Universiteit Utrecht
Mathematisch Instituut

Boedapestlaan 6
3584 CD Utrecht

Inleiding Kansrekening en Statistiek Uitwerkingen Hertentamen, 2015-16

(1) We werpen N zuivere dobbelstenen. Hierbij is N een stochastische variabele met $P(N = n) = 2^{-n}$ voor $n \geq 1$. Zij S het aantal geworpen ogen.

(a) Bepaal $P(S = 4 \,|\, N = 2)$. (0.5 punten)

(b) Bepaal $P(S = 4)$. (1 punt)

(c) Bepaal $P(N = 2 \,|\, S = 4)$. (0.5 punten)

Uitwerking (a) Zij X_i het aantal ogen dat optreedt bij de $i$de worp. Dan

$$\begin{aligned} P(S = 4 \,|\, N = 2) &= P(X_1 = 1, X_2 = 3) + P(X_1 = 3, X_2 = 1) + P(X_1 = 2, X_2 = 2) \\ &= 3/36 = 1/12 \end{aligned}$$

Uitwerking (b)

$$P(S = 4) = \sum_{n=1}^{4} P(S = 4 \,|\, N = n)P(N = n).$$

Nu,

$$P(S = 4 \,|\, N = 1)P(N = 1) = P(X_1 = 4)P(N = 1) = 1/12,$$

en

$$P(S = 4 \,|\, N = 2)P(N = 2) = \frac{1}{12}\frac{1}{4} = \frac{1}{48},$$

en

$$\begin{aligned} P(S = 4 \,|\, N = 3) &= P(X_1 = 1, X_2 = 1, X_3 = 2) + P(X_1 = 1, X_2 = 2, X_3 = 1) \\ &+ P(X_1 = 2, X_2 = 1, X_3 = 1) \\ &= \frac{3}{216}, \end{aligned}$$

dus

$$P(S = 4 \,|\, N = 3)P(N = 3) = \frac{3}{216}\frac{1}{8} = \frac{1}{576}.$$

$$P(S = 4 \,|\, N = 4)P(N = 4) = P(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 1)P(N = 4) = \frac{1}{20736}.$$

We krijgen

$$P(S = 4) = \frac{1}{12} + \frac{1}{48} + \frac{1}{576} + \frac{1}{20736} = \frac{2197}{20736}.$$

Uitwerking (c) Door Bayes regel

$$P(N = 2 \,|\, S = 4) = \frac{P(S = 4 \,|\, N = 2)P(N = 2)}{P(N = 4)} = \frac{\frac{1}{48}}{\frac{2197}{20736}} = \frac{20736}{105456} = \frac{432}{2197}.$$

(2) Zij $X_1, X_2, \ldots$ een i.i.d. (onafhankelijk gelijk verdeeld) rij Bernoulli stochasten met parameter $0 < p < 1$.

(a) Laat zien dat $\lim_{n\to\infty} P\Big(X_1 + \cdots + X_n < 3pn\Big) = 1$. (1punt)

(b) Laat zien dat $P\Big(\Big|X_1 + \cdots + X_n - np\Big| \geq np\Big) \leq \frac{1-p}{np}$. (0.5 punten)

(c) Laat zien dat $\lim_{n\to\infty} P\Big(X_1 + \cdots + X_n - np \leq 0\Big) = 1/2$. (0.5 punten)

Uitwerking (a): Merk eerst op dat $E(X_i) = p$. We gebruiken de wet van de grote aantallen met $\epsilon = 2p$,

$$\begin{aligned}\lim_{n\to\infty} P\Big(X_1 + \cdots + X_n \geq 3pn\Big) &= \lim_{n\to\infty} P\Big(\frac{X_1 + \cdots + X_n}{n} - p \geq 2p\Big)\\ &\leq \lim_{n\to\infty} P\Big(\Big|\frac{X_1 + \cdots + X_n}{n} - p\Big| \geq 2p\Big) = 0.\end{aligned}$$

Dus

$$\lim_{n\to\infty} P\Big(X_1 + \cdots + X_n < 3pn\Big) = 1 - \lim_{n\to\infty} P\Big(X_1 + \cdots + X_n \geq 3pn\Big) = 1.$$

Uitwerking (b): Merk op dat de stochast $Y_n = X_1 + \cdots + X_n$ binomiaal (n, p) verdeeld is met $E(Y_n) = np$ en $\mathrm{Var}(Y_n) = np(1-p)$. Door Chebyshev's ongelijkheid, geldt

$$P\Big(\Big|X_1 + \cdots + X_n - np\Big| \geq np\Big) = P\Big(\Big|Y_n - E(Y_n)\Big| \geq np\Big) \leq \frac{np(1-p)}{n^2p^2} = \frac{1-p}{np}.$$

Uitwerking (c): We gebruiken de centrale limiet stelling,

$$\lim_{n\to\infty} P\Big(X_1 + \cdots + X_n - np \leq 0\Big) = \lim_{n\to\infty} P\Big(\frac{X_1 + \cdots + X_n - np}{\sqrt{np(1-p)}} \leq 0\Big) = \Phi(0) = 1/2,$$

waarbij $\Phi(x)$ de standaard normale verdelingsfunctie is.

(3) Stel dat de simultane kansdichtheid van X en Y gegeven wordt door

$$f(x,y) = \begin{cases} \dfrac{2xe^{-y}}{y^2}, & \text{als } 0 < x < y,\ 0 < y < \infty; \\ 0, & \text{elders.} \end{cases}$$

(a) Laat zien dat Y exponentieel verdeeld is met parameter $\lambda = 1$. (0.5 punten)
(b) Zij $Z = 2Y^2$, bepaal de mediaan van Z. (0.75 punten)
(c) Bepaal de kansdichtheid van Z. (0.75 punten)

Uitwerking (a): Voor $0 < y < \infty$,

$$f_Y(y) = \int_0^y \frac{2xe^{-y}}{y^2}\,dx = \frac{x^2e^{-y}}{y^2}\Big|_0^y = e^{-y}.$$

Merk op dat $f_Y(y) = 0$ voor $y \leq 0$, dus y is exponentieel verdeeld met parameter 1.

Uitwerking (b): Omdat Z een continue stochast is, dan is de mediaan m uniek en voldoet aan

$$\frac{1}{2} = P(Z \leq m) = P(2Y^2 \leq m) = P(Y \leq \sqrt{\frac{m}{2}}) = 1 - e^{-\sqrt{\frac{m}{2}}}.$$

Dus $m = 2\Big(\ln 2\Big)^2 \approx 0.96$.

Uitwerking (c): De funtie $r(z) = 2z^2$ is continu en stijgend op $(0,\infty)$. Verder $r^{-1}(z) = \sqrt{\frac{z}{2}}$ en $\dfrac{dr^{-1}(z)}{dz} = \dfrac{\sqrt{2}}{4\sqrt{z}}$. Dus voor $z \in (0,\infty)$,

$$f_Z(z) = f_Y(r^{-1}(z))\frac{dr^{-1}(z)}{dz} = \frac{\sqrt{2}}{4\sqrt{z}}\,e^{-\sqrt{\frac{z}{2}}}$$

(4) Zij $X_1, X_2, \ldots, X_n$ een rij van onafhankelijke gelijk verdeelde continue stochasten met kansdichtheid,

$$f_{X_i}(x) = f(x) = \begin{cases} (p+1)\,x^p, & \text{als } 0 < x < 1, \\ 0, & \text{anders }, \end{cases}$$

met $p > -1$ onbekend.

(a) Bepaal de maximum likelihood schatting van p gebaseerd op de dataset

$$x_1, x_2, \ldots, x_n,$$

afkomstig van de stochasten $X_1 X_2, \ldots, X_n$. (1 punt)

(b) Zij $Y_i = \ln X_i$ voor $i = 1, 2, \ldots, n$, en $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Laat zien dat $\overline{Y}_n$ een zuivere schatter voor $q = -(1+p)^{-1}$ is. (0.5 punten)

Uitwerking(a): De likelihood functie wordt gegeven door

$$L(p) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) = (p+1)^n (x_1 \cdots x_n)^p,$$

en de loglikelihood wordt gegeven door

$$\ell(p) = \ln(L(p)) = n \ln(p+1) + p \sum_{i=1}^{n} \ln x_i.$$

Dus

$$\ell'(p) = \frac{n}{p+1} + \sum_{i=1}^{n} \ln x_i,$$

en

$$\ell'(p) = 0 \Leftrightarrow p = -1 - \frac{n}{\sum_{i=1}^{n} \ln x_i}.$$

Omdat $\ell''(p) = \frac{-n}{(p+1)^2} < 0$, zien we dus dat $-1 - \frac{n}{\sum_{i=1}^{n} \ln x_i}$ de maximum likelihood schatting is voor p gebaseerd of de data.

Uitwerking(b): We moeten bewijzen dat $E(\overline{Y}_n) = -(1+p)^{-1}$. Omdat de stochasten $X_1, \ldots, X_n$ gelijk verdeeld zijn, dan door lineariteit van de verwachtingswaarde en met behulp van partieel integratie zien we dat

$$E(\overline{Y}_n) = E(X_1) = \int_0^1 (p+1)x^p \ln x \, dx = -(1+p)^{-1}.$$

Dus $\overline{Y}_n$ is een zuivere schatter voor q.

(5) Zij $X_1, X_2, \ldots, X_n$ een i.i.d. (onafhankelijk identiek verdeeld) rij met X_i uniform verdeeld op $[0, \theta]$, met θ onbekend. Beschouw de steekproefgrootheid $M = \max(X_1, \ldots, X_n)$.

(a) Voor $\alpha \in (0,1)$, laat zien dat $L_\alpha = \left(\frac{\alpha}{2}\right)^{1/n}\theta$ en $R_\alpha = \left(1 - \frac{\alpha}{2}\right)^{1/n}\theta$ voldoen aan

$$P(M \leq L_\alpha) = P(M \geq R_\alpha) = \frac{\alpha}{2}.$$

(0.75 punten)

(b) Laat zien dat

$$P\Big(\Big(\frac{2}{2-\alpha}\Big)^{1/n} M < \theta < \Big(\frac{2}{\alpha}\Big)^{1/n} M\Big) = 1 - \alpha.$$

(0.75 punten)

(c) Zij $x_1, x_2, \ldots, x_n$ een dataset afkomstig van $X_1, X_2, \ldots, X_n$. Bepaal een 95% betrouwbaarheidsinterval voor $\ln(\theta)$. (0.5 punten)

(d) Hoe groot moet n zijn zodat het 95% betrouwbaarheidsinterval voor $\ln(\theta)$ in onderdeel (c) ten hoogste lengte 0.1 heeft. (0.5 punten)

Uitwerking (a): Merk op dat dat de kansdichtheid van X_i is gegeven door

$$f_\theta(x) = f_{X_i}(x) = \begin{cases} \dfrac{1}{\theta} & \text{als } x \in [0, \theta] \\ 0 & \text{elders} \end{cases}$$

en de verdelingsfunctie is gegeven door

$$F_\theta(x) = F_{X_i}(x) = \begin{cases} \dfrac{x}{\theta} & \text{als } x \in [0,\theta] \\ 0 & \text{als } x < 0 \\ 1 & \text{als } x > \theta. \end{cases}$$

Door onafhankelijkheids, hebben we voor $x \in [0,\theta]$,

$$P(M \le x) = \prod_{i=1}^{n} P(X_i \le x) = \frac{x^n}{\theta^n}.$$

Dus

$$P\Big(M \le \Big(\frac{\alpha}{2}\Big)^{1/n}\theta\Big) = \frac{\alpha}{2},$$

en

$$P\Big(M \ge \Big(1-\frac{\alpha}{2}\Big)^{1/n}\theta\Big) = 1 - P\Big(M \le \Big(1-\frac{\alpha}{2}\Big)^{1/n}\theta\Big) = \frac{\alpha}{2}.$$

Uitwerking (b): Uit (a) zien we dat

$$P\Big(\Big(\frac{\alpha}{2}\Big)^{1/n}\theta < M < \Big(1-\frac{\alpha}{2}\Big)^{1/n}\theta\Big) = 1-\alpha$$

Nu geldt

$$\Big(\frac{\alpha}{2}\Big)^{1/n}\theta < M < \Big(1-\frac{\alpha}{2}\Big)^{1/n}\theta \Leftrightarrow \Big(\frac{2}{2-\alpha}\Big)^{1/n}M < \theta < \Big(\frac{2}{\alpha}\Big)^{1/n}M,$$

dus

$$P\Big(\Big(\frac{2}{2-\alpha}\Big)^{1/n}M < \theta < \Big(\frac{2}{\alpha}\Big)^{1/n}M\Big) = 1-\alpha.$$

Uitwerking (c): Laat $\alpha = 0.05$. Uit onderdeel (b), hebben we

$$P\Big(\Big(\frac{2}{2-0.05}\Big)^{1/n}M < \theta < \Big(\frac{2}{0.05}\Big)^{1/n}M\Big) = 0.95.$$

Omdat $g(x) = \ln(x)$ is een stijgende functie, dan

$$P\Big(\frac{1}{n}\ln(1.0256) + \ln(M) < \ln(\theta) < \frac{1}{n}\ln(40) + \ln(M)\Big) = 0.95.$$

Laat $m = \max(x_1, \dots, x_n)$, dan is $\Big(\frac{1}{n}\ln(1.0256) + \ln(m), \frac{1}{n}\ln(40) + \ln(m)\Big)$ een 95% betrouwbaarheidsinterval voor θ.

Uitwerking (d): We willen

$$0.1 = \Big(\frac{1}{n}\ln(40) + \ln(m)\Big) - \Big(\frac{1}{n}\ln(1.0256) + \ln(m)\Big) = \frac{3.664}{n}.$$

Omdat n geheel is, dit leidt tot $n \ge 37$.