Lösningar till tentamen i Matematisk statistik D (MVE055) och IT (MSN620), 18 oktober 2005

1. A: Bil släpper ut för mycket miljöfarliga ämnen. M: Bil klassas som miljöfarlig. Med dessa beteckningar: $P(A) = 0.25$, $P(M|A) = 0.99$, $P(M|A^c) = 0.17$. Sökt:

$$P(A|M) = \frac{P(M|A) \cdot P(A)}{P(M|A) \cdot P(A) + P(M|A^c) \cdot P(A^c)} = \frac{0.99 \cdot 0.25}{0.99 \cdot 0.25 + 0.17 \cdot 0.75} = 0.66$$

2. a) p=P(lyckas på försök i)=0.8. X=antal försök han behöver för att lyckas; $X \sim Ge(p)$. Sökt:

$$P(X = 3) = 0.2^2 \cdot 0.8 = 0.032 \qquad E(X) = \frac{1}{p} = \frac{1}{0.8} = 1.25$$

b) P(Inget körkort)=P(Misslyckas på uppkörningen på alla 5 försök). A_i: Misslyckas på försök i. Sökt:

$$P(A_1 \cap \ldots \cap A_5) = P(A_1) \cdot \ldots \cdot P(A_5) =$$

$$= (1 - 0.8)(1 - (0.8 + 0.02))(1 - (0.8 + 0.04))(1 - (0.8 + 0.06))(1 - (0.8 + 0.08)) \simeq 9.7 \cdot 10^{-5}$$

3. a) Vi har följden $\{a_k\}_{k=0}^{\infty}$ med $a_k = 2^k$. Genererande funktion:

$$A(x) = \sum_{k=0}^{\infty} 2^k x^k = \frac{1}{1 - 2x}$$

b) Här har vi $\{b_k\}_{k=0}^{\infty}$ med $b_k = (-3)^{k+1}$. Genererande funktion:

$$B(x) = \sum_{k=0}^{\infty} (-3)^{k+1} x^k = -3 \sum_{k=0}^{\infty} (-3)^k x^k = -\frac{3}{1 + 3x}$$

c)

$$A(X) + B(X) = \sum_{k=0}^{\infty} 2^k x^k + \sum_{k=0}^{\infty} (-3)^{k+1} x^k = \sum_{k=0}^{\infty} (2^k + (-3)^{k+1}) x^k$$

Vilket motsvarar en talföljd $\{c_k\}_{k=0}^{\infty} = a_0 + b_0, a_1 + b_1, \ldots = -2, 11, -23, 89, \ldots$

4. a) $m_{ij} = E$(antal steg tills man för första gången kommer till j då man startar i i). Sökt: m_{14}.

$$m_{ij} = 1 + \sum_{k=1}^{4} p_{ik} \cdot m_{kj}$$

Vi får ekvationssystemet

$$\begin{aligned} m_{14} &= 1 + \frac{1}{4} m_{14} + \frac{3}{4} m_{24} \\ m_{24} &= 1 + \frac{1}{3} m_{14} + \frac{1}{3} m_{24} + \frac{1}{3} m_{34} \\ m_{34} &= 1 + m_{44} \\ m_{44} &= 0 \end{aligned}$$

Detta reduceras till

$$\begin{cases} \frac{3}{4} m_{14} - \frac{3}{4} m_{24} = 1 \\ -\frac{1}{3} m_{14} + \frac{2}{3} m_{24} = \frac{4}{3} \end{cases} \iff \begin{cases} m_{14} = \frac{20}{3} \\ m_{24} = \frac{16}{3} \end{cases}$$

Vi har $m_{14} = 20/3 \simeq 6.67$.

b) En Markovkedja är en sekvens av stokastiska variabler $X_1, X_2, X_3, \ldots$ med följande typ av beroende:

$$P(X_{i+1} = x | X_1 = x_1, X_2 = x_2, \ldots, X_i = x_i) = P(X_{i+1} = x | X_i = x_i)$$

Villkoret kallas markovegenskapen och innebär att sannolikheten att vi förflyttar oss till tillstånd *i+1* när vi står i tillstånd i inte beror på vilka tillstånd vi besökte tidigare i kedjan: “vad som ska hända i framtiden beror bara på nuet”.

5. a) Typ I-fel: Det fel som görs när vi felaktigt förkastar nollhypotesen H_0.
$P(\text{Typ I-fel}) = P(H_0 \text{ förkastas } | H_0 \text{ sann}) = \alpha$

Typ II-fel: Det fel som görs när vi felaktigt inte förkastar H_0.
$P(\text{Typ II-fel}) = P(H_0 \text{ förkastas inte } | H_0 \text{ falsk}) = \beta$

b) Framför allt garderar man sig mot typ I-fel, då α väljs som nivån på testet. För att få ett litet β, alltså hög styrka på testet, måste vi kontrollera stickprovets storlek. Om vi inte lyckas förkasta H_0 har vi inte visat något med vårt test: H_0 kan vara sann men behöver inte vara det. Lyckas vi förkasta H_0 drar vi slutsatsen att H_1 gäller.

6.

$$H_0 : p \geq 0.9$$
$$H_1 : p < 0.9$$

X=antal lampor som hade den önskade livslängden. $X \sim Bin(p)$. Vi tittar på $p = 0.9$. $X \overset{\text{appr}}{\sim} N(100p, \sqrt{100p(1-p)})$ eftersom $p > 0.5$ och $100(1-p) > 5$ (ett striktare krav är $np(1-p) \geq 10$ vilket inte är uppfyllt här; egentligen borde vi räkna exakt). Vi har teststorheten

$$u = \frac{\frac{x}{100} - 0.9}{\sqrt{0.9(1-0.9)/100}} = -1$$

där motsvarande stokastiska variabel

$$U = \frac{\frac{X}{100} - 0.9}{\sqrt{0.9(1-0.9)/100}} \overset{\text{appr}}{\sim} N(0,1) \quad \text{om } H_0 \text{ sann}$$

α väljs till 0.05 till exempel.

$$\text{P-värde} = P(U < -1) \approx \Phi(-1) = 1 - \Phi(1) = 0.1587 > \alpha$$

Vi kan inte förkasta H_0. Vi lyckades inte visa att lampfabrikanten har fel i sitt påstående.

7. Vi har observationer $x_1, \ldots, x_{10}$ av $X_1, \ldots, X_{10}$ där X_i= nikotinhalt i cigarett i från fabrikat A och $y_1, \ldots, y_8$ av $Y_1, \ldots, Y_8$ där Y_j= nikotinhalt i cigarett j från fabrikat B. $X_i \sim N(\mu_x, \sigma)$ och $Y_j \sim N(\mu_y, \sigma)$. Vi har $\bar{x} = 3.1$, $\bar{y} = 2.7$, $s_x = 0.5$ och $s_y = 0.7$. Sammanvägd variansskattning:

$$s^2 = \frac{9s_x^2 + 7s_y^2}{9+7} = 0.355$$

Hjälpvariabel:

$$U = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{s\sqrt{\frac{1}{10} + \frac{1}{8}}} \sim t(16)$$

$$P(-c < U < c) = 0.95 \Leftrightarrow P\left(\bar{X} - \bar{Y} - c \cdot s\sqrt{\frac{1}{10} + \frac{1}{8}} < \mu_x - \mu_y < \bar{X} - \bar{Y} + c \cdot s\sqrt{\frac{1}{10} + \frac{1}{8}}\right) = 0.95$$

där $c = 2.12$ fås ur t-tabell för $F(c) = 0.975$. Vi får intervallet

$$I_{\mu_x-\mu_y} = \left(3.1 - 2.7 \mp 2.12 \cdot \sqrt{0.355}\sqrt{\frac{1}{10} + \frac{1}{8}}\right) = (-0.2, 1.0)$$

Vi kan inte visa någon skillnad i nikotinhalt mellan fabrikaten.

8. Vi har en s.v. X där $E(X) = 5$ och $Var(X) = 2$ och $Y = 3X + 1$. Detta ger $E(Y) = 3 \cdot 5 + 1 = 16$ och $Var(Y) = 3^2 \cdot 2 = 18$.

$$\begin{aligned}
Kov(X,Y) &= E(XY) - E(X) \cdot E(Y) = \\
\Big/ E(XY) &= E(X(3X+1)) = E(3X^2 + X) = 3E(X^2) + E(X) = \\
&= 3(Var(X) + (E(X))^2) + E(X) = 3(2 + 5^2) + 5 = 86 \Big/ \\
&= 86 - 5 \cdot 16 = 6
\end{aligned}$$

Korrelationen ρ fås via

$$\rho = \frac{Kov(X,Y)}{D(X) \cdot D(Y)} = \frac{16}{\sqrt{2} \cdot \sqrt{18}} = 1$$

Vilket också kan inses direkt eftersom Y är en linjärkombination av X.

9. a) Vi tittar på sekvensen $HHHHH$ och har därmed indikatorvariabeln

$$I_i = \begin{cases} 1, & \text{om vi har sekvensen i pos } i, \ldots, i+4 \\ 0, & \text{annars} \end{cases}$$

Antal gånger vi får sekvensen under 100 myntkast beskrivs av

$$X = \sum_{i=1}^{96} I_i$$

Vi har också $E(I_i) = (1/2)^5$. Vi får $E(X) = 96 \cdot E(I_i) = 96 \cdot (1/2)^5 = 3$

b) Vi har sekvenserna $S = HHHHH$ och $S' = TTTTT$.
$E(X_{S'} + X_S) = E(X_{S'}) + E(X_S) = 6$

10. Vi har X och Y oberoende med $E(X) = \mu_x$ och $E(Y) = \mu_y$. Sökt: fördelningen för $Z = \min\{X, Y\}$.

$$\begin{aligned}
F_Z(z) &= P(Z \leq z) = P(\min\{X,Y\} \leq z) = 1 - P(\min\{X,Y\} \geq z) = \\
&= P(X \geq z, Y \geq z) = 1 - P(X \geq z) \cdot P(Y \geq z)
\end{aligned}$$

Vi har

$$P(X \geq z) = \int_z^{\infty} \frac{1}{\mu_x} \cdot e^{\frac{x}{\mu_x}}\, dx = e^{-\frac{z}{\mu_x}}$$

Detta ger

$$\begin{aligned}
F_Z(z) &= 1 - e^{-\frac{z}{\mu_x}} \cdot e^{-\frac{z}{\mu_y}} = 1 - e^{-z\left(\frac{1}{\mu_x} + \frac{1}{\mu_y}\right)} \\
f_Z(z) &= \frac{d}{dz} F_Z(z) = \underbrace{\left(\frac{1}{\mu_x} + \frac{1}{\mu_y}\right)}_{\lambda} \cdot e^{-z\left(\frac{1}{\mu_x} + \frac{1}{\mu_y}\right)}
\end{aligned}$$

Uttrycket för $f_Z(z)$ är täthetsfunktionen för en exponentialfördelad stokastisk variabel. Vi har alltså väntevärdet

$$E(Z) = \frac{1}{\lambda} = \left(\frac{1}{\mu_x} + \frac{1}{\mu_y}\right)^{-1}$$