Toets Kansrekening voor INF en BIT
Module 4 (Data en Informatie - 201300190) Vrijdag 13 juni 2014, 13.45-15.45 uur

Deze toets bestaat uit 5 opgaven. Het formuleblad en de tabellen van de Poisson- en de standaardnormale verdeling zijn toegevoegd. Een gewone, niet programmeerbare rekenmachine is toegestaan (geen GR).

1. Geef van elk van de volgende beweringen aan of deze (in zijn algemeenheid) ***waar*** of ***niet waar*** is en geef een **korte** argumentatie voor je antwoord.

a. Als van de werklozen 60% man is en 30% van allochtone afkomst, dan is het percentage allochtone mannen onder de werklozen dus 18%.
b. Twee elkaar uitsluitende gebeurtenissen A en B zijn onderling onafhankelijk.
c. Indien van tweeverdieners bekend is dat het inkomen van een getrouwde man, in duizenden Euro's, (bij benadering) N(25, 64)- verdeeld en voor een getrouwde vrouw N(15, 36)-verdeeld is, dan is het gezamenlijk inkomen van een willekeurig tweeverdienersechtpaar N(40, 100)-verdeeld.

2. De simultane kansfunctie $P(X = x \text{ en } Y = y)$ wordt gegeven door de volgende tabel:

y x \	0	1	2
0	0.15	0.05	0.20
1	0.05	0.20	0.05
2	0.20	0.05	0.05

Bepaal

a. $E(X)$ en $var(X)$.
b. De correlatiecoëfficiënt $\rho(X, Y)$ en geef aan wat de gevonden waarde betekent voor de samenhang van X en Y.
c. De kansverdeling van $Z = X + Y$.
d. $E(X|Y = 2)$.

3. De klantenservice van een energiebedrijf is gedurende kantooruren bereikbaar en wordt dan gemiddeld 2 keer per minuut door een klant gebeld. We nemen aan dat X = "het aantal telefoontjes in een willekeurige minuut" een Poisson-verdeling heeft.
Als we n opeenvolgende minuten beschouwen en X_i = "het aantal telefoontjes in minuut i", dan nemen we aan dat $X_1, \ldots, X_n$ onderling onafhankelijk zijn en allen bovengenoemde Poisson-verdeling hebben

a. Bereken de kans dat in één minuut minstens 3 telefoontjes binnenkomen.
b. Bereken de kans dat er in 4 minuten minstens 7 telefoontjes binnenkomen.
c. Benader m.b.v. de normale verdeling de kans dat er binnen één uur (60 minuten) minstens 105 telefoontjes binnenkomen.

4. In opdracht van een discussieprogramma doet een opiniepeiler onderzoek naar de standpunten van Nederlanders over politieke aangelegenheden. De opiniepeiler gebruikt een aselecte steekproef van n = 400 Nederlanders om onder meer te onderzoeken of de kiezers van mening zijn dat de PVV terecht buiten het bestuur van de gemeenten is gehouden (Volgens de PVV zelf is er sprake van "weren").
We noemen p de fractie van alle Nederlanders die deze mening is toegedaan en X is het aantal van deze personen in de steekproef.

a. Druk $E(X)$ en $var(X)$ uit in p.
b. Welke **benaderende** verdeling kunnen we het beste gebruiken voor het berekenen van kansen m.b.t. X, als n = 400 en p = 0.01. Geef het type verdeling en de parameter(s).
c. Bereken of benader $P(X \geq 50)$, voor het geval dat $\boldsymbol{p} = \mathbf{0.10}$ (en weer n = 400).

5. Een verzekeringsbedrijf heeft een *call center*, dat werkt met een automatische voorsortering van klanten: via een keuzemenu, waarbij vragen moeten worden beantwoord en gegevens moeten worden ingetoetst, wordt de klant naar een ter zake kundige medewerker geleid.
De keuzemenutijd X, de benodigde tijd in minuten tot het eerste contact met de medewerker, wordt gemodelleerd als een exponentieel verdeelde variabele met $\lambda_1 = ½$.
De bedieningsduur Y (in minuten) door de medewerker wordt verondersteld exponentieel verdeeld te zijn met parameter $\lambda_2 = 1$.
We veronderstellen dat de keuzemenutijd X en de bedieningsduur Y o.o. zijn.
$V = X + Y$ is de verblijfstijd van een klant in het systeem.

- **a.** Bereken $P(X > 2\mu_X | X > \mu_X)$
- **b.** Bepaal de kansdichtheid van V, met behulp van de convolutie-integraal.
- **c.** Bereken ook $E(V)$ en $var(V)$.
- **d.** Twee klanten komen tegelijk het telefoonsysteem binnen: hun keuzemenutijden noemen we X_1 en X_2. Bepaal de **verdelingsfunctie** van het maximum van deze 2 keuzemenutijden (het maximum is het tijdstip waarop beiden klaar zijn).

Normering: cijfer = $1 + \frac{aantal\ punten}{31} \times 9$

1			2				3			4			5				Tot
a	b	c	a	b	c	d	a	b	c	a	b	c	a	b	c	d	
1	1	1	3	3	1	2	1	1	3	1	2	3	2	2	1	3	31

Formuleblad Kansrekening voor INF en BIT t.b.v. toetsen

Verdeling	$E(X)$	$var(X)$
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 $\quad P(X = x) = \frac{e^{-\mu}\mu^x}{x!}, x = 0, 1, 2, ...$	μ	μ
Exponentieel	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Uniform op (a, b)	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$

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

Uitwerkingen:

Opgave 1

a. Niet waar: de kans op beide gebeurtenissen is alleen als een product 0.60×0.30 te berekenen als de gebeurtenissen “man” en “van allochtone afkomst” onderling onafhankelijk zijn. Dat zal in zijn algemeenheid niet het geval zijn.

b. Niet waar: elkaar uitsluiten impliceert $P(AB) = 0$. Dan kunnen A en B alleen o.o. zijn als $P(AB) = P(A)P(B) = 0$, ofwel, als $P(A) > 0$ en $P(B) > 0$, dan zijn A en B afhankelijk.

c. Niet waar: de inkomens X en Y van een man en zijn vrouw zijn waarschijnlijk afhankelijk ($cov(X,Y) \neq 0$), dus geldt **niet** dat $var(X + Y) = var(X) + var(Y)$
(*de afhankelijkheid wordt in de praktijk veroorzaakt doordat bijv. beide partners eenzelfde opleiding(sniveau) en dus salarisniveau hebben, de percentages deeltijdwerk samenhangen e.d. Terecht wordt door sommige studenten opgemerkt dat de verdeling bij een homopaar zeker niet geldt…*)

Opgave 2

a. De kansverdeling van X: $P(X = 0) = 0.4$ en $P(X = 1) = P(X = 2) = 0.3$
Of in een tabel:

x	0	1	2	Totaal
$P(X = x)$	0.4	0.3	0.3	1

$E(X) = \sum_x xP(X = x) = 0 + 1\times0.3 + 2\times0.3 = 0.9$
$E(X^2) = \sum x^2P(X = x) = 0 + 1\times0.3 + 4\times0.3 = 1.5$
$var(X) = E(X^2) - (EX)^2 = 1.5 - 0.81 = 0.69$

b. Y heeft dezelfde verdeling als X, dus $E(X) = E(Y) = 0.9$ en $var(X) = 0.69$
$E(XY) = \sum\sum xyP(X = x \text{ en } Y = y) = 1\times1\times0.2 + 1\times2\times0.05 + 2\times1\times0.05 + 2\times2\times0.05 = 0.6$
$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X\sigma_Y} = \frac{E(XY)-EX\cdot EY}{\sigma_X\sigma_Y} = \frac{(0.6 - 0.9\times0.9)}{\sqrt{0.69}\sqrt{0.69}} = -\frac{0.21}{0.69} \approx -0.304$
X en Y zijn dus licht (tot matig) negatief gecorreleerd.

c. $P(X + Y = 0) = 0.15$, $P(X + Y = 1) = 0.05 + 0.05 = 0.10$, $P(X + Y = 2) = 0.60$, $P(X + Y = 3) = 0.10$ en $P(X + Y = 4) = 0.05$.

(*Of maak een tabel voor $Z = X + Y$*:

z	0	1	2	3	4	Totaal
$P(X = x)$	0.15	0.1	0.6	0.1	0.05	1

d. $P(X = 0|Y = 2) = \frac{P(X=0 \text{ en } Y=2)}{P(Y=2)} = \frac{0.20}{0.30} = \frac{2}{3}$ en $P(X = 1|Y = 2) = P(X = 2|Y = 2) = \frac{0.05}{0.30} = \frac{1}{6}$
Dus $E(X|Y = 2) = \sum_x xP(X = x|Y = 2) = 0 \times \frac{2}{3} + 1 \times \frac{1}{6} + 2 \times \frac{1}{6} = \frac{1}{2}$

Opgave 3

a. $P(X \geq 3) = 1 - P(X \leq 2) =$ 0.323 (zie tabel)

b. Y = “het aantal telefoontjes in 4 minuten” is Poisson verdeeld met verwachting $\mu = 4\times2 = 8$
$P(Y \geq 7) = 1 - P(X \leq 6) = 1 - 0.313 = 68.7\%$ (zie Poisson tabel met $\mu = 8$).
$S = \sum X_i$ is bij benadering $N(\mu, \sigma^2)$ met $\mu = 60\times 2 = 120$ en $\sigma^2 = 60 \times 2 = 120$
$P(S \geq 105) = P(S \geq 105) = P(Z \geq (105 - 120)/\sqrt{120}) \approx P(Z \geq -1.37) = P(Z \leq 1.37) =$ 91.47%
(eventueel met continuïteitscorrectie (hoeft niet standaard bij Poisson, wel bij binomiaal), dus $P(S \geq 105) \overset{\text{c.c.}}{=} P(S \geq 104.5) = \ldots = 92.14\%$)

Opgave 4

a. $E(X) = np = 400p$ en $var(X) = np(1-p) = 400p(1-p)$

b. Voor $n = 400$ en $p = 0.01$ kan de Poisson verdeling met $\mu = 400\times0.01 = 4$ (< 5) als benadering gebruikt worden. (*voor een normale benadering is np te klein!*)

c. Zie onderdeel a. voor μ en σ^2, nu met p = 0.10: de N(40, 36)-verdeling kan als benaderende verdeling voor X gebruikt worden (controle: n > 25, $np = 40 > 5$ en $n(1-p) = 360 > 5$). Dus:

$$P(X \geq 50) \overset{\text{c.c.}}{=} P(X \geq 49.5) = P\left(\frac{X-40}{6} \geq \frac{49.5-40}{6}\right) \approx 1 - \Phi(1.58) = 5.71\%$$

Opgave 5

a. $E(X) = \frac{1}{\lambda_1} = 2$

$P(X > 2\mu_X | X > \mu_X) = P(X > 4 | X > 2) = P(X > 2)$ (wegens geheugenloosheid)

$$= \int_2^\infty \frac{1}{2} e^{-x/2} dx = -e^{-x/2} \Big|_2^\infty = e^{-1} \approx 36.8\%$$

b. $f_V(v) = \int_0^v \frac{1}{2} e^{-x/2} e^{-(v-x)} dx = e^{-v} e^{x/2} \Big|_0^v = e^{-v/2} - e^{-v}$, voor $v \geq 0$ (en $f_V(v)$ = 0 als v < 0)

c. $E(V) = E(X_1) + E(X_2)$ = 2 + 1 = 3

$var(V) = var(X_1) + var(X_2)$ = 4 + 1 = 5 wegens o.o.-heid van X_1 en X_2 (en $var(X) = \frac{1}{\lambda^2}$)

d. $Z = max(X_1, X_2)$

$F(z) = P(max(X_1, X_2) \leq z) = P(X_1 \leq z$ en $X_2 \leq z) = P(X_1 \leq z) \times P(X_2 \leq z) = \left(1 - e^{-\frac{z}{2}}\right)^2$

voor $z \geq 0$