Avd. Matematisk statistik

KTH Matematik

TENTAMEN I 5B1504 MATEMATISK STATISTIK GRUNDKURS FÖR E (gamlingar)
TISDAGEN DEN 16 DECEMBER 2003 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 onsdagen den 21 januari 2004 på Matematisk statistiks anslagstavla i entréplanet, Lindstedtsvägen 25, rakt fram innanför porten och tentamen kommer att finnas tillgänglig på elevexpeditionen under sju veckor efter tentamensdagen.

Uppgift 1

a) För en viss region gäller följande: 20% av befolkningen är rökare. Sannolikheten att en rökare får lungcancer är 10 gånger så hög som att en icke-rökare får lungcancer. Antag att sannolikheten att få lungcancer för en slumpmässigt vald person är 0.006. Vad är då sannolikheten att en person får lungcancer *givet* att personen i fråga är rökare? (5 p)

b) Beräkna väntevärde och varians för produkten

$$X_1 \cdot X_2 \cdots X_{10}$$

där alla X_i:na är oberoende och $R(0, 2)$.

(5 p)

Uppgift 2

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 antalet komponenter som fortfarande fungerar vid tiden $t = 6$, dvs att 25 av de 50 komponenterna fungerar. (5 p)

Uppgift 3

En kemisk mätmetod uppges ge mätvärden som kan ses som utfall av oberoende stokastiska variabler, $N(m, \sigma)$-fördelade, där m är det sanna, okända värdet och där σ uppges vara 0.3. En kemist har gjort 4 mätningar och fått genomsnittsvärdet $\bar{x} = 37.15$.

a) Beräkna med utnyttjande av förutsättningarna ovan ett 95%-igt konfidensintervall för m. (3 p)

b) Kemisten blir dock misstänksam mot det uppgivna σ-värdet och skattar σ ur sina 4 värden med resultatet $s = 0.95$. Beräkna ett 95%-igt konfidensintervall för m utan att utnyttja det uppgivna σ-värdet 0.3. (3 p)

c) Kemisten önskar nu få klarhet i om det uppgivna σ-värdet var fel. Beräkna därför ett 95%-igt konfidensintervall för σ och undersök med hjälp av detta om det uppgivna σ-värdet kan förkastas. Det skall klart framgå om σ-värdet 0.3 förkastas eller ej. (4 p)

Uppgift 4

Vitt brus $(X(t); t \in R)$ med $EX(t) = 3$ och $S_X(\omega) = 5$ påverkar ett mekaniskt system. Processen $(Y(t); t \in R)$ beskriver systemets rörelser och är definierad genom sambandet

$$Y''(t) + 2Y'(t) + 4Y(t) = X(t),\ t \in R$$

Beräkna väntevärde, spektraltäthet och kovariansfunktion för $(Y(t); t \in R)$ (10 p)

Uppgift 5

I ett elektriskt system beskrivs de kraftiga störningarna av en Poissonprocess $\{N(t);\ t \geq 0\}$ med intensitet λ.

Beräkna $P(N(1) < N(2) < N(3))$. (10 p)

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I 5B1504 TISDAGEN DEN 16 DECEMBER 2003 KL 8.00–13.00

Uppgift 1

a) Inför beteckningarna $R =$ "personen är rökare" resp. $C =$ "personen får cancer". Enligt texten är $P(R) = 0.2$, $P(C) = 0.006$ och $P(C|R) = 10 \cdot P(C|R^*)$.
Lagen om total sannolikhet ger

$$P(C) = P(C|R)P(R) + P(C|R^*)P(R^*)$$

dvs

$$0.006 = P(C|R) \cdot 0.2 + \frac{1}{10}P(C|R) \cdot (1 - 0.2) = 0.28 \cdot P(C|R)$$

och således

$$P(C|R) = \frac{0.006}{0.28} \approx 0.021.$$

b)

$$\begin{aligned} E(X_1X_2\cdots X_{10}) = &\ \{\text{oberoende}\} \ = E(X_1)E(X_2)\cdots E(X_{10}) \\ &= \{X_i \in R(0,2),\ \text{se FS}\ \} = 1^{10} = 1. \end{aligned}$$

$$\begin{aligned} V(X_1X_2\cdots X_{10}) &= E\left((X_1X_2\cdots X_{10})^2\right) - \left(E(X_1)E(X_2)\cdots E(X_{10})\right)^2 = \{\text{oberoende}\} \\ &= E(X_1^2)E(X_2^2)\cdots E(X_{10}^2) - 1^2 \\ &= \left(V(X_1) + E(X_1)^2\right)\left(V(X_2) + E(X_2)^2\right)\cdots\left(V(X_{10}) + E(X_{10})^2\right) - 1 \\ &= \{X_i \in R(0,2), \text{ se FS}\} = \left(\frac{1}{3} + 1\right)^{10} - 1 = \left(\frac{4}{3}\right)^{10} - 1 \approx 16.76. \end{aligned}$$

Uppgift 2

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 $\mathrm{Bin}(50, p)$ 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 3

a) Enligt λ-metoden (FS 11.1) får vi konfidensintervallet

$$\bar{x} \pm \lambda_{0.025}\frac{\sigma}{\sqrt{n}} = 37.15 \pm 1.96 \cdot \frac{0.3}{\sqrt{4}} \approx 37.15 \pm 0.3$$

b) Enligt t-metoden (FS 11.2) får vi konfidensintervallet

$$\bar{x} \pm t_{0.025}(4-1)\frac{s}{\sqrt{n}} = 37.15 \pm 3.18 \cdot \frac{0.95}{\sqrt{4}} \approx 37.15 \pm 1.5$$

c) Enligt χ^2-metoden (FS 11.4) får vi för σ konfidensintervallet

$$\left(s\sqrt{\frac{n-1}{\chi^2_{0.025}(n-1)}}, s\sqrt{\frac{n-1}{\chi^2_{0.975}(n-1)}}\right) =$$

$$= \left(0.95\sqrt{\frac{3}{\chi^2_{0.025}(3)}}, 0.95\sqrt{\frac{3}{\chi^2_{0.975}(3)}}\right) \approx (0.54, 3.51).$$

Eftersom 0.3 ligger utanför detta intervall förkastas $H_0 : \sigma = 0.3$ på nivån 5%.

Uppgift 4

För testfunktionen $x(t) = e^{i\omega t}$ fås utsignalen $y(t) = H(\omega)e^{i\omega t}$, dvs

$$H(\omega)(i\omega)^2 e^{i\omega t} + 2H(\omega)i\omega e^{i\omega t} + 4H(\omega)e^{i\omega t} = e^{i\omega t}$$

som ger $H(\omega) = 1/(2i\omega + 4 - \omega^2)$. Alltså får vi $m_Y = E(Y(t)) = H(0)E(X(t)) = 3/4 = \underline{0.75}$.

$$S_Y(\omega) = |H(\omega)|^2 S_X(\omega) = \frac{5}{4\omega^2 + (4-\omega^2)^2}$$

Med hjälp av inverstranformering erhålls ur formelsamlingen 16 2:a tranformparet med $a = 1$ och $b = \sqrt{3}$

$$r_Y(\tau) = \underline{5\frac{e^{-|\tau|}}{4 \cdot 4}\left(\cos(\sqrt{3}\tau) + \frac{\sin(\sqrt{3}|\tau|)}{\sqrt{3}}\right)}$$

Uppgift 5

Vi har

$$P(N(1) < N(2) < N(3)) = (\text{processen är växande}) = P(N(2) - N(1) > 0,\ N(3) - N(2) > 0)$$

$$= (\text{oberoende tillskott}) = P(N(2) - N(1) > 0)P(N(3) - N(2) > 0) = \underline{\left(1 - e^{-\lambda}\right)^2}.$$