TENTAMEN: Matematisk statistik och diskret matematik D (MVE055/MSG810)
Tid och plats: Tisdagen den 25 augusti 2009, kl. 08.30–12.30, VV-salar.
Jour: Marcus Isaksson, tel 0708-527663. Besöker tentamenssalen ca kl 10.00, 11.30.
Tillåtna hjälpmedel: Chalmersgodkänd räknare och Beta.
Betygsgränser: 3: 12 poäng, 4: 18 poäng, 5: 24 poäng. Maximalt antal poäng är 30.

1. (3p) Låt X vara likformigt fördelad på intervallet $[0, 1]$ och inför händelserna $A = \{X \leq 0.5\}$ och $B = \{X > 0.3\}$.

   a) Vad är $\mathbf{P}(B)$?
   b) Vad är $\mathbf{P}(A \cap B)$?
   c) Ange en händelse $C \subseteq [0, 1]$ sådan att A och C är oberoende och $\mathbf{P}(C) = 0.5$.

   Lösning:

   a) $\mathbf{P}(B) = 1 - \mathbf{P}(X \leq 0.3) = 1 - 0.3 = 0.7$.
   b) $\mathbf{P}(A \cap B) = \mathbf{P}(0.3 \leq X < 0.5) = 0.2$
   c) T.ex. $C = \{0.25 \leq X \leq 0.75\}$.

2. (3p) Vad är $\mathbf{E}[X]$ om

   a) $X = Y + Z$ där $Y \sim \mathrm{Bin}(100, 0.5)$ och $Z \sim \mathrm{Bin}(30, 0.4)$.
   b) $X = (Z - 3)^2$ där Z är normalfördelad med väntevärde 3 och varians 7.
   c) X har frekvensfunktion $f_X(x) = \frac{|5-x|}{30}$, $x = 0, 1, \ldots, 10$.

   Lösning:

   a) $\mathbf{E}[X] = \mathbf{E}[Y + Z] = \mathbf{E}[Y] + \mathbf{E}[Z] = 100 \cdot 0.5 + 30 \cdot 0.4 = 62$
   b) $\mathbf{E}[X] = \mathbf{E}[(Z - 3)^2] = \mathbf{E}[(Z - \mathbf{E}[Z])^2] = \mathbf{Var}\, Z = 7$
   c) Då f_X är symmetrisk kring 5 är $\mathbf{E}[X] = 5$.

3. (4p) Det är inte ovanligt att nya LCD-skärmar levereras med defekta pixlar. En defekt pixel kan t.ex. lysa konstant rött eller inte lysa alls.

   En tillverkare av 8.9-tumsskärmar som består av 1024×600 pixlar placerade i ett rutmönster påstår sig i snitt ha 0.5 defekta pixlar per skärm. Tillverkaren garanterar dessutom att skärmar med minst 3 defekta pixlar byts ut utan kostnad.

   Låt oss anta att pixlar på nytillverkade skärmar är defekta med samma sannolikhet oberoende av varandra.

   a) Hur stor andel av de nytillverkade skärmarna har minst 3 defekta pixlar om tillverkarens påstående stämmer?
   b) För att undersöka om tillverkarens påstående stämmer plockar du ut 10 nytillverkade skärmar och räknar antalet defekta pixlar. Bestäm ett 99%-igt konfidensintervall för antalet defekta pixlar i snitt per skärm om de 10 skärmarna totalt hade 7 defekta pixlar.

   Lösning:

   a) Sannolikheten att en pixel är defekt är $p = \frac{0.5}{1024 \cdot 600} = \frac{1}{1228800}$. Låt X vara antalet defekta pixlar på en skärm. Då är $X \sim \mathrm{Bin}(n, p)$ där $n = 1024 \cdot 600 = 614400$. Sannolikheten för att en skärm har minst 3 defekta pixlar blir:

$$\mathbf{P}(X \geq 3) = 1 - \mathbf{P}(X \leq 2) = 1 - \sum_{x=0}^{2} f_X(x) = 1 - \sum_{x=0}^{2} \binom{n}{x} p^x (1-p)^{n-x} \approx 0.014388 \approx 1.4\%$$

b) Anta nu att p (sannolikheten att en pixel är defekt) är okänd och låt X vara antalet observerade defekta pixlar på de 10 skärmarna. Då är $X \sim \text{Bin}(n, p)$ där $n = 10 \cdot 1024 \cdot 600 = 6144000$. Vi får skattningen $\hat{p} = 7/6144000 \approx 1.1393 \cdot 10^{-6}$ och det (approximativa) konfidensintervallet

$$\hat{p} \pm z_{0.005}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx 1.14 \cdot 10^{-6} \pm 1.11 \cdot 10^{-6}$$

där vi använt $z_{0.005} \approx 2.58$. Det förväntade antalet defekta pixlar per skärm blir därför med 99% konfidens:

$$1024 \cdot 600 \cdot p \approx 0.7 \pm 0.682$$

4. (3p) Pelle har konstruerat en fusktärning som ger sexor med 25% sannolikhet och en till utseendet identisk normal tärning som ger sexor i $\frac{1}{6}$ av kasten. Tyvärr har han råkat blanda ihop tärningarna och vet inte vilken som är vilken. För att försöka avgöra det väljer han ut en tärning på måfå och kastar den 10 gånger. Tre av kasten blir sexor. Vad är sannolikheten att han valde ut fusktärningen?

   Lösning:
   Inför händelserna $F = \{\text{fusktärningen valdes}\}$ och $T = \{\text{tre av kasten blir sexor}\}$. Då är

$$\begin{aligned}
\mathbf{P}(F) &= \frac{1}{2} \\
\mathbf{P}(T|F) &= \binom{10}{3} 0.25^3 \cdot 0.75^7 \approx 0.25028 \\
\mathbf{P}(T|F^C) &= \binom{10}{3}\left(\frac{1}{6}\right)^3 \cdot \left(\frac{5}{6}\right)^7 \approx 0.15505
\end{aligned}$$

   Totala sannolikhetslagen ger $\mathbf{P}(T) = \mathbf{P}(T|F)\mathbf{P}(F) + \mathbf{P}(T|F^C)\mathbf{P}(F^C)$.
   Det som söks är

$$\mathbf{P}(F|T) = \frac{\mathbf{P}(T|F)\mathbf{P}(F)}{\mathbf{P}(T)} = \frac{0.25028 \cdot 0.5}{0.25028 \cdot 0.5 + 0.15505 \cdot 0.5} \approx 0.61749 \approx 61.7\%$$

5. (3p) Du har glömt bort den fyrsiffriga portkoden till ditt trapphus och börjar därför i panik trycka in slumpmässiga siffror i hopp om att förr eller senare råka trycka rätt. Låt oss för enkelhets skull anta att varje knapptryckning du utför är en likformigt fördelad siffra $(0, 1, \ldots, 9)$ och oberoende av tidigare tryckningar. Vad är det förväntade antalet knapptryckningar tills dörren öppnar sig om den okända koden är 7272

   Knappsatsen i fråga är konstruerad så att låset öppnar sig så fort de fyra senast intryckta siffror motsvarar den rätta koden, dvs om rätt kod är 7272 så öppnar sig låset om du t.ex. knappar in sekvensen 17272 eller 12467277272.

   Lösning:
   Inför tillstånden $0, 7, 72, 727, 1$ där tillstånd 1 står för att dörren har öppnats, tillstånd 727 för att vi just knappat in 727, tillstånd 72 för att vi just knappat in 72, tillstånd 7 för att vi just knappat in 7 men inte 727 och tillstånd 0 för att inget av ovanstående gäller.

   Vi har då en Markovkedja med 1 som absorberande tillstånd. Låt
   $m_i = \mathbf{E}[\text{antal steg tills tillstånd 4 nås om vi börjar i tillstånd } i]$. Då är

$$\begin{cases} m_1 = 0 \\ m_{727} = 1 + \frac{1}{10}m_1 + \frac{1}{10}m_7 + \frac{8}{10}m_0 \\ m_{72} = 1 + \frac{1}{10}m_{727} + \frac{9}{10}m_0 \\ m_7 = 1 + \frac{1}{10}m_{72} + \frac{1}{10}m_7 + \frac{8}{10}m_0 \\ m_0 = 1 + \frac{1}{10}m_7 + \frac{9}{10}m_0 \end{cases} \iff$$

$$\iff \begin{bmatrix} -0.8 & -0.1 & 0 & 1 \\ -0.9 & 0 & 1 & -0.1 \\ -0.8 & 0.9 & -0.1 & 0 \\ 0.1 & -0.1 & 0 & 0 \end{bmatrix} \begin{bmatrix} m_0 \\ m_7 \\ m_{72} \\ m_{727} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \Rightarrow m_0 = 10100$$

6. (4p) En tomatimportör har fått en ny leverans av 1200 kg tomater. I syfte att uppskatta mängden tomater väger han ett stickprov av 100 tomater och får stickprovsmedelvärdet $\bar{x} = 87.2$ g och stickprovsstandardavvikelsen $s = 7.8$ g. Beräkna

   a) ett approximativt 90%-igt konfidensintervall för de levererade tomaternas medelvikt.
   b) ett approximativt 90%-igt konfidensintervall för antalet tomater i leveransen.

   Lösning:

   a) Genom centrala gränsvärdessatsen kan vi anta att stickprovsmedelvärdet är normalfördelat. Konfidensintervallet för dess väntevärde μ ges därför av:

   $$\bar{x} \pm t_{0.05}\frac{s}{\sqrt{100}} = 87.2 \pm 1.660\frac{7.8}{10} \approx 87.2 \pm 1.3 = [85.9, 88.5]$$

   b) Antalet tomater är $N = \frac{1.2 \cdot 10^6}{\mu}$. Enligt 6a) har vi med 90% konfidens att $85.9 \leq \mu \leq 88.5$. Men,

   $$85.9 \leq \mu \leq 88.5 \Rightarrow 85.9 \leq \frac{1.2 \cdot 10^6}{N} \leq 88.5 \Rightarrow \frac{1.2 \cdot 10^6}{88.5} \leq N \leq \frac{1.2 \cdot 10^6}{85.9} \Rightarrow 13559 \leq N \leq 13970$$

   Alltså är $[13559, 13970]$ ett 90%-igt konfidensinterval för N.

7. (3p) Vid manuell tidtagning med stoppur under ett sprinterlopp kan mätfelet (upmätt tid minus verklig tid) antas vara normalfördelat med väntevärde 0.00 sekunder och standardavvikelse 0.05 sekunder.

   a) Vad är sannolikheten att absolutbeloppet av mätfelet överstiger 0.02 sekunder?
   b) Anta istället att 10 personer mäter tiden på samma sätt men oberoende av varandra. Vad är sannolikheten att absolutbeloppet av mätfelet överstiger 0.02 sekunder om man använder medelvärdet av deras mätningar som uppmätt tid?

   Lösning:

   a) Låt $Z \sim \text{Norm}(0, 0.05)$ vara mätfelet.

   $$\begin{aligned} \mathbf{P}(|Z| > 0.02) &= \mathbf{P}(|Z/0.05| > 0.4) = 2\mathbf{P}(Z/0.05 < -0.4) = \\ &= 2\mathbf{P}(Z < -2) = 2\Phi(-0.4) \approx 0.68916 \approx 69\% \end{aligned}$$

   b) Låt $Z_1, \ldots, Z_{10} \sim \text{Norm}(0, 0.05)$ vara de oberoende mätfelen. Medelvärdet $Z = \frac{Z_1 + \ldots + Z_{10}}{10}$ är då också normalfördelat med väntevärde $\mathbf{E}[Z] = \frac{\mathbf{E}[Z_1] + \ldots + \mathbf{E}[Z_{10}]}{10} = 0$ och variansen $\mathbf{Var}\, Z = \frac{\mathbf{Var}\, Z_1 + \ldots + \mathbf{Var}\, Z_{10}}{10^2} = \frac{0.05^2}{10}$, dvs standardavvikelsen $\sigma_Z = \frac{0.05}{\sqrt{10}} \approx 0.01581$. Alltså,

   $$\begin{aligned} \mathbf{P}(|Z| > 0.02) &= \mathbf{P}(|Z/0.01581| > 1.265) = 2\mathbf{P}(Z/0.05 < -1.265) = \\ &= 2\mathbf{P}(Z < -2) = 2\Phi(-1.265) \approx 0.20587 \approx 20.6\% \end{aligned}$$

8. (4p) Ett spelbolag erbjuder 3.1 gånger insatsen om du satsar på att Arsenal vinner kvällens fotbollsmatch och samtidigt 6.2 gånger insatsen om du satsar på att Colorado Avalanche vinner morgondagens hockeymatch. (Om du t.ex. satsar 100kr på Arsenal får du alltså tillbaks 310kr om Arsenal vinner och ingenting annars. Din vinst blir alltså 210kr om Arsenal vinner och -100kr annars)

   Du bedömer att sannolikheten att Arsenal vinner är 70% och sannolikheten att Colorado vinner är 35% samt att dessa båda händelser är oberoende. Under antagandet att du har gjort en korrekt bedömning bestämmer du dig för att satsa totalt 100kr och funderar på hur du ska fördela dessa. Anta att du satsar a kr på Arsenal och $100 - a$ kr på Colorado.

   a) Beräkna väntevärdet av din vinst och visa att väntevärdet inte beror på a.
   b) Beräkna variansen av din vinst och bestäm det värde på a som minimerar variansen.

   Lösning:

   a) Låt X resp Y vara vinst per satsad kr på Arsenal resp. Colorado. Då är
   $X = \begin{cases} 2.1 \text{ m.s. } 0.7 \\ -1 \text{ m.s. } 0.3 \end{cases}$ och $Y = \begin{cases} 5.2 \text{ m.s. } 0.35 \\ -1 \text{ m.s. } 0.65 \end{cases}$.
   Vi får $\mathbf{E}[X] = 2.1 \cdot 0.7 - 1 \cdot 0.3 = 1.17$ och $\mathbf{E}[Y] = 5.2 \cdot 0.35 - 1 \cdot 0.65 = 1.17$
   Din vinst är $aX + (100 - a)Y$ och den förväntade vinsten blir alltså
   $\mathbf{E}[aX + (100 - a)Y] = 117$, vilket ej beror på a.
   b) Varianserna blir $\mathbf{Var}[X] = \mathbf{E}[(X - 1.17)^2] = 0.93^2 \cdot 0.7 + 2.17^2 \cdot 0.3 \approx 2.0181$ och
   $\mathbf{Var}[Y] = \mathbf{E}[(Y - 1.17)^2] = 4.03^2 \cdot 0.35 + 2.17^2 \cdot 0.65 \approx 8.7451$.
   Variansen för din vinst blir alltså $\mathbf{Var}[aX + (100 - a)Y] =$
   $a^2 \mathbf{Var}\, X + (100 - a)^2 \mathbf{Var}\, Y = a^2 2.0181 + (100 - a)^2 8.7451 = 10.763a^2 - 1749a + 87451$
   vilket minimeras då derivatan m a p a är 0, dvs då $21.526a - 1749 = 0$, dvs
   $a \approx 81.25 \approx 81$.

9. (3p) En stokastisk variabel Z sägs vara Skellamfördelad med parametrar a och b om Z har samma fördelning som differensen $X - Y$ av två oberoende Poissonfördelade variabler X, Y med väntevärde a respektive b. Vad är den momentgenererande funktionen för en Skellamfördelad variabel med parametrar 3 och 1.5? Du kan utnyttja att den momentgenererande funktionen för en Poissonfördelad variabel med väntevärde λ är $m(t) = e^{\lambda(e^t - 1)}$.

   Lösning:
   Låt X och Y vara oberoende och Poissonfördelade med väntevärde 3 respektive 1.5. Då är $Z = X - Y$ Skellamfördelad med parametrar 3 och 1.5?

$$m_Z(t) = \mathbf{E}[e^{tZ}] = \mathbf{E}[e^{t(X-Y)}] = \mathbf{E}[e^{tX}e^{-tY}] = \mathbf{E}[e^{tX}]\mathbf{E}[e^{-tY}] = m_X(t)m_y(-t) =$$
$$= e^{3(e^t-1)}e^{1.5(e^{-t}-1)} = e^{4.5+3e^t+1.6e^{-t}}$$

Lycka till!