Uitwerking Tentamen Kansrekening en Statistiek (153008) voor INF en TEL
Donderdag 6 april 2006, 13.30-16.30 uur

Opgave 1

a. X = “het aantal commissieleden uit partij A” is hypergeometrisch verdeeld, omdat het hier gaat om “kiezen van leden zonder terugleggen”.
Gevr: $P(X \geq 1) = 1- P(X=0)$

$$=1- \binom{6}{0}\binom{14}{5} \Big/ \binom{20}{5} = 1 - \frac{14}{20} \times \frac{13}{19} \times \frac{12}{18} \times \frac{11}{17} \times \frac{10}{16} = 1 - \frac{3003}{23256} \approx 87.1\%$$

b. X= “het IQ van willekeurige INF-student” die N(118,10)-verdeeld is.
Extra veronderstelling is dus dat de normale verdeling geldt.

$$P(X > 140) = P\left(\frac{X-118}{10} > \frac{140-118}{10}\right) = 1 - P(Z \leq 2.20) = 1-0.9861 = 1.39\%$$

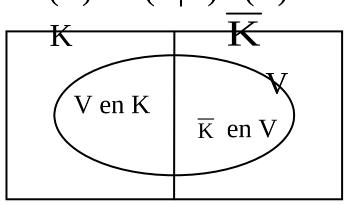
c. X= “het aantal lekke banden in de tunnel bij 10000 passerende auto’s” is B(10000,0.00002)- verdeeld, omdat het krijgen van een lekke band van de 10000 auto’s (onafhankelijke) Bernoulli-pogingen zijn.
X is bij benadering Poisson verdeeld met $\mu = np = 0.2$, dus met de Poisson-tabel:
$P(X > 2) = 1- P(X \leq 2) = 1- 0.999 = 0.1\%$

d. X = “het aantal krasloten t/m de eerste die prijs geeft voor Jan” en
Y = “het aantal krasloten t/m de eerste die prijs geeft voor Klaas”.
X en Y zijn beide geometrisch verdeeld met p = 0.05 (Bernoulli pogingen t/m eerste succes) en X en Y zijn o.o.

Gevraagd: $P(X > 10 \text{ en } Y > 10) = P(X > 10) \times P(Y > 10) = 0.95^{10} \times 0.95^{10} = 35.8\%$

Opgave 2

a. Definieer K = “borstkanker” en V = “vermoeden van borstkanker na röntgenfoto”
Dan is gegeven $P(K) = 0.05$, $P(V|K) = 0.90$ en $P(V|\overline{K}) = 0.02$
$P(V) = P(V|K)\, P(K) + P(V|\overline{K})\, P(\overline{K}) = 0.90 \times 0.05 + 0.02 \times 0.95 = 6.4\%$



b. $P(K|V) = P(K \text{ en } V)/P(V) = P(V|K)\, P(K)/ P(V) = 0.045 / 0.064 \approx 70.3\%$

c. X is B(100, 0.064)- verdeeld, dus
$E(X) = np = 6.4$
$\text{var}(X) = np(1-p) = 5.98$

d. Y is, gegeven X = 10, B(10, 0.7)-verdeeld, dus

$$P(Y = 7|\, X=10) = \binom{10}{7} 0.7^7 0.3^3 \approx 26.7\%$$

$E(Y \mid X=10) = np = 7$

e. $P(Y = 7) = \sum P(X= i \text{ en } Y = 7) = \sum P(Y = 7 \mid X= i) \times P(X= i)$

$$= \sum_{i=7}^{100} \binom{i}{7} 0.7^7 0.3^{i-7} \times \binom{100}{i} 0.064^i 0.936^{100-i}$$

Maar ook (zie f): $P(Y=7) = \binom{100}{7} 0.045^7 0.955^{93} \approx 8.3\%$

f. $P(V \text{ en } K) = P(K)P(V|K) = 0.045$ (zie a), maar ook (zie b):

P(V en K) = P(V)P(K|V) = 0.064×0.703 ≈ 0.045.
Dus Y , het aantal vrouwen onder de 100, waarbij borstkanker wordt vastgesteld is B(100,0.045)-verdeeld (onafhankelijk van elkaar wordt bij elk van de 100 vrouwen met kans 4.5% borstkanker geconstateerd: 100 Bernoulli-pogingen).

Opgave 3

a. E(X+Y) = E(X) +E(Y) = 5 + 1 =6
E(X) = $1/\lambda_1$ = 5 dus $\lambda_1 = 1/5$ en E(Y) = $1/\lambda_2$ = 5 dus $\lambda_2 = 1$.
Omdat X en Y o.o. zijn, geldt Var(X+Y) = var(X) + Var(Y) = $1/\lambda_1^2 + 1/\lambda_2^2$ = 25 +1 = 26.

b. $\rho(X, Z) = \frac{\text{cov}(X, X+Y)}{\sigma_X * \sigma_{X+Y}} = \frac{\text{cov}(X,X)+\text{cov}(X,Y)}{\sqrt{\text{var}(X)} * \sqrt{\text{var}(X+Y)}} = \frac{25+0}{5*\sqrt{26}} \approx 0.98$

c. $f_Z(z) = \int_{-\infty}^{\infty} f_{X_1}(x) f_{X_2}(z-x)dx = \int_{x=0}^{x=z} \frac{1}{5}e^{-x/5} e^{-(z-x)}dx = \int_{x=0}^{x=z} \frac{1}{5}e^{-z}e^{+\frac{4}{5}x}dx = \frac{1}{4}e^{-z}e^{+\frac{4}{5}x}\Big|_{x=0}^{x=z} = \frac{1}{4}(e^{-z/5} - e^{-z})$
(voor $z \geq 0$ en $f_Z(z) = 0$ voor $z < 0$)

d. Als we aannemen dat $Z_1, \ldots, Z_n$ o.o. zijn (alle met de in c bepaalde verdeling),
dan is volgens de CLS (n > 25!) wordt de kansverdeling van $\sum_{i=1}^{100} Z_i$ benaderend door de normale verdeling: N($n\mu$, $n\sigma^2$) = N(100×6, 100×26) = N(600, 2600)-verdeling.

Opgave 4

a. $f_X(x) = \int f(x,y)dy = \int_{y=0}^{2} xydy = \frac{1}{2}xy^2\Big|_{y=0}^{y=2} = 2x$ voor $0 \leq x \leq 1$ en

Evenzo: $f_Y(y) = \int f(x,y)dy = \int_{x=0}^{1} xydx = \frac{1}{2}x^2y\Big|_{x=0}^{x=1} = \frac{1}{2}y$ voor $0 \leq y \leq 2$

b. $E(X) = \int x f_X(x)dx = \int_{x=0}^{1} x \cdot 2xdx = \frac{2}{3}x^3\Big|_{x=0}^{x=1} = \frac{2}{3}$

$E(Y) = \int y f_Y(y)dy = \int_{x=0}^{1} y \cdot \frac{1}{2}ydy = \frac{1}{6}y^3\Big|_{y=0}^{y=2} = \frac{4}{3}$

$E(XY) = \iint xyf(x,y)dxdy = \int_{x=0}^{1}\int_{y=0}^{2} x^2y^2dy = \frac{1}{3}x^3\Big|_{x=0}^{x=1} \times \frac{1}{3}y^3\Big|_{y=0}^{y=2} = \frac{8}{9}$

(Of na onderdeel c: *E(XY)* =*E(X)*×*E(Y)* =8/9)

c. Zijn *X* en *Y* o.o.? Motiveer uw antwoord.
X en Y zijn o.o. omdat voor alle *x* en *y* geldt dat $f(x,y) = f_X(x) \times f_Y(y)$
Voor $0 \leq x \leq 1$ en $0 \leq y \leq 2$ geldt: $xy = 2x \times \frac{1}{2}y$

d. $P(X > Y) = \iint_{x>y} f(x,y)dxdy = \int_{x=0}^{1}\int_{y=0}^{x} xydydx = \int_{x=0}^{1} x \cdot \frac{1}{2}y^2\Big|_{y=0}^{y=x}dx = \int_{x=o}^{1} \frac{1}{2}x^3dx = \frac{1}{8}x^4\Big|_{x=0}^{x=1} = \frac{1}{8}$

Opgave 5

a. Met de rekenmachine vinden we: $\bar{x}$ = 54.1 en
$s^2 \approx 33.66$ ($s \approx 5.80$)

b. T is een schatter van θ dan is de verwachte kwadratische fout: $E(T-\theta)^2$.
$\overline{X}$ is een zuivere schatter van μ, dus de verwachte kwadratische fout is:
$E(\overline{X}-\mu)^2 = \text{var}(\overline{X}) = \sigma^2/n$

Voor n = 10 is de verwachte kwadratische fout dus kleiner dan voor n = 5: voor n = 10 is $\overline{X}$ dus een betere schatter van μ.

c. Kansmodel: levensduren X_1 t/m X_{10} en de te voorspellen levensduur X_{11} zijn onderling onafhankelijk en alle $N(\mu,\sigma^2)$-verdeeld, met onbekende μ en σ^2

Dus het voorspellingsinterval is

95%-VI(X_{11}) = $(\overline{x}-cs\sqrt{1+1/n},\ \overline{x}+cs\sqrt{1+1/n}) \approx (40.4, 67.8)$

Hierin zijn n =10, $\overline{x}$ = 54.1 en $s \approx 5.80$

en c uit de t_{10-1}–tabel, zodat $P(T_9 < c)$= 0.975 , dus c = 2.26

d. 95%-BI(σ)= $\left(\sqrt{\dfrac{(n-1)s^2}{c_2}}, \sqrt{\dfrac{(n-1)s^2}{c_1}}\right) \approx (4.0, 10.6)$

Hierin zijn n =10, s^2 = 33.66 en

c_1 = 2.70 en c_2 = 19.0, omdat $P(\chi^2_{10-1} \le c_1) = 0.025$ en $P(\chi^2_9 \le c_2) = 0.975$.

e. 1. Kansmodel: levensduren X_1 t/m X_{10} zijn o.o. en alle $N(\mu,\sigma^2)$-verdeeld

2. Toets $H_0 : \mu = 57$ (of $\mu \geq 57$) tegen $H_1 : \mu < 57$ (met $\alpha = 0.05$)

3. Toetsingsgrootheid T = $\dfrac{\overline{X}-57}{S/\sqrt{10}}$ is onder H_0 t_9-verdeeld.

4. Waargenomen waarde: t = $\dfrac{54.1-57}{5.80/\sqrt{10}} \approx -1.58$

5. Linkseenzijdig Kritiek Gebied : T ≤ c, zodat $P(T \le c|H_0)$= 0.05, dus c = -1.83

6. t = -1.58 ligt niet in het KG dus H_0 *niet* verwerpen:

7. Conclusie: met onbetrouwbaarheid van 5% is met deze meetgegevens statistisch niet aangetoond dat de claim van een levensduur van minstens 57000 km verworpen moet worden.

(Beslissing met (rechter)overschrijdingskans P(T ≤ -1.58) = 1- P(T ≤ 1.58) ligt tussen 5% en 10%, dus groter dan 5% = α: H_0 niet verwerpen.)