Toets Kansrekening voor INF en BIT
Module 4 (Data en Informatie - 201300180) Dinsdag 16 juni 2015, 8.45-11.00 uur

Deze toets bestaat uit 6 opgaven, het formuleblad en tabellen van de binomiale en de standaardnormale verdeling. Een gewone, niet programmeerbare rekenmachine is toegestaan (geen GR).

1. Een firma doet een bod om een groot project binnen te halen. Het management van de firma schat in dat de kans om het project te krijgen 60% is. Na indiening van het bod kan de beoordelende instantie die het project toewijst om extra informatie vragen. Uit het verleden is bekend dat voor 75% van de gehonoreerde aanvragen om extra informatie is gevraagd en dat bij de niet gehonoreerde aanvragen in 40% van de gevallen om extra informatie is verzocht.
Er wordt om extra informatie gevraagd. Bereken met behulp van dit gegeven de kans dat de aanvraag door de firma voor het project gehonoreerd wordt. Definieer daartoe eerst een aantal relevante gebeurtenissen en geef de gegeven kansen weer in termen van (voorwaardelijke) kansen op die gebeurtenissen.

2. Persoon H beweert helderziend te zijn. Om dit te testen, worden hem 10 dozen getoond met in elke doos een hermetisch afgesloten flesje. De flesjes zijn willekeurig gevuld met olie of met water (steeds met gelijke kans). Persoon H moet per doos zeggen of er olie of water in het flesje zit.
X het aantal goede antwoorden.
Als we aannemen dat dat H niet helderziend is en per doos willekeurig antwoordt, bepaal dan $P(X \geq 8)$, $E(X)$ en $var(X)$.

3. De simultane kansfunctie $P((X = x \text{ en } Y = y)$ van de stochastische variabelen X en Y wordt gegeven in nevenstaande tabel.

x \ y	0	1	4
-1	0.04	0.10	0.15
0	0.16	0.05	0.10
1	0.20	0.10	0.10

a. Bepaal de marginale verdelingen van X en van Y.
b. Bereken $cov(X, Y)$.
c. Bereken $P(X^2 + Y = 1)$
d. Bereken $E(Y|X = 1)$

4. De (continue) stochastische variabele X bezit een kansdichtheid f die gegeven wordt door

$$f(x) = \begin{cases} cx^2\ , & \text{voor } 0 \leq x \leq 1 \\ 0\ , & \text{elders} \end{cases}$$

a. Toon aan dat $c = 3$.
b. Bereken $E(X)$ en $var(X)$.
c. Bepaal de kansdichtheid van $Y = X^2$.

5. De dienst computerfaciliteiten van een grote universiteit wil onderzoeken of er onder studenten voldoende belangstelling is voor het aanschaffen van laptops in het kader van het pc-privé-project voor studenten. Via een aselecte steekproef van 100 uit de 60000 studenten wil men de fractie p bepalen van studenten, die serieuze belangstelling tonen voor een concreet laptop-aanbod.
X is het aantal belangstellenden in de steekproef en het geschatte fractie is dan $\frac{X}{100}$.

a. Waarom kunnen we bij trekkingen zonder terugleggen (van studenten uit de populatie) toch de binomiale verdeling als model voor X gebruiken?

b. Als we aannemen dat $p = 0.20$, bereken dan de kans dat het gemeten percentage in de steekproef minstens 5% hiervan afwijkt: bereken of benader $2 \cdot P(X \geq 25|p = 0.20)$.

6. Aan een loket komt een tweetal typen klanten om zich te laten bedienen. De bijbehorende bedieningsduren X en Y kunnen we modelleren als o.o. exponentieel verdeelde stochastische variabelen met parameter $\lambda = 1$ respectievelijk $\lambda = 2$.

a. Bepaal $E(X + Y)$ en $var(X + Y)$.

b. Bepaal $\rho(X, X + Y)$.

c. Van beide typen komen op zekere dag 100 klanten aan het loket, waarvan we de bedieningsduren modelleren als een aselecte steekproef $X_1, \ldots, X_{100}$ van X (dus 100 o.o. en $Exp(\lambda = 1)$-verdeelde bedieningsduren) en een aselecte steekproef $Y_1, \ldots, Y_{100}$ van Y.
Welke kansverdeling, met parameter(s), heeft de totale bedieningsduur van de 200 klanten bij benadering? Geef aan welke eigenschappen je daarbij gebruikt.

Normering: cijfer = $1 + \frac{aantal\ punten}{36} \times 9$

1	2	3				4			5		6			Tot
		a	b	c	d	a	b	c	a	b	a	b	c	
4	4	2	3	2	2	2	2	3	1	4	2	2	3	36

Formuleblad Kansrekening voor INF en BIT t.b.v. toetsen in module 4

Verdeling	$\boldsymbol{E(X)}$	$\boldsymbol{var(X)}$
Geometrisch	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Hypergeometrisch	$n \cdot \frac{R}{N}$	$n \cdot \frac{R}{N} \cdot \frac{N-R}{N} \cdot \frac{N-n}{N-1}$
Poisson $P(X = x) = \frac{e^{-\mu}\mu^x}{x!}, x = 0, 1, 2, \ldots$	μ	μ
Uniform op (a, b)	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponentieel	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Erlang $f_X(x) = \frac{\lambda(\lambda x)^{n-1}e^{-\lambda x}}{(n-1)!}, x \geq 0$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$

$$var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} var(X_i) + \sum_{i} \sum_{\neq j} cov(X_i, X_j)$$

Uitwerkingen:

Opgave 1

Noteer G = "aanvraag gehonoreerd" en E = "extra informatie gevraagd";
dan $P(G) = 0.60$, $P(E|G) = 0.75$ en $P(E|\overline{G}) = 0.40$.
Gevraagd: $P(G|E)$. Er geldt (met de regel van Bayes):

$$P(G|E) = \frac{P(G\cap E)}{P(E)} = \frac{P(E|G)P(G)}{P(E|G)P(G)+P(E|\overline{G})P(\overline{G})} = \frac{0.75\cdot 0.60}{0.75\cdot 0.60+0.40\cdot 0.40} = \frac{0.45}{0.61} = 0.74.$$

Opgave 2

$X \sim B\left(10, \frac{1}{2}\right)$, dus met de tabel: $P(X \geq 8) = 1 - P(X \leq 7) = 1 - 0.945$ = 5.5%

of: $P(X \geq 8) = P(X = 8) + P(X = 9) + P(X = 10) = \left[\binom{10}{8} + \binom{10}{9} + \binom{10}{10}\right] \times \left(\frac{1}{2}\right)^{10} = \frac{56}{1024} \approx 5.47\%$

$E(X) = np = 10 \cdot \frac{1}{2} = 5$ en
$var(X) = np(1-p) = 2.5$

Opgave 3

y / x	0	1	4	$P(X = x)$
-1	0.04	0.10	0.15	0.29
0	0.16	0.05	0.10	0.31
1	0.20	0.10	0.10	0.40
$P(Y = y)$	0.40	0.25	0.35	1

a. De marginale verdelingen van X en van Y zijn toegevoegd aan de tabel:

b. $cov(X,Y) = E(XY) - EX \cdot EY$, waarin
$E(X) = \sum_x xP(X = x)$
$= -1 \cdot 0.29 + 1 \cdot 0.40 = 0.11$
en $E(Y) = 1 \cdot 0.25 + 4 \cdot 0.35 = 1.65$
$E(XY) = \sum\sum x \cdot y \cdot P(X = x \text{ en } Y = y)$
$= -1 \cdot 1 \cdot 0.1 + (-1) \cdot 4 \cdot 0.15 + 1 \cdot 1 \cdot 0.1 + 1 \cdot 4 \cdot 0.1 = -0.2$
Dus $cov(X,Y) = -0.2 - 0.11 \cdot 1.65 = -0.3815$

c. $P(X^2 + Y = 1) = P(X = -1 \text{ en } Y = 0) + P(X = 0 \text{ en } Y = 1) + P(X = 1 \text{ en } Y = 0)$
$= 0.04 + 0.05 + 0.20 = 0.29$

d. $P(Y = 0|X = 1) = \frac{P(X=1 \text{ en } Y=0)}{P(Y=0)} = \frac{0.20}{0.40} = \frac{1}{2}$, Evenzo $P(Y = 1|X = 1) = P(Y = 4|X = 1) = \frac{1}{4}$, dus
$E(Y|X = 1) = \sum_y y \cdot P(Y = y|X = 1) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = 1.25$

Opgave 4

a. $\int_{-\infty}^{\infty} f(x)dx = 1 \quad \Leftrightarrow \quad \int_0^1 cx^2 dx = \left[\frac{1}{3}cx^3\right]_{x=0}^{x=1} = \frac{1}{3}c = 1 \quad \Leftrightarrow \quad c = 3$

b. $E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 x \cdot 3x^2 dx = \left[\frac{3}{4}x^4\right]_{x=0}^{x=1} = \frac{3}{4}$

$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 x^2 \cdot 3x^2 dx = \left[\frac{3}{3}x^5\right]_{x=0}^{x=1} = \frac{3}{5}$

Dus $var(X) = E(X^2) - (EX)^2 = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \frac{3}{80}$ $(= 0.0375)$

c. $F_Y(y) = P(X^2 \leq y) \overset{y>0}{=} P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$

$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{1}{2\sqrt{y}}f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}}f_X(-\sqrt{y})$, nog steeds met $y > 0$.

Dus omdat $f_X(x) = 3x^2$ voor $0 < x \leq 1$ geldt voor $0 < \sqrt{y} \leq 1$:

$f_Y(y) = \frac{1}{2\sqrt{y}} \cdot 3(\sqrt{y})^2 + 0 = \frac{3}{2}\sqrt{y}$, als $0 < y \leq 1$

(En $f_Y(y) = 0$ anders)

Opgave 5

a. Voor relatief kleine steekproeven uit grote populaties is de hypergeometrische verdeling bij benadering gelijk aan de binomiale verdeling. (Vuistregel $60000 = N \geq 5n^2 = 5 \cdot 100^2 = 50000$)

b. Voor $p = 0.2$ is X volgens de CLS bij benadering $N\big(np, np(1-p)\big) = N(20{,}16)$-verdeeld, dus:

$$2 \cdot P(X \geq 25 | p = 0.20) \overset{c.c.}{=} 2 \cdot P(X \geq 24.5) \overset{CLS}{\approx} 2 \cdot P\left(Z \geq \frac{24.5-20}{\sqrt{16}}\right) = 2\big(1 - \Phi(1.125)\big)$$
$$= 2(1 - 0.8697) \approx 16.1\%$$

(*het gemiddelde van* $\Phi(1.12)$ *en* $\Phi(1.13)$ *uit de* $N(0{,}1)$*-tabel is gebruikt*)

Opgave 6

a. $E(X+Y) = E(X) + E(Y) = \frac{1}{1} + \frac{1}{2} = 1.5$ en wegens o.o.-heid van X en Y:

$var(X+Y) = var(X) + var(Y) = \frac{1}{1^2} + \frac{1}{2^2} = 1.25$

b. $\rho(X, X+Y) = \frac{cov(X,X+Y)}{\sigma_X \cdot \sigma_{X+Y}} = \frac{cov(X,X) + cov(X,Y)}{\sqrt{var(X)} \cdot \sqrt{var(X+Y)}} = \frac{var(X) + 0}{\sqrt{1} \cdot \sqrt{1.25}} = \frac{1}{\sqrt{1.25}} \approx 0.894$

c. $X_1 + \cdots + X_{100}$ is volgens de CLS ($n > 25$) bij benadering $N\left(n \cdot \frac{1}{\lambda}, n \cdot \frac{1}{\lambda^2}\right) = N(100{,}100)$-verdeeld.

Evenzo $Y_1 + \cdots + Y_{100} \sim N(50, 25)$.

Omdat alle variabelen o.o. zijn, is de som (van de twee sommen) ook weer normaal verdeeld:

De totale som van bedieningsduren $(X_1 + \cdots + X_{100}) + (Y_1 + \cdots + Y_{100})$ is dus bij benadering $N(100 + 50{,}100 + 25)$-verdeeld.