Matematisk statistik TMS063
Tentamen 2018-05-30

Tid: 8:30-12:30 **Tentamensplats:** SB
Hjälpmedel: Bifogad formelsamling och tabell samt Chalmersgodkänd räknare.
Kursansvarig: Olof Elias
Telefonvakt/jour: Olof Elias, 0762026293. Till salen ca 9.30 och 11.30
Betygsgränser: För betyg **3**, **4** resp. **5** krävs minst **12**, **18** resp. **24** poäng.
För att bli godkänd på kursen krävs godkänt på tentan **och** godkänt resultat på båda flervariabel-duggorna.
Till varje uppgift skall fullständig lösning lämnas!

---

1. Låt $A \subseteq B \subseteq C$ vara tre händelser med $P(A) = 0.2, P(B) = 0.4, P(C) = 0.6$.

   (a) Beräkna sannolikheten av att alla tre händelser inträffar. (**2 poäng)**

   (b) Beräkna sannolikheten av att exakt två händelser inträffar. (**1 poäng)**

   (c) Beräkna sannolikheten av att exakt en av händelserna inträffar. (**1 poäng)**

   (Ledning: Rita ett Venn-diagram av händelserna.)
   Lösning:

   (a) Eftersom $A \subseteq B \subseteq C$ så innebär det att om A inträffar så måste även B och C inträffa. Vilket innebär att sannolikheten av att alla tre händelser inträffar ges av sannolikheten att A inträffar. Alternativt så observerar man:

   $$P(\text{ A,B,C inträffar}) = P(A \cap B \cap C) = P(A).$$

   (b) Samma resonemang ger att

   $$P(\text{ Exakt två av händelserna inträffar }) = P(B \backslash A^c) = P(B) - P(A) = 0.2.$$

   (c) Med hjälp av samma resonemang som tidigare så ser man följande. Om C inträffar men inte B, så kan inte A inträffa heller eftersom $A \subseteq B$. Detta ger oss att

   $$P(\text{Exakt en av } A, B, C \text{ inträffar }) = P(C \cap B^c) = P(C) - P(B) = 0.2.$$

2. En mackapär består av 30 likadana elektriska komponenter som fungerar oberoende av varandra. Livslängden (i år) för varje enskild komponent kan antas vara exponentialfördelad med parameter $\lambda = 1/5$.

   (a) Beräkna sannolikheten att en given komponent håller i mer än 10 år? **(2 poäng)**

   (b) Vad är sannolikheten att minst 1 av komponenterna lever längre än 10 år? (**2 poäng)**

(c) Använd centrala gränsvärdessatsen för att approximativt beräkna sannolikheten att den sammanslagna livslängden av alla komponenter är längre än 150 år? (**2 poäng**)

Lösning

(a) Livslängden är exponentialfördelad med parameter $\lambda = 1/5$ vilket ger att om vi låter X vara livslängden för en given komponent så söks
$$P(X \geq 10) = \mathrm{e}^{-10/5} = \mathrm{e}^{-2}.$$

(b) Om vi låter Y vara antalet komponenter som lever längre än 10 år så är $Y \sim \mathrm{Bin}(30, \mathrm{e}^{-2})$. Vilket innebär att vi söker
$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - (1 - \mathrm{e}^{-2})^{30}$$

(c) Låt $W = \sum_{i=1}^{30} X_i$ vara den totala livslängden, då har vi
$$\mathrm{E}[W] = 30\,\mathrm{E}[X_1] = 30{\cdot}5 = 150, \mathrm{Var}(W) = 30\,\mathrm{Var}(X_1) = 30{\cdot}5^2 = 750,$$
vilket ger att det vi söker ges av
$$\begin{aligned} P(W \geq 150) &= P\left(\frac{W - E[W]}{\sqrt{\mathrm{Var}(W)}} \geq \frac{150 - 150}{\sqrt{750}}\right) \\ &\overset{CLT}{\approx} P(Z \geq 0) = 0.5 \end{aligned}$$

3. En Poisson process har egenskapen att den har så kallade oberoende ökningar.

Detta innebär att om vi har fyra tidspunkter $t_1 < t_2 < t_3 < t_4$ så är de stokastiska variablerna $N(t_4) - N(t_3)$ och $N(t_2) - N(t_1)$ **oberoende** och **Poissonfördelade** med parameter $\lambda(t_4 - t_3)$ respektive $\lambda(t_2 - t_1)$.

(a) Beräkna sannolikheten
$$P\left(N(4) - N(3) = 0,\ N(2) - N(1) = 0\right).$$
(**2 poäng**)

(b) Vad är den gemensamma tätheten för variablerna $N(4) - N(3)$ och $N(2) - N(1)$. (**2 poäng**)

Lösning:

(a) Eftersom $N(4) - N(3)$ och $N(2) - N(1)$ är oberoende stokstiska variabler så ges sannolikheten av
$$\begin{aligned} &P(N(4) - N(3) = 0, N(2) - N(1) = 0) = \\ &P(N(4) - N(3) = 0)P(N(2) - N(1) = 0) = \mathrm{e}^{-\lambda}\mathrm{e}^{-\lambda} \end{aligned}$$
där den sista likheten fås av att $N(4) - N(3) \sim \mathrm{Po}(\lambda)$ och $N(2) - N(1) \sim \mathrm{Po}(\lambda)$.

(b) Eftersom $N(4)-N(3)$ och $N(2)-N(1)$ är oberoende så fås den gemensamma tätheten av produkten av de enskilda fördelningarna. Låt $X=N(4)-N(3)$ och låt $Y=N(2)-N(1)$ då fås den gemensamma tätheten $f_{X,Y}$ av:

$$f_{X,Y}(x,y)=f_X(x)f_Y(y)=\left(\frac{\lambda^x}{x!}\mathrm{e}^{-\lambda}\right)\left(\frac{\lambda^y}{y!}\mathrm{e}^{-\lambda}\right)=\frac{\lambda^{x+y}}{x!y!}\mathrm{e}^{-2\lambda}$$

4. Låt

$$f(x|\alpha)=\frac{\alpha}{x^{\alpha+1}}, x\geq 1,$$

vara en täthet för en kontinuerlig stokastisk variabel X där $\alpha>1$ är en okänd parameter. Låt

$$x=(1.8513, 1.0100, 1.0147, 1.2724, 1.0771, 1.9173, 1.3335, 1.0298, 1.0807, 1.0139)$$

vara ett observerat stickprov från fördelningen. Beräkna Maximum likelihood-skattaren och momentmetodsskattaren av α.
(Beräkningshjälp: $\bar{x}=1.2601, \sum_{i=1}^{10}\log(x_i)=2.0151, n=10$)
(2 poäng för momentmetoden och 4 poäng för maxmimum likelihoodmetoden. Totalt 6 poäng.)
Lösning:
Vi börjar med momentmetoden. Låt $x=(x_1,x_2,...,x_n)$ vara ett observerat stickprov från X, eftersom vi bara har en parameter vi vill skatta räcker det med att betrakta ekvationen

$$\bar{x}=E[X]=\int_1^\infty x\frac{\alpha}{x^{\alpha+1}}\mathrm{d}x=\alpha\int_1^\infty\frac{1}{x^\alpha}\mathrm{d}x=\alpha\frac{1}{\alpha-1}.$$

Om vi löser ekvationen för α får vi

$$\hat{\alpha}=\frac{\bar{x}}{\bar{x}-1}.$$

Sätter vi in våra värden får vi att $\hat{\alpha}=4.8452$.
Nu vänder vi oss till maximumlikehood-metoden. Låt x vara som ovan och bilda likelihoodfunktionen

$$L(\alpha|x)=\prod_{i=1}^n f(x_i|\alpha)=\prod_{i=1}^n\frac{\alpha}{x_i^{\alpha+1}}=\alpha^n\left(\prod_1^n\frac{1}{x_i}\right)^{\alpha+1}.$$

Log-likelihoodfunktionen ges därefter av

$$l(\alpha|x)=\log(L(\alpha|x))=n\log(\alpha)-(\alpha+1)\sum_1^n\log(x_i)$$

och dess första och andra-derivata ges av

$$l'(\alpha|x)=\frac{n}{\alpha}-\sum_1^n\log(x_i),$$
$$l''(\alpha|x)=-\frac{n}{\alpha^2}.$$

Eftersom $l''(\alpha|x) < 0$ så är l konkav vilket innebär att l har sitt maximum där derivatan är noll:

$$\frac{n}{\alpha} - \sum_1^n \log(x_i) = 0 \Leftrightarrow \alpha = n/\sum_1^n \log(x_i)$$

dvs. maximum-likelihood skattaren ges av

$$\alpha^* = n/\sum_1^n \log(x_i).$$

Sätter vi in våra värden får vi i detta fall

$$\alpha^* = 4.9626.$$

5. Lenny och Carl singlar slant med varandra, visar den krona bjuder Carl på nästa runda medan om den visar klave så bjuder Lenny. Eftersom Lenny och Carl är så goda vänner så vill dom försäkra sig om att myntet dom kastar är rättvist, dvs att proportionen av krona och klave är lika, och ber Moe testa det.

   Moe gör 100 slantsinglingar och skriver ner om den visar krona eller klave. Efter hundra kast har myntet visat krona 43 gånger och klave 57 gånger. Hjälp Moe med att testa hypotesen om myntet är rättvist:

$$H_0 : p = 0.5$$
$$H_1 : p \neq 0.5,$$

   genom att ställa upp ett konfidensintervall för proportionen p med signifikansnivå $\alpha = 0.05$ och avgör om myntet är rättvist eller ej.
   (5 poäng)
   Lösning: Låt $\hat{p} = \frac{43}{100}$ vara den observerade proportionen av antalet krona (det går lika bra att betrakta proportionen av antalet klave). Konfidensintervallet ges av

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.43 \pm 1.96\sqrt{(0.43 \cdot 0.57)/100} = [0.333, 0.527].$$

   Eftersom $0.5 \in [0.333, 0.527]$ så kan vi inte förkasta H_0 vilket innebär att med en signifikansnivå på $\alpha = 0.05$ så är myntet rättvist.

6. Låt $x = (x_1, x_2, ..., x_n)$ vara ett observerat stickprov från $N(\mu, 4)$, med $\bar{x} = 1.37,\ n = 20$. Testa hypotesen

$$H_0 : \mu \leq 1.1$$
$$H_1 : \mu > 1.1$$

   med genom att beräkna p-värdet och förkasta med lämpligt vald signifikansnivå. **(5 poäng)**
   Lösning: Låt $\alpha = 0.05$ och låt $X = (X_1, ..., X_n)$ vara ett stickprov från $N(\mu, 4)$, då gäller att

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Vidare, låt

$$z_{\text{obs}} = \frac{\bar{x} - 1.1}{2/\sqrt{20}} = 0.6$$

vara den motsvarande observerade kvantiteten. Eftersom $H_0 : \mu \leq 1.1$ så ges p-värdet av

$$p = P(Z > z_{\text{obs}}) = P(Z > 0.6) = 1 - 0.73 = 0.27.$$

Då $p > \alpha$ så förkastar vi inte H_0.