{cases}$$

Aangezien F_X monotoon niet-dalend en rechts-continu is, volgt direct dat F deze eigenschappen ook heeft. Aangezien $\lim_{z\to\infty} F_X(z) = 1$ volgt uit de rekenregels voor limieten dat

$$\lim_{z\to\infty} F(z) = \lim_{z\to\infty} \frac{F_X(z) - F_X(\alpha)}{1 - F_X(\alpha)} = \frac{\lim_{z\to\infty} F_X(z) - F_X(\alpha)}{1 - F_X(\alpha)} = 1.$$

Tenslotte is het duidelijk dat $\lim_{z\to-\infty} F(z) = 0$. Uit een stelling uit de college aantekeningen volgt nu dat F de verdelingsfunctie van een kansvariabele Z is.

(b) We kunnen Z interpreteren als de levensduur van een accu, gegeven dat we weten dat de accu al α uren werkzaam is (in het bijzonder kunnen we aannemen dat $\alpha \geq 0$). De verwachting $\mathbb{E}Z$ is daarom de verwachte totale levensduur onder de aanname dat de accu minstens α uren werkt. Uit deel (a) volgt dat de kansdichtheid f_Z van Z gegeven wordt door

$$f_Z(z) = F'(z) = \begin{cases} \frac{f_X(z)}{1-F_X(\alpha)} & \text{als } z \geq \alpha \\ 0 & \text{als } z < \alpha. \end{cases}$$

Merk op dat

$$F_X(\alpha) = \int_0^\alpha x e^{-x^2/2}\, dx = -e^{-x^2/2}\Big|_{x=0}^{x=\alpha} = 1 - e^{-\alpha^2/2}$$

en dus

$$\frac{1}{1-F_X(\alpha)} = e^{\alpha^2/2}.$$

Nu volgt met behulp van partiële integratie (met '$f(z) = z$, $g'(z) = ze^{-z^2/2}$', zodat '$f'(z) = 1$, $g(z) = -e^{-z^2/2}$')

$$\begin{aligned}
\mathbb{E}Z &= \int_{-\infty}^{\infty} z f_Z(z)\, dz \\
&= e^{\alpha^2/2} \int_\alpha^\infty z^2 e^{-z^2/2}\, dz \\
&= e^{\alpha^2/2} \Big[-ze^{-z^2/2}\Big|_\alpha^\infty + \int_\alpha^\infty e^{-z^2/2}\, dz \Big] \\
&= e^{\alpha^2/2} \Big[\alpha e^{-\alpha^2/2} + \sqrt{2\pi}\,\mathbb{P}(Y \geq \alpha) \Big] \\
&= e^{\alpha^2/2} \Big[\alpha e^{-\alpha^2/2} + \sqrt{2\pi}(1 - \mathbb{P}(Y < \alpha)) \Big] = \alpha + e^{\alpha^2/2}\sqrt{2\pi}(1 - F_Y(\alpha)),
\end{aligned}$$

waarbij Y standaard normaal verdeeld is.

Vraag 3 [5 punten]

Laat X en Y onafhankelijke kansvariabelen zijn die beide Bernoulli verdeeld zijn met parameter p. Definieer

$$Z = ((X+1)^Y)^{X+2} + X(1-X)^{Y+1}.$$

Bepaal alle p zodat $\mathbb{E}Z \geq 3$.

Uitwerking:

Aangezien X waarden in $\{0,1\}$ aanneemt, geldt

$$X(1-X)^{Y+1} = 0.$$

Merk op dat

$$Z = 1 \Leftrightarrow (X,Y) \in \{(0,0),(0,1),(1,0)\}$$

en

$$Z = 8 \Leftrightarrow (X,Y) = (1,1).$$

Nu volgt direct dat

$$\begin{aligned}\mathbb{E}Z &= 1 \cdot \mathbb{P}(Z = 1) + 8 \cdot \mathbb{P}(Z = 8)\\ &= \mathbb{P}((X, Y) \in \{(0, 0), (0, 1), (1, 0)\}) + 8\mathbb{P}(X = 1, Y = 1)\\ &= (1 - p)^2 + 2p(1 - p) + 8p^2 = 1 - 2p + p^2 + 2p - 2p^2 + 8p^2 = 7p^2 + 1.\end{aligned}$$

Daarom geldt $\mathbb{E}Z \geq 3$ dan en slechts dan als $\sqrt{\frac{2}{7}} \leq p \leq 1$.

Vraag 4 [6 punten]

Wiskundestudenten Guiseppe en Chiara zitten in lockdown in hun studentenhuis in Turijn en vervelen zich te pletter. Om de tijd te verdrijven spelen ze een spelletje. Guiseppe vult een doos met m ballen met de nummers 1 tot en met m. Chiara krijgt de doos en mag n keer blind een bal uit de doos trekken om het totaal aantal ballen m in de doos te raden. Na iedere trekking moet zij de bal weer terug in de doos doen. Chiara gebruikt als schatter van m het grootste getal T_n dat zij na n trekkingen op een bal gezien heeft.

(a) Bereken de verdelingsfunctie van T_n en laat zien dat

$$\mathbb{P}(T_n \leq t) \to \begin{cases} 0 & \text{als } t < m \\ 1 & \text{als } t \geq m \end{cases}$$

voor $n \to \infty$.

(b) Laat zien dat T_n een consistente schatter van m is.

Uitwerking:

(a) Zij X_i het nummer op de als i-de getrokken bal. De X_i zijn onafhankelijk en identiek verdeeld en er geldt voor alle $1 \leq i \leq n$

$$\mathbb{P}(X_i = j) = \frac{1}{m}, \qquad j = 1, \ldots, m.$$

Zij $t \in \mathbb{R}$ en $\lfloor t \rfloor \in \mathbb{Z}$ het grootste gehele getal dat kleiner of gelijk aan t is, d.w.z.,

$$\lfloor t \rfloor = \sup\{r \in \mathbb{Z} \ : \ r \leq t\}.$$

Uit de onafhankelijkheid van de X_i volgt dat (het onderstaande staat ook als stelling in de college aantekeningen)

$$\begin{aligned}\mathbb{P}(T_n \leq t) &= \mathbb{P}(\max\{X_1, \ldots, X_n\} \leq t)\\ &= \mathbb{P}\Big(\bigcap_{i=1}^{n}\{X_i \leq t\}\Big) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq t) = (\mathbb{P}(X_1 \leq t))^n.\end{aligned}$$

Merk nu op dat $\mathbb{P}(X_1 \leq t) = 1$ als $t \geq m$, $\mathbb{P}(X_1 \leq t) = 0$ voor $t < 1$ en

$$\mathbb{P}(X_1 \leq t) = \frac{1}{m}\sum_{i=1}^{m} 1_{\{i \leq t\}} = \frac{\lfloor t \rfloor}{m} < 1$$

voor $1 \leq t < m$. We concluderen dat

$$F_{T_n}(t) = \begin{cases} 0 & \text{als } t < 1 \\ \left(\frac{\lfloor t \rfloor}{m}\right)^n & \text{als } 1 \leq t < m \\ 1 & \text{als } t \geq m. \end{cases}$$

Aangezien

$$\lim_{n\to\infty} x^n = \begin{cases} 0 & \text{als } 0 \le x < 1 \\ 1 & \text{als } x = 1 \end{cases}$$

volgt de tweede uitspraak direct.

(b) Voor elke $\varepsilon > 0$ geldt, aangezien $T_n \le m$,

$$\mathbb{P}(|T_n - m| \le \varepsilon) = \mathbb{P}(m - \varepsilon \le T_n \le m + \varepsilon) = \mathbb{P}(T_n \ge m - \varepsilon) = 1 - \mathbb{P}(T_n < m - \varepsilon).$$

Uit onderdeel (a) volgt nu direct

$$\lim_{n\to\infty} \mathbb{P}(|T_n - m| \le \varepsilon) = 1.$$

Aangezien $\varepsilon > 0$ willekeurig was, is T_n een consistente schatter van m.

Vraag 5 [6 punten]

President Loekie Sjenkie van Blauwe-Rusjes-Land heeft in een toespraak verkondigd dat de inwoners van zijn land immuun zijn voor het coronavirus en dat ze zonder zorgen in een stadion naar een voetbalwedstrijd kunnen gaan kijken. Dr. Semenov vertrouwt het niet en besluit de data van de intensive care in zijn ziekenhuis te analyseren. Hij weet dat de patiënten normaal gesproken arriveren volgens een Poisson proces met intensiteit $\lambda = 1$ per uur en vermoedt dat deze intensiteit zich door de virusuitbraak verdubbeld heeft. Hij toetst daarom

$$H_0 : \lambda = 2 \qquad \text{v.s.} \qquad H_1 : \lambda = 1$$

Hij bekijkt het aantal patiënten X dat in de afgelopen 3 uur is aangekomen en gebruikt als toetscriterium

$$\text{verwerp } H_0 \text{ als } X \le n,$$

waarbij $n \in \mathbb{N} \cup \{0\} = \{0, 1, 2, \ldots\}$.

(a) Beschrijf de type I fout en de type II fout van de toets in woorden. Is de toets op een juiste manier opgezet?

(b) Voor Dr. Semenov is de maximaal acceptabele kans op een type I fout gelijk aan 1%. Welke waarde voor n kan hij het beste kiezen?

(c) Maak een afschatting van de type II fout met behulp van Chebyshev's ongelijkheid als $n = 3$.

Uitwerking:

(a) Bij een type I fout concludeert Dr. Semenov foutief dat een virusuitbraak plaats heeft gevonden (terwijl er geen sprake is van een virusuitbraak) en bij een type II fout concludeert Dr. Semenov foutief dat geen virusuitbraak plaats heeft gevonden. Aangenomen dat we de volksgezondheid de hoogste prioriteit geven, is de type I fout de ergere fout en dus is de toets correct opgezet.

(b) Onder H_0 nemen we aan dat $\lambda = 2$. Uit een stelling uit de college aantekeningen weten we dat X onder deze aanname Poisson verdeeld is met parameter $2 \cdot 3 = 6$. De kans op een type I fout is daarom

$$\mathbb{P}_{H_0}(X \le n) = \sum_{k=0}^{n} \mathbb{P}_{H_0}(X = k) = \sum_{k=0}^{n} \frac{e^{-6} 6^k}{k!}.$$

Aangezien

$$\mathbb{P}_{H_0}(X \le 0) = e^{-6} \approx 0.00248 < 0.01$$

en

$$\mathbb{P}_{H_0}(X \le 1) = e^{-6}(1+6) \approx 0.01735 > 0.01$$

kan Dr. Semenov het beste $n = 0$ kiezen.

(c) De kans op een type II fout is

$$\mathbb{P}_{H_1}(X > n) = \mathbb{P}_{H_1}(X > 3) = \mathbb{P}_{H_1}(X \ge 4).$$

Onder H_1 is X Poisson verdeeld met parameter $1 \cdot 3 = 3$. Uit een stelling uit de college aantekeningen weten we dat $\mathbb{E}_{H_1}(X) = 3$ en $\mathrm{Var}_{H_1}(X) = 3$. Uit Chebyshev's ongelijkheid volgt direct

$$\mathbb{P}_{H_1}(X \ge 4) = \mathbb{P}_{H_1}(X - 3 \ge 1) \le \mathbb{P}_{H_1}(|X - 3| \ge 1) \le \frac{\mathrm{Var}_{H_1}(X)}{1^2} = 3.$$

In dit geval geeft Chebyshev's ongelijkheid een triviale afschatting.

Vraag 6 [6 punten]

Zij X een kansvariabele met verdelingsfunctie

$$F_X(x) = \begin{cases} 1 - \left(\frac{\kappa^*}{x}\right)^{\mu^*} & \text{als } x \ge \kappa^* \\ 0 & \text{elders,} \end{cases}$$

waarbij $\kappa^* > 0$, $\mu^* > 0$ onbekende parameters zijn. Bepaal een meest aannemelijke schatter $(\hat{\kappa}, \hat{\mu})$ voor (κ^*, μ^*) op basis van een steekproef $X_1, \ldots, X_n$ uit de verdeling van X.

Hint: bepaal eerst een meest aannemelijke schatter voor κ^* onder de aanname dat μ^* bekend is.

Uitwerking:

Merk als eerste op dat X een continue kansvariabele is met kansdichtheid f_X gegeven door

$$f_X(x) = F_X'(x) = \mu^*(\kappa^*)^{\mu^*} \frac{1}{x^{\mu^*+1}}$$

voor $x \ge \kappa^*$ en $f_X(x) = 0$ elders. Zij $x_1, \ldots, x_n$ een realisatie van de steekproef. De aannemelijkheid $L(\cdot, \cdot | x_1, \ldots, x_n) : (0, \infty) \times (0, \infty) \to [0, \infty)$ wordt gegeven door

$$\begin{aligned} L(\kappa, \mu | x_1, \ldots, x_n) &= \prod_{i=1}^n f_{X_i}(x_i \mid \kappa, \mu) \\ &= \prod_{i=1}^n \left[\mu \kappa^\mu \frac{1}{x_i^{\mu+1}} 1_{\{\kappa \le x_i\}}\right] = \mu^n \kappa^{\mu n} \left(\frac{1}{\prod_{i=1}^n x_i}\right)^{\mu+1} 1_{\{\kappa \le \min_{1 \le i \le n} x_i\}}. \end{aligned}$$

Merk op dat, voor iedere vaste $\mu > 0$, de aannemelijkheid $L(\kappa, \mu | x_1, \ldots, x_n)$ maximaal is als $\kappa = \min_{1 \le i \le n} x_i$. Aangezien κ niet van μ afhangt vinden we direct $\hat{\kappa}(x_1, \ldots, x_n) =$

$\min_{1\le i\le n} x_i$ als meest aannemelijke schatting van κ^*. Om $\hat{\mu}$ te bepalen bekijken we de log-aannemelijkheid voor $\kappa = \hat{\kappa}$, d.w.z.,

$$\ell(\hat{\kappa}, \mu | x_1, \ldots, x_n) = n\log(\mu) + \mu n \log(\hat{\kappa}) - (\mu+1)\sum_{i=1}^n x_i.$$

Merk op dat

$$\frac{\partial \ell}{\partial \mu}(\hat{\kappa}, \mu | x_1, \ldots, x_n) = \frac{n}{\mu} + n\log(\hat{\kappa}) - \sum_{i=1}^n x_i = 0$$

dan en slechts dan als

$$\mu = \frac{1}{\frac{1}{n}\sum_{i=1}^n x_i - \log(\hat{\kappa})}$$

en bovendien geldt

$$\frac{\partial^2 \ell}{\partial \mu^2}(\hat{\kappa}, \mu | x_1, \ldots, x_n) = -\frac{n}{\mu^2} < 0$$

voor alle $\mu > 0$. Hieruit volgt dat de log-aannemelijkheid (voor $\kappa = \hat{\kappa}$) in

$$\hat{\mu}(x_1, \ldots, x_n) = \frac{1}{\bar{x}_n - \log(\min_{1\le i\le n} x_i)}$$

een uniek maximum aanneemt en daarmee is dit punt de meest aannemelijke schatting van μ^*. Aangezien $x_1, \ldots, x_n$ een willekeurige realisatie van de steekproef was, vinden we

$$(\hat{\kappa}, \hat{\mu}) = \left(\min_{1\le i\le n} X_i, \frac{1}{\bar{X}_n - \log(\min_{1\le i\le n} X_i)}\right)$$

als meest aannemelijke schatter van (κ^*, μ^*).

Bonusvraag [2 punten]

Zij $S \subset \mathbb{R}^2$ gegeven door

$$S = \{(x,y) \ : \ |x| + |y| \le 1\}$$

en zij (X, Y) uniform verdeeld in S. Bepaal $\mathbb{P}(-X < Y)$.

Uitwerking:

Merk op dat S een ruit in $\mathbb{R}^2$ is en dat de lijn $y = -x$ deze ruit in twee even grote delen snijdt (maak een plaatje!). De gevraagde kans is daarom $\frac{1}{2}$.