Lösningar till tentamen i Matematisk statistik och diskret matematik IT (MVE050).
Den 21 december 2006.

1. a) $P(B|A) \overset{\text{def.}}{=} \frac{P(B\cap A)}{P(A)} \overset{\text{ober.}}{=} \frac{P(B)P(A)}{P(A)} = P(B) = 0.5$
   b) $P(B|A) \overset{\text{def.}}{=} \frac{P(B\cap A)}{P(A)} \overset{\text{disj.}}{=} \frac{P(\emptyset)}{P(A)} = 0$
   c) Bayes sats ger $P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{0.2\cdot 0.5}{0.25} = 0.4$

2. a) X: antal anrop under 1:a tiondels sekunden.
   I snitt sker $20/10 = 2$ anrop på en tiondels sekund, så $X \sim Poi(2)$. Alltså:
   $P(X \geq 5) = 1 - P(X \leq 4) \approx 1 - 0.9473 = 5.27\%$
   b) Y: antal sekunder till första anropet
   $Y \sim \text{Exp}(20)$ ger $\text{E}[Y] = 1/20 = 0.05$
   Svar: 0.05s

3. a) $\hat{p} = \frac{24}{100} = 0.24$
   Approximativt 95%-igt konfidensintervall: $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx 0.24 \pm 0.084$
   b) Nej, $\frac{1}{6}$ ligger i intervallet.

4. a) $m_X(t) = \text{E}[e^{tX}] = \sum_x e^{tx} f_X(x) = \sum_{x=0}^{\infty} e^{tx} 2^{-x-1} = \frac{1}{2}\sum_{x=0}^{\infty}\left(\frac{e^t}{2}\right)^x = \frac{1}{2}\frac{1}{1-\frac{e^t}{2}} =$ Obs:
   $= \frac{1}{2-e^t}\ ,\ t < \ln 2$
   summan är konvergent $\iff \frac{e^t}{2} < 1 \iff t < \ln 2$.
   b) $m'_X(t) = \frac{e^t}{(2-e^t)^2}$ ger $\text{E}[X] = m'_X(0) = 1$

5. a) $c = \frac{1}{36}$ ger $\sum_x \sum_y f_{XY}(x,y) = 1$ och $f_{XY}(x,y) \geq 0$
   b) $f_X(x) = \sum_y f_{XY}(x,y) = \begin{cases} 3/12 & \text{om } x = 1 \\ 4/12 & \text{om } x = 2 \\ 5/12 & \text{om } x = 3 \end{cases}$
   Symmetri ger $f_Y(y) = f_X(x)$.
   c) $\text{E}[X] = \text{E}[Y] = \frac{13}{6}$
   $\text{E}[XY] = \frac{168}{36}$
   $\text{Cov}(X,Y) = \text{E}[XY] - \text{E}[X]\,\text{E}[Y] = \frac{168}{36} - \frac{169}{36} = -\frac{1}{36}$

6. a) $\text{E}[4X - 2Y] = 4\,\text{E}[X] - 2\,\text{E}[Y] = 4\cdot 3 - 2\cdot 2 = 12 - 4 = 8$ $\text{Var}[4X - 2Y] = \text{Var}[4X] + \text{Var}[-2Y] = 16\,\text{Var}[X] + 4\,\text{Var}[Y] = 16\cdot 6^2 + 4\cdot 5^2 = 676$, där första likheten följer av att X och Y är oberoende. Alltså är $\sigma_{4X-2Y} = \sqrt{676} = 26$
   b) X, Y är normalfördelade så $4X - 2Y$ är också det.
   Alltså är $4X - 2Y \sim \text{Norm}(8, 26)$ och $\frac{4X-2Y-8}{26} \sim \text{Norm}(0,1)$. Så:

$$\begin{aligned} P[4X < 2Y] &= P[4X - 2Y < 0] = P[\tfrac{4X-2Y-8}{26} < \tfrac{-8}{26}] = \phi(-8/26) \approx \\ &\approx \phi(-0.3077) \approx 0.38 \end{aligned}$$

7. a) Tillstånd 2 och 3 är absorberande, ty har vi en gång hamnat i tillstånd 2 eller 3 kommer vi aldrig därifrån.

b) $P = \begin{bmatrix} 0 & 1/4 & 1/4 & 1/2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/4 & 1/2 & 1/4 & 0 \end{bmatrix}$

Låt vektorn $\mathbf{u_i}$ vara fördelningen efter i steg. Då är:

$\mathbf{u_0} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}$

$\mathbf{u_1} = \mathbf{u_0}P = \begin{bmatrix} 1/4 & 1/2 & 1/4 & 0 \end{bmatrix}$

$\mathbf{u_2} = \mathbf{u_1}P = \begin{bmatrix} 0 & 9/16 & 5/16 & 2/16 \end{bmatrix}$

Svar: 9/16

c) Låt $q_i = P$(kedjan absorberas i tillstånd 2 då man startar i tillstånd i).

Vi får: $\begin{cases} q_1 = \frac{1}{4}q_2 + \frac{1}{4}q_3 + \frac{1}{2}q_4 \\ q_2 = 1 \\ q_3 = 0 \\ q_4 = \frac{1}{4}q_1 + \frac{1}{2}q_2 + \frac{1}{4}q_3, \end{cases}$, vilket ger $q_4 = \frac{9}{14}$.

8. a) $$\begin{aligned} \mathrm{E}[\hat{\mu}] &= \mathrm{E}[X_1 - X_2 + X_3 - X_4 + \cdots + X_n] = \\ &= \mathrm{E}[X_1] - \mathrm{E}[X_2] + \mathrm{E}[X_3] - \mathrm{E}[X_4] + \cdots + \mathrm{E}[X_n] \end{aligned}$$

Men $\mathrm{E}[X_i] = \mathrm{E}[X]$ och n är udda så alla väntevärden utom det sista tar ut varandra parvis. Alltså är:

$\mathrm{E}[\hat{\mu}] = \mathrm{E}[X] = \mu$

b) $$\begin{aligned} \mathrm{Var}\,\hat{\mu} &= \mathrm{Var}[X_1 - X_2 + X_3 - X_4 + \cdots + X_n] \overset{ober.}{=} \\ &= \mathrm{Var}\,X_1 + \mathrm{Var}\,X_2 + \mathrm{Var}\,X_3 + \mathrm{Var}\,X_4 + \cdots + \mathrm{Var}\,X_n = n\,\mathrm{Var}\,X = n\sigma^2 \end{aligned}$$

c) Nej! Dess varians blir ju bara större och större när vi ökar storleken på stickprovet. En jämförelse med $\bar{X}$ som är väntevärdesriktig och har $\mathrm{Var}\,\bar{X} = \sigma^2/n$ visar tydligt hur katastrofalt dålig $\hat{\mu}$ är!

9. a) 90%-igt konfidensintervall för μ: $\bar{X} \pm t_{0.05}\frac{S}{\sqrt{10}}$, där $t_{0.05} \approx 1.833$

I vårt fall är $\bar{X} = 7.691$ och $S = 0.1367$, vilket ger konfidensintervallet: $7.69 \pm 0.08 = [7.61, 7.77]$

b) Sannolikheten att det stokastiska intervallet $\bar{X} \pm t_{0.05}\frac{S}{\sqrt{10}}$ innehåller det korrekta värdet på μ är 90%. Om X är antalet gånger av 100 som detta händer så är $X \sim \mathrm{Bin}(100, 0.9)$ och $\mathrm{E}[X] = 90$.