Tentamenskrivning för TMS063, Matematisk Statistik.
Onsdag fm den 1 juni, 2016, Eklandagatan 86.
Examinator: Marina Axelson-Fisk. Tel: 070-2288113.
Tillåtna hjälpmedel: typgodkänd miniräknare, tabell- och formelhäfte (utdelas).

****

Tentan är på totalt 50 poäng: 40 poäng på matstat-delen, 10 poäng på flervariabel-delen. För att bli godkänd krävs godkänt på båda delarna. Betygsgränser för betyg 3, 4 och 5 är 40%, 60% och 80%, respektive. Betygen på de båda delarna vägs samman. **Fullständiga och välmotiverad lösningar skall ges till varje uppgift.**

Del I: Matematisk Statistik

1. En kortlek består av 52 kort, med fyra färger och 13 valörer i varje färg.
   (a) En kåk i poker är när 3 kort har en gemensam valör, och två kort har en annan gemensam valör (tex 3 kungar och två ess). Om man får fem kort, vad är sannolikheten att få kåk? (4p)
   (b) Vad är sannolikheten att kåken består av 3 ess och två kungar? (2p)

2. En urna innehåller 3 röda och 2 blåa bollar. En boll dras. Om bollen är röd behålls den och en ny boll dras. Om bollen är blå läggs den tillbaka och ytterligare en röd boll läggs i urnan, och ytterligare en boll dras.
   (a) Vad är sannolikheten att båda bollarna är röda? (2p)
   (b) Om den andra bollen är röd, vad är sannolikheten att den första bollen var blå? (5p)

3. Låt X och Y vara två oberoende stokastiska variabler med $X \sim N(3, 4)$ och $Y \sim Bin(5, 0.5)$. Beräkna väntevärde och varians för den stokastiska variabeln $Z = 3X - 2Y$. (5p)

4. Antag att X och Y är kontinuerliga stokastiska variabler med gemensam täthetsfunktion $f_{XY}(x, y)$. För vilken/vilka av följande tätheter är X och Y oberoende. Motivering krävs.
   (a) $f_{XY}(x, y) = 4x^2y^3$. (2p)
   (b) $f_{XY}(x, y) = \frac{1}{2}(x^3y + xy^3)$. (2p)
   (c) $f_{XY}(x, y) = 6e^{-3x-2y}$. (2p)

5. Antag att bakgrundsstrålningen i Göteborg följer en kontinuerlig fördelning med täthetsfunktion
$$f(x) = x\theta e^{-\frac{1}{2}\theta x^2}, x \geq 0.$$
   För ett stickprov från fördelningen,
$$1.2 \quad 4.1 \quad 2.5 \quad 3.2 \quad 0.9 \quad 1.3 \quad 2.2 \quad 1.3$$
   ange en Maximum Likelihood-skattning av θ. (6p)

6. En statistiker drar 27 datum slumpmässigt och kontrollerar beläggningen på ett givet hotell under dessa datum. Det visar sig att stickprovsvariansen är $s^2 = 5.86$. Antag att antal uthyrda rum är normalfördelad och ange ett 95%-konfidensintervall för den teoretiska variansen σ^2. (4p)

Var god vänd!

7. Vi låter X och Y vara stokastiska variabler för tjäran (i mg) i röken från cigaretter, med respektive utan filter. Antag att $X \sim N(\mu_X, \sigma_X^2)$ och $Y \sim N(\mu_Y, \sigma_Y^2)$. Antag att vi fick följande stickprov

Med filter x_i	0.9	1.1	0.1	0.7	0.3	0.9	0.8	1.0	0.4		
Utan filter y_i	1.5	0.9	1.6	0.5	1.4	1.9	1.0	1.2	1.3	1.6	2.1

Testa på signifikansnivå $\alpha = 0.05$ om det är någon skillnad mellan cigarett-typerna. (6p)

Del II: Flervariabelanalys

8. Ange definitionsområdet D_f för funktionen $f(x, y) = \ln(2 + xy)$. Skissa även området D_f och de tre nivåkurvor till $f(x, y)$ som går genom $(-1, 1), (-1, -2)$ respektive $(1, 0)$. (3p)

9. Massan av en kropp K med densiteten $\delta(x, y, z)$ kan beräknas genom en trippelintegral;

$$\iiint_K \delta(x, y, z)\, dV$$

Bestäm massan av den kropp K som begränsas nedåt av konen $z = \sqrt{x^2 + y^2}$ och uppåt av sfären $x^2 + y^2 + z^2 = 4$ då kroppens densitet ges av $\delta(x, y, z) = z$ (4p)

10. Beräkna kurvintegralen $\int_{\mathcal{C}} \sqrt{y}\, ds$ då $\mathcal{C}$ är den kurva i rummet som beskrivs av parametriseringen $\mathbf{r}(t) = \mathbf{i} + t^2\mathbf{j} + 2t\mathbf{k}$, $0 \leq t \leq 2$. (3p)

Lycka till!

Lösningar

Del I: Matematisk Statistik

1. (a) Första valören kan väljas på $\binom{13}{1} = 13$ olika sätt. I valören väljer vi 3 kort på $\binom{4}{3} = 4$ olika sätt. Den andra valören kan sedan väljas på $\binom{12}{1} = 12$ sätt, och två kort i valören väljs på $\binom{4}{2} = 6$ olika sätt. Med andra ord kan en kåk väljas på

$$13 \cdot 4 \cdot 12 \cdot 6 = 3\,744$$

sätt. För sannolikheten för kåk behöver vi dividera antalet med totala antalet sätt att välja 5 kort, dvs $\binom{52}{5} = 2\,598\,960$ sätt. Vi får

$$P(\text{kåk}) = \frac{37\,44}{2\,598\,960} = 0.00144$$

(b) Formuleringen i den här uppgiften var otydligt och därför ger jag rätt för två olika tolkningar. Man kan antingen tolka det som

Fall I: Vad är sannolikheten att få en kåk med tre ess och två kungar?
Fall II: Givet att vi har fått en kåk, vad är sannolikheten att den består av tre ess och två kungar?

Båda uträkningar ger rätt och skiljer sig bara i vad vi skriver i nämnaren. I täljaren har vi antal sätt att få kåk på med tre ess och två kungar. Det finns 1 sätt att välja valören 'ess', men de tre essen kan väljas på $\binom{4}{3} = 4$. På samma sätt kan valören 'kung' väljas på endast 1 sätt, men de två kungarna på $\binom{4}{2} = 6$ sätt.

Fall I. Sannolikheten blir

$$P(3 \text{ kungar och } 2 \text{ ess}) = \frac{4 \cdot 6}{2\,598\,960} = 0.00000923$$

Fall II: Nu är det givet att vi redan har en kåk, så betingning ger oss

$$P(3 \text{ ess o } 2 \text{ kungar}|\text{kåk}) =$$

$$= \frac{P(\text{kåk med 3 ess och två kungar})}{P(\text{kåk})} = \frac{4 \cdot 6/2\,598\,960}{13 \cdot 4 \cdot 12 \cdot 6/2\,598\,960} = 0.0064$$

2. Låt R_1 = "första röd", R_2 = "andra röd", B_1 = "första blå" och B_2 = "andra blå". Vi får

(a)

$$P(R_1 \cap R_2) = P(R_1)P(R_2|R_1) = \frac{3}{5} \cdot \frac{2}{4} = \frac{6}{20} = 0.3.$$

(b) Bayes formel ger

$$P(B_1|R_2) = \frac{P(R_2|B_1)P(B_1)}{P(R_2)} = \frac{P(R_2|B_1)P(B_1)}{P(R_2|R_1)P(R_1) + P(R_2|B_1)P(B_1)}$$

och

$$P(R_2|B_1) = \frac{4}{6},\ P(B_1) = \frac{2}{5},\ P(R_2|R_1) = \frac{2}{4},\ P(R_1) = \frac{3}{5}$$

så vi får

$$P(B_1|R_2) = \frac{\frac{4}{6} \cdot \frac{2}{5}}{\frac{2}{4} \cdot \frac{3}{5} + \frac{4}{6} \cdot \frac{2}{5}} = \frac{8}{17} = 0.47$$

3. Väntevärde och varians av $Z = 3X - 2Y$ där $X \sim N(3,4)$ och $Y \sim Bin(5, 0.5)$. Först har vi att

$$E[X] = 3, \quad V(X) = 4, \quad E[Y] = 5 \cdot 0.5 = 2.5, \quad V(X) = 5 \cdot 0.5(1 - 0.5) = 1.25$$

$$\begin{aligned} E[Z] &= E[3X - 2Y] = 3E[X] - 2E[Y] = 3 \cdot 3 - 2 \cdot 2.5 = 4 \\ V(Z) &= V(3X - 2Y) = 9V(X) + 4V(Y) = 9 \cdot 4 + 4 \cdot 1.25 = 41 \end{aligned}$$

4. X och Y är oberoende om $f_{XY}(x,y) = f_X(x) f_Y(y)$.

   (a) Oberoende: $f_{XY}(x,y)$ kan separeras till en produkt $f_X(x) = ax^2$ och $f_Y(y) = by^3$ där $ab = 4$.

   (b) Inte oberoende: $f_{XY}(x,y)$ kan inte separeras till en produkti x och y.

   (c) Oberoende: $f_{XY}(x,y)$ kan separeras i en produkt $f_X(x) = ae^{-3x}$ och $f_Y(y) = be^{-2y}$ där $ab = 6$.

5. Likelihoodfunktionen är

$$L(\theta) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} x_i \theta e^{-\frac{1}{2}\theta x_i^2}$$

Log-likelihoodfunktionen blir

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \left(\ln x_i + \ln \theta - \frac{1}{2}\theta x_i^2 \right) = \sum_{i=1}^{n} \ln x_i + n \ln \theta - \frac{1}{2}\theta \sum_{i=}^{n} x_i^2$$

och vi får

$$\frac{dl(\theta)}{d\theta} = 0 \Leftrightarrow \frac{dl(\theta)}{d\theta} = \frac{n}{\theta} - \frac{1}{2}\sum_{i=1}^{n} x_i^2 = 0 \Leftrightarrow \hat{\theta} = \frac{2n}{\sum_{i=1}^{n} x_i^2}$$

Från stickprovet har vi att $n = 8$ och $\sum_{i=1}^{n} x_i^2 = 43.77$ så $\hat{\theta} = 0.366$.

6. Vi har att $n = 27$ och $s^2 = 5.86$, och vet att ett konfidensintervall av σ^2 ges av

$$\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}.$$

Från χ^2_{26}-tabell får vi

$$\begin{aligned} \chi^2_{26,0.025} &= 41.92 \\ \chi^2_{26,0.975} &= 13.84 \end{aligned}$$

Vi får

$$\begin{aligned} \frac{26 \cdot 5.86}{41.92} &= 3.63 \\ \frac{26 \cdot 5.86}{13.84} &= 11.01 \end{aligned}$$

så ett 95%-konfidensintervall blir

$$3.63 \le \sigma^2 \le 11.01$$

7. Vi vill alltså testa

$$H_0 : \mu_X - \mu_Y = 0 \ \text{ mot } \ H_1 : \mu_X - \mu_Y \neq 0$$

och eftersom varianserna är okända använder vi teststatistikan

$$T_0 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}}$$

Vi har att

$$\begin{aligned} n_X &= 9 & n_Y &= 11 \\ \bar{X} &= 0.69 & \bar{Y} &= 1.36 \\ \sum_{i=1}^{9} x_i^2 &= 5.22 & \sum_{i=1}^{11} y_i &= 22.54 \end{aligned}$$

$$\begin{aligned} s_X^2 &= \frac{1}{9-1}\left(5.22 - 9 \cdot 0.69^2\right) = 0.1186 \\ s_Y^2 &= \frac{1}{11-1}\left(22.54 - 11 \cdot 1.36^2\right) = 0.2085 \end{aligned}$$

så vi får att

$$T_0 = \frac{0.69 - 1.36}{\sqrt{\frac{0.1186}{9} + \frac{0.2085}{11}}} = -3.74$$

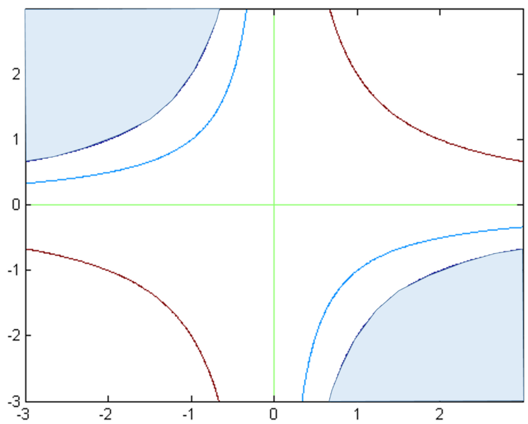
och $T_0 \sim t_\gamma$ där γ skattas med

$$\hat{\gamma} = \frac{\left(\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}\right)^2}{\left(\frac{\left(\frac{s_X^2}{n_X}\right)^2}{n_X - 1} + \frac{\left(\frac{s_Y^2}{n_Y}\right)^2}{n_Y - 1}\right)} = \frac{\left(\frac{0.1186}{9} + \frac{0.2085}{11}\right)^2}{\left(\frac{\left(\frac{0.1186}{9}\right)^2}{9-1} + \frac{\left(\frac{0.2085}{11}\right)^2}{11-1}\right)} = 17.91 \approx 17$$

Med $\alpha = 0.05$ får vi $t_{17,0.975} = 2.11$ och eftersom $|T_0| > 2.11$ förkastar vi H_0. Det verkar vara skillnad mellan cigarett-sorterna.

Del II: Flervariabelanalys

8. $f(x,y) = \ln(2+xy)$ är definierad då $2+xy>0$ så $D_f = \{(x,y)\in\mathbb{R}^2 : xy > -2\}$. Definitionsområdet och de tre nivåkurvorna är markerade i följande figur (D_f är hela planet utom det som är färgat med ljusblått):



9. Övergång till sfäriska koordinater ger;

$$\iiint_K \delta(x,y,z)\,dV = \iiint_K z\,dxdydz =$$

$$\int_0^2 \left(\int_0^{\pi/4} \left(\int_0^{2\pi} \rho\cos\phi\rho^2\sin\phi\,d\theta\right)d\phi\right)d\rho = 2\pi\left[\frac{1}{2}\sin^2\phi\right]_0^{\pi/4}\left[\frac{1}{4}\rho^4\right]_0^2 = 2\pi$$

10.

$$\int_{\mathcal{C}} \sqrt{y}\,ds = \int_0^2 \sqrt{t^2}\sqrt{0^2+(2t)^2+2^2}\,dt = \int_0^2 2t\sqrt{t^2+1}\,dt =$$

$$= \left[\frac{2}{3}\left(t^2+1\right)^{3/2}\right]_0^2 = \frac{2}{3}\left(5\sqrt{5}-1\right)$$