Chalmers tekniska högskola
Matematisk statistik

TMS051 Matematisk statistik och simuleringsteknik, Z

Tentamen 30 mars 2005 f V

Tillåtna hjälpmedel är räknedosa utan lagrad information om kursen, Beta samt kursens formel- och tabellsamling.

Examinator är Tommy Norberg, ankn 3528 eller 0730 794209. Tommy är just nu i Asien och återkommer till Sverige den 10:e april.

Jour är

Maximalt antal tentamenspoäng är 30, av dessa krävs 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a. Lösningar går att ladda ner från kurshemsidan. Rättningsprotokoll anslås i första våningen i Matematisk centrum, Eklandagatan 86. Granskning av tentan kan under terminstid göras i mottagningsrummet på entréplanet (våning 2) i Matematisk centrum må-fr 12^{30}–13.

Svar skall motiveras om ej annat sägs i uppgiften.

Uppgifter

1. Antag att du i din hand har en välblandad standard kortlek om 52 kort, varav 13 med ”färgen” spader. Du lägger först ett kort på bordet och sedan ett till. Båda korten läggs med baksidan upp så att man inte kan se deras färg.

   (a) Hur stor är sannolikheten att det första kortet har färgen spader? (1 p)

   (b) Hur stor är sannolikheten att det andra kortet har färgen spader? (1 p)

   (c) Är händelserna ”första kortets färg är spader” och ”andra kortets färg är spader” oberoende? (2 p)

   Ingen motivering krävs på deluppgifterna (a) och (b). Däremot är det viktigt att deluppgift (c) motiveras korrekt.

2. Denna uppgift och nästa handlar om en ”binary symmetric communication channel”. En sådan sänder en bit (som kan anta värdena $i = 0, 1$) åt gången och definieras av att
$$P(M_1|S_0) = P(M_0|S_1) = p$$
   Här betecknar M_i händelsen att biten i mottages och S_i händelsen att biten i sänds. Beräkna $P(S_1|M_1)$ om $P(S_1) = 0.5$. (3 p)

3. Visa, för den binära kommunikationskanalen som beskrivs i ovanstående uppgift, att $P(S_1|M_1) > P(S_1) \Leftrightarrow 2p < 1$. (4 p)

4. (a) Kan funktionen $f(x)$, given av att
$$\frac{f(x)}{c} = \begin{cases} \dfrac{x-0.2}{0.15} & \text{då} \quad 0.2 \le x \le 0.35 \\[2ex] \dfrac{0.7-x}{0.35} & \text{då} \quad 0.35 \le x \le 0.7 \end{cases}$$

för något c (och i så fall vilket) fungera som täthet för någon stokastisk variabel. (Svaret är, som du förstår av nästa fråga, ja, men du måste motivera det korrekt.) (2 p)

(b) Låt den stokastiska variabeln X ha denna täthet. Beräkna $P(X \geq 0.35)$.(2 p)

5. Låt X, Y vara bivariat normalfördelade med $E[X] = 2$, $E[Y] = 1$, $\mathrm{Var}[X] = 4$, $\mathrm{Var}[Y] = 1$ och $\rho = 0.5$. Sätt $U = X - 2Y$. Vilken fördelning har U? (4 p)

6. Man gjorde 7 mätningar av en exponentialfördelad tid. Därvid erhölls:

$$1.226 \quad 0.765 \quad 3.090 \quad 2.457 \quad 0.667 \quad 0.335 \quad 3.340$$

(tidsenhet: timmar). Beräkna trolighetsskattningen av den 75:e percentilen. Ge svaret med 2 korrekta decimaler. (4 p)

7. I en simuleringsstudie gjordes $n = 10\,000$ oberoende simuleringar av en slumpmässig storhet x med okänt väntevärde μ. Därvid erhölls $\sum x = 27\,400$ och $\sum x^2 = 85\,075$. Hur stort är ”felet” i skattningen $\hat{\mu} = 2.74$ av μ? Räkna ut det med 95% konfidens. (3 p)

8. Efter reparation av ett mätinstrument skulle det kalibreras m.a.p väntevärdesriktighet medelst jämförelse med ett väntevärdesriktigt instrument. Man gjorde $n_1 = 9$ mätningar och erhöll medelvärdet $\bar{x}_1 = 3.57$ samt variansen $s_1^2 = 0.538$ med det just reparerade instrumentet och i ytterligare $n_2 = 6$ mätningar av samma storhet med det väntevärdesriktiga instrumentet erhölls $\bar{x}_2 = 4.47$ och $s_2^2 = 0.621$.

(a) Finns det anledning att misstänka att de båda instrumenten hade olika standardavvikelser i denna mätning? (2 p)

(b) Finns det anledning att misstänka att det just reparerade instrumentet ej ger väntevärdesriktigta mätresultat? (2 p)

Oberoende normalfördelade data kan förutsättas. I deluppgift (a) är det lämpligt med konfidensen 90% och i deluppgift (b) 95%.

Lycka till!

1. Låt A_i, för $i = 1, 2$, beteckna händelsen att kort i är spader.
   (a) $P(A_1) = \frac{13}{52} = \frac{1}{4}$
   (b) $P(A_2) = \frac{1}{4}$ (för alla korten i leken gäller ju att sannolikheten att det är spader är $13/52 = 1/4$ om man inte vet något om de övriga)
   (c) Nej, ty $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1) = \frac{13}{52}\frac{12}{51} \neq \frac{1}{4}\frac{1}{4} = P(A_1)P(A_2)$

2. Bayes formel ger:

$$P(S_1|M_1) = \frac{P(S_1)P(M_1|S_1)}{P(S_1)P(M_1|S_1) + P(S_0)P(M_1|S_0)} = \frac{1-p}{1-p+p} = 1-p$$

3. $P(S_1|M_1) > P(S_1) \Leftrightarrow P(S_1 \cap M_1) > P(S_1)P(M_1) \Leftrightarrow P(M_1|S_1) > P(M_1) = P(S_1)P(M_1|S_1) + P(S_0)P(M_1|S_0) \Leftrightarrow P(S_0)P(M_1|S_1) > P(S_0)P(M_1|S_0) \Leftrightarrow 1-p > p \Leftrightarrow 2p < 1$, qed

4. (a) $f(x) \geq 0$ och integrabel, så den måste vara en täthet för något positivt c. Vi ser att $\int_{-\infty}^{\infty} \frac{f(x)}{c}\, dx = \int_{0.2}^{0.7} \frac{f(x)}{c}\, dx = \frac{0.5 \cdot 1}{2} = 0.25 \Rightarrow c = 4$ ($\frac{f(x)}{c}$ är ju en triangel med basen $0.15 + 0.35 = 0.5$ och höjden 1)
   (b) Den sökta sannolikheten är $1 - \int_{-\infty}^{0.35} f(x)\, dx = 1 - \int_{0.2}^{0.35} f(x)\, dx = 1 - \frac{0.15 \cdot 4}{2} = 1 - 0.3 = 0.7$ (restriktionen av $f(x)$ till intervallet $[0.2, 0.35]$ är en triangel med basen 0.15 och höjden c)

5. Linjärkombinationer av bivariat normalfördelade variabler är normalfördelade, så U är $\mathrm{N}(\mu, \sigma)$. Vi får vidare att $\mu = E[U] = E[X] - 2\,E[Y] = 2 - 2 \cdot 1 = 0$, samt $\sigma^2 = \mathrm{Var}[U] = \mathrm{Kov}[U,U] = \mathrm{Kov}[X - 2Y, X - 2Y] = \mathrm{Kov}[X,X] - 2\,\mathrm{Kov}[X,Y] - 2\,\mathrm{Kov}[Y,X] + 4\,\mathrm{Kov}[Y,Y] = \mathrm{Var}[X] + 4\,\mathrm{Var}[Y] - 4\,\mathrm{Kov}[X,Y] = \mathrm{Var}[X] + 4\,\mathrm{Var}[Y] - 4\,\rho\sqrt{\mathrm{Var}[X]\,\mathrm{Var}[Y]} = 4 + 4 \cdot 1 - 4 \cdot \frac{1}{2} \cdot \sqrt{4 \cdot 1} = 4$. Således är $U \sim \mathrm{N}(0, 2)$

6. Modellen är att data är $\exp(\lambda)$-fördelade. Då är tätheten

$$f(x) = \lambda\, e^{-\lambda x} \quad \text{för} \quad x \geq 0$$

Den 75:e percentilen $x_{0.75}$ uppyller

$$0.75 = \int_0^{x_{0.75}} \lambda\, e^{-\lambda x}\, dx = 1 - e^{-\lambda x_{0.75}} \quad \Rightarrow \quad x_{0.75} = -\frac{\ln 0.25}{\lambda} = \frac{\ln 4}{\lambda}$$

Således gäller att trolighetsskattningen av $x_{0.75}$ ges av

$$\hat{x}_{0.75} = \frac{\ln 4}{\hat{\lambda}}$$

där $\hat{\lambda}$ är trolighetsskattningen av λ. Denna fås medelst maximering av trolighetsfunktionen

$$L(\lambda) = \prod_i f(x_i) = \lambda\, e^{-\lambda x_1} \cdot \ldots \cdot \lambda\, e^{-\lambda x_n} = \lambda^n\, e^{-\lambda \sum_i x_i}$$

Här betecknar $x_1, \ldots, x_n$ de uppmätta tiderna. Tricket, som man ska känna till, i den här typen av maximi-problem, är att man istället ska maximera logaritmen. Därför bildar vi

$$\mathcal{L}(\lambda) = n \ln \lambda - \lambda \sum_i x_i$$

Derivera, sätt derivatan till noll och lös ut λ:

$$\frac{d\mathcal{L}}{d\lambda} = \frac{n}{\lambda} - \sum_i x_i = 0 \quad \Rightarrow \quad \frac{n}{\lambda} = \sum_i x_i \quad \Rightarrow \quad \lambda = \frac{n}{\sum_i x_i} = \frac{1}{\bar{x}}$$

Loggade trolighetsfunktioner av denna typ är konkava, så detta måste vara en maximipunkt. (Inget poängavdrag om argumentation för att detta är en maximipunkt saknas.) Således gäller att $\hat{\lambda} = \frac{1}{\bar{x}} \Rightarrow \hat{x}_{0.75} = \bar{x} \ln 4$. Återstår att räkna ut att

$$\bar{x} = \frac{11.88}{7} = 1.6971\ldots \quad \Rightarrow \quad \hat{x}_{0.75} = 2.35$$

7. Börja med att beräkna $\bar{x} = \frac{1}{n} \sum x = 2.74$ och $(n-1)s^2 = \sum x^2 - \frac{1}{n} \left(\sum x\right)^2 = 9999$ $\Rightarrow s^2 = 1 \Rightarrow s = 1$. Konfidensintervallet blir $\mu = \bar{x} \pm 1.96 s/\sqrt{n} \approx \bar{x} \pm 2s/\sqrt{n} = 2.74 \pm 0.02$. Den som svarar $\pm 1.96 \cdot 10^{-2} = \pm 0.0196$ får naturligtvis också rätt

8. (a) Under nollhypotesen $H_0 : \sigma_1 = \sigma_2$ är kvoten $\frac{s_1^2/n_1}{s_2^2/n_2}$, vars evaluerade värde är $1.846840\ldots$, $F(\mathrm{df}_1, \mathrm{df}_2)$-fördelad, där $\mathrm{df}_i = n_i - 1 = 8$ resp 5. I tabellen i formelsamlingen ser vi att det för en $F(5,8)$-fördelad variabel f gäller att $P(f \leq 0.208) = P(f \geq 3.688) = 0.05$. Notera nu att $0.208 \leq 1/1.846840\ldots = 0.541 \leq 3.688$ ($\frac{s_2^2/n_2}{s_1^2/n_1}$ är ju $F(5,8)$-fördelad). Data tyder alltså inte alls på att antagandet $\sigma_1 = \sigma_2$ är oriktigt

   (b) Eftersom vi inte har någon anledning att misstänka att $\sigma_1 \neq \sigma_2$, "poolar" vi ihop de två variansskattningarna till en enligt

$$s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 0.5699\ldots \quad \Rightarrow \quad s = 0.755$$

   Totala antalet frihetsgrader i denna skattning är $n_1 + n_2 - 2 = 13$ och en titt i motsvarande t-tabell ger oss 0.975-kvantilen $t_{0.025} = 2.160$. Således gäller att konfidensintervallet

$$\mu_1 - \mu_2 = \bar{x}_1 - \bar{x}_2 \pm t_{0.025} s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = -0.90 \pm 0.86$$

   har konfidensen 95%. Då origo ej ingår i detta intervall drar vi slutsatsen att påståendet $\mu_1 \neq \mu_2$ har minst konfidensen 95%. Risken att det är fel är alltså högst 5%. Ett annat sätt att lösa denna uppgift är att testa $H_0 : \mu_1 = \mu_2$ mot alternativet $H_1 : \mu_1 \neq \mu_2$ på nivån 5%. Teststatistikan $\frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ är $t(13)$-fördelad och ur

$$\frac{|\bar{x}_1 - \bar{x}_2|}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 2.262 \geq 2.160$$

   ser vi att testet förkastar H_0. Det finns alltså anledning att misstänka att det just reparerade instrumentet ej ger väntevärdesriktiga resultat.