

TMS056 Matematisk statistik V2, 7.5 hp, 16 januari em V

Anvisningar

Tentamen består av en teoridel med 3 uppgifter och en problemdel med 5 uppgifter. Teoriuppgifterna ska besvaras först. Inga hjälpmedel 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, samt de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Riskkostnader, kostnads-nyttokalkyler och värdet av information (eller 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 alla svar skall motiveras såvida ej annat sägs. Ställ upp modell när det behövs och ange de för uträkningarna nödvändiga förutsättningarna. Skriv förståeligt (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 normalt 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.

— Teoridel —

1. I en studie av risken för utbrott av vattenburen smitta (d.v.s smitta spriden genom dricksvattnet) på Europeanivå redogjordes för 86 kända utbrott under åren 1990–2005. Ange en lämplig fördelning för antalet utbrott under ett år. Ange även approximativa värden på dess parametrar samt hur man beräknar ett estimat av risken (sannolikheten) för minst ett utbrott under året. (3 p)
2. Vad definierar en återkomstnivå? Redogör för hur man räknar ut 100-årsnivån x_{100} i följande POT-modell: larm, betecknade T , sker enl en Poissonprocess med intensitet $\lambda_T = 0.15$ st/år, och styrkan X av dessa är oberoende och Paretofördelade, så att

$$P(X > x|T) = \left(\frac{x_T}{x}\right)^\alpha \quad \text{för } x > x_T$$

där $x_T = 10$ och $\alpha = 7.3$. (4 p)

3. Vad menas med ett stickprov på en stokastisk variabel X ? Hur definieras stickprovets varians? Visa att denna väntevärdesriktigt skattar X 's varians. (4 p)

— Problemdel —

4. Härled en formel för simulering av observationer ifrån Paretofördelningen. (3 p)
5. I en urvalsundersökning svarade 2231 ja på en viss fråga. Urvalet bestod av 10 000 individer slumpmässigt valda i en väldigt stor population. Punkt- och intervallskatta proportionen individer som skulle svarat ja på frågan i hela populationen. Konfidensgrad ska vara ca 95%. (4 p)
6. Punkt- och intervallskatta med konfidensgraden 0.99 väntevärdet μ om stickprovet består av följande 6 normalfördelade observationer

3.46 6.21 5.62 4.05 2.63 4.98 (4 p)

7. I en undersökning av förorenad mark indelades marken i 7 lika stora delområden. Två av dessa valdes på måfå. I varje gjordes 5 mätningar av koncentrationen arsenik i slumpvis utvalda positioner. Efter logaritmering av data (vilket är vanligt att man gör då det är koncentrationer som mäts) erhöles för det ena delområdet: $\bar{x} = 3.6$, $s_x^2 = 2.25$; och för det andra: $\bar{y} = 3.0$, $s_y^2 = 1.75$. Finns det någon anledning att misstänka att koncentrationen arsenik skiljer sig åt i de två valda delområdena? (Antag normalfördelning.)(4 p)
8. För en potentiellt förorenad tomt gäller att samhällets kostnad är ca 10 Mkr om man ej sanerar den trots att den är förorenad. Det kostar ca 1 Mkr att rena den. Man bedömer att sannolikheten $P(F)$ att tomten verkligen är förorenad är ca 0.3. (a) Gör en kostnads-nyttokalkyl för att se om det är lönsamt att sanera tomten. Man tar nu ett stickprov av jordhalter för att om möjligt försäkra sig om huruvida tomten är förorenad eller ej. Låt D beteckna händelsen att förorening detekteras. Detta behöver emellertid inte innebära att tomten verkligen är förorenad, för som bekant är statistiska analyser behäftade med fel. Här ska vi anta att $P(D|F) = P(D'|F') = 0.9$. Antag att förorening ej detekterades. (b) Gör om kostnads-nyttokalkylen för att se om det nu är lönsamt att sanera tomten. (4 p)

Lycka till med uppgifterna!

Svar eller lösningar till vissa uppgifter den 16/1-08

1. Poissonfördelning brukar fungera bra i den här typen av sammanhang. Intensiteten är $\lambda \approx 86/16 = 43/8 = 5.375$. Sannolikheten för minst ett utbrott beräknas ur $\sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1 - e^{-\lambda}$
2. En återkomstnivå, t.ex för tiden 100 år, är den nivå som i genomsnitt uppnås en gång på 100 år. En gång på $1/\lambda_T \approx 6$ år kommer processen upp till nivån $x_T = 10$. En gång på 100 år inträffar $X > x_{100}$ om $\lambda_T P(X > x_{100}|T) = 1/100$. Stoppa in att $P(X > x_{100}|T) = (x_T/x_{100})^\alpha$ och lös ut x_{100} ur ekvationen.
3. Ett stickprov på X är n oberoende stokastiska variabler $X_1, \dots, X_n$ — alla med samma fördelning som X . Se i kurslitteraturen hur stickprovets varians definieras och hur beviset av att denna väntevärdesriktigt skattar X :s varians genomförs.
4. Enl känd sats är det bara att lösa x ur $P(X > x) = u$, där u är ett slumpstal. I uppgift 2 ses att $P(X > x) = (x_T/x)^\alpha$. Svaret blir $x = x_T/u^{1/\alpha}$
5. $p = 0.223 \pm 1.96\sqrt{0.2231(1 - 0.2231)/10000} = 0.223 \pm 0.008 = (0.215, 0.231)$
6. $\mu = 4.492 \pm 4.032 * 1.3557/\sqrt{6} = 4.492 \pm 2.232 = (2.260, 6.724)$
7. Uppgiften består i att göra ett konfidensintervall för differensen $\mu_x - \mu_y$ och dra rätt slutsats beroende på om 0 tillhör konfidensintervallet eller ej. Då inget är sagt om konfidensgraden väljer vi 0.95. Formeln i rutan på sida 303 ger $df = 7.88$ som avrundas ner till 7. Ur t -tabell avläses $t_{0.025} = 2.365$ och konfidensintervallet blir således $\mu_x - \mu_y = 3.6 - 3.0 \pm 2.365\sqrt{2.25/5 + 1.75/5} = 0.6 \pm 2.1$. Origo ligger alltså väl inne i intervallet, så det finns inget som tyder på olika väntevärden på logskalan. Data ger oss således ingen anledning till att tro att väntevärdena skiljer sig åt.
8. (a) Vi ser direkt att saneringsalternativet är lönsamt (ty $10 \cdot 0.3 = 3 > 1$). (b) Bayes formel ger $P(F|D') = 0.3 \cdot 0.1 / (0.3 \cdot 0.1 + 0.7 \cdot 0.9) = 0.03/0.66 \approx 0.046$. Nu inses att sanering ej är lönsamt, eftersom $0.046 \cdot 10 = 0.46 \not> 1$.

TMS056 Matematisk statistik V2, 5 p, 22 augusti 2007 em VV

Anvisningar

Tentamen består av en teoridel med 3 uppgifter och en problemdel med 5 uppgifter. Teoriuppgifterna ska besvaras först. Inga hjälpmedel 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, samt de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Riskkostnader, kostnads-nyttokalkyler och värdet av information (eller 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 alla svar skall motiveras såvida ej annat sägs. Ställ upp modell när det behövs och ange de för uträkningarna nödvändiga förutsättningarna. Skriv förståeligt (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 normalt 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.

— Teoridel —

1. I en förorenad mark-tillämpning bedömer man att sannolikheten att marken är förorenad är ca $2/3$. Mätapparaturen som används för att analysera jordprover kan tyvärr ge felaktigt utslag. Känt är att den detekterar förorening i ca 8 fall av 10 om provet verkligen är förorenat och i ca 3 fall av 10 om provet ej är det. Anta att du har ett prov i vilket förorening har detekterats. Hur stor är då sannolikheten att provet verkligen är förorenat. (4 p)
2. På omstående sida ser du ett histogram över längder kallade TRC61z.
 - (a) Skatta med hjälp av histogrammet sannolikheten p att längden är ≥ 4 . (2 p)
 - (b) Skatta med hjälp av histogrammet förväntad längd μ . (2 p)
3. En i bayesianska sammanhang känd täthet är betatätheten, som definieras av uttrycket

$$f(\theta|\alpha, \beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \quad \text{för } 0 < \theta < 1$$

och $\alpha, \beta > 0$. Härled ett uttryck för betatäthetens mod. (4 p)

— Problemdel —

4. I ett försök kan tre händelser A, B, C inträffa. Känt är att

$$\begin{aligned} P(A) &= P(B) = P(C) = 0.40 \\ P(A \cap B) &= P(A \cap C) = P(B \cap C) = 0.15 \\ P(A \cap B \cap C) &= 0.10 \end{aligned}$$

Låt x vara antalet händelser som inträffar. Beräkna x 's massfunktion (pmf). Ett korrekt ritat Venn-diagram duger gott som argument för beräkningarna. (4 p)

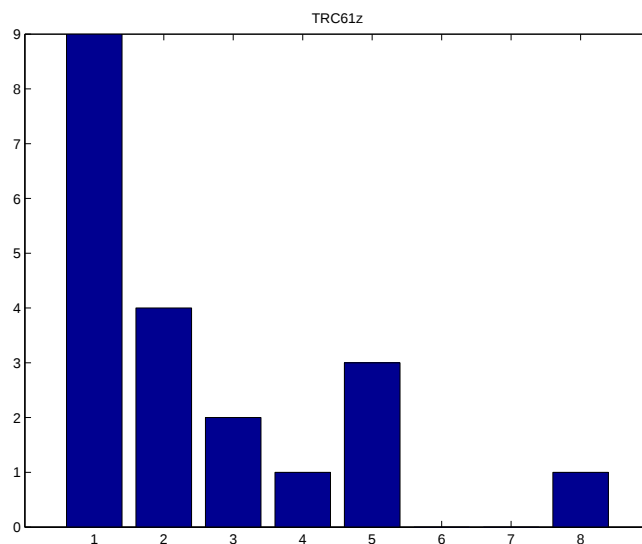
5. Du singlar en slant 100 gånger. Ungefär hur stor är sannolikheten att du erhåller mellan 40 och 60 klave. Räkna approximativt och använd gärna en känd tumregel. (3 p)

6. Beräkna ett 95%-konfidsintervall för din skattning av förväntad längd i uppgift 2. Du får lov att räkna som om data vore normalfördelade, fast de som du ser av histogrammet inte är det. Vilken sats är det som gör detta möjligt? (3+1=4 p)
7. I en risktillämpning behöver man veta den så kallade 50-årsnivån x_{50} . Man har data för totalt 10 år och med hjälp av denna skattade man parametern α i POT-modellen

$$P(X > x|T) = \left(\frac{5}{x}\right)^\alpha \quad \text{för } x > 5$$

samt intensiteten λ_T för överskridande av nivån 5. Därvid erhöles $\hat{\alpha} = 7.0$ och $\hat{\lambda}_T = 4.3$ st/år. Använd dessa värden till att skatta x_{50} . (3 p)

8. Man mätte en hållfasthet hos ett förstärkt material och erhöles i $n_1 = 5$ mätningar medelvärde $\bar{x}_1 = 15.3$ och standardavvikelsen $s_1 = 3.37$. Man gjorde sedan $n_2 = 4$ kontrollmätningar på motsv icke-förstärkt material och erhöles då medelvärde $\bar{x}_2 = 13.5$ och standardavvikelsen $s_2 = 2.95$. Stämmer det att dessa mätningar visar med en rimlig statistisk signifikans att det förstärkta materialet har en högre hållfasthet än det icke-förstärkta? Beräkna ett 95% konfidsintervall för differensen i hållfasthet och svara därefter på frågan. Anta normalfördelade observationer. (4 p)



Lycka till med uppgifterna!

1. Bayes formel ger

$$P(C|D) = \frac{P(C)P(D|C)}{P(D)} = \frac{P(C)P(D|C)}{P(C)P(D|C) + P(C')P(D|C')} \\ \approx \frac{\frac{2}{3} \cdot \frac{8}{10}}{\frac{2}{3} \cdot \frac{8}{10} + \frac{1}{3} \cdot \frac{3}{10}} = \frac{16}{16 + 3} = \frac{16}{19}$$

2. Vi har totalt $n = 20$ observationer. Av dessa är $f = 5 \geq 4$, så skattningen av den sökta sannolikheten p är $\hat{p} = 5/20 = 0.25$. Man räknar sedan ut att $\sum x = 9 \cdot 1 + 4 \cdot 2 + \dots = 50$ så medelvärdet är $\bar{x} = 50/20 = 2.5$ och $\hat{\mu} = 2.50$.
3. Moden är det θ som maximerar tätheten och då räcker det att känna till alla faktorer som innehåller θ . Vi får att den logaritmerade tätheten är

$$(\alpha - 1) \ln \theta + (\beta - 1) \ln(1 - \theta)$$

Sätts derivatan av detta till 0 erhålls ekvationen

$$\frac{\alpha - 1}{\theta} - \frac{\beta - 1}{1 - \theta} = 0$$

som löses av

$$\theta = \frac{\alpha - 1}{\alpha + \beta - 2}$$

Här bör man försäkra sig om att detta är en maximipunkt, vilket garanterat är fallet om $\alpha, \beta > 1$.

4. Vi får "the hard way" att

$$p(3) = P(A \cap B \cap C) = 0.10 \\ p(2) = P(A \cap B \cap C') + P(A \cap B' \cap C) + P(A' \cap B \cap C) = 3P(A \cap B \cap C') \\ = 3(P(A \cap B) - P(A \cap B \cap C)) = 3(0.15 - 0.10) = 0.15 \\ p(1) = P(A \cap B' \cap C') + P(A' \cap B \cap C') + P(A' \cap B' \cap C) = 3P(A \cap B' \cap C') \\ = 3P(A \cap (B \cup C)') = 3(P(A) - P(A \cap (B \cup C))) = 3(P(A) - P(A \cap B \cup A \cap C)) \\ = 3(P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C)) \\ = 3(0.40 - 0.15 - 0.15 + 0.10) = 3 \cdot 0.20 = 0.60 \\ p(0) = P(A' \cap B' \cap C') = 1 - P(A \cup B \cup C) \\ = 1 - P(A) - P(B) - P(C) + P(A \cap B) + P(A \cap C) + P(B \cap C) - P(A \cap B \cap C) \\ = 1 - 3 \cdot 0.40 + 3 \cdot 0.15 - 0.10 = 0.15$$

Här är, som synes $p(1)$ och $p(0)$ svåra, men tar man den ena, så tar man den andre m.h.a sambandet $p(0) + p(1) + p(2) + p(3) = 1$. Lättare är nog att successivt innifrån och ut beräkna sannolikheterna för de olika områdena i ett Venn-diagram med tre händelser A, B, C . Då noteras först att $P(A \cap B \cap C) = 0.10$, sedan att det finns tre händelser av typen $P(A \cap B \cap C') = 0.15 - 0.10 = 0.05$, att det finns tre händelser av typen $P(A \cap B' \cap C') = 0.40 - 2 \cdot 0.05 - 0.10 = 0.20$. Sedan ser man lätt att

$$p(3) = 0.10 \quad p(2) = 3 \cdot 0.05 = 0.15 \quad p(1) = 3 \cdot 0.20 = 0.60$$

samt

$$p(0) = 1 - (0.10 + 0.15 + 0.60) = 1 - 0.85 = 0.15$$

5. Antalet klave är $(100, 0.5)$ -fördelat. denna fördelning har väntevärdet $\mu = 100 \cdot 0.5 = 50$ och variansen $\sigma^2 = 100 \cdot 0.5 \cdot 0.5 = 25$, så $\sigma = 5$. Enl cgs är $\text{bin}(100, 0.5) \approx N(50, 5)$, så den sökta sannolikheten är ungefär $P(-2 \leq Z \leq 2) \approx 0.95$, där Z betecknar en $N(0, 1)$ -variabel. Ett plus till dem som känner till tumregeln och tämligen direkt svarar ca 0.95.

6. Vi har redan konstaterat att $n = 20$, $\sum x = 50$ och räknar nu ut att $\sum x^2 = 9 \cdot 1^2 + 4 \cdot 2^2 + \dots = 198$, så $s^2 = (198 - 50^2/20)/19 = 3.8420 = 1.96^2$. Ur $t(19)$ -tabell fås $t_{0.025} = 2.093$, så konfidenstervallat är $\mu = \bar{x} \pm t_{0.025} s/\sqrt{n} = 2.50 \pm 0.92$. Det är centrala gränsvärdesatsen som gör det möjligt att beräkna konfidenstervall på detta sätt, men observera att egentligen bör man ha ca 30 observationer för att approximationen i resultatet skall bli bra.

7. Vi ska lösa x_{50} ur

$$\lambda_T P(X > x_{50}) = \frac{1}{50} \Leftrightarrow \left(\frac{5}{x_{50}}\right)^\alpha = \frac{1}{50\lambda_T} \Leftrightarrow \frac{5}{x_{50}} = \left(\frac{1}{50\lambda_T}\right)^{1/\alpha}$$

så

$$x_{50} = 5 (50\lambda_T)^{1/\alpha}$$

och skattningen av 50-årsnivån är alltså

$$\hat{x}_{50} = 5 (50 \cdot 4.3)^{1/7} = 10.77$$

8. Notera att $se_1 = 3.37/\sqrt{5} = 1.5071$, $se_2 = 2.95/\sqrt{4} = 1.475$. Detta ger

$$df = \frac{(1.5071^2 + 1.475^2)^2}{\frac{1.5071^4}{4} + \frac{1.475^4}{3}} = \frac{19.7759}{1.2898 + 1.5778} = 6.90$$

så antalet frihetsgrader i skattningen $\sqrt{1.5071^2 + 1.475^2} = 2.109$ av standardfelet är $\nu = 6$. I t -tabell ser vi att $t_{0.025}(5) = 2.447$, så det sökta 95%-konfidenstervallat är

$$15.3 - 13.5 \pm 2.447 \cdot 2.109 = 1.8 \pm 5.16$$

eller

$$(-3.36, 6.96)$$

Vi ser $0 \in$ intervallet. Det stämmer alltså inte att det förstärkta materialet med en rimlig statistisk signifikans har en annan hållfasthet än det icke-förstärkta. Annorlunda uttryckt, i denna undersökning har man ej kunnat säkerställa statistiskt att det förstärkta materialet har annan hållfasthet än det icke-förstärkta. Men man ville ju visa att det förstärkta materialet har högre hållfasthet än det icke-förstärkta. Rätt är att då testa $H_0 : \mu_1 = \mu_2$ mot $H_1 : \mu_1 > \mu_2$ Teststatistikan, som har värdet

$$t = \frac{15.3 - 13.5}{2.109} = 0.853$$

är inte större än $t_{0.10}(6) = 1.440$. Så inte ens på 10%-nivån kan nollhypotesen förkastas.

TMS056 Matematisk statistik V2, 5 p, 1 juni 2007 fm V

Anvisningar

Tentamen består av en teoridel med 3 uppgifter och en problemdel med 5 uppgifter. Teoriuppgifterna ska besvaras först. Inga hjälpmedel 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, samt de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Riskkostnader, kostnads-nyttokalkyler och värdet av information (eller 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 alla svar skall motiveras såvida ej annat sägs. Ställ upp modell när det behövs och ange de för uträkningarna nödvändiga förutsättningarna. Skriv förståeligt (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.

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

— Teoridel —

1. (a) Låt X vara en kontinuerligt fördelad stokastisk variabel med fördelningsfunktion $F(x)$ och täthet $f(x) = F'(x)$, låt U vara likformigt fördelad på det öppna enhetsintervallet $(0, 1)$ och definiera

$$\tilde{X} = F^{-1}(U)$$

där $F^{-1}(u)$ är inversen till $F(x)$. Visa att $\tilde{X}$ och X har samma fördelning. (3 p)

- (b) Paretofördelningens täthet är

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

Här är $x_T > 0$ och $\alpha > -1$. Härled en algoritm för simulering av observationer med denna täthet. (2 p)

2. I en vetenskaplig tidningsartikel redogjordes för en jämförelse av hållfasthet mellan en standardkvalitet av betong och samma kvalitet men med vissa elastiska tillsatser inblandade.

- (a) Vid test av likhet mot alternativet att den nya blandningen är bättre erhöles P -värdet 7.2%. Vad innebär detta? (2 p)
- (b) Ett 95% konfidensintervall för den nya blandningens hållfasthet är 8.17 ± 0.27 i någon enhet. Vad innebär detta? (2 p)

I båda fallen vill jag att du föreställer dig att du förklarar för en kompis som inte läst någon grundläggande kurs i matematisk statistik. Och var kortfattad.

3. Efter raset i Munkedal har Vägverket påbörjat en genomgång av rasrisker utmed våra viktigaste vägar. För en relativt kort, men intensivt trafikerad vägsträcka i utkanten av en större stad, har man bedömt att rasrisken på årsbasis är ca 1 på 50, vilket anses väl högt. Att återuppbygga en raserad väg beräknas kosta 420 Mkr, alltså 420 millioner kr. Utöver detta beräknas samhällets och alla andra inblandades kostnader vid ett ras till ca 480 Mkr. Om det kostar ca x Mkr att förstärka vägen så att rasrisken reduceras till 1 på 1000, hur stor får då x maximalt vara för att det i kostnads-nyttomening ska vara samhällsekonomiskt lönsamt att förstärka vägen? (3 p)

— Problemdel —

4. Ett ofta rimligt antagande är att om A är en ovanlig händelse (som alltså inträffar då och då, men relativt sällan), så gäller att antalet gånger som A inträffar under en tidsperiod är $\text{Poi}(\lambda)$ -fördelat. Anta att man under de fem senaste åren sammanlagt har observerat 18 olyckor utmed en viss vägsträcka.

(a) Punktskatta λ med trolighetsmetoden. (2 p)

(b) Punktskatta sannolikheten p att det under nästa år inträffar minst 2 olyckor. (2 p)

5. De stokastiska variablerna U, V är kontinuerligt fördelade på enhetskvadraten $(0, 1) \times (0, 1)$ med täthet

$$f(u, v) = \begin{cases} c & \text{då } 0 < u < 0.8, 0 < v < 0.6 \\ c & \text{då } 0.8 \leq u < 1, 0.6 \leq v < 1 \\ 0 & \text{för övrigt} \end{cases}$$

(a) Beräkna $P(U < 0.8)$ och $P(U < 0.8, V > 0.6)$. (2 p)

(b) Är U, V oberoende? (2 p)

I (b) skall argumenteringen vara matematisk.

6. Antag att du har observerat ett stickprov av storlek 16 på en stokastisk variabel X , som är normalfördelad med väntevärde 2 och standardavvikelse 4.

(a) Hur stor är sannolikheten att stickprovets medelvärde är positivt? (2 p)

Antag att du erhöll $\sum x = 5.92$, $\sum x^2 = 240.105$.

(b) Är det då rimligt att påstå att data tyder på att väntevärdet är > 0 ? (2 p)

I (a) är det tillåtet att basera svaret på en välkänd tumregel.

7. Av 160 långtidstestade slumpvis utvalda säkringar bröt 16 strömmen, trots att belastningen ej översteg säkringens märkström som var 10 Ampere. Punkt- och intervallskatta proportionen säkringar i den aktuella populationen som under likartad belastning felaktigt bryter strömmen. (3 p)

8. En potentiellt förorenad tomt delades in i 2 delområden, kallade A och B . I båda gjordes ett antal mätningar av koncentrationen av det förorenande ämnet. Därvid erhöles:

Område A : 3.20 5.38 5.63 5.42 6.58

Område B : 5.20 5.99 2.40

Beräkna ett 99% konfidensintervall för skillnaden av koncentrationen i de två områdena. Antag oberoende normalfördelade observationer. (3 p)

Lycka till!

3. Riskminskningen som investeringen ger upphov till är

$$R_0 - R_1 = \left(\frac{1}{50} - \frac{1}{1000} \right) (420 + 480) = 17.1 \text{ Mkr}$$

så det är vad förstärkningen av vägen får kosta. Formulering av uppgiften blev något olycklig. Jag borde skrivit "rasrisken är ca 1 på 50" istället för "rasrisken på årsbasis är ca 1 på 50".

4. (a) Vi vet att under
- $n = 5$
- år observerades
- $\sum_i x_i = 18$
- olyckor och att antalet olyckor under ett givet år är
- $\text{Poi}(\lambda)$
- , som har massfunktionen

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad \text{för } x = 0, 1, \dots$$

Så trolighetsfunktion är

$$L(\lambda) = \prod_i f(x_i) = e^{-n\lambda} \lambda^{\sum_i x_i} \frac{1}{x_1!} \dots \frac{1}{x_n!}$$

$$\Rightarrow \mathcal{L}(\lambda) = -n\lambda + \left(\sum_i x_i \right) \ln \lambda - \sum_i \ln x_i! \Rightarrow \mathcal{L}'(\lambda) = -n + \frac{\sum_i x_i}{\lambda}$$

och vi ser att $\mathcal{L}'(\lambda) = 0$ löses av $\lambda = \frac{1}{n} \sum_i x_i = \bar{x}$. Trolighetsskattningen av olycksbenägenheten λ är följaktligen $\hat{\lambda} = 18/5 = 3.6$ st per år. (1p har de fått som utan motivering räknat ut att $\hat{\lambda} = 18/5 = 3.6$. Ytterligare 1p har de fått som nöjaktigt härlett trolighetsskattningen.)

- (b) Trolighetsskattningen av sannolikheten

$$p = P(X \geq 2) = 1 - P(X < 2) = 1 - \left(e^{-\lambda} \frac{\lambda^0}{0!} + e^{-\lambda} \frac{\lambda^1}{1!} \right) = 1 - e^{-\lambda} (1 + \lambda)$$

$$\text{är } \hat{p} = 1 - e^{-\hat{\lambda}} (1 + \hat{\lambda}) = 1 - 4.6 e^{-3.6} = 0.874.$$

5. (a) Rita figur! $\int_0^1 \int_0^1 f(u, v) du dv = 1$ ger att $0.8 \cdot 0.6c + 0.2 \cdot 0.4c = 1 \Rightarrow c = 1/0.56 = 25/14 \approx 1.786$. Vidare är $P(U < 0.8, V > 0.6) = 0$ och $P(U < 0.8) = 0.8 \cdot 0.6/0.56 = 0.48/0.56 = 6/7 \approx 0.857$.
- (b) Man ser även att $P(V > 0.6) = 0.2 \cdot 0.4/0.56 = 1/7 \approx 0.143$, och ur $P(U < 0.8, V > 0.6) \neq P(U < 0.8) P(V > 0.6)$ dras slutsatsen att U, V är beroende.
6. (a) $\bar{x}$ är normalfördelad med väntevärde 2 och standardavvikelse $4/\sqrt{16} = 1$, så $P(\bar{x} > 0) = 1 - \Phi((0 - 2)/1) = 1 - \Phi(-2) \approx 1 - 0.025 = 0.975$ lite grovt räknat enl en känd tumregel.
- (b) Vi får att $\bar{x} = 5.92/16 = 0.370$ och $s = \sqrt{(240 \cdot 105 - 5.92^2/16)/15} = \sqrt{15.8610} = 3.9826$. I syfte att svara på frågan testar jag $H_0: \mu \leq 0$ mot $H_a: \mu > 0$ och konstaterar att

$$t = \frac{\bar{x} - 0}{s/\sqrt{n}} = \frac{0.370}{3.9826/\sqrt{16}} = 0.372$$

inte är särskilt signifikant (P -värdet är 0.357). I tabell ser vi ju att $t_{0.10}(15) = 1.341$ och $0.372 \not\geq 1.341$. Så testet förkastar inte ens på nivån 10% och det finns därför ingen anledning att påstå att data tyder på att $\mu > 0$. Ett alternativt sätt att lösa uppgiften är att beräkna vad som ibland kallas LCL95: $0.370 - 1.753 \cdot 3.9826/\sqrt{16} = -1.375$ och ur faktumet $\text{LCL95} < 0$ dra slutsatsen att data ej tyder på att $\mu > 0$.

7. Naturligt är att välja konfidensgraden 0.95 om ej annat sägs och vi får

$$p = \frac{16}{160} \pm 1.96 \cdot \sqrt{\frac{\frac{16}{160} \cdot \frac{160-16}{160}}{160}}$$

$$= 0.10 \pm 0.046$$

$$\in (0.054, 0.146)$$

Det är inte fel (kanske t.o.m bättre) att räkna med 2 istället för 1.96 och erhålla (den statistiska) felmarginalen ± 0.047 .

8. Data medför

$$\begin{array}{llll} n_A = 5 & \bar{x}_A = 5.242 & s_A = 1.2411 & s_A/\sqrt{n_A} = 0.555 \\ n_B = 3 & \bar{x}_B = 4.530 & s_B = 1.8865 & s_B/\sqrt{n_B} = 1.0891 \end{array}$$

Formeln för antalet frihetsgrader på sida 322 i D&F ger $df = 3.07$, så vi räknar med $\nu = 3$ frihetsgrader. Ur tabell fås att $t_{0.005}(3) = 5.841$, så konfidensintervallet för differensen av väntevärdena är

$$\mu_A - \mu_B = 0.712 \pm 5.841\sqrt{0.555^2 + 1.0891^2} = 0.712 \pm 7.140$$