Uitwerking tentamen Kansrekening en Statistiek voor INF/TEL d.d. 24-04-03

1. a. P("4 harten")= $\binom{13}{4} \Big/ \binom{52}{4} = \frac{13}{52} \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49}$

   Dus P("4 kaarten van 1 kleur")= 4 × P("4 harten") ≈ 0.0106

   b. E(2X-Y)=2E(X) –E(Y)= 20-4=16 en var(2X-Y)= 4var(X) +var(Y)= 36 + 4 =40
   Dus 2X-Y ~ N(16, 40)

   c. $\overline{X}$ is een zuivere schatter van μ. Het criterium voor het bepalen welke schatter beter is, is de verwachte kwadratische fout $E(\overline{X} - \mu)^2 = var(\overline{X}) = \sigma^2/n$. Voor n=100 is deze kleiner dan voor n=10 dus $\overline{X}$ is een betere schatter voor n=100.

2. a. Als de gebeurtenissen O= "fiets (steeds) Op slot" en G= "Fiets Gestolen" worden gedefinieerd, dan is gegeven: $P(\overline{O})= 0.1$, $P(G|\overline{O})= 0.15$ en P(G|O)= 0.05. Dus:
   $P(G)= P(GO) + P(G\overline{O})= P(G|O)P(O) + P(G|\overline{O})P(\overline{O})= 0.05\times(1-0.1) + 0.15\times0.1= 0.06$

   b. $P(\overline{O}|G) = P(G\overline{O})/P(G) = P(G|\overline{O})P(\overline{O})/P(G) = 0.15\times0.1/ 0.06 =25\%$

   c. Van de 10000 staan er 9000 op slot (=evt verwachte aantal op slot staande fietsen), die, onafhankelijk van elkaar, elk met kans 0.05 gestolen worden: X is B(9000, 0.05)
   (Of: kans op "op slot staan en gestolen" is p= P(GO)=0.045 dus X is B(10000, 0.045))

   d. Het verwachte aantal uitbetalingen E(X)=np=450. Dus de verwachte winst is de premieopbrengst – vaste kosten – verwachte uitbetaling =
   10000×40 – 150000 - E(X)×400 = 70000 Euro

3. a. $E(X)= \lambda^{-1} = 4$ dus $\lambda=1/4$

   $$P(X > EX) = \int_4^\infty f_X(x)dx = \int_4^\infty \frac{1}{4}e^{-x/4}dx = -e^{-x/4}\Big|_{x=4}^{x=\infty} = e^{-1} \approx 36.8\%$$

   b. $$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)dx = \int_{x=0}^{x=z} \frac{1}{4}e^{-x/4}e^{-(z-x)}dx = e^{-z}\int_{x=0}^{x=z}\frac{1}{4}e^{3x/4}dx$$

   $$= e^{-z}\times\frac{1}{4}\times\frac{4}{3}e^{3x/4}\Big|_{x=0}^{x=z} = e^{-z}\frac{1}{3}[e^{3z/4}-1]$$ (voor $z \ge 0$ en $f_Z(z)=0$ voor $z < 0$)

   c. $$\rho(X, X+Y)= \frac{cov(X, X+Y)}{\sigma_X\,\sigma_{X+Y}} = \frac{cov(X,X)+cov(X,Y)}{\sqrt{var(X)}\sqrt{var(X)+var(Y)}} = \frac{var(X)+0}{\sqrt{16\times(16+1)}} = \sqrt{\frac{16}{17}}$$

   Dus $\rho \approx 0.97$: een sterke positieve samenhang tussen wachttijd en verblijfstijd.

4. $95\% - BI(p) = (\hat{p} - c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + c\sqrt{\frac{\hat{p}(1-\hat{p})}{n}})$, waarin n =3500, $\hat{p} = \frac{x}{n} = \frac{85}{3500}$ en
c =1.96 uit de $N(0,1)$-tabel zodat $P(Z<c)$= 0.975.
Dus $95\% - BI(p) = (0.0192, 0.0294)$
Statistisch zou de kans op overlijden $\frac{59}{3500} \approx 0.0169$ moeten zijn: deze waarde ligt buiten het 95%-BI(p), dus er is een statistisch significant verschil.
(Of: met een betrouwbaarheidsinterval voor het verwachte aantal overleden vrouwen in een groep van 3500 van deze vrouwen: $95\% - BI(3500p) = (0.0192 \times 3500, 0.0294 \times 3500)$ = (67, 103): 59 ligt niet in dit interval dus daarmee is een substantiële afwijking aangetoond.)
b. Het aantal suïcidegevallen X is B(3500, $\frac{5}{3500}$)-verdeeld dus bij benadering Poisson-verdeeld met verwachtingswaarde μ = np = 5 Dus $P(X \geq 14) = 1 - P(X \leq 13) = 1 - 0{,}999 = 0.001$

5. 1. Kansmodel:de winstpercentages $X_1, ..., X_{25}$ zijn o.o. en alle (bij benadering) $N(\mu, \sigma^2)$-verdeeld, met onbekende μ en σ^2.
2. Toets $H_0 : \mu = 2.00$ tegen $H_1 : \mu < 2.00$ (met $\alpha = 0.05$).
3. Toetsingsgrootheid T = $\frac{\overline{X} - 2}{S / \sqrt{25}}$ is onder H_0 t_{24}-verdeeld.
4. Waargenomen waarde: t = $\frac{1.90 - 2}{0.18 / \sqrt{25}} \approx -2.78$
5. Linkseenzijdig Kritieke Gebied : $T \leq c$, Zodat $P(T \leq c | H_0) = 0.05$ dus c= -1.71
6. t = -2.78 ligt in het KG dus H_0 verwerpen.
7. Met onbetrouwbaarheid van 5% is statistisch aangetoond dat de winstgevendheid in deze bedrijfstak lager is dan dat de overheid veronderstelt.

5. en 6. met overschrijdingskans $P(T \leq -2.78) \approx 0.5\% < 5\% = \alpha$, dus H_0 verwerpen.