Tentamen Kanstat
WI1275TA
13 april 2011, 18:30 – 21 uur

Bij dit examen is het gebruik van een (evt. grafische) rekenmachine toegestaan.
Een formuleblad wordt bijgeleverd.
Normering: **1.** 9 pt, **2.** 9 pt, **3.** 8 pt, **4.** 10 pt, **5.** 11 pt, **6.** 13 pt.
Een antwoord moet voorzien zijn van een berekening, toelichting en/of motivatie.
Dit alles goed leesbaar en in goed Nederlands (of Engels).

---

1. Vijf vriendinnen (Ada, Bea, Cora, Dora en Eva) gaan lootjes trekken.

a. Bereken de kans dat Bea zichzelf trekt.

b. Bereken de kans dat Eva zichzelf trekt als gegeven is dat Ada zichzelf *niet* trekt.

c. Bereken de kans dat precies drie van de vijf vriendinnen zichzelf trekken.

2. Stel X heeft een exponentiële verdeling met parameter λ (met $\lambda > 0$).

a. Alleen in dit onderdeel: stel $\lambda = 0.25$; bereken dan $\mathrm{P}(X \geq 5)$.

b. Bereken $\mathrm{E}[X^2]$. (N.B. ingewikkeld integreerwerk niet nodig!)

c. Bereken het derde kwartiel $q_{0.75}$.

3. De stochast U heeft een uniforme verdeling op $(0, 1)$.

a. Bereken de (cumulatieve) verdelingsfunctie van $X = \lceil 3U \rceil$, waarbij $\lceil .. \rceil$ staat voor afronden naar het dichtstbijzijnde gehele getal.

b. Bereken de verdelingsfunctie G van $Y = -\ln U$.

4. X en Y hebben een gezamenlijke verdeling volgens de volgende tabel, waarin $p(a, b) = \mathrm{P}(X = a, Y = b)$:

$p(a,b)$		b = 0	1	2	3	$p_X(a)$
a	-1	0.05	0.15	0.05	0.10	0.35
	0	0.00	0.05	0.15	0.05	0.25
	2	0.05	0.20	0.05	0.10	0.40
$p_Y(b)$		0.10	0.40	0.25	0.25	1

a. Bereken $\mathrm{P}(X = 2 \,|\, Y = 2)$ en $\mathrm{P}(Y = 2 \,|\, X \neq 2)$.

b. Bereken $\mathrm{E}[XY]$.

c. Geef de kansmassafunctie van $S = X + Y$.

5. Uit het bevolkingsregister van Fantasia van de afgelopen vijftig jaar valt op te maken dat het aantal kinderen dat in maart geboren wordt gemiddeld genomen gelijk is aan 1600 en goed beschreven wordt met een Poissonvariabele. Dit nemen we als model aan. In maart 2011 zijn er 1655 kinderen geboren. Jumpballkenners verklaren dit hoge aantal door het veroverenen van de wereldbeker jumpball door Fantasia in juni 2010 (negen maanden eerder ;-).

a. Formuleer een (eenzijdige) toets waarmee nagegaan kan worden of hier sprake is van een (95 %) significante afwijking ten opzichte van het gemiddelde.

Voor de laatste twee onderdelen kun je gebruik maken van de volgende eigenschap:
als X een Poissonverdeling heeft met als parameter μ een 'groot' getal N, dan heeft $Y = \dfrac{X - N}{\sqrt{N}}$ bij benadering een standaard normale verdeling:
als $N \to \infty$, dan $\mathrm{P}(Y \leq a) \longrightarrow \mathrm{P}(Z \leq a)$, waarbij $Z \sim \mathcal{N}(0, 1)$.

b. Bereken (een benadering van) de p-waarde van de toets.

c. Bereken het kritieke gebied.

d. Ga na of de nulhypothese wel of niet wordt verworpen. (Vergeet niet een argument te geven.)

6. Gegevens uit het zwangerschapsonderzoek van Weinberg en Gladen.
Regel 1: perioden tot zwanger worden
Regel 2: aantal (rokende) vrouwen die zoveel perioden nodig hadden

perioden	1	2	3	4	5	6	7	8	9	10	11	12
frequentie f_i	198	107	55	38	18	22	7	9	5	3	6	6

In totaal deden er 474 vrouwen mee.
We nemen aan: elke vrouw heeft elke maand opnieuw dezelfde kans om zwanger te geraken. Daaruit volgt dat Y, het aantal perioden dat een vrouw nodig heeft om zwanger te geraken, een geometrische verdeling heeft met parameter p.

a. We schatten p via de relatieve frequentie:

$$\hat{p} = \frac{\text{aantal vrouwen dat in 1 maand zwanger was}}{\text{totaal aantal vrouwen}}$$

en we noemen de bijbehorende schatter T_1. Toon aan dat T_1 zuiver is voor p.

b. Hoe kun je met behulp van de gevonden waarde van $\hat{p}$ de kans $\mathrm{P}(Y > 2)$ schatten? (Bedenk dat de verdeling van Y gegeven is.)
Geef de waarde van de schatting.
Noem de bijbehorende schatter T_2.
Geef de functie g zodat $T_2 = g(T_1)$.
(N.B. Twee vragen beantwoorden!)

c. Wat is de schatting van $\mathrm{P}(Y > 2)$ op grond van de 'empirische' kans?
Noem de hier relevante schatter S_2.

d. Ga na van beide schatters voor $\mathrm{P}(Y > 2)$ na of ze zuiver zijn.

Uitwerkingen

1a Voer in de gebeurtenissen A: Ada trekt zichzelf, B: Bea trekt zichzelf, enz. Uiteraard is dan $\mathrm{P}(B) = 1/5$.

1b $\mathrm{P}(E \,|\, A^c) = \dfrac{\mathrm{P}(E \cap A^c)}{\mathrm{P}(A^c)} = \dfrac{\mathrm{P}(E)\,\mathrm{P}(A^c \,|\, E)}{\mathrm{P}(A^c)} = \dfrac{1/5 \cdot 3/4}{4/5} = \dfrac{3}{16}$, want als gegeven is dat E zich voordoet, dan volgt dat Ada met gelijke kansen A,B, C of D trekt.
Alternatief: Voer in de gebeurtenis F: Ada trekt Eva. Dan geldt:
$\mathrm{P}(E \,|\, A^c) = \mathrm{P}(E \cap F \,|\, A^c) + \mathrm{P}(E \cap F^c \,|\, A^c)$. De eerste van deze twee kansen is natuurlijk nul. De tweede is gelijk aan $\mathrm{P}(F^c \,|\, A^c) \cdot \mathrm{P}(E \,|\, F^c \cap A^c) = \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{16}$.

1c Voor elke drietal, zeg A,B en E is de kans dat zij zichzelf trekken en de andere twee niet zichzelf trekken gelijk aan $\dfrac{1}{5}\cdot\dfrac{1}{4}\cdot\dfrac{1}{3}\cdot\dfrac{1}{2} = \dfrac{1}{120}$. Er zijn $\binom{5}{3} = 10$ drietallen, dus de kans dat precies een drietal zichzelf trekt is gelijk aan $\frac{10}{120} = \frac{1}{12}$.

2a $\mathrm{P}(X \geq 5) = 1 - F(5) = 1 - (1 - e^{-0.25 \cdot 5}) = 0.2865$.

2b Via het formuleblad: $\mathrm{E}[X^2] = \mathrm{Var}(X) + (\mathrm{E}[X])^2 = \dfrac{1}{\lambda^2} + \left(\dfrac{1}{\lambda}\right)^2 = \dfrac{2}{\lambda^2}$.

2c Op te lossen: $F(q) = 0.75$, waarbij $F(x) = 1 - e^{-\lambda x}$. Dit is niet zo moeilijk:

$$1 - e^{-\lambda q} = 0.75 \Leftrightarrow e^{-\lambda q} = 0.25 \Leftrightarrow -\lambda q = \ln(0.25) = -\ln 4 \Leftrightarrow q = q_{0.75} = \frac{\ln 4}{\lambda}$$

3a X is discreet, en neemt de waarden 0,1,2,3 aan met kansen $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{3}$ resp. $\frac{1}{6}$; bijvoorbeeld; $\mathrm{P}(X) = 3 = \mathrm{P}\big(3U \geq 2.5 = \frac{5}{2}\big) = \mathrm{P}\big(U \geq \frac{5}{6}\big) = \frac{1}{6}$.
De verdelingsfunctie F van X wordt dan een 'trapfunctie' met sprongen in 0,1,2 en 3:
$F(x) = 0$ als $x < 0$, $F(x) = \frac{1}{6}$ als $0 \leq x < 1$, $F(x) = \frac{1}{2}$ als $1 \leq x < 2$, $F(x) = \frac{5}{6}$ als $2 \leq x < 3$, en $F(x) = 1$ als $x \geq 3$.

3b Ten eerste bepaal ik het waardenbereik van Y: aangezien $0 < U < 1$ volgt $-\infty < \ln U < 0$, dus $0 < -\ln U < \infty$.
Daaruit volgt: $G(y) = 0$ als $y \leq 0$.
Voor $y > 0$ volgt verder: $G(y) = \mathrm{P}(Y \leq y) = \mathrm{P}(-\ln U \leq y) = \mathrm{P}(\ln U \geq -y) = \mathrm{P}(U \geq e^{-y}) = \mathrm{P}(e^{-y} \leq U \leq 1) = 1 - e^{-y}$, want U is uniform op $(0, 1)$.

4a

$$\mathrm{P}(X = 2 \,|\, Y = 2) = \frac{\mathrm{P}(X = 2, Y = 2)}{\mathrm{P}(Y = 2)} = \frac{0.05}{0.25} = 0.2$$

$$\mathrm{P}(Y = 2 \,|\, X \neq 2) = \frac{\mathrm{P}(Y = 2, X \neq 2)}{\mathrm{P}(X \neq 2)} = \frac{0.05 + 0.15}{0.35 + 0.25} = 0.333$$

4b Via de formule $\mathrm{E}[XY] = \sum_{j,k} j \cdot k \cdot \mathrm{P}(X = j, Y = k) = \ldots = 0.65$. (er zijn maar zes termen $\neq 0$)

4c Simpelweg een kwestie van alle mogelijkheden langsgaan. Wellicht is het handig om eerst een tabel te maken met de waarden van $X + Y$:

$p(a,b)$		b			
		0	1	2	3
a	-1	0.05	0.15	0.05	0.10
	0	0.00	0.05	0.15	0.05
	2	0.05	0.20	0.05	0.10

$a+b$		b			
		0	1	2	3
a	-1	-1	0	1	2
	0	0	1	2	3
	2	2	3	4	5

Waaruit de onderstaande kanstabel snel volgt:

a	-1	0	1	2	3	4	5
$\mathrm{P}(S = a)$	0.05	0.15	0.10	0.30	0.25	0.05	0.10

5a Toetsingsgrootheid X: aantal kinderen dat geboren wordt in maart;
aanname: $X \sim Pois(\mu)$, μ onbekend.
Nulhypothese (H_0): $\mu = 1600$, alternatieve hypothese: $\mu > 1600$.
H_0 bij waargenomen waarde x *verwerpen* als $\mathrm{P}(X \geq x \,|\, H_0) < 0.05$.

5b p-waarde: $\mathrm{P}(X \geq x \,|\, H_0) = \mathrm{P}(X \geq 1655 \,|\, \mu = 1600) =$
$$\mathrm{P}\left(\frac{X - 1600}{\sqrt{1600}} \geq \frac{1655 - 1600}{\sqrt{1600}}\right) \approx \mathrm{P}\left(Z \geq \frac{1655 - 1600}{\sqrt{1600}} = 1.375\right) \approx 0.085.$$
Het wordt (nog) niet gevraagd, maar met een p-waarde groter dan 0.05 zal de nulhypothese *niet* verworpen worden.

5c We moeten (de kleinste) c berekenen waarvoor $\mathrm{P}(X \geq c \,|\, \mu = 1600) \leq 0.05$. net als in het vorige onderdeel:
$$\mathrm{P}(X \geq c \,|\, \mu = 1600) \approx \mathrm{P}\left(Z \geq \frac{c - 1600}{\sqrt{1600}}\right) = \mathrm{P}\left(Z \geq \frac{c - 1600}{40}\right)$$
Dit wordt < 0.05 zodra $\dfrac{c - 1600}{40} \geq z_{0.05} = 1.645$, omgerekend: $c - 1600 \geq 1.645 \cdot 40 = 65.8$. Aangezien c geheel is volgt dat $c \geq 1666$, en het kritieke gebied wordt $\{1666, 1667, 1668, \ldots\}$.

5d Het antwoord is al gegeven in onderdeel **b**. Een ander argument om H_0 niet te verwerpen: de waargenomen waarde 1655 ligt *niet* in het kritieke gebied.

6a $\mathrm{P}(Y=1)=p$.

$$T_1=\frac{\text{aantal } Y_i \text{ gelijk aan } 1}{474};\quad \text{schatting } \hat{p}=\frac{198}{474}\approx 0.42.$$

In feite $T_1=\dfrac{R_1+R_2+\ldots+R_n}{n}$ waarbij $R_i=\begin{cases}1, & \text{als } Y_i=1\\ 0, & \text{als } Y_i>1\end{cases}$
Dus $T_1=\frac{1}{n}X$, waarbij $X\sim\mathrm{Bin}(n,p)$, met $p=\mathrm{P}(Y_i=1)$,
en dan uiteraard $\mathrm{E}[T_1]=\frac{1}{n}\mathrm{E}[X]=\mathrm{P}(Y_i=1)$.

6b Y heeft een Geom(p)-verdeling, dus $\mathrm{P}(Y>2)=(1-p)^2$ — eventueel via $\mathrm{P}(Y>2)=1-\mathrm{P}(Y\le 2)$ — , en dat kan worden geschat met $(1-\hat{p})^2\approx 0.34$.
Dus we kunnen schrijven $T_2=g(T_1)=(1-T_1)^2$.

6c

$$S_2=\frac{\text{aantal } Y_i>2}{474};\quad \text{schatting } p^*=\frac{474-198-107}{474}\approx 0.36.$$

6d Met hetzelfde argument als in **a**: S_2 is zuiver voor $\mathrm{P}(Y>2)$.
$T_2=g(T_1)=(1-T_1)^2$; $g'(x)=2>0$, dus g is convex. De ongelijkheid van Jensen geeft dan dat $\mathrm{E}[T_2]=\mathrm{E}[g(T_1)]>g(\mathrm{E}[T_1])=g(p)=(1-p)^2$, dus $\mathrm{E}[T_2]>(1-p)^2=\mathrm{P}(Y>2)$, waarmee is aangetoond dat T_2 een positieve bias heeft t.o.v. $\mathrm{P}(Y>2)$.