Uitwerking Tentamen Kansrekening en Statistiek voor INF (191530082), 7 april 2011

Opgave 1

a. (10 trekkingen zonder terugleggen uit 100 loten waarvan 5 van Gait, dus X is hypergeometrisch verdeeld)

$P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{\binom{95}{10}}{\binom{100}{10}} = 1 - \frac{95}{100} \times \frac{94}{99} \times \ldots \times \frac{86}{91} \approx 41.6\%$

b. (10 trekkingen met terugleggen, dus X is B(10, 0.05)-verdeeld)

$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.95^{10} \approx 40.1\%$

Opgave 2

a. Kansverdeling van X zie tabel

x \ y	0	1	2	$P(X=x)$
0	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{13}{40}$
1	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{14}{40}$
2	$\frac{1}{10}$	$\frac{1}{8}$	$\frac{1}{10}$	$\frac{13}{40}$
$P(Y=y)$	$\frac{13}{40}$	$\frac{14}{40}$	$\frac{13}{40}$	1

E(X) = 1 (symmetrie) en

var(X) = E(X^2) – (EX)^2

$= \left(0 \times \frac{13}{40} + 1 \times \frac{14}{40} + 4 \times \frac{13}{40}\right) - 1^2 = \frac{13}{20}$

b. De covariantie van X en Y:

cov(X,Y) = E(XY) - EX×EY

E(XY) = ∑∑ xy P(X=x en Y=y)

$= 1\times1\times\frac{1}{10} + 1\times2\times\frac{1}{8} + 2\times1\times\frac{1}{8} + 2\times2\times\frac{1}{10} = 1$

Cov(X,Y) = 1 - 1×1 = 0, dus ook ρ = 0

c. X en Y zijn afhankelijk, want (bijvoorbeeld):

$\frac{1}{10}$ = P(X = 0 en Y = 0) ≠ P(X = 0)P(Y = 0) = $(\frac{13}{40})^2$.

(Uit cov(X,Y) = 0 volgt niet (automatisch) dat X en Y o.o. zijn).

d. P(Y=0 | X=1) = P(Y=0 en X=1)/ P(X=1) = (1/8) / (14/40) = 5/14

Evenzo P(Y=1 | X=1) = 4/14 en P(Y=2 | X=1) = 5/14 . Dus E(Y | X = 1) = 1 (symmetrie)

Opgave 3

a. $E(X^k) = \int_{-\infty}^{\infty} x^k f(x)dx = \int_0^1 x^k\, dx = \left[\frac{1}{k+1}x^{k+1}\right]_{x=0}^{x=1} = \frac{1}{k+1}$

b. $var(X) = E(X^2) - (EX)^2 = \frac{1}{2+1} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$

c. $F_Y(y) = P(Y \leq y) = P(-2\ln(X) \leq y) = P\left(X \geq e^{-\frac{1}{2}y}\right) = 1 - P\left(X \leq e^{-\frac{1}{2}y}\right) = 1 - F_X\left(e^{-\frac{1}{2}y}\right)$

$f_Y(y) = \frac{d}{dy}F_Y(y) = +\frac{1}{2}e^{-\frac{1}{2}y} f_X\left(e^{-\frac{1}{2}y}\right)$

Omdat $f_X(x) = 1$ als $0 \leq x \leq 1$, is $f_X\left(e^{-\frac{1}{2}y}\right) = 1$ als $y \geq 0$, dus:

$f_Y(y) = \frac{1}{2}e^{-\frac{1}{2}y}$, voor $y \geq 0$, dus Y is exponentieel verdeeld met λ = ½ en $E(Y)$ = 1/λ = 2

d. $\mu = E(X_i) = \frac{1}{2}$ en $\sigma^2 = var(X_i) = \frac{1}{12}$, dus $S = \sum_{i=1}^{300} X_i$ is volgens de CLS bij benadering N(nμ, nσ^2)-verdeeld. n = 300, dus N(150, 25). De benaderde kans is (**geen** continuïteitscorrectie!):

$P(S \geq 158) = P\left(\frac{S-150}{5} \geq \frac{158-150}{5}\right) \approx P(Z \geq 1.6) = 1 - P(Z \leq 1.6) = 5.48\%$ (hierin is Z ~ N(0, 1))

Opgave 4

a. Als we onafhankelijkheid van de levensduren aannemen, is X + Y N(5+5, 2.5^2 +2.5^2) verdeeld.

Dus $P(X + Y > 12) = P\left(Z > \frac{12-10}{\sqrt{12.5}}\right) \approx 1 - P(Z \leq 0.57) = 1 - 0.7157 \approx 28.4\%$

b. X+Y is Erlang verdeeld met n = 2 en $\lambda = \frac{1}{5}$ (want $E(X)$ = 1/λ =5): kansdichtheid op formuleblad

Dus $P(X + Y > 12) = \int_{12}^{\infty} \frac{1}{25} x e^{-\frac{1}{5}x}\, dx = \frac{1}{5}x \times -e^{-\frac{1}{5}x}\Big|_{x=12}^{x\to\infty} + \int_{12}^{\infty} \frac{1}{5}e^{-\frac{1}{5}x}\, dx = \frac{12}{5}e^{-\frac{12}{5}} + \left[-e^{-\frac{1}{5}x}\right]_{x=12}^{x\to\infty}$

$= \frac{12}{5}e^{-\frac{12}{5}} + e^{-\frac{12}{5}} \approx 30.8\%$

Opgave 5

a. X is B(n, p)-verdeeld (als de leerlingen aselect met terugleggen uit de populatie worden getrokken: of bij benadering binomiaal, indien een relatief kleine steekproef zonder terugleggen aselect uit een grote populatie wordt genomen)

b. De schatter van p is $\hat{p} = \frac{X}{n}$. Deze is zuiver, want $E(\hat{p}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = \frac{1}{n}np = p$

c. De verwachte kwadratische fout is $E(\hat{p} - p)^2 = var(\hat{p}) = \frac{p(1-p)}{n}$

Hieruit volgt: naarmate *n* groter is, is de verwachte kwadratische fout kleiner en de schatter dus beter.

d. In de formule $\hat{p} \pm c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ voor het 95%-betrouwbaarheidsinterval is c = 1.96 zodat P(Z ≤ c) = 0.975
en $\hat{p} \approx 0.25$

Voor de halve intervalbreedte moet gelden $c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq 0.05$, dus $\frac{c}{0.05}\sqrt{\hat{p}(1-\hat{p})} \leq \sqrt{n}$ ofwel

$n \geq \left(\frac{1.96}{0.05}\right)^2 \times \frac{1}{4} \times \frac{3}{4} = 288.12$. Dus *n* = 289 (of hoger)

e. X is voor $p = ¼$ bij benadering N(192× ¼ , 192× ¼ × ¾) = N(48, 36)-verdeeld

$P(X \geq 60) = P(X \geq 59.5) = P\left(\frac{X-48}{\sqrt{36}} \geq \frac{59.5-48}{\sqrt{36}}\right) = 1 - P(Z \leq 1.92)$ = 1- 0.9726 = 2.74%

f. De in e. berekende kans is de overschrijdingskans voor de toets op H_0: $p \leq 0.25$ tegen H_1: $p > 0.25$

Opgave 6

a. Kansmodel: de responsietijden $X_1,\ldots, X_{16}$ zijn onafhankelijk en alle (bij benadering) $N(\mu,\sigma^2)$-verdeeld, met onbekende μ en σ^2: We gebruiken dus de formule

90%-BI(μ)= $\left(\bar{x} - c\frac{s}{\sqrt{n}}, \bar{x} + c\frac{s}{\sqrt{n}}\right)$, waarin $n = 16$, $\bar{x} = 15.10$, s = 5.06 en

c uit de t_{15}-tabel, zodat $P(T_{15} < c) = 0.95$, dus $c = 1.75$

90%-BI(μ) = $\left(\bar{x} - c\frac{s}{\sqrt{n}}, \bar{x} + c\frac{s}{\sqrt{n}}\right) \approx (12.9, 17.3)$.

b. Niet correct: een betrouwbaarheidsinterval betreft het <u>gemiddelde</u> van alle mogelijke responsietijden, niet de responsietijd van één opdracht. De gegeven interpretatie heeft betrekking op een voorspellingsinterval.

c. 1. Model: de responsietijden $X_1, \ldots, X_{16}$ zijn o.o. en normaal verdeeld met onbekende μ en onbekende σ (We voeren dus de t-toets op μ uit)

2. Toets H_0: $\mu \leq 13$ tegen H_1: $\mu > 13$ met $\alpha = 0.05$

3. Toetsingsgrootheid $T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{\bar{X}-13}{S/\sqrt{16}}$ is t_{15}-verdeeld als μ = 13 waar is

4. Het is een rechtseenzijdige toets, dus deze luidt: "**als $T \geq 1.75$, dan H_0 verwerpen**", want uit de t_{15}-tabel volgt $P(T_{15} \leq 1.75) = 0.95$

5. Waargenomen t $= \frac{15.10-13}{5.05/\sqrt{16}} = 1.66$

6. Statistische beslissing: t = 1.66 < 1.75 , dus H_0 ***niet*** verwerpen

7. Conclusie: bij een onbetrouwbaarheid van 5% is er onvoldoende statistisch bewijs voor de bewering dat de verwachte responsietijd groter dan 13 is.

*Alternatief voor 4. en 6. is het bepalen van de overschrijdingskans $P(T_{15} \geq 1.66)$: deze ligt tussen 5% en 10%, dus de **overschrijdingskans** > $\boldsymbol{\alpha} = 0.05 \rightarrow H_0$ **niet** verwerpen)*