

Petter Mostad
Tillämpad matematik och statistik
Matematiska Vetenskaper, Chalmers

**Lösningssförslag för
MVE051/MSG810 Matematisk statistik och diskret matematik**

Tenta 14 Januari 2017, 8:30 - 12:30

1. Frågan är formulerat lite otydligt: Under de vanligaste poker-reglerna kan esset representera antingen 1 eller 14, dock inte samtidigt. Man har då, i varje färj, 10 möjliga sekvenser av efterföljande värden, totalt 40 möjliga straight flush händer. Som uppgiften är formulerat används "deuce-to-seven low" regler, där esset bara kan representera 14. Man har då 9 möjliga sekvenser av efterföljande värden i varje färj, totalt 36 möjliga straight flush händer. I första fall får man en sannolikhet

$$\frac{40}{\binom{52}{5}} = \frac{40 \cdot 5!}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48} = 1,539077 \cdot 10^{-5}$$

i andra fall en sannolikhet

$$\frac{36}{\binom{52}{5}} = \frac{36 \cdot 5!}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48} = 1,385169 \cdot 10^{-5}$$

Båda svaren har godkänts.

2. Om man drar utan återläggning så kan sannolikheten beräknas som

$$\frac{8}{19} \cdot \frac{7}{18} \cdot \frac{6}{17} \cdot \frac{5}{16} \cdot \frac{5}{15} \cdot \frac{4}{14} = 0,001719986$$

Dock var uppgiften otydlig i det att man skulle kunna tänkas dra med återläggning. Då blir sannolikheten

$$\left(\frac{8}{19}\right)^2 \cdot \left(\frac{6}{19}\right)^2 \cdot \left(\frac{5}{19}\right)^2 = 0,001224337$$

Detta svaret har också godkänts.

3. Genom att använda definitionerna får man

$$\begin{aligned} F_Y(y) &= \mathbf{P}(Y \leq y) \\ &= 1 - \mathbf{P}(Y \geq y) \\ &= 1 - \mathbf{P}(\min(X_1, \dots, X_n) \geq y) \\ &= 1 - \mathbf{P}((X_1 \geq y) \cap (X_2 \geq y) \cap \dots \cap (X_n \geq y)) \\ &= 1 - \mathbf{P}(X_1 \geq y) \mathbf{P}(X_2 \geq y) \cdots \mathbf{P}(X_n \geq y) \\ &= 1 - (1 - \mathbf{P}(X_1 \leq y)) \cdot (1 - \mathbf{P}(X_2 \leq y)) \cdots (1 - \mathbf{P}(X_n \leq y)) \\ &= 1 - (1 - F_x(y))^n \end{aligned}$$

och

$$\begin{aligned}
 F_Z(z) &= \mathbf{P}(Z \leq z) \\
 &= \mathbf{P}(\max(X_1, \dots, X_n) \leq z) \\
 &= \mathbf{P}((X_1 \leq z) \cap (X_2 \leq z) \cap \dots \cap (X_n \leq z)) \\
 &= \mathbf{P}(X_1 \leq z) \mathbf{P}(X_2 \leq z) \cdot \dots \cdot \mathbf{P}(X_n \leq z) \\
 &= F_X(z)^n
 \end{aligned}$$

4. (a) Om A är en stokastisk variabel med exponentialfördelning med väntevärde 3 så är täthetsfunktionen $f_A(x) = \frac{1}{3}e^{-x/3}$, och därmed är fördelningsfunktionen

$$F_A(x) = \int_0^x \frac{1}{3}e^{-t/3} dt = [-e^{-t/3}]_0^x = 1 - e^{-x/3}$$

Om motsvarande B är en stokastisk variabel med exponentialfördelning med väntevärde 5, så får vi

$$F_B(x) = 1 - e^{-x/5}$$

Om X är en stokastisk variabel som motsvarar levnadslängden till komponenten så kan vi skriva

$$\begin{aligned}
 F_X(x) &= \mathbf{P}(X \leq x) = \mathbf{P}(X \leq x \mid \text{typ A}) \mathbf{P}(\text{typ A}) + \mathbf{P}(X \leq x \mid \text{typ B}) \mathbf{P}(\text{typ B}) \\
 &= (1 - e^{-x/3}) \cdot 0.5 + (1 - e^{-x/5}) \cdot 0.5 = 1 - 0.5 \cdot (e^{-x/3} + e^{-x/5})
 \end{aligned}$$

- (b) Vi kan använda Bayes formel:

$$\begin{aligned}
 \mathbf{P}(\text{typ B} \mid X \geq t) &= \frac{\mathbf{P}(X \geq t \mid \text{typ B}) \mathbf{P}(\text{typ B})}{\mathbf{P}(X \geq t)} \\
 &= \frac{(1 - F_B(t)) \cdot 0.5}{1 - F_X(t)} \\
 &= \frac{e^{-t/5} \cdot 0.5}{0.5 \cdot (e^{-t/3} + e^{-t/5})} \\
 &= \frac{1}{e^{t/5-t/3} + 1} \\
 &= \frac{1}{e^{-2t/15} + 1}
 \end{aligned}$$

5. (a)

$$\begin{aligned}
 E(X) &= 1 \cdot 0.2 + 2 \cdot 0.4 + 2.5 \cdot 0.4 = 2 \\
 \text{Var}(X) &= E(X^2) - E(X)^2 = 1^2 \cdot 0.2 + 2^2 \cdot 0.4 + 2.5^2 \cdot 0.4 - 2^2 = 0.3
 \end{aligned}$$

- (b)

$$m_X(t) = E(e^{tX}) = e^t \cdot 0.2 + e^{2t} \cdot 0.4 + e^{2.5t} \cdot 0.4$$

(c)

$$\begin{aligned}m'_X(t) &= e^t \cdot 0.2 + 2e^{2t} \cdot 0.4 + 2.5e^{2.5t} \cdot 0.4 \\m'_X(0) &= 0.2 + 2 \cdot 0.4 + 2.5 \cdot 0.4 = 2\end{aligned}$$

6. (a) Vi får

$$E(X) = \int_0^1 x \cdot 3x^2 dx = \left[\frac{3}{4}x^4 \right]_0^1 = \frac{3}{4} = 0,75$$

och

$$\begin{aligned}Var(X) &= E(X^2) - E(X)^2 = \int_0^1 x^2 \cdot 3x^2 dx - \left(\frac{3}{4}\right)^2 \\&= \left[\frac{3}{5}x^5 \right]_0^1 - \frac{9}{16} = \frac{3}{5} - \frac{9}{16} = \frac{3}{80} = 0,0375\end{aligned}$$

(b) En sfär med radius X har volymen

$$V = \frac{4}{3}\pi X^3$$

och därmed får vi

$$E(V) = \frac{4}{3}\pi E(X^3) = \frac{4}{3}\pi \int_0^1 x^3 \cdot 3x^2 dx = \frac{4}{3}\pi \left[\frac{3}{6}x^6 \right]_0^1 = \frac{2\pi}{3} = 2,0944$$

7. Vi använder genererande funktioner. För x_1, x_2, x_3, x_4 får vi de respektiva genererande funktionerna

$$\begin{aligned}P_1(x) &= 1 + x + x^2 + x^3 + x^4 = \frac{1 - x^5}{1 - x} \\P_2(x) &= x^6 + x^7 + x^8 + \dots = \frac{x^6}{1 - x} \\P_3(x) &= 1 + x + x^2 + x^3 + \dots = \frac{1}{1 - x} \\P_4(x) &= 1 + x + x^2 + x^3 + \dots = \frac{1}{1 - x}\end{aligned}$$

Produkten blir

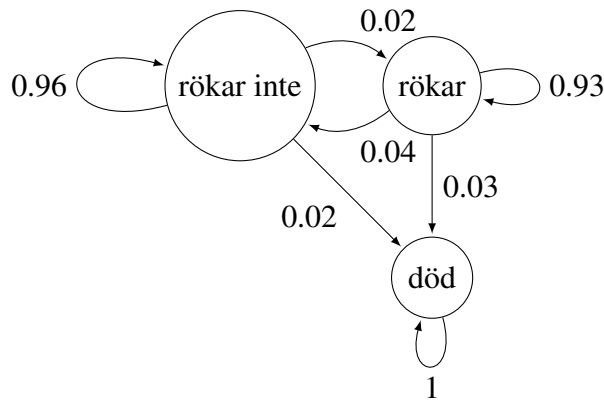
$$P_1(x)P_2(x)P_3(x)P_4(x) = \frac{(1 - x^5)x^6}{(1 - x)^4} = (x^6 - x^{11}) \sum_{n=0}^{\infty} \binom{n+3}{3} x^n$$

Vi får därmed att koefficienten framför x^{21} i denna serien blir

$$\binom{15+3}{3} - \binom{10+3}{3} = \frac{18!}{15!3!} - \frac{13!}{10!3!} = 530$$

som blir lika med antalet heltalslösningar av de ursprungliga ekvationerna.

8. (a) Vi kan illustrera med



(b) När vi listar tillstånden i ordningen “rökar inte”, “rökar” och “död” får vi

$$\begin{bmatrix} 0.96 & 0.02 & 0.02 \\ 0.04 & 0.93 & 0.03 \\ 0 & 0 & 1 \end{bmatrix}$$

(c) För det första beror sannolikheten att dö generellt på ålder. Det är alltså inte så att det finns en fixerad sannolikhet varje år att dö oberoende av hur gammal man är. Ett annat sätt att säga detta på är att förväntad återstående levnadstid beror på hur gammal man är. Detta är den viktigaste orsaken till att man inte bör använda Markovkädjor för denna sortens modeller. Även sannolikheten för att gå mellan tillstånden “rökar” och “rökar inte” beror i praktiken på ålder, och på tidigare rökning. Det är också så att effekten av rökning på sannolikheten för att dö kvarstår i viss mån även efter att man slutat röka. Tillståndet “rökar inte” har inget “minne” av om man rökt tidigare eller inte. Även detta gör Markovkädjor till en olämplig modell i detta fallet.

(d) Om vi skriver övergångsmatrisen på formen

$$\begin{bmatrix} 0.96 & 0.02 & 0.02 \\ 0.04 & 0.93 & 0.03 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} Q & R \\ 0 & 1 \end{bmatrix}$$

får vi att den fundamentala matrisen blir

$$\begin{aligned} N &= (I - Q)^{-1} = \begin{bmatrix} 0.04 & -0.02 \\ -0.04 & 0.07 \end{bmatrix}^{-1} = \frac{1}{0.04 \cdot 0.07 - 0.02 \cdot 0.04} \begin{bmatrix} 0.07 & 0.02 \\ 0.04 & 0.04 \end{bmatrix} \\ &= \frac{1}{0.002} \begin{bmatrix} 0.07 & 0.02 \\ 0.04 & 0.04 \end{bmatrix} = \begin{bmatrix} 35 & 10 \\ 20 & 20 \end{bmatrix} \end{aligned}$$

Förväntad levnadstid om man inte röker motsvarar summan av värdena i första linje, i.e., $35 + 10 = 45$ år, medan förväntad levnadstid om man röker motsvarar summan av värdena i andra linje, i.e., $20 + 20 = 40$ år.

En alternativ lösning baserar sig på att lösa ett ekvationssystem: Låt m_1 och m_2 beteckna förväntad återstående levnadslängd (i.e., förväntad antal steg till absorption) för icke-rökare respektive rökare. Då får vi ekvationssystemet

$$\begin{aligned}m_1 &= 1 + 0.96m_1 + 0.02m_2 \\m_2 &= 1 + 0.04m_1 + 0.93m_2\end{aligned}$$

Vi skriver om till systemet

$$\begin{aligned}0 &= 1 - 0.04m_1 + 0.02m_2 \\0 &= 1 + 0.04m_1 - 0.07m_2\end{aligned}$$

Addition av de två ekvationerna ger

$$0 = 2 - 0.05m_2$$

som ger $m_2 = 2/0.05 = 40$. Om vi sätter detta in i första ekvationen får vi

$$0 = 1 - 0.04m_1 + 0.02 \cdot 40$$

som ger $m_1 = 1.8/0.04 = 45$.

9. Låt p ange sannolikheten för vinst i spelet, och låt $p_0 = 0.1$. Hypoteserna blir

$$\begin{aligned}H_0 &: p = p_0 \\H_1 &: p < p_0\end{aligned}$$

Estimatorn är $\hat{p} = \frac{6}{94} = 0.06383$. Teststatistikan blir

$$T = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} = \frac{\frac{6}{94} - 0.1}{\sqrt{0.1(1 - 0.1)/94}} = -1.169$$

Då vi har en en-sidig test så blir förkastningsområdet blir alla T som är mindre än -1.64, när signifikansnivåen är 95%. Vi förkastar alltså inte nollhypotesen.

Enligt tabellen för standard normalfördelning blir p-värdet 0.121 när test-statistikan har värdet -1.17.

10. (a) Vi måste anta att observationerna är stickprov (alltså oberoende), och att populationerna är normalfördelade.

- (b) Tyvärr kan uppgiften läsas som att observationerna från population B är antingen 6.5, 4.3, och 1.1, eller 6, 5, 4.3, och 1.1. Under beskrivs detta som alternativ 1 och alternativ 2:

Alternativ 1:

Medelvärde och stickprovsvarians för population A: $\bar{x}_A = 5.425$ och $s_A^2 = 2.849167$.

Medelvärde och stickprovsvarians för population B: $\bar{x}_B = 3.966667$ och $s_B^2 = 7.373333$.

Pooled variance:

$$s_p^2 = \frac{3 \cdot s_A^2 + 2 \cdot s_B^2}{3 + 2} = \frac{3 \cdot 2.849167 + 2 \cdot 7.373333}{3 + 2} = 4.658833$$

Ett 95%-ig konfidensintervall för $\mu_A - \mu_B$ blir:

$$\bar{x}_A - \bar{x}_B \pm t_{0.025, 4+3-2} \sqrt{s_p^2 \left(\frac{1}{4} + \frac{1}{3} \right)} = 5.425 - 3.966667 \pm 2.571 \cdot \sqrt{2.717653} = 1.45833 \pm 4.238372$$

alltså $[-2.78, 5.70]$, och ett 90%-ig konfidensintervall blir

$$\bar{x}_A - \bar{x}_B \pm t_{0.05, 4+3-2} \sqrt{s_p^2 \left(\frac{1}{4} + \frac{1}{3} \right)} = 5.425 - 3.966667 \pm 2.015 \cdot \sqrt{2.717653} = 1.45833 \pm 3.321789$$

alltså $[-1.86, 4.78]$.

Alternativ 2:

Medelvärde och stickprovsvarians för population A: $\bar{x}_A = 5.425$ och $s_A^2 = 2.849167$.

Medelvärde och stickprovsvarians för population B: $\bar{x}_B = 4.1$ och $s_B^2 = 4.486667$.

Pooled variance:

$$s_p^2 = \frac{3 \cdot s_A^2 + 3 \cdot s_B^2}{3 + 3} = \frac{3 \cdot 2.849167 + 3 \cdot 4.486667}{3 + 3} = 3.667917$$

Ett 95%-ig konfidensintervall för $\mu_A - \mu_B$ blir:

$$\bar{x}_A - \bar{x}_B \pm t_{0.025, 4+4-2} \sqrt{s_p^2 \left(\frac{1}{4} + \frac{1}{4} \right)} = 5.425 - 4.1 \pm 2.447 \cdot \sqrt{1.833958} = 1.325 \pm 3.313818$$

alltså $[-1.99, 4.64]$, och ett 90%-ig konfidensintervall blir

$$\bar{x}_A - \bar{x}_B \pm t_{0.05, 4+4-2} \sqrt{s_p^2 \left(\frac{1}{4} + \frac{1}{4} \right)} = 5.425 - 4.1 \pm 1.943 \cdot \sqrt{1.833958} = 1.325 \pm 2.631283$$

alltså $[-1.31, 3.96]$.