

TMS051: Matematisk statistik för Z, del A

Tentamen 26 augusti 2004 f V

Detta är även omtentamen på TMS050 Matematisk statistik för Z, del A.

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

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p.

Jour och examinator är Tommy Norberg (ankn 3528 eller 0730 794209).

Observera att svar skall motiveras i samtliga 8 uppgifter om ej annat sägs.

Uppgifter

1. Svaren på de tre frågorna i denna uppgift ska ej motiveras.
 - (a) Antag att $P(A) = 0.63$, och att $P(B) = 0.41$. Kan A och B vara ömsesidigt uteslutande? (1 p)
 - (b) Är funktionen $f(x) = 3e^{-x/3}$ täthet för någon kontinuerlig stokastisk variabel med utfallsrum $x > 0$? (1 p)
 - (c) Man vill med ett statistiskt test påvisa att medeldrifttiden, θ , för en viss enhet i en produktionsanläggning är mindre än $\theta_0 = 32.3$ sekunder. Ska man då som nollhypotes välja $H_0 : \theta \leq \theta_0$ eller $H_0 : \theta \geq \theta_0$? (1 p)
2. Låt A_1, A_2 vara två händelser i ett utfallsrum. Visa, utgående ifrån sannolikhetsaxiomen, att
 - (a) $P(A_1 \setminus A_2) = P(A_1) - P(A_1 \cap A_2)$ (2 p)
 - (b) $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$ (2 p)
3.
 - (a) Antag att du gör ett kast med en symmetrisk sex-sidig tärning. Beräkna väntevärde μ och varians σ^2 för resultatet som vi kan kalla X . (2 p)
 - (b) Antag att du gör 30 oberoende kast med en symmetrisk sex-sidig tärning. Låt $X_1, \dots, X_{30}$ vara de erhållna resultaten. Beräkna väntevärde och varians för medelvärde $\bar{X} = (X_1 + \dots + X_{30})/30$. (2 p)
4. Antag att man gör oberoende upprepningar av ett fysikaliskt försök i vilket en händelse A antingen inträffar eller inte inträffar. Antag att man håller på tills dess att A inträffat en gång. Ett exempel på ett sådant experiment är att man tar oberoende prover av en viss stålqualität och testar om de håller för en viss standardbelastning ända tills man hittar en provstång som ej klarar lasten. Antag att $P(A) = 0.1$. Låt X vara antalet försök som görs. Beräkna $P(X = 5)$. (3 p)

(Vänd!)

5. Antag att X är exponentialfördelad med väntevärde 100. Beräkna undre och övre kvartilen samt medianen i X :s fördelning. Antag att man gör många oberoende observationer av X . Ange (utan att motivera) ett intervall på linjen med egenskapen att ca 50% av observationerna kommer att hamna i det. (4 p)
6. Försöket att kasta en tärning 30 gånger och räkna ut medelantalet prickar upprepades 56 gånger. Medelvärde av de 56 försöksresultaten blev $\bar{x} = 3.548$ och variansen blev $s^2 = 0.0911$. Punkt- och intervallskatta det teoretiska medelvärdet μ av 30 kast. Önskad konfidsgrad är 95%. (4 p)
7. I en enkätundersökning av partisympatier visade det sig att av de 1000 röstberättigade som svarade sympatiserade 7.3% med ett visst parti. Antag att urvalet av enkätdeltagare är gjort så att varje röstberättigad har samma chans att komma med oberoende av alla andra. Beräkna ett konfidsintervall med ca 95% konfidsgrad för proportionen röstberättigande som sympatiserar med partiet ifråga. (4 p)
8. Punkt- och intervallskatta den teoretiska variansen σ^2 av 30 tärningskast. Tag data från uppgift 6. (4 p)

TMS051: Matematisk statistik för Z, del A

Tentamen 15 januari 2004 f V

Detta är även omtentamen på TMS050 Matematisk statistik för Z, del A.

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

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p.

Jour och examinator är Tommy Norberg (ankn 3528 eller 0730 794209).

Observera att svar skall motiveras i samtliga 8 uppgifter om ej annat sägs.

Uppgifter

1. I en urna finns 3 vita, 4 blåa och 5 gröna kulor. Man drar på måfå 3 st utan återläggning. Bestäm sannolikheten att man erhåller 1 kula av varje färg. (4 p)
2. Låt y vara koncentrationen av någon giftig tungmetall i ett jordprov. Låt x vara resultatet av en mätning av y . Provet betraktas som förorenat om $y \geq c$, där c är ett av Naturvårdsverket specificerat tröskelvärde. Låt C beteckna händelsen $y \geq c$. Låt vidare d vara den s k detekteringsnivån (obs att typiskt är $d \neq c$). Händelsen $x \geq d$ betecknas D . Ett rimligt krav på händelserna C och D är att de är positivt beroende i betydelsen $P(C|D) \geq P(C)$. Antag nu att $P(D|C') = 0.15$ och $P(D'|C) = 0.20$. Är då C och D positivt beroende? (4 p)
3. Eskil och Nils spelar ett spel som är sådant att $P(\text{Eskil vinner}) = 1 - P(\text{Nils vinner}) = 1/3$. Om Eskil satsar 1 kr, hur mycket tycker du att Nils ska satsa för att spelet ska bli rättvist? (4 p)
4. Beräkna väntevärde och varians för X om X 's fördelningsfunktionen är

$$F(x) = 1 - \left(\frac{u}{x}\right)^\alpha \quad \text{för } x \geq u$$

där $u = 1$ och $\alpha = 3/2$. (4 p)

5. Visa att mgf för Geo(p)-fördelningen är $m(t) = pe^t/(1 - qe^t)$ för $t < -\ln q$ där $q = 1 - p$. Beräkna sedan fördelningens väntevärde μ och varians σ^2 . (4 p)
6. Antag att du har oberoende observationer $x_1, \dots, x_n$ från någon fördelning med väntevärde μ och varians σ^2 . Visa att observationernas medelvärde $\bar{x}$ resp varians s^2 är väntevärdesriktiga skattningar av sina teoretiska motsvarigheter. (4 p)

(Vänd!)

7. I en simulering av ett produktionssystem visade det sig att första gången systemet gick ner skedde innan 40 timmars drift i $f = 27$ fall. Antalet simuleringar som gjordes var $n = 100$. Punkt- och intervallskatta sannolikheten p att produktionssystemet går ner innan 40 timmar. Konfidsensgraden bör vara ca 0.95. (3 p)
8. Under en tidsperiod omfattande 100 minuter registreras 22 678 radioaktiva sönderfall. Man kan räkna med att apparaturen registrerar samtliga sönderfall. Gör ett 95% konfidsensintervall för sönderfallsintensiteten, d.v.s. det förväntade antalet sönderfall per minut. (3 p)

– Lycka till –

TMS051: Matematisk statistik för Z, del A

Tentamen 25 oktober 2003 em V

Detta är även omtentamen på TMS050 Matematisk statistik för Z, del A.

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

För betyget 3 krävs 12 p, för 4:a 18 p och för 5:a 24 p av totalt 30 p.

Jour är

Examinator är Tommy Norberg (ankn 3528 eller 0730 794209).

Observera att svar skall motiveras i samtliga 8 uppgifter om ej annat sägs.

Uppgifter

1. En godsterminal har 8 lastbryggor. För var och en av dessa gäller ett av följande: lastbryggan är i funktion och ledig, lastbryggan är i funktion och upptagen, lastbryggan är ej i funktion. Beräkna m.h.a den klassiska sannolikhetsdefinition sannolikheten för händelsen att

(a) totalt 7 lastbryggor är i funktion, varav 3 är upptagna och 4 lediga, (2 p)

(b) totalt 7 st lastbryggor är i funktion. (2 p)

2. Låt händelserna A , A^c och B ha strikt positiva sannolikheter. Då gäller att

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}$$

Visa detta. (4 p)

3. (a) Kan funktionen $f(x) = c \sin x$ för något värde på konstanten c fungera som täthet för en stokastisk variabel X med utfallsrum $0 \leq x \leq \pi$? (1 p)

(b) Vad är isåfall c 's exakta värde? (1 p)

(c) Vad är isåfall $\mu = E[X]$ och $\sigma = \sqrt{\text{Var}[X]}$? (2 p)

4. Man avser kasta ett symmetriskt mynt 100 gånger och räkna antalet gånger det landar med krona upp. Låt f vara detta antal.

(a) Vilken fördelning har f ? (2 p)

(b) Beräkna approximativt $P(40 \leq f \leq 60)$. (2 p)

5. I en övervakningsanläggning sker larm enl en Poissonprocess med intensiteten λ st per tidsenhet. Låt W vara tiden tills larmet går första gången. Vilken fördelning har W ? Vilket väntevärde har W ? (4 p)

(Vänd!)

6. Antag att de stokastiska variablerna X, Y har tätheten

$$f(x, y) = \frac{x}{4\sqrt{y}} \quad \text{för } 0 \leq x \leq 2, 0 \leq y \leq 1$$

- (a) Beräkna $\text{Cov}[X, Y]$. (1 p)
- (b) Beräkna $P(X \leq 1)$. (1 p)
- (c) Antag att man har utfört ett slumpmässigt försök i vilket resultatet är (X, Y) och att man har fått reda på att $Y > 0.5$. Inget mer är känt om försöksresultatet. Är då den betingade sannolikheten för händelsen $X \leq 1$ skild från eller lika med sannolikheten du skulle räkna ut i uppgift 6b? (1 p)
7. För att skaffa sig erfarenhet av en ny metod att mäta ytnoggrannhet gjorde man 30 st oberoende mätningar av en speciell testyta. Stickprovet fick standardavvikelsen $s = 3.562$. Bestäm under normalfördelningsantagande ett uppåt begränsat konfidensintervall för mätmetodens standardavvikelse σ med konfidensgrad 0.95. (3 p)
8. Antag att man har utfört n st oberoende försök i vilka försöksresultatet är $N(\mu, \sigma)$ -fördelat. Låt $\bar{x}, s$ vara stickprovets medelvärde resp standardavvikelse.

- (a) Visa att intervallskattningen

$$\bar{x} - t_{\alpha/2, n-1}s/\sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2, n-1}s/\sqrt{n}$$

har konfidensen $1 - \alpha$ (kvantilen $t = t_{\alpha/2, n-1}$ definieras av att $P(T > t) = \alpha/2$ för $T \sim t(n-1)$). Det är tillåtet att utgå ifrån ett lämpligt resultat i formelsamlingen. (3 p)

- (b) Om normalfördelningsantagandet skulle vara tveksamt, går det då ändå att säga något om intervallets konfidens? (1 p)

1. (a) Nej ty $P(A) + P(B) > 1$
 (b) Nej ty $\int_0^\infty f(x) dx \neq 1$
 (c) $H_0 : \theta \geq \theta_0$ ty nollhypotesen ska vara komplementet till alternativhypotesen H_1 som ska vara det man vill påvisa
2. (a) $P(A_1) = P(A_1 \cap A_2 \cup A_1 \cap A_2') = P(A_1 \cap A_2) + P(A_1 \cap A_2')$ och $P(A_1 \setminus A_2) = P(A_1 \cap A_2')$
 $\Rightarrow P(A_1 \setminus A_2) = P(A_1) - P(A_1 \cap A_2)$
 (b) $P(A_1 \cup A_2) = P(A_1 \cup A_2 \setminus A_1) = P(A_1) + P(A_2 \setminus A_1) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$
3. (a) Vi har utfallen 1, 2, 3, 4, 5, 6. Symmetrin ger att alla har sannolikheten $\frac{1}{6}$. Således blir $\mu = E[X] = 1 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = 3.5$ och $E[X^2] = 1^2 \cdot \frac{1}{6} + \dots + 6^2 \cdot \frac{1}{6} = \frac{91}{6} = 15\frac{1}{6}$, vilket ger $\sigma^2 = \text{Var}[X] = 15\frac{1}{6} - 3.5^2 = \frac{35}{12} = 2\frac{11}{12} \approx 2.92$
 (b) $E[\bar{X}] = E[X] = 3.5$ och $\text{Var}[\bar{X}] = \frac{1}{30} \text{Var}[X] \approx 0.09722$
4. $P(X = 5) = 0.9^4 \cdot 0.1 = 0.06561$ ty $X = 5 \Leftrightarrow A'$ inträffar i de fyra första försöken och A i det femte
5. X har tätheten $f(x) = \frac{1}{100} e^{-x/100}$. Första kvartilen q_1 definieras av att $\int_0^{q_1} f(x) dx = 0.25$. Man visar lätt att detta leder till att $q_1 = -100 \ln 0.75 \approx 28.77$. Analogt fås att medianen $m = -100 \ln 0.5 \approx 69.31$ och att tredje kvartilen $q_3 = -100 \ln 0.25 \approx 138.63$. Intervallen $(0, 69.31]$, $(28.77, 138.63]$ och $(69.31, \infty)$ har alla egenskapen att ca 50% av observationerna hamnar i dem. Det spelar ingen roll om gränserna inkluderas eller ej.
6. Punktskattningen av μ är $\hat{\mu} = 3.548$ och intervallskattningen är $3.548 \pm 2.004 \cdot \sqrt{0.0911/56} = 3.548 \pm 0.081$ eller $(3.467, 3.629)$. Här är 2.004 tabellvärdet i 2.5%-kolumnen (eller 97.5%-kolumnen) i t -tabellen med 55 frihetsgrader. I många tabeller finns bara värdet för 40 resp 60 frihetsgrader. Man kan då välja att interpolera eller att helt enkelt ta tabellvärdet svarande mot 40 frihetsgrader. Båda sätten ger full poäng. Full poäng får även den som istället använder normalfördelningskvantilen 1.96, även om jag fördrar att man tar sin kvantil ur t -tabellen. Man är mer på den säkra sidan då.
7. Det sökta konfidensintervallet är $0.073 \pm 1.96 \cdot \sqrt{0.073(1 - 0.073)/1000} = 0.073 \pm 0.016$ eller $(5.7, 8.9)\%$
8. Känt är att $(n-1)s^2/\sigma^2$ är χ^2 -fördelad med $n-1$ frihetsgrader, i detta fall 55 st. Formelsamlingens χ^2 -tabell anger bara kvantilerna svarande mot 50 resp 60 frihetsgrader och dessa skiljer sig åt ganska mycket så det är klart lämpligt att interpolera. Vi får då att 2.5%-kvantilen är $\approx 0.3 \cdot 32.3574 + 0.7 \cdot 40.4817 = 38.014$ och att 97.5%-kvantilen är $\approx 0.3 \cdot 59.3417 + 0.7 \cdot 71.4202 = 67.797$. Konfidensintervallet fås nu enl $38.014 < (n-1)s^2/\sigma^2 < 67.797 \Leftrightarrow (n-1)s^2/67.797 < \sigma^2 < (n-1)s^2/38.014$ eller med siffror insatta $0.0739 < \sigma^2 < 0.1318$.

$$1. \frac{\binom{3}{1} \binom{4}{1} \binom{5}{1}}{\binom{12}{3}} = \frac{3 \cdot 4 \cdot 5}{220} = \frac{3}{11} \approx 0.273$$

2. Genom att följa följande sekvens av ekvivalenser baklänges ser vi att svaret på frågan är "ja, händelserna är positivt beroende": $P(C|D) \geq P(C) \Leftrightarrow P(C \cap D) \geq P(C)P(D) \Leftrightarrow P(D|C) \geq P(D) = P(C)P(D|C) + P(C')P(D|C') \Leftrightarrow P(D|C) - P(C)P(D|C) \geq P(C')P(D|C') \Leftrightarrow P(D|C) \geq P(D|C') \Leftrightarrow P(D|C) + P(D|C') \leq 1 \Leftrightarrow 0.35 \leq 1$. Alternativ lösning: Notera först att

$$P(C|D) = \frac{P(C)P(D|C)}{P(C)P(D|C) + P(C')P(D|C')} = \frac{0.8P(C)}{0.65P(C) + 0.15}$$

Visa nu att $P(C|D) \geq P(C)$ genom att visa att

$$\frac{0.8x}{0.65x + 0.15} \geq x \quad \text{för } 0 \leq x \leq 1$$

Detta kan t.ex göras genom att man noterar att likhet gäller då $x = 0, 1$, samt att funktionen

$$f(x) = \frac{0.8x}{0.65x + 0.15} - x$$

har $f''(x) < 0$ för $0 < x < 1$ och $f'(0) > 0$.

3. Det enda rimliga kriteriet för ett rättvist spel är att båda spelarna vinner ungefär lika mycket i det långa loppet, så vi ansätter att de båda spelarnas förväntade vinst ska vara lika stor. Då

$$\frac{1}{3}x = \frac{2}{3} \cdot 1$$

ty med sannolikheten $1/3$ vinner Eskil de x kr som Nils satsar och med sannolikheten $2/3$ vinner Nils den krona Eskil satsar. Vi ser att Nils bör satsa $x = 2$ kr.

4. Insättning av $u = 1$ och $\alpha = 3/2$ ger

$$F(x) = 1 - x^{-3/2} \quad \text{för } x \geq 1$$

Börja med att härleda X 's täthet

$$f(x) = F'(x) = \frac{3}{2}x^{-5/2} \quad \text{för } x \geq 1$$

Vi får nu att

$$E[X] = \int_1^\infty x f(x) dx = \int_1^\infty \frac{3}{2}x^{-3/2} dx = 3$$

$$E[X^2] = \int_1^\infty x^2 f(x) dx = \int_1^\infty \frac{3}{2}x^{-1/2} dx = \infty$$

Variansen existerar alltså inte. (Det är inte fel att svara att variansen är oändlig.)

5. Se bevisen av satserna 3.4.1 (sida 62–63) och 3.4.3 (sida 64–65) i Milton & Arnold.
6. Se bevisen av satserna 7.1.1 (sida 226) och 7.1.3 (sida 782–783) i Milton & Arnold.
7. $p = \hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n} = 0.27 \pm 0.087$
8. Vi har att $n = 100$, $\bar{x} = 226.78$ och att $s \approx \sqrt{226.78} \approx 15.0592$ (ty för Poissonfördelade variabler gäller $\mu = \lambda$ och $\sigma = \sqrt{\mu}$). Det sökta konfidensintervallet blir $\lambda = 226.78 \pm 1.96 \cdot 15.0592/\sqrt{100} = 226.78 \pm 2.95$. Konfidensgraden är ca 0.95.

1. (a) $\binom{8}{3} \binom{5}{4} / 3^8 = 280/6561 = 0.0427$. Välj först ut de 3 bryggor som ska vara upptagna, sedan ur de 5 återstående de 4 som ska vara lediga. Detta kan ske på $\binom{8}{3} \binom{5}{4} = 280$ sätt. Det totala antalet möjligheter är $3^8 = 6561$.
- (b) $\binom{8}{7} \cdot 2^7 / 3^8 = 1024/6561 = 0.156$. Välj först ut de 7 bryggor som ska vara i funktion och notera att för var och en av dessa finns 2 möjligheter. Även här är det totala antalet möjligheter 3^8 .
2. $P(A|B) = P(A \cap B)/P(B)$, $P(A \cap B) = P(A)P(B|A)$ och $P(B) = P(B \cap A) + P(B \cap A^c) = P(A)P(B|A) + P(A^c)P(B|A^c)$ ger direkt den sökta formeln.
3. (a) Ja, ty $\sin x \geq 0$ då $0 \leq x \leq \pi$.
- (b) $c = 1/2$, ty $\int_0^\pi \sin x \, dx = 2$.
- (c) $\mu = (1/2) \int_0^\pi x \sin x \, dx = \pi/2 \approx 1.57$, $\sigma^2 = (1/2) \int_0^\pi x^2 \sin x \, dx - \mu^2 = \pi^2/2 - 2 - \pi^2/4 = \pi^2/4 - 2 \Rightarrow \sigma = \sqrt{\pi^2/4 - 2} \approx 0.684$. Dessa integraler är inte så svåra att bestämma om man använder Betas tabell över obestämda integraler. I den ser man att $\int x \sin x \, dx = \sin x - x \cos x$ och att $\int x^2 \sin x \, dx = 2 \cos x + 2x \sin x - x^2 \cos x$.
4. (a) Man kan gott räkna med att varje kast utgör ett oberoende försök och att sannolikheten att det landar med krona upp är $1/2$. Då måste ju $f \sim \text{Bin}(100, 1/2)$.
- (b) Enl c.g.s är f approx N-fördelad med $\mu = 100 \cdot 1/2 = 50$ och $\sigma = \sqrt{100 \cdot 1/2 \cdot 1/2} = 5$. Således gäller att den sökta sannolikheten är ca 0.95, för det är ju så att ca 95% av utfallen av en N-fördelad variabel hamnar inom $\pm 2\sigma$ från μ .
5. Att tätheten är $f(t) = \lambda e^{-\lambda t}$ för $t > 0$ och därför $W \sim \text{Exp}(\lambda)$ följer av att $P(W > t) = P(N(t) = 0) = e^{-\lambda t}$. Här betecknar $N(t)$ antalet impulser fram t.o.m tidpunkten $t > 0$. Nu följer att $E[W] = \int_0^\infty t \lambda e^{-\lambda t} \, dt = 1/\lambda$.
6. (a) $\text{Cov}[X, Y] = 0$, ty X, Y är oberoende. Detta inses av att utfallsrummet är rektangulärt ($0 \leq x \leq 2$, $0 \leq y \leq 1 \Leftrightarrow (x, y) \in [0, 2] \times [0, 1]$) och att tätheten kan faktoriseras i en del som bara innehåller x och en del som bara innehåller y .
- (b) Notera först att X 's täthet är $f(x) = \int_0^1 [x/(4\sqrt{y})] \, dy = x/2$. Härur följer $P(X \leq 1) = \int_0^1 (x/2) \, dx = 1/4$.
- (c) Lika stor p.g.a oberoendet.
7. Ur formelsamlingen (eller Beta) fås att $(n-1)s^2/\sigma^2 \sim \chi^2(n-1)$. Givet är att $n = 30$ och $s = 3.562$. Ur χ^2 -tabell ($n-1 = 29$ frihetsgrader) fås $P(c \leq (n-1)s^2/\sigma^2) = 0.95$ om $c = 17.7084$. Värdena på s , n och c sätts in i $c \leq (n-1)s^2/\sigma^2 \Leftrightarrow \sigma^2 \leq (n-1)s^2/c \Leftrightarrow \sigma \leq s\sqrt{(n-1)/c}$. Det sökta konfidensintervallet är alltså $\sigma \leq 3.562\sqrt{29/17.7084} = 4.56$.
8. (a) Låt $\bar{X}$ och S vara de stokastiska variablerna som svarar mot stickprovets medelvärde $\bar{x}$ och dess standardavvikelse s . Det sökta resultatet följer ifrån ekvivalensen
$$\bar{x} - t_{\alpha/2, n-1} s / \sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2, n-1} s / \sqrt{n} \Leftrightarrow -t_{\alpha/2, n-1} \leq \frac{\bar{x} - \mu}{s / \sqrt{n}} \leq t_{\alpha/2, n-1}$$
och faktumet $(\bar{X} - \mu)/(S/\sqrt{n}) \sim t(n-1)$ (som hittas i formelsamlingen eller i Beta).
- (b) Ja, om n inte är för litet, ty då kommer intervallskattningen p.g.a c.g.s att ha konfidensen $\approx 1 - \alpha$.