MVE090 Matematisk statistik Z

Tentamen 18 januari 2007 fm M

Även omtentamen på TMS051 Matematisk statistik och simuleringsteknik Z, del B.

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

Examinator är Tommy Norberg, ankn 3528 eller 0730 79 42 09. Tommy går att nås per mobiltelefon under tentamen. Ingen jour således.

Maximalt antal tentamenspoäng är 30, av dessa krävs 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a. Lösningar till tentamensproblemen går att ladda ner från kurshemsidan. Rättningsprotokoll anslås i MV:F, plan 2.

Svar skall motiveras om ej annat sägs i uppgiften.

Uppgifter

1. För en viss typ av rökdetektorer gäller under normala driftsförhållanden att den betingade sannolikheten för larm är 0.99 givet att det brinner och 0 givet att det ej brinner. Detektorn kan emellertid felfungera. I så fall gäller att den betingade sannolikheten för larm är 0.5 oberoende av ifall det brinner eller ej. Sannolikheten för felfunktion är 0.02, sannolikheten för brand är 0.01 och dessa två händelser är oberoende av varandra. Hur stor är (a) sannolikheten att detektorn larmar, givet att det brinner resp inte brinner och (b) sannolikheten att det brinner, givet att detektorn larmar? (4 p)

2. Beräkna (a) förväntad tid till fel och (b) sannolikheten att fel uppträder inom 1 år, om intensiteten för fel är konstant och lika med $\lambda = 2.74$ fel per 1 000 dygn. (3 p)

3. Visa att $EX = 1/p$ om $X \sim \text{Geo}(p)$. (4 p)

4. Antag att den stokastiska variabeln X är kontinuerligt fördelad med fördelningsfumnktion $F(x)$ och täthet $f(x) = F'(x)$. Visa att då är den stokastiska variabeln $U = F(X)$ likformigt fördelad på enhetsintervallet $(0, 1)$. (4 p)

5. (a) Man ämnar i en undersökning tillfråga 1000 slumpvis utvalda högstadieelever om de har smakat starksprit eller ej. Man tror att den verkliga proportionen högstadieelever som smakat starksprit är ungefär 30%. Ungefär hur stor är sannolikheten att fler än 330 elever svarar ja på frågan? (2 p)

   (b) Då man i ett jordprov mäter koncentrationen arsenik vet man (1) att man mäter väntevärdesriktigt och (2) att felet är additivt och normalfördelat med standardavvikelse ≈ 0.1 mg per kg torrsubstans. I den här typen av mätningar hämtar man hem från provplatsen en stor mängd jord (flera liter) som representerar den volym man vill undersöka. Därefter blandas jorden väl. Man tar sedan ut ett antal små prover som analyseras. Det är i analysen av dessa som man får N(0, 0.1)-fördelade mätfel. Ungefär hur många oberoende prover behövs minst för att medelvärdets standardavvikelse ska bli mindre än 0.005? (2 p)

6. 100 slumpmässigt utvalda gymnasieelever utfrågades om sina studievanor. På en viss fråga svarade 23 elever ja. Kan man på basis av detta dra någon slutsats om hela populationen gymnasieelever och i så fall vad? Ge minst ett par olika exempel. (4 p)

7. I en studie av transistorers livslängd erhölls i en standardiserad tidsenhet: 2.25, 0.89, 0.27, 0.23, 0.64. Den här typen av transistorer tros ha exp(λ)-fördelade livslängder med $\lambda \approx 1$.

   (a) Momentskatta λ.

   (b) ML-skatta λ.

   Obs att i uppgiften ingår att härleda teoretiska uttryck för de två skattningarna. (4 p)

8. I ett jordprov gjordes upprepade mätningar av halten arsenik. Följande halter erhölls:

   11.93 11.94 11.89 12.07 12.04 12.07 11.85

   Enhet: mg per kg torrsubstans. Antag att detta är oberoende N(μ, σ)-observationer.

   (a) Punkt och intervallskatta μ. Konfidensgrad: 95%. (2 p)

   (b) Testa nollhypotesen $H_0 : \sigma \geq 0.1$ mot alternativet $H_1 : \sigma < 0.1$. Nivå: 10%. (2 p)

   Räknehjälp: $\sum x = 83.79$, $\sum x^2 = 1003.0146$.

Lycka till!

Svar eller lösningar till MVE090 den 18/1-07

1. Låt B, F och L beteckna händelsrena "det brinner", "detektorn felfungerar" resp "detektorn larmar." Vi vet att $P(B) = 0.01$, $P(F) = 0.02$ och att B, F är oberoende. Vi får även reda på att $P(L|F', B) = 0.99$, $P(L|F', B') = 0$ samt att $P(L|F, B) = P(L|F, B') = 0.5$ (L och B är således betingat oberoende givet F). Sökt är $P(L|B)$ och $P(L|B')$ i (a), och $P(B|L)$ i (b). Här gäller att

$$\begin{aligned}P(B \cap L) &= P(B \cap L \cap F) + P(B \cap L \cap F') \\ &= P(B \cap F)P(L|F, B) + P(B \cap F')P(L|F', B) \\ &= P(B)P(F)P(L|F, B) + P(B)P(F')P(L|F', B) = 0.009802\end{aligned}$$

Analogt,

$$P(B' \cap L) = P(B')P(F)P(L|F, B') + P(B')P(F')P(L|F', B') = 0.0099$$

Så (a)

$$\begin{aligned}P(L|B) &= \frac{P(B \cap L)}{P(B)} = \frac{0.009802}{0.01} \approx 0.980 \\ P(L|B') &= \frac{P(B' \cap L)}{P(B')} = \frac{0.0099}{0.99} = 0.01\end{aligned}$$

Vidare

$$P(L) = P(B \cap L) + P(B' \cap L) = 0.019702$$

och (b)

$$P(B|L) = \frac{P(B \cap L)}{P(L)} = \frac{0.009802}{0.019702} \approx 0.498$$

Trots att rökdetektorn verkar väldigt bra ($P(L|B) = 0.98$, $P(L|B') = 0.01$) är alltså ungefär hälften av larmen falska.

2. Konstant felintensitet innebär att tiden till fel är exponentialfördelad. Därför
   (a) $\mu = 1/\lambda \approx 365$ dygn eller 1år.
   (b) $1 - e^{-2.74 \cdot 10^{-3} \cdot 365} \approx 1 - e^{-1} = 0.632$

3. Se läroboken, s 64.

4. $U = F(X)$ tar sina värden ur enhetsintervallet och har fördelningsfunktionen

$$F_U(u) = P(U \leq u) = P(F(X) \leq u) = P(X \leq F^{-1}(u)) = F\left(F^{-1}(u)\right) = u$$

Således gäller att tätheten är $f_U(u) = F'_U(u) = 1$ för $u \in (0, 1)$. QED

5. (a) Antalet ja-svarare är approx bin(1000, 0.3) $\approx$ N($\mu = 300$, $\sigma = \sqrt{1000 \cdot 0.3 \cdot 0.7} \approx 14.5$). Den sökta sannolikheten är alltså grovt lika med sannolikheten att en N-variabel antar ett värde större än $\mu + 2\sigma$. Som bekant är denna sannolikhet ungefär 2.5%, eller 0.025.

(b) Ur $0.1/\sqrt{n} = 0.005$ fås $n = (0.1/0.005)^2 = 400$.

6. Först kan man ju konstatera att det går fint att skatta proportionen ja-svarare i hela populationen. Vi kallar den p och dess skattning är $\hat{p} = 0.23$. Sedan kan man konstatera att förutsättningarna för c.g.s är uppfyllda. Så vi kan bilda konfidensintervall för p. Väljes konfidensgraden till ca 95% blir "den statistiska felmarginalen" $\pm 1.96 \cdot \sqrt{0.23 \cdot 0.77/100} \approx \pm 0.08$. Vill man istället ha en undre gräns för p som gäller med en viss konfidens, t.ex 95%, så blir denna $0.23 - 1.645 \cdot \sqrt{0.23 \cdot 0.77/100} \approx 0.16$. Analogt är 0.30 en övre gräns för p som gäller med konfidensen ca 95%. Etc.

7. Både ML-skattningen och momentskattningen är $\widehat{\lambda} = 1/\bar{x} = 1/0.856 \approx 1.17$. Så till härledningarna: (a) Momentskattningen fås ur $E[X] = 1/\lambda$, så $\bar{x} = 1/\widehat{\lambda} \Rightarrow \widehat{\lambda} = 1/\bar{x}$ och (b) trolighetsfunktionen är $L(\lambda) = \prod_i \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_i x_i}$, så $\ln L(\lambda) = n \ln \lambda - \lambda \sum_i x_i$ och via lösandet av $d \ln L(\lambda)/d\lambda = 0$ fås att trolighetsfunktionen är max då $\lambda = 1/\bar{x}$. Även ML-skattniongen är alltså $\widehat{\lambda} = 1/\bar{x}$.

8. Vi får $\bar{x} = 83.79/7 = 11.970$, $(n-1)s^2 = 1003.0146 - 83.79^2/7 = 0.0483 \Rightarrow s^2 = 0.0483/6 = 0.00805$ och $s = \sqrt{0.00805} \approx 0.0897$. Antalet frihetsgrader är 6 och 0.025-kvantilen i $t(6)$-fördelningen är 2.447. Det i (a) sökta konfidensintervallet är således

$$\mu = 11.970 \pm 0.0897 \cdot 2.447/\sqrt{7} = 11.970 \pm 0.083$$

(b) Testresultatets P-värde är ≤ 0.10 och H_0 kan förkastas på nivån 10%, om $(n-1)s^2/0.1^2 < 2.204$ (variabeln $(n-1)s^2/\sigma^2$ är ju $\chi^2(6)$-fördelad och 2.204 är 0.90-kvantilen i denna fördelning). Vi räknar ut att $(n-1)s^2/0.1^2 = 4.83$. Nollhypotesen H_0 kan alltså ej förkastas på nivån 10%.

Chalmers tekniska högskola
Matematisk statistik

MVE090 Matematisk statistik Z

Tentamen onsdag 25 oktober 2006 fm

Detta är även omtentamen på TMS051 Matematisk statistik och simuleringsteknik Z, del B.

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

Examinator är Tommy Norberg, ankn 3528 eller 0730 79 42 09.

Maximalt antal tentamenspoäng är 30, av dessa krävs 12 för godkänt betyg och 18 resp 24 för 4:a och 5:a. Lösningar till tentamensproblemen går att ladda ner från kurshemsidan. Rättningsprotokoll anslås i MV:F, plan 2.

Svar skall motiveras om ej annat sägs i uppgiften.

Uppgifter

1. Brandlarm med automatlarm till Räddningstjänsten är populärt att installera. Dock är många larm av undermålig kvalitet. De både larmar när det inte borde larmas (vilket inte är så farligt, men dock kostar utryckningspengar) och missar att larma när det brinner (vilket inte alls är bra). Låt B vara händelsen att det brinner en viss given vecka och låt D beteckna händelsen att larmet går. Antag att de brandlarm som är installerade i ett visst industriområde karakteriseras av felsannolikheterna $P(D|B') = 0.05$ och $P(D'|B) = 0.10$. Om det i genomsnitt brinner i en byggnad försedd med larm ca 1 gång per år, så att $P(B) \approx 0.02$, vad är då (a) $P(D)$ och (b) $P(B|D)$? (4 p)

2. I en motor finns fem packningar. Motorn fungerar tillfredställande så länge minst 3 av packningarna håller tätt. Känt är att sannolikheten att en packning håller tätt under motorns driftstid är ca 0.80 och man tror sig veta att händelserna huruvida packningarna håller tätt eller ej är oberoende. Beräkna sannolikheten att motorn fungerar tillfredsställande under hela driftstiden. Svaret ska ges med åtminstone 2 decimalers noggranhet. (4 p)

3. Låt T vara en kontinuerlig icke-negativ stokastisk variabel. Då kan T fungera som modell för en slumpmässig funktionstid. Antag att T:s tillförlitlighetsfunktion är

$$R(t) = e^{-Z(t)}, \qquad t \geq 0$$

för någon icke-negativ växande funktion $Z(t)$ sådan att $Z(0) = 0$. Visa att då är T:s felbenägenhet $z(t) = Z'(t)$. (4 p)

Vänd!

4. Utfallet X, Y av ett visst försök har den bivariata tätheten

$$f(x, y) = c$$

för $x \geq 0$, $y \geq 0$ sådana att $0 \leq x + y \leq 1$.

   (a) Beräkna $E[X]$. (2 p)

   (b) Beräkna $\mathrm{Cov}[X, Y]$. (2 p)

5. Låt $X_1, \ldots, X_n$ vara ett stickprov på X och låt $E[X] = \mu$, $\mathrm{Var}[X] = \sigma^2$. (a) Visa att $\bar{X}$ skattar μ väntevärdesriktigt, och (b) beräkna $\mathrm{Var}[\bar{X}]$. (4 p)

6. Antag att du har oberoende $\mathrm{Poi}(\lambda)$-observationer $x_1, \ldots, x_n$. Härled ett uttryck för hur ML-skattningen av λ ska beräknas. (3 p)

7. I $n = 10$ oberoende mätningar av tillverkningstiden för en produkt erhölls $\sum x = 71.4$ (enhet: timmar) och $\sum x^2 = 531.07$. Man vill skaffa sig en uppfattning om hur lång den förväntade tillverkningstiden i värsta fall är. Bestäm därför ett uppåt begränsat konfidensintervall för denna med konfidensgrad 0.99. Det finns skäl att antaga att observationerna är normalfördelade. (4 p)

8. Man ville undersöka om halten alluminium i två jordprov tagna på femton meters avstånd är positivt korrelerad. Man slumpade därför ut fyra par av positioner med 15 meters inbördes avstånd. I varje sådant par mättes aluminiumhalten i båda positionerna. Följande data erhölls:

$$(12.2,\ 12.8),\ (13.8,\ 12.3),\ (10.7,\ 11.0),\ (12.8,\ 12.3)$$

   Enhet: mg/kg torrsubstans. Antag att detta är oberoende observationer av en bivariat normalfördelning med korrelation ρ. Testa på nivån 10% nollhypotesen $H_0 : \rho \leq 0$ mot alternativet $H_1 : \rho > 0$. (3 p)

Lycka till!

Svar eller lösningar till MVE090 den 25/10-06

1. (a) Med "lagen om total sannolikhet" fås $P(D) = P(B)P(D|B)+P(B')P(D|B') = 0.02 \cdot 0.90 + 0.98 \cdot 0.05 = 0.067$

   (b) Med Bayes formel fås $P(B|D) = \frac{P(B)P(D|B)}{P(D)} = \frac{0.018}{0.067} = 0.269$

2. Antalet packningar, X, som fungerar hela driftstiden är $\text{Bin}(n, p)$-fördelat med $n = 5$ och $p = 0.8$, så den sökta sannolikheten är

$$P(X \geq 3) = \binom{5}{3} 0.8^3 \cdot 0.2^2 + \binom{5}{4} 0.8^4 \cdot 0.2^1 + \binom{5}{5} 0.8^5$$
$$= 0.2048 + 0.4096 + 0.32768 \approx 0.942$$

3. Ur $F(t) = 1 - R(t)$ fås $f(t) = F'(t) = -R'(t) = Z'(t)e^{-Z(t)}$, vilket implicerar för felbenägenheten $z(t) = \frac{f(t)}{R(t)} = \frac{Z'(t)e^{-Z(t)}}{e^{-Z(t)}} = Z'(t)$, vsv

4. Rita figur över utfallsrummet $\{(x, y) : x \geq 0, y \geq 0, 0 \leq x + y \leq 1\}$ och dra slutsatsen att $c = 2$ ur faktumet att utfallsrummet har arean $\frac{1}{2}$. Annars

$$1 = \int_0^1 \int_0^{1-x} c\, dy\, dx = c \int_0^1 (1 - x)\, dx = c\frac{1}{2} \quad \Rightarrow \quad c = 2$$

(ty då $x \in [0, 1]$ är fixerat gäller $0 \leq y \leq 1 - x$). Vi får nu att $f_X(x) = \int_0^{1-x} 2\, dy = 2(1-x)$ för $0 \leq x \leq 1$ och, analogt, $f_Y(y) = \int_0^{1-y} 2\, dx = 2(1-y)$ för $0 \leq y \leq 1$. Att variablerna har samma marginalfördelning kan man faktiskt sluta sig till ur figuren över utfallsrummet. Vi får vidare att $\mu_Y = \mu_X = \int_0^1 x \cdot 2(1 - x)\, dx = \frac{1}{3}$. Dessa väntevärden kan även räknas ut så här: $\mu_Y = \mu_X = \int_0^1 \int_0^{1-x} 2x\, dxdy = \int_0^1 2x(1 - x)\, dx = 1 - \frac{2}{3} = \frac{1}{3}$. Dessutom har vi att $E[XY] = \int_0^1 \int_0^{1-x} xy \cdot 2\, dydx = \int_0^1 x(1 - x)^2\, dx = \frac{1}{12}$. Således gäller att $\text{Cov}[X, Y] = \frac{1}{12} - \frac{1}{3} \cdot \frac{1}{3} = -\frac{1}{36}$.(Att kovariansen och därmed även korrelationen är negativ är rimligt med tanke på att då $x \in [0, 1]$ är fixerat gäller $0 \leq y \leq 1 - x$.)

5. Notera först att $E\left[\bar{X}\right] = E\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\sum_{i=1}^n E\left[X_i\right] = \frac{1}{n} n\mu = \mu$ (E är ju linjär). Detta visar väntevärdesriktigheten i (a). För att se påståendet i (b), notera $\text{Var}\left[\bar{X}\right] = \text{Var}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n^2}\text{Var}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n^2}\sum_{i=1}^n \text{Var}\left[X_i\right] = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$. I den 3:e likheten utnyttjas att variablerna är oberoende.

6. Poissontätheten är $f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$ för $x = 0, 1, \ldots$. Så trolighetsfunktionen är $L(\lambda) = \prod_i e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = e^{-n\lambda}\lambda^{\sum_i x_i}\prod_i \frac{1}{x_i!}$. Dess loggade motsvarighet är $\mathcal{L}(\lambda) = -n\lambda + (\sum_i x_i)\ln\lambda - \sum_i \ln(x_i!)$. Derivera och sätt derivatan till noll. Då ses att $n = \frac{\sum_i x_i}{\lambda}$, vilket medför att $\lambda = \frac{\sum_i x_i}{n} = \bar{x}$. ML-skattningen av λ är således lika med medelvärdet $\bar{x}$, som vi ju redan vet är en bra skattning av väntevärdet μ. Detta var kanske inte så förvånande med tanke på att $\mu = \lambda$.

7. Vi räknar först ut att $\bar{x} = 7.14$ och $s^2 = 2.3638 = 1.537^2$. Vi behöver $t_{0.01}(9) = 2.821$ och får $\mu = 7.14 + 2.821 \cdot 1.537/\sqrt{10} = 7.14 + 1.37 = 8.51$.

8. Börja med att beräkna $\sum x = 49.5$, $\sum y = 48.4$, $\sum x^2 = 617.61$, $\sum y^2 = 587.42$ och $\sum xy = 601.04$. Härur fås $S_{xx} = 617.61 - 49.5^2/4 = 5.0475$, $S_{yy} = 587.42 - 48.4^2/4 = 1.7800$ och $S_{xy} = 2.09 - 49.5 \cdot 48.4/4 = 2.0900$ och vi ser att $\hat{\rho} = \frac{2.0900}{\sqrt{5.0475 \cdot 1.78}} = 0.6973$. Obs värde av teststatistikan $\frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}}$ är $\frac{0.6973\sqrt{4-2}}{\sqrt{1-0.6973^2}} = 1.3756$, vilket ska vara större än $t_{0.10}(2) = 1.886$ om H_0 kan förkastas på nivån 10%. Vi ser att H_0 ej kan förkastas.