Chalmers tekniska högskola
Matematisk statistik

TMS051 Matematisk statistik och simuleringsteknik Z, del A

Tentamen 23 oktober 2004 e V

Även omtentamen på TMS050 Matematisk statistik och simuleringsteknik Z, del A.

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

Examinator är Tommy Norberg, ankn 3528 eller 0730 794209. Någon ansvarig lärare besöker skrivningssalen efter kl 15 och vid behov ca 16^{30}.

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

Svar skall motiveras om ej annat sägs i uppgiften.

Fyll i den bifogade enkäten när du är klar med uppgifterna nedan. Ju fler svar, desto lättare kan vi förbättra kursen!

Uppgifter

1. Man önskar kontrollera ett parti omfattande N hårddiskar genom att slumpmässigt välja ut n st och kontrollera funktionen hos dessa. Om det i verkligheten är så att D av hårddiskarna är defekta, hur stor är då sannolikheten att man hittar minst en av dessa? Härled en allmän formel och evaluera denna för fallet då $N = 20$, $D = 2$ och $n = 4$. (3 p)

2. Låt A och B vara händelser sådana att $P(A) = 0.2$ och $P(B) = 0.6$. Vad blir $P(A \cup B)$, om (a) A och B är ömsesidigt uteslutande, (b) A och B är oberoende, (c) $P(A \mid B) = 0.4$ och (d) $A \subseteq B$? Svaren behöver ej motiveras. (4 p)

3. Oplanerade stopp i en viss produktionsanläggning kan antas vara Poi(λ), där man av erfarenhet vet att $\lambda \approx 4$ st/år. Beräkna sannolikheten att man under ett kvartal får minst två stopp. (3 p)

4. Visa att m.g.f för N(μ, σ)-fördelningen är $m(t) = e^{\mu t + \sigma^2 t^2/2}$. (4 p)

5. (a) För vilket $c > 0$ gäller att funktionen
$$f(x) = \begin{cases} c/x^{\alpha+1} & \text{då } x \geq 1 \\ 0 & \text{för övrigt} \end{cases}$$
är en täthet för alla parametervärden $\alpha > 0$? (b) För vilka $\alpha > 0$ existerar väntevärdet m.a.p denna fördelning? (4 p)

6. Visa (a) att S^2 är en väntevärdesriktig estimator av σ^2, samt (b) att S ej skattar σ väntevärdesriktigt. (Obs att i (b) går det bra att använda resultatet från (a) även om du inte lyckades genomföra beviset av det.) (4 p)

(Vänd!)

7. I syfte att kontrollera kvaliteten hos en viss typ av plaster inköpta av en underleverantör, testades 10 slumpmässigt utvalda stickprov i en klimatugn för att snabba på nedbrytningsprocesserna. Efter 100 timmar i klimatugnen mätte man upp varje provs draghållfasthet. Följande värden erhölls:

   32.28 32.61 35.91 29.43 32.10 30.39 32.52 42.21 29.36 24.92

   Enhet: N/m^2. Antag att mätningarna är utan systematiskt fel och att variationerna i dem följer en normalfördelning. (a) Beräkna ett nedåt begränsat konfidensintervall för förväntad draghållfasthet μ. Låt konfidensgraden vara 0.95. (b) Gör om beräkningen, men låt nu konfidensgraden vara 0.99. (c) Kvalitetskravet för plasten ifråga är $\mu > 29$ N/m^2. Bedöm så noggrant du kan hur stor risken är att kvalitetskravet ej är uppfyllt? (Obs $n = 10$, $\sum x = 321.73$ och $\sum x^2 = 10537.2761$.) (5 p)

8. Beräkna ett konfidensintervall för standardavvikelsen σ i draghållfasthetsmätningarna i ovanstående uppgift. Önskad konfidensgrad är 0.95. (3 p)

Lycka till!

1. Totala antalet oordnade val av n enheter ur en urna med N st är $\binom{N}{n}$. Av dessa innehåller $\binom{N-D}{n}$ inga defekta. Den sökta sannolikheten är således
$$p(N,D,n) = 1 - \frac{\binom{N-D}{n}}{\binom{N}{n}}$$
Obs att detta gäller om $n \le N - D$. Annars är $p(N,D,n) = 0$. Medelst insättning fås
$$p(20,2,4) = 1 - \frac{\binom{18}{4}}{\binom{20}{4}} = 1 - \frac{18 \cdot 17 \cdot 16 \cdot 15}{20 \cdot 19 \cdot 18 \cdot 17} = \frac{7}{19} \approx 0.368$$
Termen $\frac{18\cdot17\cdot16\cdot15}{20\cdot19\cdot18\cdot17}$ antyder att det finns ett annat sätt att lösa uppgiften. Drar man en efter en, så är i första dragningen P[ingen defekt i dragning 1] $= \frac{18}{20}$. I andra dragningen får vi P[ingen defekt i dragning 2 | en felfri är dragen] $= \frac{17}{19}$, osv. Sannolikheten för ingen defekt i 4 dragningar blir därför $\frac{18}{20} \cdot \frac{17}{19} \cdot \frac{16}{18} \cdot \frac{15}{17}$ och sannolikheten för minst en defekt
$$1 - \frac{18}{20} \cdot \frac{17}{19} \cdot \frac{16}{18} \cdot \frac{15}{17}$$
Med ett generellt resonemang av denna typ fås utrycket
$$p(N,D,n) = 1 - \frac{N-D}{N} \cdot \frac{N-D-1}{N-1} \cdot \ldots \cdot \frac{N-D-n+1}{N-n+1}$$
Det är inte svårt att se att båda formlerna för $p(N,D,n)$ är ekvivalenta. (1 p dras för de som nöjer sig med att räkna ut den sökta sannolikheten utan att härleda formlen.)

2. (a) $P(A \cup B) = 0.2 + 0.6 = 0.8$, (b) $P(A \cup B) = 0.2 + 0.6 - 0.2 \cdot 0.6 = 0.68$, (c) $P(A \cup B) = 0.2 + 0.6 - 0.6 \cdot 0.4 = 0.56$ och (d) $P(A \cup B) = 0.2 + 0.6 - 0.2 = 0.6$. Deluppgift (c) är felkonstruerad, ty det kan omöjligt inträffa att $P(A \cap B) = 0.6 \cdot 0.4 = 0.24 > P(A)$. Detta påpekades av KS372 och LE628.

3. Ett kvartal är $s = 0.25$ år. Antalet stopp under ett kvartal är därför Poi(1) (ty $\lambda s \approx 1$) och sannolikheten för minst två stopp, som ju är ett minus sannolikheten för högst ett stopp, blir ungefär
$$1 - \left(e^{-1} \frac{1^0}{0!} + e^{-1} \frac{1^1}{1!} \right) = 1 - 2e^{-1} \approx 1 - 0.736 = 0.264$$

4. N(μ, σ)-tätheten är $f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$ för $-\infty < x < \infty$ (se formelsamlingen eller Beta). Således är
$$m(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} \, dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2 + t\mu + t\sigma z} \, dz$$
$$= e^{t\mu + t^2\sigma^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z - t\sigma)^2} \, dz = e^{\mu t + \sigma^2 t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \, dy = e^{\mu t + \sigma^2 t^2/2}$$
Den andra likheten ovan fås m.h.a substitutionen $z = (x - \mu)/\sigma$. För att se den tredje, kvadratkomplettera i exponenten och bryt ut konstanta faktorer. Den fjärde likheten följer sedan från substitutionen $y = z - t\sigma$.

5. (a) Konstanten c ges av att $\int_1^{\infty} f(x) \, dx = 1$. Vi får att $c = \alpha$, eftersom $\int_1^{\infty} \frac{c}{x^{\alpha+1}} \, dx = \frac{c}{\alpha}$.
(b) Väntevärdet m.a.p denna fördelning är $\mu = \int_1^{\infty} x \frac{\alpha}{x^{\alpha+1}} \, dx = \int_1^{\infty} \frac{\alpha}{x^{\alpha}} \, dx < \infty$ precis då $\alpha > 1$. Den sista frågan hade man kunnat besvara även om man inte visste c:s exakta värde. (Varje del av denna uppgift värderas till 2 p.)

6. (a) Att S^2 skattar σ^2 vämtevärdesriktigt, kan man se så här: Notera först att

$$(n-1)S^2 = \sum_i (X_i - \bar{X})^2 = \sum_i X_i^2 - n\bar{X}^2$$

Härur fås

$$(n-1)E[S^2] = E[(n-1)S^2] = E\left[\sum_i X_i^2 - n\bar{X}^2\right] = \sum_i E[X_i^2] - nE[\bar{X}^2]$$

Men $E[X_i^2] = \text{Var}[X_i] + E[X_i]^2 = \sigma^2 + \mu^2$ och $E[\bar{X}^2] = \text{Var}[\bar{X}] + E[\bar{X}]^2 = \sigma^2/n + \mu^2$. Vi får därför att

$$(n-1)E[S^2] = n(\sigma^2 + \mu^2) - n(\sigma^2/n + \mu^2) = (n-1)\sigma^2$$

ur vilket $E[S^2] = \sigma^2$ följer. (b) Om $E[S] = \sigma$, så skulle det följa att

$$\text{Var}[S] = E[S^2] - E[S]^2 = \sigma^2 - \sigma^2 = 0$$

Det är bara konstanter som har variansen 0. Alltså skulle $S^2 = c$ för något reellt $c \geq 0$. Men detta kan bara inträffa då $\sigma^2 = 0$ och i så fall är $c = 0$. Vi sluter oss till att S ej kan skatta $\sigma > 0$ väntevärdesriktigt. (Varje del av denna uppgift värderas till 2 p.)

7. Börja med att räkna ut

$$\bar{x} = \frac{1}{n}\sum x = 32.173,\ s^2 = \frac{1}{n-1}\left(\sum x^2 - n\bar{x}^2\right) = 20.69520 \text{ och } s = \sqrt{s^2} = 4.5492$$

Ur t-tabell (9 frihetsgrader) fås $t_{0.05} = 1.833$ och $t_{0.01} = 2.821$ och man kan nu räkna ut (a) $\mu > \bar{x} - t_{0.05}s/\sqrt{n} = 29.54$ N/m^2 och (b) $\mu > \bar{x} - t_{0.01}s/\sqrt{n} = 28.11$ N/m^2. Utgå i båda fallen ifrån

$$\frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1) \text{ och } P\left(\frac{\bar{X} - \mu}{s/\sqrt{n}} \leq t_\alpha\right) = 1 - \alpha$$

($3 = 2 + 1$ p ges till de som räknar ut båda konfidensintervallen korrekt.) (c) Risken att utsagan $\mu > 29$ N/m^2 är falsk är mindre än 5% (ty $\mu > 29.54 \Rightarrow \mu > 29$) och större än 1% (ty $\mu > 28.11 \not\Rightarrow \mu > 29$). (1 p ges till de som svarar att risken ligger mellan 1% och 5%.) Man kan "pin-pointa" denna risk ytterligare genom att ansätta

$$\bar{x} - t_\alpha s/\sqrt{n} = 29 \Rightarrow t_\alpha = \frac{\bar{x} - 29}{s/\sqrt{n}} = 2.206 \Rightarrow \alpha = 0.027$$

Den sökta risken är alltså ca 2.7%. Men detta går ju inte att räkna ut med sådana hjälpmedel en standard Z-teknolog har med sig till tentamen. Fast obs att man kan komma en bit på väg genom att jämnföra $t_\alpha = 2.206$ med $t_{0.025} = 2.262$, alternativt räkna ut ett 97.5% konfidensintervall. Av $t_\alpha < t_{0.025}$ sluter vi oss till att $\alpha > 0.025$. Den sökta risken är alltså större än 2.5%. (Ytterligare 1 p ges till de som kommer fram till detta.)

8. Ur faktumet $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$ som hämtas ur formelsamlingen eller Beta, härledes det allmänna uttrycket för ett konfidensintervall för σ^2:

$$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}$$

Här är kvantilerna $\chi^2_{n-1,1-\alpha/2} = \chi^2_{9,0.975} = 2.7004$ och $\chi^2_{n-1,\alpha/2} = \chi^2_{9,0.025} = 19.023$. Insättning och uträkning (inkl kvadratrotsdragning av alla tre leden) ger resultatet $3.13 \leq \sigma \leq 8.31$. Den som gör ett uppåt begränsat intervall m.h.a formeln $\sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha}}$ och erhåller $\sigma \leq 7.48$ (obs att $\chi^2_{9,0.95} = 3.3251$) får inget poängavdrag. Man bör däremot inte beräkna ett nedåt begränsat intervall då det är tveksamt varför ett sådant kan vara av intresse. Avdrag med 1 p m.a.o till de som levererar ett nedåt begränsat intervall. Avdrag med 1 p även till de som hämtar fel kvantiler ur tabell och därför får fel konfidensgrad.