Tentamen Inleiding Kansrekening en Statistiek

3 juli 2014

- Elke opgave dient **op een apart blad** ingeleverd te worden.
- Zet **op elk blad** je naam en studentnummer.
- Het gebruik van rekenmachines, het boek, aantekeningen of andere hulpmiddelen is absoluut niet toegestaan. Het is i.h.b. ook absoluut niet toegestaan een telefoon o.i.d. op tafel te hebben.
- Geef telkens een beknopt bewijs van je antwoord of een berekening/toelichting, tenzij duidelijk in de opgave staat aangegeven dat dit niet hoeft.
- **NOTA BENE:** het gebruiken van aparte bladen voor elke opgave met daarop telkens je naam en studentnummer is 10 van de 100 punten waard.

Opgave 1. (a: 6pt, b: 6pt, c: 6pt, d: 6pt, e: 6pt).
Voor één of andere $n \in \mathbb{N}$, nemen de twee stochasten X, Y waarden aan in $1, \dots, n$ en hun gezamelijke kansdichtheid wordt gegeven door:

$$\mathbb{P}(X = x, Y = y) = \begin{cases} \frac{c}{n(n+1)} & \text{als } 1 \leq y \leq x \leq n, \\ 0 & \text{anders} \end{cases}$$

Hier is $c > 0$ een nader te bepalen constante.

(a) Laat zien dat $c = 2$.

(b) Laat zien dat voor elke $1 \leq k \leq n$ geldt dat $\mathbb{P}(X = k) = \frac{2k}{n(n+1)}$ en $\mathbb{P}(Y = k) = \frac{2(n-k+1)}{n(n+1)}$.

(c) Bewijs of weerleg: X en Y zijn onafhankelijk.

(d) Geef de cumulatieve verdelingsfunctie F_X van X.

(e) Bereken $\mathbb{E}X$.
(*Hint:* je mag zonder bewijs de bekende formule $\displaystyle\sum_{k=1}^{n} k^2 = \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n$ voor de som van de kwadraten van de eerste n getallen gebruiken.)

Z.O.Z.

Opgave 2. (a: 7pt, b: 7pt, c: 8pt, d: 8pt).
Stel $X_1, \dots, X_n$ zijn i.i.d. met een Bernouilli verdeling met parameter $0 < p < 1$.

(a) Bewijs of weerleg: $T = \overline{X}$ is een zuivere schatter van p.

(b) Bewijs of weerleg: $T = (\frac{1}{2}X_1 + \frac{1}{2}X_2)^2$ is een zuivere schatter van p^2.

(c) Laat zien dat het interval

$$\left(\frac{(2\overline{X} + \frac{z_{\alpha/2}^2}{n}) - z_{\alpha/2}\sqrt{\frac{4\overline{X}(1-\overline{X})}{n} + \frac{z_{\alpha/2}^2}{n^2}}}{2(1 + \frac{z_{\alpha/2}^2}{n})}, \frac{(2\overline{X} + \frac{z_{\alpha/2}^2}{n}) + z_{\alpha/2}\sqrt{\frac{4\overline{X}(1-\overline{X})}{n} + \frac{z_{\alpha/2}^2}{n^2}}}{2(1 + \frac{z_{\alpha/2}^2}{n})} \right)$$

een betrouwbaarheidsinterval is voor p met significantieniveau (bij benadering) α, gebaseerd op de normale benadering/centrale limietstelling.
(*Hint:* begin door op te merken dat $\mathbb{P}\left(-z_{\alpha/2} < \sqrt{n} \cdot \frac{\overline{X}-p}{\sqrt{p(1-p)}} < z_{\alpha/2}\right) \approx 1 - \alpha$.)

In een ver land is opium volledig legaal. De lokale medici weten sinds jaar en dag dat opiumverslaafden i.h.a. jong overlijden. De producenten van de opium stellen echter in een verklaring in de media:

> "Dat opiumverslaafden gemiddeld eerder overlijden betekent niet noodzakelijk dat opium de oorzaak van dit vroege overlijden is. Er zou bijvoorbeeld een bepaald gen kunnen zijn dat zowel zorgt dat je een grotere kans hebt om vroeg te overlijden als ook ervoor zorgt dat je een grotere kans hebt om aan opium verslaafd te raken. Het zou best kunnen dat opium in feite juist goed voor de gezondheid is."

(De opiumproducenten hebben mogelijk wat opgestoken van de beroemde statisticus/geneticus R.A. Fischer.) In de hoop de hypothese van de opiumproducten te ontkrachten, bedenken de medici (die mogelijk iets hebben opgestoken van een bekend Fins experiment over rokers) het volgende experiment. Ze zullen n ééneiige tweelingen (dus $2n$ personen in totaal) opsporen waarvan de ene verslaafd is aan opium en de andere niet. Vervolgens zullen ze wachten tot van elk paar tenminste één van beiden is overleden. Het ligt voor de hand om dit experiment te modelleren met n i.i.d. Bernouilli toevalsvariabelen $X_1, \dots, X_n$ met onbekende parameter p. Hier is dan $X_i = 1$ als de verslaafde uit de i-de tweeling het eerst overlijdt en $X_i = 0$ anders. We gaan ervan uit dat de kans dat ze precies tegelijk overlijden verwaarloosbaar is. Als de verklaring van de opiumproducenten waar is moet gelden dat $p \leq \frac{1}{2}$.
De medici willen de nulhypothese $H_0 : \{p \leq \frac{1}{2}\}$ toetsen tegen de alternatieve hypothese $H_1 : \{p > \frac{1}{2}\}$, op basis van de normale benadering (centrale limiet stelling) met (bij benadering) significantie niveau α.

(d) Stel een éénzijdige hypothesetoets op om H_0 te toetsen tegen H_1 met (bij benadering) significantieniveau α, gebaseerd op de normale benadering/centrale limiet stelling. Gebruik de toetsingsgrootheid $T := \overline{X}$.
Voor welke waarden van T wordt H_0 verworpen?
(Een antwoord van de vorm "verwerp H_0 als $\overline{X} > c$" volstaat.)

Nota bene: in het boek van Dekking et al. zou de nulhypothese $H_0 : \{p_0 = \frac{1}{2}\}$ getoetst worden tegen $H_1 : \{p > \frac{1}{2}\}$.

Z.O.Z.

Opgave 3. (a: 6pt, b: 6pt, c: 6pt, d: 6pt, e: 6pt)
Twee muizen, een mannetjesmuis en een vrouwtjesmuis, bevinden zich in een oneindig lange gang als in het plaatje hieronder.

De vrouwtjesmuis bevindt zich vier kamers naar rechts van de mannetjesmuis. Elke seconde gaat het mannetje met kans a een kamer naar rechts en met kans $1-a$ een kamer naar links, terwijl het vrouwtje met kans b een kamer naar rechts gaat en met kans $1-b$ een kamer naar links. We zijn geïnteresseerd in de kans dat de beide muizen zich ooit in dezelfde kamer zullen bevinden.

Laat hiertoe eerst E_{2N} de eventualiteit zijn dat de afstand tussen de muizen nul bereikt voordat hij $2N$ bereikt (met $N > 2$.) Laat d_i de kans zijn op E_{2N} als de beginafstand gelijk is aan i.

(a) Laat zien dat de kans op E_{2N} gelijk is aan de waarde van d_4 in het stelsel:

$$\begin{aligned}
&d_0 = 1, \quad d_{2N} = 0,\\
&d_1 = d_3 = \cdots = d_{2N-1} = 0,\\
&d_i = (1-a)b \cdot d_{i+2} + ((1-a)(1-b) + ab) \cdot d_i + a(1-b) \cdot d_{i-2} \qquad \text{voor } i = 2, 4, \ldots, 2(N-1)
\end{aligned}$$

(Een korte toelichting volstaat.)

(b) Zet nu $p_i = d_{2N-2i}$ voor $i = 0, \ldots, N$. Laat zien dat de p_i voldoen aan het bekende stelsel:

$$\begin{aligned}
&p_0 = 0, \quad p_N = 1,\\
&p_i = (1-p) \cdot p_{i-1} + p \cdot p_{i+1} \quad (\text{voor } i = 1, \ldots, N-1),
\end{aligned}$$

waar $p := \frac{a(1-b)}{(1-a)b + a(1-b)}$.

(c) Stel nu eerst dat $a = b$. Laat zien dat in dit geval

$$d_i = \frac{N - i/2}{N} \qquad \text{voor } i = 0, 2, 4, \ldots, 2N,$$

en $d_i = 0$ voor oneven i.
(*Hint*: Helemaal afleiden van de gambler's ruin is niet nodig mits je de gebruikte resultaten uit de handout duidelijk vermeldt.)

(d) Stel nu dat $a \neq b$. Laat zien dat in dat geval

$$d_i = \frac{1 - \left(\frac{(1-a)b}{a(1-b)}\right)^{N-i/2}}{1 - \left(\frac{(1-a)b}{a(1-b)}\right)^{N}} \qquad \text{voor } i = 0, 2, 4, \ldots, 2N,$$

en $d_i = 0$ voor oneven i.

(e) Leid hieruit af dat

$$\mathbb{P}(\text{de muizen komen ooit in dezelfde kamer}) = \begin{cases} 1 & \text{als } a \geq b, \\ \left(\frac{a(1-b)}{(1-a)b}\right)^2 & \text{als } a < b. \end{cases}$$

(Je mag zonder bewijs gebruiken dat $\mathbb{P}(\text{de muizen komen ooit in dezelfde kamer}) = \mathbb{P}\left(\bigcup_{N>2} E_{2N}\right)$.)