



Avd. Matematisk statistik

KTH Matematik

TENTAMEN I 5B1504 MATEMATISK STATISTIK GRUNDKURS FÖR E (gamlingar)
LÖRDAGEN DEN 28 AUGUSTI 2004 KL 8.00–13.00

Examinator: Gunnar Englund, 790 7416

Tillåtna hjälpmedel: Formel- och tabellsamling i matematisk statistik för E. Beta Mathematics Handbook. Kalkylator.

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.

Varje korrekt lösning ger 10 poäng. Gränsen för godkänt är preliminärt 20 poäng.

Resultatet anslås senast fredagen den 17 september 2004 på Matematisk statistiks anslags-tavla i entréplanet, Lindstedtsvägen 25, rakt fram innanför porten och tentamen kommer att finnas tillgänglig på elevexpeditionen under sju veckor efter tentamensdagen.

Uppgift 1

a) Den stokastiska variabeln X kan endast anta värdena 0, 1, 2, 3 och 4. Värdena 0, 1 respektive 2 antas med sannolikheterna 0.2, 0.3 respektive 0.1. Vidare gäller att X har väntevärde 1.8. Bestäm sannolikhetsfunktionen för X samt variansen för X . (5 p)

b) Volymen v av sfäriska bubblor kan beräknas genom att mäta bubblornas diameter d på en fotografisk plåt och tillämpa formeln $v = \pi d^3/6$. Antag att bestämningen av d är behäftad med slumpfel så att den kan betraktas som en stokastisk variabel med väntevärde d och standardavvikelse 0.02 mm. Bestäm med lämplig approximation volymsbestämningens precision (standardavvikelse) när den uppmätta diametern i verkligheten är 1.80 mm. (5 p)

Uppgift 2

I ett företag tillverkas rektangulära komponenter som senare i produktionen placeras 5 och 5 i rad i en produkt. Längden av en komponent är en stokastisk variabel X där X är $N(1, \sigma)$ (cm), och längderna av olika komponenter är oberoende. σ är ett mått på tillverkningsprecisionen. Toleranskravet är att den sammanlagda längden av de 5 komponenterna ligger i intervallet 5 ± 0.05 med 99.9% sannolikhet.

a) Hur stort får σ högst vara för att detta krav skall vara uppfyllt? (5 p)

b) Även om X inte antas vara normalfördelad så kan man med en ”känd” olikhet (dvs som återfinns i formelsamlingen) ge en övre gräns för hur stor standardavvikelsen σ för X kan vara för att toleranskravet skall vara uppfyllt oavsett vilken fördelning X än har, om bara $E(X) = 1$ och $D(X) = \sigma$. Bestäm en sådan övre gräns. (5 p)

Uppgift 3

I ett kemiskt försök avser man att mäta mängden kol i n prover som har vikterna $u_1, u_2, \dots, u_n$ (kända positiva tal). De exakta kolmängderna i de n proverna är $\beta u_1, \beta u_2, \dots, \beta u_n$ där β

är ett okänt tal. Med den analysmetod man använder så kan man anta att de uppmätta kolmängderna $y_1, y_2, \dots, y_n$ i de n proverna är utfall av oberoende stokastiska variabler $Y_1, Y_2, \dots, Y_n$, där Y_i är $N(\beta u_i, \omega u_i)$, $i = 1, 2, \dots, n$. Här är ω ett känt positivt tal.

a) På intuitiva grunder kom man på skattningen

$$\beta^* = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{u_i}.$$

av β . Undersök om denna är väntevärdesriktig samt beräkna dess varians. (4 p)

b) Bestäm maximum-likelihood-skattningen av β . (6 p)

Uppgift 4

Livslängden av en viss sorts komponenter kan anses vara log-normal-fördelade(m, σ), dvs om X är livslängden så är $Y = \ln X$ normalfördelad $N(m, \sigma)$. För att skatta m undersöktes livslängden hos 10 komponenter med följande resultat (i timmar):

112 161 91 111 115 63 122 153 58 99

a) Beräkna ett 95%-igt konfidensintervall för m . (7 p)

b) En intressant storhet i detta sammanhang är $a = e^{m+\sigma^2/2}$. Ge en punktskattning av a . (3 p)

Uppgift 5

Vid tillverkning av en typ av komponenter gäller att två egenskaper varierar som en tvådimensionell stokastisk variabel $\underline{X}$ som är $N_2(\underline{m}, C)$ där

$$\underline{m} = \begin{pmatrix} 10 \\ 15 \end{pmatrix} \text{ och } C = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 4 \end{pmatrix}.$$

Bestäm den betingade fördelningen för X_1 givet $X_1 + X_2 = 30$ samt beräkna $P(X_1 \geq 10 | X_1 + X_2 = 30)$. (10 p)

Uppgift 6

Två oberoende signaler matas in i en multiplicerande krets. De bägge signalerna kan beskrivas av två oberoende stationära stokastiska processer $(X(t); t \in R)$ och $(Y(t); t \in R)$ med väntevärden $m_X = 3$ och $m_Y = 2$ och med spektraltätheten

$$S_X(\omega) = S_Y(\omega) = e^{-\omega^2}, \quad \omega \in R.$$

Utprocessen $(Z(t); t \in R)$ blir alltså $Z(t) = X(t)Y(t)$. Beräkna väntevärdet m_Z och spektraltätheten $S_Z(\omega)$, $\omega \in R$ för denna utprocess. (10 p)



Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I 5B1504 LÖRDAGEN DEN 28 AUGUSTI 2004 KL 8.00–13.00

Uppgift 1

Låt $p_X(3) = a$ och $p_X(4) = b$ där vi först söker a och b . Eftersom totala sannolikhetsmassan=1 så får vi

$$1 = \sum_{k=-\infty}^{\infty} p_X(k) = \sum_{k=0}^4 p_X(k) = 0.2 + 0.3 + 0.1 + a + b$$

dvs $a + b = 0.4$. Vidare är enligt texten $E(X) = 1.8$ som ger

$$1.8 = \sum_{k=-\infty}^{\infty} k p_X(k) = 0.5 + 3a + 4b$$

dvs $3a + 4b = 1.3$. Dessa två samband ger $a = 0.3$ och $b = 0.1$. Vidare är

$$E(X^2) = \sum_{k=0}^4 k^2 p_X(k) = 0^2 \cdot 0.2 + 1^2 \cdot 0.3 + 2^2 \cdot 0.1 + 3^2 \cdot 0.3 + 4^2 \cdot 0.1 = 5.0$$

Detta ger $V(X) = E(X^2) - (E(X))^2 = 5.0 - 1.8^2 = \underline{1.76}$.

b) Låt vår mätning av d vara X . Enligt texten är $E(X) = d$ och $D(X) = 0.02$. Bestämningen av v blir då $Y = g(X)$ där $g(x) = \pi x^3/6$. Gauss' approximationsformler ger nu eftersom $g'(x) = \pi x^2/2$ att $V(Y) \approx V(X)(g'(E(X)))^2 = 0.02^2(\pi 1.80^2/2)^2$ som ger $D(Y) = \sqrt{V(Y)} \approx \underline{0.102}$.

Uppgift 2

a) Låt X_i vara längden av komponent i , $i = 1, 2, 3, 4, 5$. Enligt texten är X_i då $N(1, \sigma)$ och sammanlagda längden $S_5 = \sum_{i=1}^5 X_i$ blir då också normalfördelad eftersom X_i -na var normalfördelade och oberoende.

$$E(S_5) = E\left(\sum_{i=1}^5 X_i\right) = (\text{linjäriteten}) = \sum_{i=1}^5 E(X_i) = 5$$

$$V(S_5) = V\left(\sum_{i=1}^5 X_i\right) = (\text{oberoendet}) = \sum_{i=1}^5 V(X_i) = 5\sigma^2$$

och alltså är S_5 $N(5, \sigma\sqrt{5})$ -fördelad. Vi vill bestämma σ så att $P(|S_5 - 5| \leq 0.05) \geq 0.999$ dvs att

$$P\left(\left|\frac{S_5 - 5}{\sigma\sqrt{5}}\right| \leq \frac{0.05}{\sigma\sqrt{5}}\right) \geq 0.999.$$

Eftersom $(S_5 - 5)/\sigma\sqrt{5}$ är $N(0, 1)$ ger detta att

$$\frac{0.05}{\sigma\sqrt{5}} \geq \lambda_{0.0005} \iff \sigma \leq \frac{0.05}{\sqrt{5} \cdot 3.2905} \approx \underline{0.0068}.$$

b) $E(S_5) = 5$ och $D(S_5) = \sigma\sqrt{5}$ gäller fortfarande. Enligt formelsamlingen avsnitt 6 ger Tjebyskovs olikhet att för varje $k > 0$ gäller att

$$P(|S_5 - 5| > k\sigma\sqrt{5}) \leq \frac{1}{k^2}$$

Välj nu $k = 0.05/\sigma\sqrt{5}$ så får vi

$$P(|S_5 - 5| > 0.05) \leq \frac{5\sigma^2}{0.05^2}$$

Kravet är alltså uppfyllt så länge $5\sigma^2/0.05^2 \leq 0.001$ dvs då $\sigma^2 \leq 0.05^2 \cdot 0.0002$ som ger $\sigma \leq \underline{0.00071}$.

Uppgift 3

a) Betrakta stickprovsvariabeln svarande mot β^* , dvs

$$\beta^* = \frac{1}{n} \sum_{i=1}^n \frac{Y_i}{u_i}.$$

Vi har

$$E(\beta^*) = E\left(\frac{1}{n} \sum_{i=1}^n \frac{Y_i}{u_i}\right) = \frac{1}{n} \sum_{i=1}^n \frac{E(Y_i)}{u_i} = \frac{1}{n} \sum_{i=1}^n \frac{\beta u_i}{u_i} = \frac{1}{n} \sum_{i=1}^n \beta = \beta$$

dvs β^* är väntevärdesriktig. På liknande sätt fås

$$\text{Var}(\beta^*) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}\left(\frac{Y_i}{u_i}\right) = \frac{1}{n^2} \sum_{i=1}^n \frac{\text{Var}(Y_i)}{u_i^2} = \frac{1}{n^2} \sum_{i=1}^n \frac{\omega^2 u_i^2}{u_i^2} = \frac{\omega^2}{n}.$$

b) Eftersom

$$f_{Y_i}(y) = \frac{1}{\sqrt{2\pi\omega u_i}} e^{-\frac{1}{2}\left(\frac{y - \beta u_i}{\omega u_i}\right)^2}$$

Så fås likelihoodfunktionen

$$L(\beta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\omega u_i}} e^{-\frac{1}{2}\left(\frac{y_i - \beta u_i}{\omega u_i}\right)^2}.$$

Detta ger

$$\ln L(\beta) = \ln\left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi\omega u_i}}\right) - \sum_{i=1}^n \frac{1}{2} \left(\frac{y_i - \beta u_i}{\omega u_i}\right)^2$$

vilket medför att

$$\frac{d \ln L(\beta)}{d\beta} = 0 - \frac{1}{2} \sum_{i=1}^n 2 \left(\frac{y_i - \beta u_i}{\omega u_i}\right) \left(-\frac{u_i}{\omega u_i}\right).$$

Sätter vi $\frac{d \ln L(\beta)}{d\beta} = 0$ fås

$$\begin{aligned} \frac{1}{2} \sum_{i=1}^n 2 \left(\frac{y_i - \beta u_i}{\omega u_i} \right) \left(-\frac{u_i}{\omega u_i} \right) &= 0 \Leftrightarrow \sum_{i=1}^n \frac{y_i - \beta u_i}{u_i} = 0 \\ \Leftrightarrow \sum_{i=1}^n \frac{y_i}{u_i} &= \beta \sum_{i=1}^n 1 (= \beta n) \Leftrightarrow \beta = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{u_i}. \end{aligned}$$

Eftersom man lätt visar att extremvärdet är ett maximum så gäller att maximum-likelihood-skattningen av β är

$$\beta^* = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{u_i},$$

dvs skattningen i a-delen.

Uppgift 4

De logaritmerade data $y_i = \ln x_i$ är alltså observationer från oberoende $N(m, \sigma)$ -fördelade stokastiska variabler. De logaritmerade data blir

4.718 5.081 4.511 4.710 4.745 4.143 4.804 5.030 4.060 4.595

och härav får $\bar{y} = 4.640$ och $s_y^2 = 0.1108$. Ett 95%-igt konfidensintervall med t-metoden ges av $\bar{y} \pm t_{0.025}(9)s_y/\sqrt{10}$ vilket blir 4.640 ± 0.238 .

b) a skattas med $a^* = e^{\bar{y} + s_y^2/2}$. Detta ger $a^* = e^{4.640 + 0.1108/2} \approx 109.44$. Anm: Man kan visa att $a = E(X)$. En annan men sämre skattning är alltså $\hat{a} = \bar{x} = 108.5$.

Uppgift 5

Vi bestämmer först fördelningen för

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_1 + X_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = B \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

som blir

$$N_2(B\bar{m}, BCB^T) = N_2\left(\begin{pmatrix} 10 \\ 25 \end{pmatrix}, \begin{pmatrix} 1 & 1.5 \\ 1.5 & 6 \end{pmatrix}\right)$$

Fördelningen för Y_1 givet $Y_2 = 30$ ges av

$$N\left(m_1 + \frac{c_{12}}{c_{22}}(30 - m_2), \sqrt{c_{11} - \frac{c_{12}^2}{c_{22}}}\right) = N(11.25, \sqrt{0.625})$$

som ger att

$$P(X_1 \geq 10 | X_1 + X_2 = 30) = 1 - \Phi\left(\frac{10 - 11.25}{\sqrt{0.625}}\right) \approx \underline{0.9429}.$$

Uppgift 6

$$m_Z = E(X(t)Y(t)) = (\text{oberoendet}) = E(X(t))E(Y(t)) = 3 \cdot 2 = \underline{6}.$$

För att beräkna $S_Z(\omega)$ beräknar vi först $r_Z(\tau)$. $S_X(\omega) = e^{-\omega^2}$ ger att

$$r_X(\tau) = r_Y(\tau) = \frac{1}{\sqrt{2\pi}\sqrt{2}}e^{-\tau^2/4}$$

och vi får vidare att

$$r_Z(\tau) = E(Z(t+\tau)Z(t)) - m_Z^2 = E(X(t+\tau)Y(t+\tau)X(t)Y(t)) - 6^2 = (\text{oberoendet}) =$$

$$E(X(t+\tau)X(t))E(Y(t+\tau)Y(t)) - 36 = (r_X(\tau) + m_X^2)(r_Y(\tau) + m_Y^2) - 36 =$$

$$\left(\frac{1}{\sqrt{2\pi}\sqrt{2}}e^{-\tau^2/4} + 3^2\right)\left(\frac{1}{\sqrt{2\pi}\sqrt{2}}e^{-\tau^2/4} + 2^2\right) - 36 = \frac{1}{2\sqrt{2\pi}\sqrt{2\pi}}e^{-\frac{\tau^2}{2 \cdot 1}} + 13\frac{1}{\sqrt{2\pi}\sqrt{2}}e^{-\frac{\tau^2}{2 \cdot 2}}$$

Inverstranformering ger nu att

$$S_Z(\omega) = \frac{1}{2\sqrt{2\pi}}e^{-\frac{\omega^2}{2}} + 13e^{-\omega^2}.$$