TMS055 Matematisk statistik V2, 11 januari 2006 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. Tillåtna hjälpmedel är även de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Något om riskkostnader samt föreläsningsanteckningarna som kan laddas ner ifrån kurshemsidan. Det är tillåtet med rimligt omfattande anteckningar 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 och använd etablerad notation.

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 till uppgifterna publiceras på webben. Se kursens hemsida. Rättningsprotokoll anslås i nya MV-huset, Chalmers tvärgata 3.

— Teoridel —

1. Svaren på de fyra frågorna i denna uppgift skall ej motiveras.

   (a) I en studie av brinntider för ved av tall resp ek, som diskuterades i en tidskrift om luftföroreningar, rapporterades inledningsvis att brinntiden för tall överstiger den för ek. Lite längre in i uppsatsen fick man veta att data i testet som ledde till detta konstaterande hade P-värdet 0.032. Hur stor är den maximala risken att påståendet är felaktigt? Välj mellan alternativen 10%, 5%, 1% och 0.1%. (1 p)

   (b) Är förväntat resultat av ett kast med en symmetrisk sex-sidig tärning 3? (1 p)

   (c) Låt $X \sim \mathrm{N}(\mu, \sigma)$. Är $E[X] = \mu$? (1 p)

2. (a) Visa att den 5:e percentilen i $\exp(\lambda)$-fördelningen är $x_{0.05} = -(\ln 0.95)/\lambda$. (2 p)

   (b) Visa att felintensiteten för en $\exp(\lambda)$-fördelad stokastisk variabel är lika med λ. (2 p)

3. Antag att du har 5 oberoende observationer av något fenomen. Du vill kolla om det verkar vara rimligt att i den analys som ska göras utgå ifrån att observationerna är normalfördelade. Redogör på högst en A4-sida hur man kan göra en sådan kontroll. (4 p)

— Problemdel —

4. För en brusig binär kommunikationslänk gäller att en sänd etta felaktigt tas emot som nolla i ca 5% av fallen och att en sänd nolla felaktigt tas emot som etta i ca 2.5% av fallen. Man räknar med att proportionen ettor som sänds är ca 50%. Antag att en nolla togs emot. Hur stor är den betingade sannolikheten att det var en nolla som sändes? (4 p)

5. Den s.k generaliserade Paretofördelning har fördelningsfunktionen

$$F_\gamma(x) = 1 - \left(1 + \gamma\frac{x}{\sigma}\right)^{-1/\gamma} \quad \text{för} \;\; x > 0 \text{ sådana att } 1 + \gamma\frac{x}{\sigma} > 0$$

Då $\gamma = 0$ är $F(x) = \lim_{\gamma\to 0} F_\gamma(x) = 1 - e^{-x/\sigma}$ för $x > 0$. Hur kan man göra för att simulera observationer från denna fördelning? Det räcker att svara på fallet $\gamma \neq 0$. (4 p)

6. Under 240 minuter registreras 38057 radioaktiva sönderfall. Räknaren registrerar samtliga sönderfall. Punkt- och intervallskatta sönderfallsintensiteten λ (tidsenhet: minut). Önskad konfidensgrad: 95%. (Tips: Ant sönderfall per tidsenhet är Poi(λ) $\approx$ N(μ, σ), där $\mu = \lambda$ och $\sigma = \sqrt{\lambda}$.)(4 p)

7. I en analys av en kontinuerlig belastningsprocess fann man att den i medel översteg larmnivån $x_T = 10$ MPa, $\hat{\lambda}_T = 0.46$ gånger per dygn. Man ansatte att styrkan x MPa av ett godtyckligt överskridande följer en Paretofördelning med täthet

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

Vid ett tillfälle registrerade man följande 6 överskridanden:

14.60 10.56 10.11 10.67 11.92 12.17

Beräkna trolighetsskattningen av α. (4 p)

8. (Forts av föregående uppgift) Man bedömmer att varje överskridande av nivån 12.5 MPa kostar ca 100 kSEK. Ungefär hur stor är riskkostnaden om $\alpha \approx 7.8$? (3 p)

Lycka till!

Svar eller kortfattade lösningar till TMS055 Matematisk statistik V2 den 11/1-06

1. (a) 5%

   (b) Nej

   (c) Ja

2. (a) $x_{0.05}$ definieras av att sannolikheten för en observation $\leq x_{0.05}$ är lika med 0.05. Men då är sannolikheten för en observation $> x_{0.05}$ lika med 0.95. Således gäller att

$$\int_{x_{0.05}}^{\infty} \lambda e^{-\lambda x} \, dx = 0.95 \quad \Leftrightarrow \quad e^{-\lambda x_{0.05}} = 0.95 \quad \Leftrightarrow \quad \lambda x_{0.05} = -\ln 0.95 \quad \Leftrightarrow \quad x_{0.05} = \frac{-\ln 0.95}{\lambda}$$

   (b) För felintensiteten $z(t)$ gäller

$$z(t) = \frac{f(t)}{1 - F(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

3. Jag skulle göra en s.k "probability plot" eller en QQ-plot ("quantile-quantile plot"). Läs själv i läroboken hur man gör en QQ-plot. Det ingår i kraven för full poäng på uppgiften att man med hyfsad noggranhet beskriver hur en sådan görs. Sedan skulle jag, om jag ej kände mig säker på hur plotten ska tolkas, simulera för att skaffa mig en uppfattning om hur variationerna i QQ-plottar ser ut om modellen är den riktiga. I princip gäller ju att punkterna i QQ-plotten approximativt ska ligga utefter en rät linje. Det svåra är att avgöra om punkterna ligger för bra utmed en rät linje eller för dåligt. Det kan nämligen bli fel åt båda hållen.

4. Låt S och M beteckna det som sänds resp det som tas emot. Vi får reda på att $P(M = 0|S = 1) \approx 0.05$, att $P(M = 1|S = 0) \approx 0.025$ och att $P(S = 1) \approx P(S = 0) \approx 0.50$. Bayes formel ger

$$\begin{aligned} P(S = 0|M = 0) &= \frac{P(S = 0)P(M = 0|S = 0)}{P(S = 0)P(M = 0|S = 0) + P(S = 1)P(M = 0|S = 1)} \\ &\approx \frac{0.50 \cdot (1 - 0.025)}{0.50 \cdot (1 - 0.025) + 0.50 \cdot 0.05} = \frac{0.4875}{0.5125} \approx 0.951 \end{aligned}$$

5. Enl simulerarnas huvudsats (inversmetoden) gäller att

$$u = \left(1 + \gamma \frac{x}{\sigma}\right)^{-1/\gamma} \quad \Leftrightarrow \quad u^{-\gamma} = 1 + \gamma \frac{x}{\sigma} \quad \Leftrightarrow \quad \frac{u^{-\gamma} - 1}{\gamma} = \frac{x}{\sigma} \quad \Leftrightarrow \quad x = \sigma \frac{u^{-\gamma} - 1}{\gamma}$$

   Här är u ett slumptal och x den simulerade observationen.

6. Vi har $n = 240$ observationer och att $\sum x = 38057$, vilket medför att $\bar{x} = 38057/240 = 158.57$. Härur fås att $\lambda = 158.6 \pm 1.96 \cdot \sigma/\sqrt{240}$. Svårigheten är att vi inte känner σ. Men enl ledningen är $\sigma = \sqrt{\lambda} \approx \sqrt{158.57} \approx 12.59$. Således gäller $\lambda = 158.6 \pm 1.96 \cdot 12.59/\sqrt{240} = 158.6 \pm 1.6$.

7. Genom att argumentera som i Paretohäftet (vilket krävs, men som jag inte vill repetera här) inses att trolighetsskattningen av α ska beräknas ur data $x_1, \ldots, x_n$ enl formeln

$$\hat{\alpha} = 1 + \frac{n}{\sum_i \ln(x_i/x_T)}$$

   Insättning av $x_T = 10$ och $\sum_i \ln(x_i/x_T) = 0.881$ ger $\hat{\alpha} = 7.81$.

8. Riskkostnaden är $\lambda_T \cdot P(X > 12.5) \cdot 100$, där $P(X > 12.5) = \int_{12.5}^{\infty} f(y) \, dy = \left(\frac{12.5}{x_T}\right)^{1-\alpha} \approx 1.25^{-6.8} = 0.2193$. Vi ser att riskkostnaden är ungefär $0.46 \cdot 0.2193 \cdot 100 = 10.1$ kSEK.

TMS055 Matematisk statistik V2, 17 aug e 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. Tillåtna hjälpmedel är även de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Något om riskkostnader samt föreläsningsanteckningarna som kan laddas ner ifrån kurshemsidan. Det är tillåtet med anteckningar 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 och använd etablerad notation.

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 till uppgifterna publiceras på webben. Se kursens hemsida. Tentamensresultat publiceras på anslagstavla i källaren i Matematiskt centrum.

— Teoridel —

1. Svaren på de fyra frågorna i denna uppgift skall ej motiveras.
   (a) Antag att $P(A) = 0.63$, och att $P(B) = 0.39$. Kan A och B vara ömsesidigt uteslutande? (1 p)
   (b) Låt X vara heltalsdelen av $6 * U + 1$, där $U \sim \mathrm{U}(0, 1)$. Vilken fördelning har X? (1 p)
   (c) Stopp i en industriell process inträffar i snitt 8 gånger per år. Man räknar med att stopp i snitt orsakar en merkostnad om ca 1.25 Mkr (miljoner kr). Hur stor är riskkostnaden? (1 p)
   (d) I en undersökning av tiden t_0 tills korrosionen gör en viss plåtkvalitet oanvändbar erhölls med den ungefärliga konfidensen 95% att $t_0 \geq 7.3$ år. Hur stor är risken att det i verkligheten gäller att $t_0 < 7.3$ år? (1 p)

2. Vid mätning av koncentrationen av en viss förorening erhölls i μg/dm^3:

   $$11.83,\ 9.68,\ 24.39,\ 6.35,\ 15.68$$

   Genom att titta på erforderliga kvantilplottar, försök att avgöra om dessa observationer kan tänkas komma ifrån en (a) normalfördelning eller (b) lognormalfördelning. (Hjälp: $\Phi(.25) = 0.6$, $\Phi(.52) = 0.7$, $\Phi(.84) = 0.8$, $\Phi(1.28) = 0.9$.) (4 p)

3. Härled uttrycket $\hat{\lambda} = 1/\bar{x}$ för trolighetsskattningen av λ då data är oberoende observationer av en $\exp(\lambda)$-fördelad tid. (4 p)

— Problemdel —

4. Mätning av miljöförorening. Antag att man i en mätning får ett resultat som indikerar att marken är förorenad. Kalla denna händelse för D (positiv detektion). P.g.a mätfel kan man trots detta inte vara säker på marken verkligen är förorenad. Låt C beteckna händelsen att marken är

förorenad; C är alltså den händelse vi vill detektera. Antag att sannolikheten för "a false positive", $P(D|C')$, är 0.19, medan sannolikheten för "a true positive", $P(D|C)$, är 0.68. Vad blir $P(C|D)$ om $P(C) = 0.36$? (4 p)

5. I en simuleringsstudie gjorde man 10 000 oberoende simuleringar av ett försök och fann att en viss sällsynt händelse inträffade 73 gånger. Skatta väntevärdesriktigt sannolikheten p för händelsen ifråga och beräkna ett uppåt begränsat konfidensintervall för p med konfidensgrad ca 99%.(3 p)

6. Antag att olyckor sker enligt en Poissonprocess med intensitet $\lambda = 0.04$ olyckor per dag. (a) Bestäm väntevärdet av antalet olyckor under en vecka. (b) Bestäm sannolikheten för minst 2 olyckor under 3 på varandra följande dagar. (c) Bestäm sannolikheten för ingen olycka under 4 dagar. (3 p)

7. I en analys av en kontinuerlig belastningsprocess fann man att den i medel översteg larmnivån $x_L = 10$ MPa, $\hat{\lambda} = 0.79$ gånger per dygn. Man ansatte att styrkan x MPa av ett godtyckligt överskridande följer en Paretofördelning med täthet

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

och ML-skattade α. Därvid erhölls $\hat{\alpha} = 7.32$. 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 2 kkr i utryckningskostnad, ett tillbud ytterligare ca 23 kkr och en katastrof ytterligare ca 75 kkr. För ett tillbud och katastrof är alltså totala kostnaden ca 25 kkr resp 100 kkr. Skatta riskkostnaden. (Om du tar formler ur något läromedel, tänk på att vi i kursen har arbetat med två olika definitioner av Paretofördelningens täthet. Så det gäller att välja rätt.) (4 p)

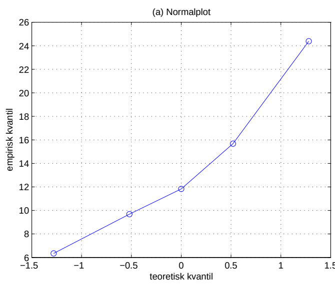
8. I ett försök att avgöra om en just inköpt tärning är symmetrisk eller ej gjordes 100 kast. Därvid erhölls 26 st sexor. Formulera ett lämpligt testproblem och lös det. (4 p)

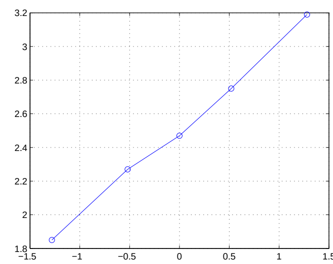
Lycka till!

Svar eller kortfattade lösningar till TMS055 Matematisk statistik V2 den 17/8-05

1. (a) Nej
   (b) Likformig på heltalen $1, 2, 3, 4, 5, 6$
   (c) 10 Mkr
   (d) 5%

2. De 5 datapunkterna ska i storleksordning jämföras med de teoretiska kvantilerna svarande mot sannolikheterna 0.1, 0.3, 0.5, 0.7 och 0.9. Enligt hjälpen är kvantilerna -1.28, -0.52, 0, 0.52, 1.28. Efter storleksordning har vi datapunkterna 6.35, 9.68, 11.83, 15.68, 24.39. Så i deluppgift (a) ska man plotta punkterna $(-1.28, 6.35), (-0.52, 9.68), (0, 11.83), (0.52, 15.68), (1.28, 24.39)$. I deluppgift (b) ska data logaritmeras innan jämförelsen görs. De loggade datapunkterna är 1.85, 2.27, 2.47, 2.75, 3.19. Så här ska punkterna $(-1.28, 1.85), (-0.52, 2.27), (0, 2.47), (0.52, 2.75), (1.28, 3.19)$ plottas. Nedan har jag plottat punkterna i (a) till vänster och de i (b) till höger:





Man ser att de loggade datapunkterna passar betydligt bättre ihop med de teoretiska normalkvantilerna. Således förefaller data vara lognormalfördelade. (Fast man ska akta sig för att dra långtgående slutsatser ur endast 5 datapunkter.)

3. Exp(λ)-fördelningens täthet är $f(x) = \lambda e^{-\lambda x}$ för $x > 0$. Trolighetsfunktionen hörande till stickprovet $x_1, \ldots, x_n$ blir således
$$L(\lambda) = \prod_i f(x_i) = \prod_i \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_i x_i}$$
Vi söker nu det λ som maximerar $\mathcal{L}(\lambda) = \log L(\lambda) = n \log \lambda - \lambda \sum_i x_i$. Vi får $\mathcal{L}'(\lambda) = n/\lambda - \sum_i x_i$ och genom att lösa λ ur ekvationen $n/\lambda - \sum_i x_i = 0$ erhålles $\lambda = n/\sum_i x_i = 1/\bar{x}$, v.s.v.

4. Bayes formel ger direkt att
$$P(C|D) = \frac{P(C)P(D|C)}{P(C)P(D|C) + P(C')P(D|C')} = \frac{0.36 \cdot 0.68}{0.36 \cdot 0.68 + (1 - 0.36) \cdot 0.19} = \frac{0.2448}{0.3664} \approx 0.668$$

5. Den väntevärdesriktiga skattningen av sannolikheten p är $\hat{p} = 73/10\,000 = 0.0073$ och konfidensintervallet är
$$p \le \hat{p} + z_{0.05} \cdot \sqrt{\hat{p}(1 - \hat{p})/n} = 0.0073 + 0.0014 = 0.0083$$
ty $z_{0.05} = 1.645$.

6. (a) $\mu_a = 7\lambda = 7 \cdot 0.04 = 0.28$ ty det går 7 dagar på en vecka
   (b) Tätheten är $p(k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$ för $k = 0, 1, \ldots$, där $\lambda = 0.04$ och $t = 3$. Den sökta sannolikheten är
$$p_b = p(2) + p(3) + \ldots = 1 - (p(0) + p(1))$$
$$= 1 - (e^{-0.04 \cdot 3} + e^{-0.04 \cdot 3}\, 0.04 \cdot 3) = 1 - (0.8869 + 0.1064) \approx 0.0067$$
   (c) Sannolikheten för ingen olycka under en dag är $e^{-\lambda}$, så sannolikheten för ingen olycka under fyra dagar (vare sig de kommer i följd eller ej) blir $p_c = (e^{-\lambda})^4 = e^{-4\lambda} = e^{-0.16} \approx 0.852$

7. Medelst integration får vi att sannolikheten för ett överskridande av nivån x är
$$1 - F(x) = \int_x^\infty f(y)\, dy = \left(\frac{x}{x_L}\right)^{-\alpha}$$
där $\alpha \approx 7.32$. Vi ser att $p(T|L) = 1 - F(x_T) = 1 - F(15) = (10/15)^{7.32} \approx 0.05141$ och att $p(K|L) =$

$1-F(x_K) = 1-F(20) = (10/20)^{7.32} \approx 0.006258$. Intensiteten för överskridande av larmnivån är $\lambda_L \approx 0.79$, för överskridande av tillbudsnivån $\lambda_T = \lambda_L p(T|L) =\approx 0.79 \cdot 0.05141 \approx 0.04061$ och för överskridande av katastrofnivån $\lambda_K p(K|L) = \lambda_K \approx 0.79 \cdot 0.006258 \approx 0.004944$. Riskkostnaden blir således

$$0.79 \cdot 2 + 0.04061 \cdot 23 + 0.004944 \cdot 75 \approx 2.89 \text{ kkr per dygn}$$

8. Man bör testa mot alternativet $H_1 : p \neq 1/6$. Den som väljer att testa mot $H_1 : p > 1/6$ skulle jag tro misstänker att tärningen har blivit manipulerat så att den ska ge fler sexor än en symmetrisk, men i uppgiften står det att man ville försöka att avgöra om den är symmetrisk eller inte. Så rätt är att testa nollhypotesen $H_0 : p = 1/6$ mot alternativet $H_1 : p \neq 1/6$. Observerat värde av teststatistikan $Z = \frac{\widehat{p}-1/6}{\sqrt{\widehat{p}(1-\widehat{p})/n}} \overset{\text{ap}}{\sim} \mathrm{N}(0,1)$ är

$$z = \frac{0.26 - 1/6}{\sqrt{0.26(1-0.26)/100}} = 2.13$$

I normalfördelningstabell ser vi att $P(Z \geq 2.13) = 1 - 0.9834 = 0.0166$, vilket ger att P-värdet $\approx 2 \cdot 0.0166 \approx 0.033$. Således gäller att vi kan förkasta H_0 på nivån 5%, men ej på nivån 1%. Med normal felrisk (5%) kan vi alltså påstå att tärningen är ej symmetrisk.

TMS055 Matematisk statistik V2, 27 maj 2005 f M

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. Tillåtna hjälpmedel är även de fyra häftena Introduktion till stokastisk simulering, Bayesiansk uppdatering av sannolikhetsskattningar, Poissonprocessen och extrema laster samt Något om riskkostnader samt föreläsningsanteckningarna som kan laddas ner ifrån kurshemsidan. Det är tillåtet med anteckningar i läroboken och i nedladdade häften. Inga andra anteckningar är tillåtna.

Kom ihåg att att alla svar skall motiveras om ej annat sägs. Ställ upp modell när det behövs och ange de för uträkningarna nödvändiga förutsättningarna. Skriv förståligt och använd etablerad notation.

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 till uppgifterna publiceras på webben. Se kursens hemsida. Tentamensresultat publiceras på anslagstavla i källaren i Matematiskt centrum.

— Teoridel —

1. Svaren på de tre frågorna i denna uppgift skall ej motiveras.
   - (a) Händelserna A och B är ömsesidigt uteslutande och har båda strikt positiv sannolikhet. Kan de vara oberoende? Svara Ja eller Nej. (1 p)
   - (b) Antag att du i en analys av ett stickprov erhöll $\mu = -0.23 \pm 0.19$ med konfidensgraden ca 99%. Nu säger din kollega: ”Detta bevisar att μ är negativ.” Hur stor är maximalt risken att hon har fel? Svara med ett %-tal. (1 p)
   - (c) Låt $X \sim \mathrm{LN}(\mu, \sigma)$. Är $E[X] = e^{\mu}$? (1 p)

2. (a) Visa att medianen i $\exp(\lambda)$-fördelningen är
$$m = \frac{\ln 2}{\lambda} \qquad (2\text{ p})$$
   (b) Härled uttrycket
$$x = -\frac{\ln u}{\lambda}$$
   för simulering av observationer ifrån $\exp(\lambda)$-fördelningen (u är ett slumptal). (2 p)

3. (a) Visa att väntevärde och varians i $\mathrm{U}(0,1)$-fördelningen är
$$\mu = \frac{1}{2} \text{ och } \sigma^2 = \frac{1}{12} \qquad (2\text{ p})$$
   Låt nu $V = a + (b-a)U$, där $U \sim \mathrm{U}(0,1)$ och $a < b$. (b) Visa ordentligt att
$$E[V] = \frac{a+b}{2} \text{ och } \mathrm{Var}[V] = \frac{(b-a)^2}{12} \qquad (2\text{ p})$$
   (c) Vilken fördelning har V? Svaret på (c) behöver ej motiveras. (1 p)

— Problemdel —

4. En ideell förening planerar att lansera ett nytt medlemslotteri. Man uppskattar sannolikheten för vinst till ca 0.024 och vinsten ska vara 100 kr. Hur mycket ska föreningen ta betalt per lott om överskottet från lotteriet ska vara ca 20%. (3 p)

5. Vid mätning av koncentrationen av en viss förorening erhölls i μg/dm^3:

$$15.61,\ 14.88,\ 14.22,\ 16.81,\ 12.32$$

Antag att dessa observationer är lognormal(μ, σ)-fördelade. Punkt- och intervallskatta μ. Konfidensgraden ska vara ca 95%. (4 p)

6. Ett år senare mätte man återigen koncentrationen av föroreningen och erhöll i samma enhet:

$$23.31,\ 25.97,\ 24.85,\ 25.03,\ 15.69$$

(Även dessa koncentrationer antas vara LN(μ, σ)-fördelade.) Jämför med data i uppgift 5 och svara på om det finns anledning att misstänka att μ har ökat? (4 p)

7. En ingenjör ska medelst provtagning skatta sannolikheten θ att ett visst gränsvärde överstigs. Hon har inte möjlighet att göra fler än 4 mätningar. Tidigare mätningar av liknaande ”sajter” säger henne att θ nog är ungefär en tredjedel. Hon önskar väga in denna information i den skattning av θ hon avser beräkna efter det att hon fått de 4 mätresultaten. Hur ska hon gå till väga om säkerheten i hennes gissning värderas till att svara mot ungefär den noggranhet hon erhåller i de 4 nya mätningarna? Antag att hon erhöll 2 överskridanden i de 4 mätningarna. Beräkna hennes à posteriori-skattning av θ. (3 p)

8. I en analys av en kontinuerlig belastningsprocess fann man att den i medel översteg larmnivån 7.5 MPa, $\hat{\lambda} = 0.1$ gånger per dygn. Man ansatte att styrkan x MPa av ett godtyckligt överskridande följer en Paretofördelning med täthet

$$f(x|T) = \frac{\alpha}{7.5}\left(\frac{x}{7.5}\right)^{-\alpha-1} \quad \text{för} \quad x > 7.5$$

och ML-skattade α. Därvid erhölls $\widehat{\alpha} = 4.64$. Skatta den nivå som i medel återkommer en gång per 250 dygn. (4 p)

Lycka till!

Lösningar (och kommentarer) till TMS055 Matematisk statistik V2 den 27/5-05

1. (a) Nej (b) 1% (c) Nej

2. Tätheten är $f(x) = \lambda e^{-\lambda x}$ för $x > 0$. (a) Medianen m fås enl

$$\int_0^m \lambda e^{-\lambda x}\,dx = 0.5 \quad\Rightarrow\quad 1 - e^{-\lambda m} = 0.5 \quad\Rightarrow\quad e^{-\lambda m} = 0.5$$

$$\Rightarrow\quad -\lambda m = \ln 0.5 = -\ln 2 \quad\Rightarrow\quad m = \frac{\ln 2}{\lambda}$$

Låt u vara ett slumptal. Enl simulerarnas huvudsats är lösningen x till ekvationen $u = P(X > x)$ en simulerad observation från X:s fördelning. Vi får

$$u = \int_x^\infty \lambda e^{-\lambda x}\,dx = e^{-\lambda x} \quad\Rightarrow\quad \ln u = -\lambda x \quad\Rightarrow\quad x = \frac{-\ln u}{\lambda}$$

3. (a) U(0, 1)-tätheten är $f(u) = 1$ för $0 \le u \le 1$. Vi får

$$\mu = \int_0^1 x\,dx = \frac{1}{2}$$

$$\sigma^2 = \int_0^1 x^2\,dx - \mu^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

(b) Låt nu $V = a + (b-a)U$. Vi får

$$E[V] = E[a + (b-a)U] = a + (b-a)E[U] = \frac{a+b}{2}$$

$$\mathrm{Var}[V] = \mathrm{Var}[a + (b-a)U] = \mathrm{Var}[(b-a)U] = (b-a)^2\mathrm{Var}[U] = \frac{(b-a)^2}{12}$$

(avdrag görs om det inte syns tydligt att man förstår alla leden ovan). (c) $V \sim \mathrm{U}(a, b)$.

4. Föreningen behöver få in $100/0.8 = 125$ kr för att kunna betala ut 100 kr för en vinst. Om en lott kostar x kr, så behöver man sälja $125/x$ lotter för att få in 125 kr. På $125/x$ lotter skall det alltså finnas en vinst. Således gäller för vinstsannolikheten p att

$$p = \frac{x}{125} \quad\Rightarrow\quad x = 125p = 125 \cdot 0.024 = 3.00$$

Alt kan man tänka så här: På n sålda lotter är föreningens intäkt xn kr och den kan förväntas betala ut $100pn$ kr. Om överskottet ska vara 20%, så måste vi ha

$$\frac{xn - 100pn}{xn} = 0.2 \quad\Rightarrow\quad 100p = 0.8x \quad\Rightarrow\quad x = \frac{100}{0.8}p = \ldots = 3.00$$

(-1 p till de som tror att det räcker att ta in 120 kr per utbetald vinst och fått $x = 2.88$ och till de som blandat ihop spelarens och föreningens synvinkel och utgått ifrån $(1-p)x - 100p = 0.2x$ och erhållit $x = 3.09$.)

5. Data logaritmeras. Därefter räknas ut att $\sum x = 13.4358$ och $\sum x^2 = 36.1581$. Vidare är $n = 5$, så

$$\bar{x} = \frac{13.4358}{5} = 2.687 \quad \text{och} \quad s = \sqrt{\frac{36.1581 - \frac{1}{5}(13.4358)^2}{4}} = 0.1165$$

Antalet frihetsgrader är $\nu = n - 1 = 4$. På motsv rad i t-tabellen avläser vi $t_{0.005} = 2.776$, så vi får

$$\mu = 2.687 \pm 2.776 \cdot 0.1165/\sqrt{5} = 2.687 \pm 0.145$$

(-1 p till de som missat att logaritmera data.)

6. Data från förra året sammanfattas i $n_1 = 5$, $\bar{x}_1 = 2.687$ och $s_1 = 0.1165$. Analogt får vi för årets data $n_2 = 5$, $\bar{x}_2 = 3.1183$ och $s_2 = 0.2079$. Glöm inte att data ska logaritmeras innan medelvärde och standardavvikelse räknas ut. Standardfelet i differensen $\bar{x}_2 - \bar{x}_1 = 0.431$ är

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{0.1162^2}{5} + \frac{0.2079^2}{5}} = 0.1065$$

Således är observerat värde på T-statistikan,

$$t = \frac{0.431}{0.1065} = 4.048$$

Antalet frihetsgrader ν fås medelst avrundning nedåt av

$$\text{df} = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}} = 6.28$$

Således gäller att $\nu = 6$ och vi ser på motsv rad i t-tabellen att P-värdet är mindre än $1 - 0.995 = 0.005$. Vi kan alltså påstå med hög säkerhet (risken att ha fel $\leq 0.5\%$) att μ har ökat under året som gått. Alt hade man kunnat beräkna ett nedåt begr konfidensintervall för differensen med, t.ex, 95% konfidens enl

$$\mu_2 - \mu_1 \geq 0.431 - 1.943 \cdot 0.1065 = 0.431 - 0.207 = 0.224$$

Om man nu påstod att μ har ökat, så är max felrisk (säkert betydligt mindre än) ca 5%. (Svårrättat eftersom det ju är uppenbart om man ser på data, att man kan misstänka att μ har ökat. Man ska naturligtvis ge ett vetenskapligt formulerat svar (alltså med max felrisk) och man ska göra en bästa jämförelse av två stickprov enl något av de två förslagen ovan. De som jämfört konfidensintervall eller testat utan att ta hänsyn till slumpmässigheten i det ena medelvärdet har förlorat 2 p och har man inte gjort klart att man har koll på felrisken, så tappar man ytterligare 1 p.)

7. Om säkerheten i hennes gissning, $\theta = 1/3$, svarar mot 4 mätningar, så fås antalet överskridanden x i dessa 4 mätningar enl

$$\frac{x}{4} = \frac{1}{3} \quad \Rightarrow \quad x = \frac{4}{3}$$

Á posteriori-skattningen blir

$$\widehat{\theta} = \frac{4/3 + 2}{4 + 4} = \frac{5}{12} = 0.417$$

(Det är viktigt att inse att antalet överskridanden i á priori-skattningen ej behöver vara ett heltal. Så de som à priori ansatt $n = 3$, $x = 1$, får finna sig i att förlora 1 p.)

8. Medelst integration fås

$$P(X > x|T) = \left(\frac{x}{7.5}\right)^{-\alpha} \quad \text{för} \quad x > 7.5$$

Under 250 dygn sker i snitt ungefär $250 \cdot \hat{\lambda} = 25$ tillbud. I snitt ska ett av dessa över 250-dygnsnivån x_{250}. Således gäller

$$P(X > x_{250}|T) = \frac{1}{25} \quad \Rightarrow \quad \left(\frac{x_{250}}{7.5}\right)^{-\alpha} = \frac{1}{25}$$

$$\Rightarrow \quad x_{250} = 7.5 \cdot 25^{1/\alpha} \quad \Rightarrow \quad \hat{x}_{250} = 7.5 \cdot 25^{1/\widehat{\alpha}} = 7.5 \cdot 25^{1/4.64} = 15.0$$

(De som trott att $P(X > x_{250}|T) = \left(\frac{x_{250}}{7.5}\right)^{1-\alpha}$ och fått $\hat{x}_{250} = 7.5 \cdot 25^{1/3.64} = 18.16$ har inte fått poängavdrag p.g.a inkonsekvensen i definitionen av Paretofördelningen i kurslitteraturen.)