Avd. Matematisk statistik

KTH Matematik

TENTAMEN I SF1903 SANNOLIKHETSLÄRA OCH STATISTIK FÖR 3-ÅRIG Media TIMEH MÅNDAGEN DEN 16 AUGUSTI 2010 KL 08.00–13.00.

Examinator: Gunnar Englund , tel. 790 74 16.

Tillåtna hjälpmedel: Läroboken. Formel- och tabellsamling i Matematisk statistik. Räknare.

Införda beteckningar skall förklaras och definieras. Resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Numeriska svar skall anges med minst två siffrors noggrannhet. Tentamen består av 6 uppgifter. Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är preliminärt 20 poäng. Möjlighet att komplettera ges för tentander med 18-19 poäng. Tid och plats för komplettering kommer att anges på kursens hemsida. Det ankommer på dig själv att ta reda på om du har rätt att komplettera.

Tentamen kommer att vara rättad inom tre veckor från skrivningstillfället och kommer att finnas tillgänglig på elevexpeditionen minst sju veckor efter skrivningstillfället.

Uppgift 1

a) För händelserna A, B och C gäller att $P(A \cap B \cap C) = 0.1$, $P(A) = 0.5$ och $P(B \mid A) = 0.4$. Beräkna $P(C \mid A \cap B)$. (5 p)

b) En kontinuerlig stokastisk variabel X har fördelningsfunktionen

$$F_X(x) = \begin{cases} 0 & \text{om } x < 0, \\ c \cdot x^2 & \text{om } 0 \leq x < 2, \\ 1 & \text{om } x \geq 2. \end{cases}$$

Bestäm konstanten c samt beräkna $E(X)$ och $D(X)$. (5 p)

Uppgift 2

Vid en kontroll av ett levererat parti togs 100 enheter ut och kontrollerades. Om högst 5 av dessa var felaktiga godkänns partiet direkt, i annat fall kontrolleras alla enheter. Partiet består av 2000 enheter.

a) Om felkvoten i partiet är 7 %, vad är sannolikheten att partiet godkänns direkt. Väl motiverade approximationer får användas. (6 p)

b) Om partiet godkänns direkt är alltså antalet kontrollerade enheter 100, i annat fall ar antalet kontrollerade enheter 2000 (hela partiet). Beräkna förväntat antal kontrollerade enheter om felkvoten i partiet är 7 %. Väl motiverade approximationer får användas. (4 p)

Uppgift 3

I den lilla banken Spargrisen kan man köpa aktier dels manuellt över disk, dels via telefon. Antag att tiden att registrera en order manuellt är exponentialfördelad med väntevärde 2 minuter, samt att tiden i minuter att registrera en telefonorder har väntevärde 1.5 och varians 2.

Under en dag registrerades 40 manuella order och 60 telefonorder. Beräkna sannolikheten att totaltiden att registrera de manuella orderna under denna dag översteg tiden att registrera telefonorderna. Antag att registreringstiderna för olika order är statistiskt oberoende. Rimliga och välmotiverade approximationer är tillåtna. (10 p)

Uppgift 4

Vid en undersökning av böjhållfasthetens beroende av bränntemperaturen hos gult tegel erhölls följande observationsmaterial på 5 tegelbitar vid temperaturen 700^o och 5 bitar vid temperaturen 800^o.

Temperatur	Böjhållfasthet				
700^o	147	140	121	138	139
800^o	193	227	201	212	207

Antag att slumpmässigheten i data kan beskrivas som normalfördelad med samma standardavvikelse vid båda temperaturerna och oberoende mellan samtliga 10 observationer.

a) Beräkna ett (exakt) 99% konfidensintervall för den systematiska skillnaden i böjhållfasthet för de två temperaturerna. (6 p)

b) Testa om en inverkan av temperaturen på böjhållfastheten kan påvisas. Välj själv en lämplig signifikansnivå. Ange den tydligt, likaså slutsatsen av testet. (4 p)

Uppgift 5

Vid livstidsprovning av elektriska komponenter sätter man n exemplar av komponenten i arbete vid en tidpunkt $t = 0$ och låter dem arbeta under uppsikt tills de upphör att fungera och man registrerar tidpunkterna för detta, dvs livslängderna $x_1, x_2, \cdots, x_n$.

Följande antagande gjordes: Olika exemplars livslängder ses som utfall av oberoende exponentialfördelade stokastiska variabler med väntevärde m.

Beräkna konfidensintervall för m med den approximativa konfidensgraden 90% i följande två fall:

a) Man håller kontinuerlig uppsikt och observerar $x_1, x_2, \cdots, x_n$. Ge ett numeriskt svar då $n = 50$ och man observerat livslängderna (ordnade i storleksordning)

0.1	0.1	0.2	0.9	1.2	1.3	1.8	1.8	2.0	2.0
2.0	2.1	2.1	2.3	2.6	2.7	3.0	3.5	3.6	3.8
3.8	3.8	4.8	5.1	5.9	7.3	7.6	7.7	8.5	8.6
8.8	9.1	12.0	12.8	13.4	13.6	14.1	14.2	14.7	16.3
16.7	16.8	16.9	20.6	22.1	25.7	26.3	32.0	33.3	40.3

Som räknehjälp kan meddelas att $\bar{x} = 9.63$. (5 p)

För att få full poäng på a-delen krävs att Du verkligen utnyttjar att data anses vara exponentialfördelade.

b) Till skillnad från i a-delen observerar man endast <u>antalet</u> komponenter som fortfarande fungerar vid tiden $t = 6$, dvs att 25 av de 50 komponenterna fungerar. (5 p)

Uppgift 6

$x_1, x_2, \cdots, x_n$ är observationer från oberoende Maxwellfördelade stokastiska variabler, dvs från variabler med täthetsfunktionen

$$f(x) = \begin{cases} \dfrac{\sqrt{2}}{\sqrt{\pi}} \dfrac{x^2}{\alpha^{3/2}} e^{-x^2/(2\alpha)} & x > 0 \\ \\ 0 & x \leq 0 \end{cases}$$

där $\alpha > 0$ är en parameter. Denna fördelning har väntevärde $\sqrt{8\alpha/\pi}$ och varians $\alpha(3 - \frac{8}{\pi})$.

a) Härled Maximum Likelihood-skattningen av α samt beräkna denna om vi fått observationerna $x_1 = 1.30, x_2 = 1.47, x_3 = 1.68$. (6 p)

b) Undersök om denna skattning är väntevärdesriktig. Din slutsats skall klart framgå och vara väl motiverad! (4 p)

För att få poäng på b-delen krävs att a-delen är i allt väsentligt korrekt löst.

Avd. Matematisk statistik

KTH Matematik

LöSNINGAR TILL
TENTAMEN I SF1903 (f d 5B2501) SANNOLIKHETSLÄRA OCH STATISTIK MÅNDAGEN DEN 16 AUGUSTI 2010 KL 08.00–13.00.

Uppgift 1

a) Vi har

$$P(C \mid A \cap B) = \{\text{def. av betingning}\} = \frac{P(C \cap A \cap B)}{P(A \cap B)} = \frac{P(A \cap B \cap C)}{P(A \cap B)}$$

$$= \frac{0.1}{P(B \mid A)P(A)} = \frac{0.1}{0.4 \cdot 0.5} = \frac{0.1}{0.2} = \underline{0.5}.$$

b) Eftersom X är kontinuerlig så gäller det att $F_X(2-) = F_X(2) = 1$, dvs att $c \cdot 2^2 = 4c = 1$. Detta ger att $\underline{c = 1/4}$. Vi har således

$$F_X(x) = \begin{cases} 0 & \text{om } x < 0, \\ \frac{1}{4}x^2 & \text{om } 0 \leq x < 2, \\ 1 & \text{om } x \geq 2, \end{cases}$$

vilket ger

$$f_X(x) = \begin{cases} \frac{1}{2}x & \text{om } 0 \leq x < 2, \\ 0 & \text{annars.} \end{cases}$$

Detta ger

$$E(X) = \int_0^2 x \cdot \frac{1}{2}x\,dx = \left[\frac{1}{6}x^3\right]_0^2 = \underline{\frac{4}{3}} \quad \text{och} \quad E(X^2) = \int_0^2 x^2 \cdot \frac{1}{2}x\,dx = \left[\frac{1}{8}x^4\right]_0^2 = 2,$$

vilket ger

$$V(X) = E(X^2) - (E(X))^2 = 2 - \frac{16}{9} = \frac{2}{9} \text{ och således } D(X) = \sqrt{\frac{2}{9}} = \underline{0.471}.$$

Uppgift 2

a) Låt X vara antalet felaktiga enheter i provgruppen. X är approximativt Bin$(100\ , p)$, ty vi tar ut 100 enheter ur en stor population. p är felkvoten, i det här fallet 0.07. Men det gäller då att X är approx Po$(100 \cdot 0.07)$=Po(7). Tabell ger nu direkt att $P(X \leq 5) \approx 0.30$.

b) Sätt Y=antalet kontrollerade enheter. $P(Y = 100) = 0.3$ enligt a) (approximativt) och således $P(Y = 2000) = 0.7$. Väntevärdet blir då $E(Y) = 100 \cdot 0.3 + 2000 \cdot 0.7 = 1430$.

Uppgift 3

Låt X_i = registreringstid för manuell order nr i, $i = 1, \ldots, 40$, $E(X_i) = 2$, $V(X_i) = 4$ enligt texten och formelsamlingen om exponentialfördelningen.

Y_i = registreringstid för telefonorder nr i, $i = 1, \ldots, 60$, $E(Y_i) = 1.5$, $V(Y_i) = 2$. $X = X_1 + \ldots +$

X_{40}, $Y = Y_1 + \ldots + Y_{60}$. Vi har

$$E(X) = E(X_1 + \ldots + X_{40}) = 40 \cdot 2 = 80$$

och p.g.a oberoendet

$$V(X) = V(X_1 + \ldots + X_{40}) = 40 \cdot 4 = 160.$$

På samma sätt fås att

$$E(Y) = 60 \cdot 1.5 = 90, \; V(Y) = 60 \cdot 2 = 120.$$

De stokastiska variablerna X_i är likafördelade och därför ger den *centrala gränsvärdessatsen* att $X \in ApprN(80, \sqrt{160})$. Samma motivering ger $Y \in ApprN(90, \sqrt{120})$.
SÖKT är sannolikheten $P(X > Y) = P(X - Y > 0)$. Med $Z = X - Y$ gäller $E(Z) = 80 - 90 = -10$, $V(Z) = 120 + 160 = 280$, ty X och Y är oberoende. Eftersom Z är en linjärkombination av approximativt normalfördelade stokastiska variabler, så fås ($\Phi(x)$ är fördelningsfunktionen för den standardiserade normalfördelningen)

$$P(X - Y > 0) = P(Z > 0) \approx 1 - \Phi\left(\frac{0 - (-10)}{\sqrt{280}}\right) =$$

$$= 1 - \Phi\left(\frac{10}{\sqrt{280}}\right) \approx 1 - \Phi(0.60) \approx 0.27.$$

SVAR: Den sökta sannolikheten ≈ 0.27.

Uppgift 4

Två oberoende stickprov med N(m_A, σ)- respektive N(m_B, σ)-fördelade observationer. Ett 99%-igt konfidensintervall för $m_A - m_B$ blir med t-metoden (FS 11.2)

$$\bar{x} - \bar{y} \pm t_{0.005}(5 + 5 - 2)s\sqrt{\frac{2}{5}}$$

där $\bar{x} = (147 + 140 + \cdots + 139)/5 = \frac{685}{5} = 137.0$ och $\bar{y} = (193 + 227 + \cdots + 207)/5 = \frac{1040}{5} = 208.0$.
Vidare får vi

$$s_A^2 = \frac{1}{5-1}\left(\sum_{i=1}^{5} x_i^2 - 5 \cdot (\bar{x})^2\right) = \frac{1}{4}\left(94215 - \frac{685^2}{5}\right) = \frac{185}{2}$$

$$s_B^2 = \frac{1}{5-1}\left(\sum_{i=1}^{5} y_i^2 - 5 \cdot (\bar{y})^2\right) = \frac{1}{4}\left(216972 - \frac{1040^2}{5}\right) = 163$$

som ger

$$s^2 = \frac{s_A^2 + s_B^2}{2} = \frac{511}{4}, \quad s = 11.30$$

och vi får intervallet till $137.0 - 208.0 \pm 3.36 \cdot 11.30\sqrt{2/5} = -71.0 \pm 24.0 = \underline{(-95.0, -47.0)}$.

b) Vi tar $H_0 : m_A = m_B$ och $H_1 : m_A \neq m_B$. Vi förkastar H_0 på signifikansnivån 1% eftersom konfidensintervallet i a-delen inte innehåller 0. Slutsatsen är alltså att inverkan av temperaturen är påvisad.

Uppgift 5

a) Vi skattar m med $m^* = \bar{x} = \frac{1}{50}\sum_{i=1}^{50} x_i = 9.63$. Motsvarande stickprovsvariabel $\bar{X}$ är approximativt $N(m, m/\sqrt{50})$ enligt CGS eftersom $E(X) = m$ och $V(X) = m^2$. Vi får då det approximativt 90%-iga konfidensintervallet

$$\bar{x} \pm \lambda_{0.05}\frac{\bar{x}}{\sqrt{50}} \approx 9.63 \pm 1.6449\frac{9.63}{\sqrt{50}} \approx \underline{9.63 \pm 2.24 = (7.39, 11.87)}.$$

b) Låt Y = antalet komponenter som fungerar vid tiden $t = 6$. Y är $\text{Bin}(50, p)$ där $p = P(X \geq 6) = e^{-6/m}$. Vi får eftersom 25 st av observationerna är mindre än 6 att $p^* = 25/50 = 1/2$ och $np(1-p) \approx np^*(1-p^*) = 12.5 > 10$ och således är normalapproximation av $\text{Bin}(50, p)$ tillåten, dvs $\text{Bin}(50, p) \approx N(50p, \sqrt{50p(1-p)})$. Detta ger att p^* är approximativt $N(p, \sqrt{p(1-p)/50})$.
Alltså är ett approximativt 90%-igt konfidensintervall för p

$$p^* \pm \lambda_{0.05}\sqrt{\frac{p^*(1-p^*)}{50}} \approx 0.5 \pm 0.117 = (0.383, 0.617).$$

Eftersom $p = e^{-6/m}$ dvs $m = -6/\ln p$ blir konfidensintervallet för m

$$\left(-\frac{6}{\ln 0.383}, -\frac{6}{\ln 0.617}\right) \approx \underline{(6.25, 12.43)}.$$

Man kan notera att konfidensintervallet i b-delen blir lite bredare och detta beror ju på att vi i b-delen utnyttjar mindre av informationen i stickprovet än i a-delen.

Uppgift 6

Vi har likelihoodfunktionen

$$L(\alpha) = \prod_{i=1}^{n} f(x_i) = \left(\sqrt{\frac{2}{\pi}}\right)^n \frac{x_1^2 x_2^2 \cdots x_n^2}{\alpha^{3n/2}} \exp\left(-\frac{1}{2\alpha}\sum_{i=1}^{n} x_i^2\right)$$

som ger

$$g(\alpha) = \ln L(\alpha) = \ln\left((2/\pi)^{n/2} x_1^2 x_2^2 \cdots x_n^2\right) - \frac{3n}{2}\ln(\alpha) - \frac{1}{2\alpha}\sum_{i=1}^{n} x_i^2.$$

$g(\alpha)$ maximeras då $g'(\alpha) = 0$ och vi har

$$g'(\alpha) = -\frac{3n}{2\alpha} + \frac{1}{2\alpha^2}\sum_{i=1}^{n} x_i^2.$$

$g'(\alpha^*) = 0$ ger att ML-skattningen blir $\alpha^* = \dfrac{1}{3n}\sum_{i=1}^{n} x_i^2$. Med de observerade värdena insatta blir $\alpha^* = \dfrac{1}{9}(1.30^2 + 1.47^2 + 1.68^2) = \underline{0.7415}$.

b) Vi får

$$E(\alpha^*) = E\left(\frac{1}{3n}\sum_{i=1}^{n} X_i^2\right) = \frac{1}{3n}\sum_{i=1}^{n} E(X_i^2) = \frac{1}{3}E(X_1^2).$$

Eftersom $V(X) = E(X^2) - (E(X))^2$ ser vi att $E(X^2) = V(X) + (E(X))^2$ som ger att

$$E(X_1^2) = V(X_1) + (E(X_1))^2 = \left(3 - \frac{8}{\pi}\right)\alpha + \left(\sqrt{\frac{8\alpha}{\pi}}\right)^2 = 3\alpha.$$

Alltså gäller att $E(\alpha^*) = \dfrac{1}{3}E(X_1^2) = \alpha$, d.v.s. $\underline{\alpha^* \text{ är en väntevärdesriktig skattning av } \alpha}$.