TMS056 Matematisk statistik V2, 5 p, 17 januari 2007 em V

Anvisningar

Tentamen består av en teoridel med 3 uppgifter och en problemdel med 5 uppgifter. Teoriuppgifterna ska besvaras först. Inga hjälpmendel är tillåtna på denna del. På problemdelen är räknedosa inkl manual, Beta och någon av de läroböcker i matematisk statistik som används på Chalmers tillåtna hjälpmedel. Det är också tillåtet att använda Tommy Norbergs Formler och tabeller till matematisk statistik på universitet och tekniska högskolor, som kan laddas ner från Norbergs hemsida. Tillåtna hjälpmedel är även de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Något om riskkostnader. Det är tillåtet med anteckningar i rimlig omfattning i läroboken och i nedladdade häften. Inga andra anteckningar är tillåtna.

Kom ihåg att att alla svar skall motiveras såvida ej annat sägs. Ställ upp modell när det behövs och ange de för uträkningarna nödvändiga förutsättningarna. Skriv förståligt (tydligt) och använd i ämnet etablerade beteckningar. Poäng dras i lösningar där detta ej är tillgodosett.

Teoridelen skall lämnas in innan några hjälpmedel får tas fram.

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p. Ev bonuspoäng från inlämningsuppgifterna sänker gränser i motsvarande grad.

Examinator är Tommy Norberg, tel. 772 3528, 0730 79 42 09. Tommy går att nås per telefon under tentamen. Ingen jour således.

Lösningar eller svar till uppgifterna publiceras på webben. Se kursens hemsida.

Rättningsprotokoll anslås i MV:F, plan 2.

— Teoridel —

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. 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)

3. (a) Varför är medelvärdet en konsistent skattning av väntevärdet? (1 p)

   (b) Är s en väntevärdesriktig skattning av σ (svara ja eller nej)? (1 p)

   (c) Skriv upp utan att motivera en formel för approximativ beräkning av ett 95% konfidensintervall för väntevärdet μ, under antagandet att antalet observationer är stort. (1 p)

— Problemdel —

4. 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)

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 $\mathrm{N}(0, 0.1)$-fördelade mätfel. Ungefär hur många oberoende prover behövs minst för att medelvärdets standardavvikelse ska bli mindre än 0.005? (2 p)

6. 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 $\exp(\lambda)$-fördelade livslängder med $\lambda \approx 1$.

   (a) Punktskatta parametern λ. (2 p)

   (b) Undersök, genom att göra en kvantil-plot, om antagandet $\exp(1)$ är rimligt. (2 p)

7. (a) Man skulle skatta en sannolikhet $p = P(A)$ med ett litet antal oberoende försök, och ville utnyttja den kunskap om p som redan fanns. Antag att man gjorde 3 försök och att A inträffade i ett av dessa. Om beta$(3, 7)$ valdes till á priori-täthet för p, vilken är då p:s á posterioritäthet? (2 p)

   (b) En viss typ av olyckor inträffar med intensiteten $\lambda = 2.74 \cdot 10^{-3}$ per dygn. Kostnaden x för en sådan olycka är Paretofördelad med parametrar $\alpha = 3$ och $u = 10$. För undvikande av missförstånd, x:s täthet är alltså

   $$f(x) = \frac{\alpha}{u}\left(\frac{u}{x}\right)^{\alpha+1} \quad \text{för} \quad x > u$$

   Enheten för x är 10 000 kr. Det minsta en olycka kan kosta är alltså 100 000 kr. Bestäm den s.k 10-årskostnaden. (2 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 $\mathrm{N}(\mu, \sigma)$-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 kortfattade lösningar till TMS056 Matematisk statistik V2 den 17/1-07

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 är väldigt bra ($P(L'|B) = 0.02$, $P(L|B') = 0.01$) är alltså ungefär hälften av larmen falska. Vad som krävs för poäng när man inte har tillgång till dosa är att man kan ställa upp sannolikheterna korrekt och medelst papper och penna räkna något så när korrekt.

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

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

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

3. (a) Därför att medelvärdets varians $\to 0$ då antalet observationer $\to \infty$
   (b) Nej
   (c) $\mu = \bar{x} \pm 2s/\sqrt{n}$ eller $\mu = \bar{x} \pm 1.96s/\sqrt{n}$

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

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. (a) ML-skattningen (och tillika momentskattningen) är $\widehat{\lambda} = 1/\bar{x} = 1/0.856 \approx 1.17$

   (b) Man ska plotta det ordnade stickprovet 0.23, 0.27, 0.64, 0.89, 2.25 mot de teoretiska kvantilerna svarande mot sannolikheterna 0.1, 0.3, 0.5, 0.7, 0.9. Dessa fås ur formeln $x_p = -\ln(1-p)$ och är 0.11, 0.36, 0.69, 1.20, 2.30. Sedan kommer den svåra bedömningen att avgöra huruvida punkterna i plotten ligger hyfsat väl utefter den räta linjen $y = x$ eller ej (detta går inte att göra utan träning). I det sistnämnda fallet är kanske exp(1)-antagandet fel.

7. (a) Vi använder att á posterioritätheten är beta$(\alpha + x,\ \beta + n - x)$ om á prioritätheten är beta$(\alpha,\ \beta)$ och $x|n, p \sim \text{bin}(n,\ p)$. Detta visades i häftet om Bayesiansk uppdatering. Således är i föreliggande fall á posterioritätheten beta(4, 9).

   (b) Notera först att

$$P(X > x) = \int_x^\infty f(y)\,dy = \left(\frac{u}{x}\right)^\alpha$$

   Omräknad till år är intensiteten $\lambda \approx 1$, så 10-årskostnaden x_{10} uppfyller ekvationen $P(X > x_{10}) = 0.1 \Rightarrow x_{10} = u \cdot 0.1^{-1/\alpha} \approx 21.5443$. D.v.s, 215 443 kr.

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

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

(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%.

Resultat:

	U	3	4	5	S:a
antal	8	2			10
%					

TMS056 Matematisk statistik V2, 5 p, 23 augusti 2006 eV

Anvisningar

Tentamen består av en teoridel med 3 uppgifter och en problemdel med 5 uppgifter. Teoriuppgifterna ska besvaras först. Inga hjälpmendel är tillåtna på denna del. På problemdelen är räknedosa inkl manual, Beta och någon av de läroböcker i matematisk statistik som används på Chalmers tillåtna hjälpmedel. Det är också tillåtet att använda Tommy Norbergs Formler och tabeller till matematisk statistik på universitet och tekniska högskolor, som kan laddas ner från Norbergs hemsida. Tillåtna hjälpmedel är även de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Något om riskkostnader. Det är tillåtet med anteckningar i rimlig omfattning i läroboken och i nedladdade häften. Inga andra anteckningar är tillåtna.

Kom ihåg att att alla svar skall motiveras såvida ej annat sägs. Ställ upp modell när det behövs och ange de för uträkningarna nödvändiga förutsättningarna. Skriv förståligt (tydligt) och använd i ämnet etablerade beteckningar. Poäng dras i lösningar där detta ej är tillgodosett.

Teoridelen skall lämnas in innan några hjälpmedel får tas fram.

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p. Ev bonuspoäng från inlämningsuppgifterna sänker gränser i motsvarande grad.

Examinator är Tommy Norberg, tel. 772 3528, 0730 79 42 09. Tommy går att nås per mobiltelefon under tentamen.

Lösningar eller svar till uppgifterna publiceras på webben. Se kursens hemsida.

Rättningsprotokoll anslås i MV:F, plan 2.

— Teoridel —

1. Visa att
$$P(C|D) = \frac{P(C)P(D|C)}{P(C)P(D|C) + P(C')P(D|C')} \qquad (4\text{ p})$$

2. Låt U, V vara två oberoende $\mathrm{N}(0,1)$-variabler, och definiera
$$X = \mu_x + \sigma_x U$$
$$Y = \mu_y + \sigma_y \Big(\rho U + \sqrt{1-\rho^2} V\Big)$$
Visa att paret X, Y har en bivariat $\mathrm{N}(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$-fördelning. (Ledning: Linjärkombinationer av oberoende normalvariabler är normalfördelade.) (4 p)

3. (a) I en undersökning av det politiska opinionsläget inför det kommande valet uppgav 46,3% att de avser att rösta röd-grönt medan 44,2% sade sig ha för avsikt att rösta borgerligt. Institutet som gjort undersökningen meddelade att ledningen var statistiskt säkerställd. Förklara med några väl valda meningar för en som inte vet speciellt mycket om matematisk statistik vad uttrycket ”statistiskt säkerställd” betyder i detta sammanhang. (2 p)

(b) Du läser i en (vetenskaplig) tidskrift att en viss teknisk storhet har uppmätts till 25.63 med standardfelet ("standard error") 3.73. Förklara kortfattat vad ordet standardfel betyder här? (2 p)

— Problemdel —

4. I idrottsföreningens lotteri har man 1 000 lotter, varav det är vinst på 50 st. Din vän köper 3 lotter. Numrera dessa 1, 2 och 3. Är sannolikheten att lott nr

   (a) 2 är en vinstlott $>$, $<$ eller $= 1/20$? (1 p)

   (b) 3 är en vinstlott, givet att lotterna 1 och 2 båda är nitar $>$, $<$ eller $= 1/20$? (2 p)

5. Beräkna väntevärde, varians och median för en stokastisk variabel X som är Paretofördelad med täthet

$$f(x) = \frac{\alpha}{x_T}\left(\frac{x_T}{x}\right)^{\alpha+1} \quad \text{för} \quad x > x_T \qquad (4 \text{ p})$$

   Parametern $\alpha > 0$. För vilka α gäller dina resultat?

6. Antag att tiden tills en viss utrustning "går ner" (alltså slutar att fungera) är exponentialfördelad med väntevärde μ. Parametern $\lambda = 1/\mu$ brukar då kallas för intensiteten. I ett test av 7 sådana utrustningar uppmättes funktionstiderna

$$76.6 \quad 4.3 \quad 3.8 \quad 1.7 \quad 30.1 \quad 4.2 \quad 22.5$$

   ML-skatta intensiteten λ. (Obs att i uppgiften ingår att härleda ML-skattningen.) (3 p)

7. I syfte att påvisa att en ny typ av betong har bättre hållfasthetsegenskaper gjordes 10 provgjutningar av standardtypen och lika många med den nya typen, och man mätte böjhållfastheten. Därvid erhölls, i någon enhet, för standardblandningen, $\bar{x} = 13.1$, $s = 2.26$ och för för den nya typen, $\bar{x} = 15.6$, $s = 1.93$. (a) Beräkna ett konfidensintervall för differensen av de två böjhållfastheterna. Konfidensgraden ska vara ca 99%. (b) Ange de för dina beräkningar nödvändiga förutsättningarna. (3+1 p)

8. En del av en gammal industritomt är förorenad av kvicksilver. Hur stor del vet man inte. Inte heller vet man något om kvicksilvrets utbredning. Om man väljer en slumpmässig punkt på tomten på måfå (d.v.s. så att alla punkter har precis samma chans att bli valda) så är sannolikheten att ett jordprov taget i punkten är förorenat lika med θ, där θ betecknar den andel av tomten som är förorenad. Antag nu att man analyserar jordproven tagna i 8 på måfå och oberoende av varandra slumpvis valda punkter. Låt x vara antalet förorenade prover. (a) Vilken fördelning har x om θ är givet? Antag vidare att man på något sätt har bestämt sig för à priorifördelningen beta$(2.25, 3.75)$ för θ, samt att 6 av de 8 proven befanns vara förorenade. (b) Vilken à posteriorifördelning har θ? (2+2 p)

Lycka till!

Svar eller kortfattade lösningar till TMS056 Matematisk statistik V2 den 23/8-06

1. Notera först att $P(C|D) = \frac{P(C\cap D)}{P(D)}$ enl def av betingad sannolikhet. Men enl samma definition är $P(C\cap D) = P(C)P(D|C)$ och ur additiviteten fås dessutom att $P(D) = P(C\cap D)+P(C'\cap D) = P(C)P(D|C) + P(C')P(D|C')$. Sätt in uttrycket för $P(C\cap D)$ och det för $P(D)$ i uttrycket för $P(C|D)$, så erhålles den sökta likheten, som kallas Bayes formel.

2. Av ledningen följer att paret X, Y har en bivariat normalfördelning, så det gäller att verifiera parametrarna i denna. Notera härvid att

$$\begin{aligned}
E[X] &= \mu_x + \sigma_x E[U] = \mu_x \\
E[Y] &= \mu_y + \sigma_y\Big(\rho E[U] + \sqrt{1-\rho^2}E[V]\Big) = \mu_y \\
\mathrm{Var}[X] &= \sigma_x^2 \mathrm{Var}[U] = \sigma_x^2 \\
\mathrm{Var}[Y] &= \sigma_y^2\rho^2\mathrm{Var}[U] + \sigma_y^2(1-\rho^2)\mathrm{Var}[V] = \sigma_y^2\rho^2 + \sigma_y^2(1-\rho^2) = \sigma_y^2
\end{aligned}$$

och

$$\begin{aligned}
\mathrm{Cov}[X,Y] &= E[(X-\mu_x)(Y-\mu_y)] = E\Big[(\sigma_x U)\,\Big(\sigma_y\Big(\rho U + \sqrt{1-\rho^2}V\Big)\Big)\Big] \\
&= E\Big[\sigma_x\sigma_y\rho U^2 + \sigma_x\sigma_y\sqrt{1-\rho^2}UV\Big] = \sigma_x\sigma_y\rho E\big[U^2\big] + \sigma_x\sigma_y\sqrt{1-\rho^2}E[UV] \\
&= \sigma_x\sigma_y\rho
\end{aligned}$$

ty $E\big[U^2\big] = \mathrm{Var}[U] + E[U]^2 = 1$ och $E[UV] = E[U]E[V] = 0$. Vi ser att väntevärden och standardavvikelser är de påstådda, samt att

$$\mathrm{Corr}[X,Y] = \frac{\mathrm{Cov}[X,Y]}{\sqrt{\mathrm{Var}[X]\mathrm{Var}[Y]}} = \frac{\sigma_x\sigma_y\rho}{\sigma_x\sigma_y} = \rho$$

Även korrelationen är alltså den påstådda.

3. (a) Resultatet i undersökningen beror ju på vilka som råkar bli tillfrågade, så det kan ju tänkas att man fick röd-grön majoritet i undersökningen trots att det i verkligheten förhåller sig tvärtom. Att ledningen är statistiskt säkerställd innebär att risken att det inte är en röd-grön majoritet i verkligheten är max ca 5%.

   (b) Standardfelet är här skattningen av standardavvikelsen i den stokastiska variabel man skattat den tekniska storheten med. Rätt får även de som svarar $\sigma/\sqrt{n}$ eller $s/\sqrt{n}$.

4. (a) Man vet ju inget om lotterna 1 och 3, så lott nr 2 måste vara en av 1000 okända varav 50 ger vinst. Sannolikheten att lott nr 2 är en vinstlott är därför $50/1000 = 1/20$. (b) Sannolikheten att lott nr 3 ger vinst givet att lotterna 1 och 2 är nitar är $50/998 = 25/499 > 1/20$, ty när 2 nitar har dragits finns 998 okända lotter kvar varav 50 ger vinst. Lott nr 3 är en av dessa 998.

5. $$\mu = E[X] = \int_{x_T}^{\infty} x\,\frac{\alpha}{x_T}\left(\frac{x_T}{x}\right)^{\alpha+1} dx = \alpha\, x_T^{\alpha}\int_{x_T}^{\infty} x^{-\alpha}\,dx = \frac{\alpha}{\alpha-1}\,x_T \quad \text{för} \quad \alpha > 1$$

$$E[X^2] = \int_{x_T}^{\infty} x^2\,\frac{\alpha}{x_T}\left(\frac{x_T}{x}\right)^{\alpha+1} dx = \alpha\, x_T^{\alpha}\int_{x_T}^{\infty} x^{-\alpha+1}\,dx = \frac{\alpha}{\alpha-2}\,x_T^2 \quad \text{för} \quad \alpha > 2$$

$$\begin{aligned}
\Rightarrow \quad \sigma^2 = \mathrm{Var}[X] = E[X^2] - E[X]^2 &= \frac{\alpha}{\alpha-2}\,x_T^2 - \left(\frac{\alpha}{\alpha-1}\,x_T\right)^2 \\
= \frac{\alpha(\alpha-1)^2 - \alpha^2(\alpha-2)}{(\alpha-1)^2(\alpha-2)}\,x_T^2 &= \frac{x_T^2}{(\alpha-1)^2(\alpha-2)} \quad \text{för} \quad \alpha > 2
\end{aligned}$$

$$0.5 = P(X > m) = \int_m^\infty \frac{\alpha}{x_T}\left(\frac{x_T}{x}\right)^{\alpha+1} dx = \alpha\, x_T^\alpha \int_m^\infty x^{-\alpha-1}\, dx = \left(\frac{x_T}{m}\right)^\alpha$$
$$\Rightarrow \quad 0.5^{1/\alpha} = \frac{x_T}{m} \quad \Rightarrow \quad m = \frac{x_T}{0.5^{1/\alpha}} = x_T\, 2^{1/\alpha}$$

6. Tätheten, uttryckt m.h.a parametern λ, är $f(x) = \lambda e^{-\lambda x}$, $x > 0$. Utfallet $x_1, \ldots, x_n$ av n st oberoende observationer har därför trolighetsfunktionen

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

ML-skattningen är det λ som maximerar trolighetsfunktionen. Logaritmera, sätt derivatan till noll och lös ut λ:

$$\mathcal{L}(\lambda) = \log L(\lambda) = n \log \lambda - \lambda \sum_i x_i$$
$$\mathcal{L}'(\lambda) = \frac{n}{\lambda} - \sum_i x_i$$
$$\mathcal{L}'(\lambda) = 0 \quad \Leftrightarrow \quad \frac{n}{\lambda} = \sum_i x_i \quad \Leftrightarrow \quad \lambda = \frac{n}{\sum_i x_i} = \frac{1}{\bar{x}}$$

ML-skattningen är således $\hat{\lambda} = 1/\bar{x}$. Till sist, $\sum_i x_i = 76.6 + 4.3 + 3.8 + 1.7 + 30.1 + 4.2 + 22.5 = 143.2 \Rightarrow \hat{\lambda} = 7/143.2 = 0.0489$.

7. Rutan på sid 321–322 i D & F talar om att vi ska räkna ut antalet frihetsgrader ν enl

$$\frac{\left(\frac{2.26^2}{10} + \frac{1.93^2}{10}\right)^2}{\frac{\left(\frac{2.26^2}{10}\right)^2}{9} + \frac{\left(\frac{1.93^2}{10}\right)^2}{9}} = 15.56 \quad \Rightarrow \quad \nu = 15$$

Konfidensintervallet ges utav

$$\mu_2 - \mu_1 = 15.6 - 13.1 \pm 2.947\sqrt{\frac{2.26^2}{10} + \frac{1.93^2}{10}} = 2.5 \pm 2.77$$

eftersom $t_{15,0.005} = 2.947$. Så till förutsättningarna: Observationerna i de två mätserierna bör vara oberoende och normalfördelade med väntevärden μ_1, μ_2 och standardavvikelser σ_1, σ_2. Om fördelningen ej är normal, så bör den ändå vara relativt symmetrisk, för då kan man luta sig mot centrala gränsvärdessatsen som säger att medelvärdena är approximativt normalfördelade.

8. (a) Givet sannolikheten θ, är $x \sim \text{bin}(8, \theta)$, eftersom de 8 provsvaren är oberoende och för alla gäller att sannolikheten att provet är förorenat är lika med θ. (b) Med Bayes formel kan man enkelt visa att om $\theta \sim \text{beta}(a, b)$ och $x|\theta \sim \text{bin}(n, \theta)$, så är $\theta|n, x \sim \text{beta}(a+x, b+n-x)$. Således gäller att θ:s à posteriorifördelning är beta$(8.25, 5.75)$.

Resultat:

	U	3	4	5	S:a
antal	8	5	1	0	14
%					

TMS056 Matematisk statistik V2, 5 p, 29 maj 2006 fm V

Anvisningar

Tentamen består av en teoridel med 3 uppgifter och en problemdel med 5 uppgifter. Teoriuppgifterna ska besvaras först. Inga hjälpmendel är tillåtna på denna del. På problemdelen är räknedosa inkl manual, Beta och någon av de läroböcker i matematisk statistik som används på Chalmers tillåtna hjälpmedel. Det är också tillåtet att använda Tommy Norbergs Formler och tabeller till matematisk statistik på universitet och tekniska högskolor, som kan laddas ner från Norbergs hemsida. Tillåtna hjälpmedel är även de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Något om riskkostnader. Det är tillåtet med anteckningar i rimlig omfattning i läroboken och i nedladdade häften. Inga andra anteckningar är tillåtna.

Kom ihåg att att alla svar skall motiveras såvida ej annat sägs. Ställ upp modell när det behövs och ange de för uträkningarna nödvändiga förutsättningarna. Skriv förståligt (tydligt) och använd etablerade beteckningar.

Teoridelen skall lämnas in innan några hjälpmedel får tas fram.

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p. Ev bonuspoäng från inlämningsuppgifterna sänker gränser i motsvarande grad.

Examinator är Tommy Norberg, tel. 772 3528, 0730 79 42 09.

Lösningar eller svar till uppgifterna publiceras på webben. Se kursens hemsida.

Rättningsprotokoll anslås i MV:F, plan 2.

— Teoridel —

1. Formulera de tre sannolikhetsaxiomen och visa att

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \qquad (4 \text{ p})$$

2. Betrakta två slumpmässiga tider T_1 och T_2. Antag att dessa är oberoende och att deras resp felintensiteter är $z_1(t) = \lambda_1$ och $z_2(t) = \lambda_2$. Låt $T = \min(T_1, T_2)$. Bestäm T:s felbenägenhet $z(t)$. (4 p)

3. Antag att du har observerat medelvärdet $\bar{x}$ och standardavvikelsen s i ett stickprov av storlek n på en variabel x, som är normalfördelad med väntevärde μ och okänd standardavvikelse σ. Härled ett konfidensintervall för μ med konfidensgraden $1 - \alpha$. (4 p)

— Problemdel —

4. En s.k binär symmetrisk överföringskanal har felsannolikheten 0.03. (D.v.s en sänd etta uppfattas som nolla i 3% av fallen och en sänd nolla uppfattas som etta i 3% av fallen.) Antag att ettor sänds i 60% av fallen och nollor i 40%. Hur stor är sannolikheten att en mottagen etta är korrekt uppfattad? Ge svaret med 3 korrekta decimaler. (4 p)

5. Låt u vara ett slumptal och låt x vara heltalsdelen av $5u+1$. Beräkna median, väntevärde och varians för den stokastiska variabel X som x är en simulerad observation utav. (3 p)

6. Här är statistiken för april månad de 10 senaste åren för antalet olyckor sådana att minst ett fordon behöver bärgning, utefter en viss typ av vägar: 15, 12, 18, 8, 20, 15, 11, 16, 17, 21. (Räknehjälp: $n = 10$, $\sum x = 153$, $\sum x^2 = 2489$.) Den totala sträckan som statistiken gäller för är 3 807 km. (a) Punkt- och intervallskatta med konfidensen ca 95% intensiteten olyckor per 1 000 km för den aktuella typen av väg. (b) Ange de för dina beräkningar nödvändiga förutsättningarna. (3+1 p)

7. I syfte att påvisa att en ny typ av betong har bättre hållfasthetsegenskaper gjordes 10 provgjutningar av standardtypen och lika många med den nya typen, och man mätte böjhållfastheten. Därvid erhölls, i någon enhet, för standardblandningen, $\bar{x} = 13.1$, $s = 2.26$ och för för den nya typen, $\bar{x} = 15.6$, $s = 1.93$. Undersök om det går att på basis av dessa mätresultat, statistiskt säkerställa att den nya typen har högre böjhållfasthet.(3 p)

8. I en analys av en kontinuerlig belastningsprocess fann man att den i medel översteg larmnivån 10 MPa, $\hat{\lambda} = 0.67$ gånger per dygn. Man ansatte att styrkan x MPa av ett godtyckligt överskridande följer en Paretofördelning med täthet

$$f(x) = \frac{\alpha - 1}{10}\left(\frac{x}{10}\right)^{-\alpha} \quad \text{för} \quad x > 10$$

och ML-skattade α. Därvid erhölls $\hat{\alpha} = 6.39$. Ett larm betraktas som allvarligt om nivån x överstiger tillbudsnivån 15 MPa och katastrofalt om den går över nivån 20 MPa. I genomsnitt kostar ett larm ca 5 kkr i utryckningskostnad, ett tillbud ytterligare ca 20 kkr och en katastrof ytterligare ca 75 kkr. För larm, tillbud och katastrof är alltså totala kostnaden ca 5, 25 resp 100 kkr. Gör nu följande två uppgifter.

   (a) Skatta 50 dygnsnivån. (2 p)

   (b) Skatta riskkostnaden. (2 p)

Svaren skall ges med två korrekta decimaler.

Lycka till!

Svar eller kortfattade lösningar till TMS056 Matematisk statistik V2 den 29/5-06

1. Sannolikhetsaxiomen räknas upp på en OH, så jag nöjer mig här med att bevisa $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Det kan göras på flera sätt. Här är en kortversion: $P(A \cup B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A) = P(A) + P(B) - P(A \cap B)$, ty $P(B) = P((B \setminus A) \cup (A \cap B)) = P(B \setminus A) + P(A \cap B)$. (Axiomen var värda 2 p och 2 p fick de som gjorde ett ordentligt matematiskt bevis av additivitetsregeln. De som endast förklarade regeln heuristiskt, t.ex m.h.a Venn-diagram, fick nöja sig med 1 p.)

2. Vi ser att de två överlevnadsfunktionerna är $R_i(t) = e^{-\lambda_i t}$ för $i = 1, 2$ (konstant felbenägenhet har ju bara exponentialfördelade tider). Notera nu att T:s överlevnadsfunktion är

$$\begin{aligned} R(t) = P(T > t) = P(\min(T_1, T_2) > t) = P(T_1 > t, T_2 > t) = P(T_1 > t)P(T_2 > t) \\ = R_1(t)R_2(t) = e^{-(\lambda_1+\lambda_2)t} \quad \Rightarrow \quad z(t) = \lambda_1 + \lambda_2 \end{aligned}$$

(Detta var tentans svåraste uppgift. Den ligger precis på gränsen till vad man som examinator möjligen kan hoppas att någon skulle kunna klara.)

3. Vi utgår ifrån att $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0,1)$. Härur följer att $\frac{\bar{x}-\mu}{s/\sqrt{n}} \sim t(n-1)$. Välj $t_{\alpha/2}$ så att $P(T > t_{\alpha/2}) = \alpha/2$, där $T \sim t(n-1)$. Då följer att $P\left(-t_{\alpha/2} \le \frac{\bar{x}-\mu}{s/\sqrt{n}} \le t_{\alpha/2}\right) = 1 - \alpha$. Men

$$-t_{\alpha/2} \le \frac{\bar{x}-\mu}{s/\sqrt{n}} \le t_{\alpha/2} \quad \Leftrightarrow \quad \bar{x} - t_{\alpha/2}s/\sqrt{n} \le \mu \le \bar{x} + t_{\alpha/2}s/\sqrt{n}$$

Således gäller att $\mu = \bar{x} \pm t_{\alpha/2}s/\sqrt{n}$ har konfidensen $1 - \alpha$. (Detta är ett standardresonemang, som man bör känna till.)

4. Låt A betyda att en etta har sänts och B att en etta har tagits emot. Enl uppgift är $P(A) = 0.60$, $P(B|A) = 0.97$, $P(A^c) = 0.40$ och $P(B|A^c) = 0.03$. Bayes formel ger att

$$P(A|B) = \frac{0.6 \cdot 0.97}{0.6 \cdot 0.97 + 0.4 \cdot 0.03} = \frac{0.582}{0.582 + 0.012} = \frac{0.582}{0.594} \approx 0.980$$

(Detta är en standardtillämpning på Bayes formel. Det ska en V-ingenjör kunna. Så är det bara.)

5. Notera först att X antar värdena $1, 2, 3, 4, 5$ med lika sannolikhet. Så $m = 3$, $\mu = \frac{1}{5}(1 + 2 + 3 + 4 + 5) = 3$ och $\sigma^2 = \frac{1}{5}\left(1^2 + 2^2 + 3^2 + 4^2 + 5^2\right) - 3^2 = 2$. (För den som kommer ihåg hur man simulerar ett tärningskast, var nog inte detta så svårt. Fast det gäller ju att ha klart för sig hur man beräknar median, väntevärde och varians.)

6. Notera först att den sökta intensiteten är $\lambda = \mu/3.807$, där μ som vanligt betecknar observationernas väntevärde. Vi börjar med att beräkna $\bar{x} = \frac{153}{10} = 15.3$ och $s^2 = \frac{1}{9}\left(2489 - \frac{153^2}{10}\right) = 16.456 = 4.06^2$. Ur t-tabell(2.5% risk, 9 frihetsgrader) får vi $t_c = 2.262$. Vi får nu att

$$\mu = 15.3 \pm 2.262 \cdot 4.06/\sqrt{10} = 15.3 \pm 2.90 \quad \Rightarrow \quad \lambda = 4.02 \pm 0.76$$

Så till förutsättningarna. De är (1) att observationerna är oberoende och likafördelade (detta är faktiskt inte självklart uppfyllt, vissa av de i studien ingående vägarna kan ju t.ex ha upprustats under studietiden; dessutom kan ett ev publicerande av olycksstatistik göra så att en del bilförare kör mer försiktigt) och (2) att observationernas fördelning är såpass symmetrisk att cgs kan tillämpas trots att vi bara har 10 observationer. Vi räknar ju som om observationerna vore normalfördelade, trots att vi inte alls vet om de är det eller ej. En indikation (som inte diskuterats på föreläsning) på att det kan vara ok att basera räkningarna på cgs är att $s/\bar{x} \approx 0.27$. Hade denna kvot varit större än ca 0.4–0.5 hade iallafall jag varit tveksam till cgs giltighet. Varför: Jo, för att fördelningen för variabeln vi mäter då är ganska osymmetrisk, vilket gör att man antagligen behöver fler än 10 observationer för att cgs ska fungera väl. (Beräkningen av konfidensintervallet ska alla klara. Det som möjligen var lite speciellt här var att man skulle dividera med 3.807, för att få rätt enhet. Jag har kommit i kontakt med V-ingenjörer som trott att data varit normalfördelade bara för att $s/\bar{x}$ är litet eller för att man haft många observationer. Gör inte det misstaget.)

7. Vi har $n_1 = 10$, $\bar{x}_1 = 13.1$, $s_1 = 2.26$ och $n_2 = 10$, $\bar{x}_2 = 15.6$, $s_2 = 1.93$ och ska testa $H_0; \mu_2 \leq \mu_1$ mot $H_a : \mu_2 > \mu_1$. Teststatistikans värde är

$$t = \frac{15.6 - 13.1}{\sqrt{\frac{2.26^2}{10} + \frac{1.93^2}{10}}} = 2.660$$

Antalet frihetsgrader ν är heltalsdelen av

$$\frac{\left(\frac{2.26^2}{10} + \frac{1.93^2}{10}\right)^2}{\frac{\left(\frac{2.26^2}{10}\right)^2}{9} + \frac{\left(\frac{1.93^2}{10}\right)^2}{9}} = 17.57$$

d.v.s $\nu = 17$. I t-tabell, 17 frihetsgrader, avläser vi $t_{0.95} = 1.740$ och $t_{0.99} = 2.567$. Av $t > t_{0.99}$ drar vi slutsatsen att H_0 kan förkastas på 1%-nivån. Det går alltså att statistiskt säkerställa att den nya typen har högre böjhållfasthet. (Även denna typ av statistisk analys ska alla klara. Särskilt när man har tillgång till läroboken. Fast det gäller ju att träna innan tentan.)

8. (a) 19.18 MPa (b) 6.05 kkr/dygn (I något av häftena framgår hur man löser den här typen av uppgifter. Att kunna beräkna en återkomstnivå m.h.a POT-tekniken, menar jag är en viktig kunskap för en V-ingenjör. Fast det är nog inte så många som kan det idag. Förhoppningsvis är det fler om några år. Även riskkostnader är det viktigt att kunna beräkna och förstå.)

Resultat:

	U	3	4	5	S:a
antal	38	31	13	1	83
%	45.8	37.3	15.7	1.2	100.0