Avd. Matematisk statistik

KTH Matematik

TENTAMEN I 5B1504 MATEMATISK STATISTIK GRUNDKURS FÖR E (gamlingar)
FREDAGEN DEN 1 APRIL 2005 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 22 april 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

Antalet uppdrag som tre kunder ger ett dataföretag under en månad beskrivs av de oberoende stokastiska variablerna X_1, X_2 respektive X_3, där X_1 är Po(25), X_2 är Po(10) och X_3 är Po(12).

Beräkna approximativt sannolikheten att den första kunden ger fler uppdrag än de två sista tillsammans. (Gjorda approximationer skall naturligtvis motiveras.) (10 p)

Uppgift 2

Avståndet till en strålningskälla beskrivs av en Rayleigh-fördelad stokastisk variabel X, det vill säga en stokastisk variabel med täthetsfunktion

$$f_X(x) = \frac{x}{b^2}\,\mathrm{e}^{-(x/b)^2/2}, \quad x \geq 0$$

där parametern $b > 0$.

a) Tio oberoende avståndsmätningar $x_1, \ldots, x_n$ sammanfattas av $\sum_{i=1}^{n} x_i^2 = 80.72$ [enhet m^2]. Tag fram Maximum-likelihoodskattningen av b. (5 p)

b) Strålningen från strålningskällan sprids över ett område av storlek $Y = \pi X^2$. Bestäm täthetsfunktionen för Y. (5 p)

Notera att a- och b-uppgiften kan/skall lösas separat, det vill säga man kan lösa den ena utan att lösa den andra.

Uppgift 3

Vid en auktion kan försäljningspriset y antas beskrivas av en normalfördelad stokastisk variabel Y där väntevärdet beror av utropspriset x enligt modellen:

$$Y(x) \text{ är } \mathrm{N}(\alpha' + \beta x, \sigma)$$

a) Baserat på följande $n = 10$ försäljningsdata

Utropspris, x_i:	900	600	1000	1100	700	1300	1300	1000	1100	1000
Försäljningspris, y_i:	1983	969	2284	2732	1337	2852	2971	1592	1813	2241

som sammanfattas av

$$\overline{x} = 1000 \qquad \overline{y} = 2077.4 \qquad S_{yy} = \sum_{i=1}^{n}(y_i - \overline{y})^2 = 3987650.4$$

$$S_{xx} = \sum_{i=1}^{n}(x_i - \overline{x})^2 = 460000 \qquad S_{xy} = \sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = 1214400$$

skatta modellens parametrar α', β och σ. (5 p)

b) En vara har utropspris $x_0 = 1100$. Dess försäljningspris $Y(x_0)$ är en stokastisk variabel med väntevärde $\alpha' + \beta x_0$. Om s är skattningen av σ kan man visa att

$$\frac{Y(x_0) - (\alpha'^* + \beta^* x_0)}{S\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}}$$

är en $t(n-2)$-fördelad stokastisk variabel.
Använd detta för att ta fram ett 95% *prediktionsintervall* för försäljningspriset $Y(x_0)$, det vill säga en observation av ett stokastiskt intervall I sådant att $\mathrm{P}\left(Y(x_0) \in I\right) = 0.95$. (5 p)

Uppgift 4

Låt $(X(t);\ t \geq 0)$ vara en standard Wienerprocess, dvs en normalprocess med väntevärdesfunktion $E(X(t)) = 0$ och kovarianskärna $C(X(s), X(t)) = \min(s,t)$. Man skall registrera processens utfall vid tidpunkterna $t = 1$ och $t = 3$ och skall sedan interpolera fram $X(2)$ med hjälp av värdet på $\left(X(1) + X(3)\right)/2$.
Man var intresserad av ”interpolationsfelet” och speciellt av

$$P\left(2 \leq X(2) \leq 3 \,\middle|\, \frac{X(1) + X(3)}{2} = 2.5\right).$$

Beräkna denna sannolikhet. (10 p)

Uppgift 5

$(X(t);\, t \in R)$ är en svagt stationär process med väntevärde $m_X = 2$ och kovariansfunktion $r_X(\tau) = e^{-|\tau|}$, $\tau \in R$ som filtreras i ett filter där utsignalen $(Y(t);\ t \in R)$ bestäms av relationen

$$Y'(t) + 3Y(t) = \int_0^\infty e^{-2v} X(t-v)\,dv.$$

Bestäm Y-processens väntevärde och kovariansfunktion samt $V(Y(3))$. (10 p)

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG TENTAMEN I 5B1504 FREDAGEN DEN 1 APRIL 2005 KL 8.00–13.00

Uppgift 1

Vi söker $P(X_1 > X_2 + X_3) = P(X_1 - (X_2 + X_3) > 0)$.
$X_2 + X_3$ är $\text{Po}(10 + 12)$=$\text{Po}(22)$ eftersom de är oberoende och Poisson-fördelade.
Vidare är $22 \geq 15$ så vi får göra approximationen $\text{Po}(22) \approx N(22, \sqrt{22})$. På samma sätt är $\text{Po}(25) \approx N(25, \sqrt{25})$.
Vi får alltså att $X_1 - (X_2 + X_3)$ är approximativt normalfördelad
$N(25 - 22, \sqrt{25 + 22})$=$N(3, \sqrt{47})$. Detta ger

$$P(X_1 > X_2 + X_3) = P\left(\frac{X_1 - (X_2 + X_3) - 3}{\sqrt{47}} > \frac{0 - 3}{\sqrt{47}}\right) \approx$$

$$\approx 1 - \Phi\left(-\frac{3}{\sqrt{47}}\right) = \Phi\left(\frac{3}{\sqrt{47}}\right) \approx \Phi(0.44) \approx \underline{0.67}.$$

Uppgift 2

Likelihoodfunktionen

$$\begin{aligned} L(b) &= f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \{\text{ober.}\} = f_{X_1}(x_1) \cdots f_{X_n}(x_n) = \frac{x_1}{b^2}\,\mathrm{e}^{-x_1^2/2b^2} \cdots \frac{x_n}{b^2}\,\mathrm{e}^{-x_n^2/2b^2} \\ &= \frac{x_1 \cdots x_n}{b^{2n}} \exp\left\{-\frac{1}{2b^2}\sum_{i=1}^{n} x_i^2\right\} \end{aligned}$$

har logaritm

$$\ln(L(b)) = \sum_{i=1}^{n} \ln(x_i) - 2n\ln(b) - \frac{1}{2b^2}\sum_{i=1}^{n} x_i^2.$$

Lösning av

$$0 = \frac{d}{db}\ln(L(b)) = -\frac{2n}{b} + \frac{1}{b^3}\sum_{i=1}^{n} x_i^2 = \frac{n}{b}\left[-2 + \frac{1}{b^2}\frac{1}{n}\sum_{i=1}^{n} x_i^2\right]$$

ger ML-skattningen

$$b^*(x_1, \ldots, x_n) = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} x_i^2} = \sqrt{\frac{80.72}{20}} = \underline{2.009}.$$

b) Med $Y = \pi X^2$ så är för $t > 0$

$$F_Y(t) = \mathrm{P}\left(Y \leq t\right) = \mathrm{P}\left(\pi X^2 \leq t\right) = \mathrm{P}\left(X \leq \sqrt{t/\pi}\right) = F_X(\sqrt{t/\pi})$$

och

$$\begin{aligned}
f_Y(t) &= \frac{d}{dt}F_Y(t) = \frac{d}{dt}F_X(\sqrt{t/\pi}) = f_X(\sqrt{t/\pi})\frac{1}{2\sqrt{\pi t}} = \frac{\sqrt{t/\pi}}{b^2}\mathrm{e}^{-\sqrt{t/\pi}^2/2b^2}\frac{1}{2\sqrt{\pi t}} \\
&= \underline{\frac{1}{2\pi b^2}\mathrm{e}^{-t/2\pi b^2}} = \frac{1}{m}\mathrm{e}^{-t/m}
\end{aligned}$$

där $m = 2\pi b^2$. Alltså, arean Y är exponentialfördelad med väntevärde $2\pi b^2$.

Uppgift 3

Med parametriseringen $Y(x_i)$ är $\mathrm{N}(\alpha + \beta(x_i - \overline{x}), \sigma)$ så skattas α, β och σ^2 med

$$\begin{aligned}
\alpha^* &= \overline{y} = 2077.4 \\
\beta^* &= S_{xy}/S_{xx} = 1214400/460000 = \underline{2.64} \\
s^2 &= \tfrac{1}{n-2}(S_{yy} - \beta^* S_{xy}) = \tfrac{1}{10-2}(3987650.4 - 2.64 \cdot 1214400) = 97704,
\end{aligned}$$

respektive. Alltså, σ skattas med $s = \underline{312.58 \text{ kronor}}$ och $\alpha' = \alpha - \beta\overline{x}$ med

$$(\alpha')^* = \alpha^* - \beta^*\overline{x} = 2077.4 - 2.64 \cdot 1000 = \underline{-562.6 \text{ kronor}}.$$

b) Låt $D_0 = S\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}$. Ur $t(n-2) = t(8)$-fördelningen fås att $t_{0.025} = 2.31$ är sådan att

$$\begin{aligned}
0.95 &= \mathrm{P}\left(-t_{0.025} \leq \frac{Y(x_0) - (\alpha'^* + \beta^* x_0)}{D_0} \leq t_{0.025}\right) \\
&= \mathrm{P}\left(-t_{0.025}D_0 \leq Y(x_0) - (\alpha'^* + \beta^* x_0) \leq t_{0.025}D_0\right) \\
&= \mathrm{P}\left((\alpha'^* + \beta^* x_0) - t_{0.025}D_0 \leq Y(x_0) \leq (\alpha'^* + \beta^* x_0) + t_{0.025}D_0\right) \\
&= \mathrm{P}\left(Y(x_0) \in (\alpha'^* + \beta^* x_0) \pm t_{0.025}S\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}}\right).
\end{aligned}$$

Alltså ett 95% prediktionsintervall ges av

$$\begin{aligned}
Y(x_0) &\in (\alpha'^* + \beta^* x_0) \pm t_{0.025}s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}} \\
&= (-562.6 + 2.64 \cdot 1100) \pm 2.31 \cdot 312.58\sqrt{1 + \frac{1}{10} + \frac{(100)^2}{460000}} \\
&= \underline{2341.4 \pm 764.74 = (1576.7,\ 3106.1) \text{ kronor}}.
\end{aligned}$$

Uppgift 4

Vi får att

$$\begin{pmatrix} X(1) \\ X(2) \\ X(3) \end{pmatrix} \text{ är } N_3\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}\right)$$

Låt $Y_1 = (X(1) + X(3))/2 = (1/2\ 0\ 1/2)(X(1)\ X(2)\ X(3))^T$ och $Y_2 = X(2)$. Vi får då $E(Y_1) = 0$, $E(Y_2) = 0$, $V(Y_2) = C(2,2) = 2$ och enligt momentsatsen

$$V(Y_1) = (1/2\ 0\ 1/2)\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}\begin{pmatrix} 1/2 \\ 0 \\ 1/2 \end{pmatrix} = 3/2$$

Vidare är (denna gång med direkt kalkyl ur kovarianskärnan som omväxling)

$$C(Y_1, Y_2) = C\left((X(1) + X(3))/2, X(2)\right) =$$

$$= \frac{1}{2}C(X(1), X(2)) + \frac{1}{2}C(X(3), X(2)) = \frac{1}{2}\min(1,2) + \frac{1}{2}\min(3,2) = \frac{1}{2} + 1 = \frac{3}{2}$$

Vi har alltså att

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \text{ är } N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 3/2 & 3/2 \\ 3/2 & 2 \end{pmatrix}\right)$$

Med hjälp av Formelsamlingen erhålls därför att $Y_2|Y_1 = x$ är

$$\mathrm{N}\left(0 + \frac{3/2}{3/2}(x - 0), \sqrt{2 - \frac{(3/2)^2}{3/2}}\right) = \mathrm{N}(x, \sqrt{0.5})$$

Speciellt ser vi att $Y_2|Y_1 = 2.5$ är $\mathrm{N}(2.5, \sqrt{0.5})$ och den sökta sannolikheten blir

$$P\left(2 \le X(2) \le 3 \middle| \frac{X(1) + X(3)}{2} = 2.5\right) =$$

$$= P\left(\frac{2 - 2.5}{\sqrt{0.5}} \le \frac{X(2) - 2.5}{\sqrt{0.5}} \le \frac{3 - 2.5}{\sqrt{0.5}} \middle| \frac{X(1) + X(3)}{2} = 2.5\right) =$$

$$= \Phi(\sqrt{0.5}) - \Phi(-\sqrt{0.5}) = 2\Phi(\sqrt{0.5}) - 1 \approx 2\Phi(0.707) - 1 \approx 2 \cdot 0.7602 - 1 = \underline{0.5205}.$$

Uppgift 6

Med testsignalen $x(t) = e^{i\omega t}$ blir utsignalen $y(t) = H(\omega)e^{i\omega t}$ där $H(\omega)$ är överföringsfunktionen. Vi får i högerledet med testsignalen $x(t) = e^{i\omega t}$

$$\int_0^\infty e^{-2v}x(t - v)dv = \int_0^\infty e^{-2v}e^{i\omega(t-v)}dv = e^{i\omega t}\int_0^\infty e^{-(2+i\omega)v}dv = e^{i\omega t}\frac{1}{2 + i\omega}.$$

Därför erhåller vi eftersom $y'(t) = H(\omega)i\omega e^{i\omega t}$ att

$$e^{i\omega t}\left(H(\omega)i\omega + 3H(\omega)\right) = e^{i\omega t}\frac{1}{2 + i\omega}$$

som ger

$$H(\omega) = \frac{1}{(2 + i\omega)(3 + i\omega)} \text{ dvs } |H(\omega)|^2 = \frac{1}{(4 + \omega^2)(9 + \omega^2)}$$

$r_X(\tau) = e^{-|\tau|},\ \tau \in R$ ger enligt formelsamligen att spektraltätheten $S_X(\omega) = \dfrac{2}{1 + \omega^2},\ \omega \in R$ som alltså med filtersatsen ger

$$S_Y(\omega) = |H(\omega)|^2 S_X(\omega) = \frac{2}{(1 + \omega^2)(4 + \omega^2)(9 + \omega^2)} = \text{(partialbråksuppdelning)} =$$

$$= \frac{1}{20}\cdot\frac{1}{3^2 + \omega^2} - \frac{2}{15}\cdot\frac{1}{2^2 + \omega^2} + \frac{1}{12}\cdot\frac{1}{1^2 + \omega^2} = \frac{1}{120}\cdot\frac{2 \cdot 3}{3^2 + \omega^2} - \frac{1}{30}\cdot\frac{2 \cdot 2}{2^2 + \omega^2} + \frac{1}{24}\cdot\frac{2 \cdot 1}{1^2 + \omega^2}$$

som i sin tur med inverstransformering ger Y-processens kovariansfunktion till

$$r_Y(\tau) = \frac{1}{120}e^{-3|\tau|} - \frac{1}{30}e^{-2|\tau|} + \frac{1}{24}e^{-|\tau|},\ \tau \in R.$$

Detta ger $V(Y(3)) = r_Y(0) = 1/120 - 1/30 + 1/24 = 1/60 \approx \underline{0.0167}$.

Vidare gäller att $m_Y = E(Y(t)) = H(0)m_X = \frac{1}{2}\cdot\frac{1}{3}\cdot 2 = \frac{1}{3} \approx \underline{0.33}$.