Tentamentsskrivning i **Matematisk Statistik TMA321**

Tid: den 13 oktober, 2018
Hjälpmedel: Typgodkänd miniräknare, egenhändigt skriven formelsamling om två A4 fram och bak (dvs 4 sidor), samt utdelade tabeller.

Tentamen består av 8 frågor om sammanlagt 50 poäng. Preliminära betygsgränser är satta till:
betyg "3": 20 till 29 poäng
betyg "4": 30 till 39 poäng
betyg "5": 40 eller fler poäng.

OBS! Alla lösningar skall vara väl redovisade och motiverade. Talen är ej ordnade efter svårighetsgrad.

1. Låt $U_1 \sim U(\{1,\ldots,n\})$ dvs U_1 är likormigt fördelad på mängden $\{1,2,\ldots,n\}$. Betingat på att $U_1 = k$, låt sedan $U_2 \sim U(\{1,\ldots,2k\})$.

   (a) Hitta den gemensamma sannolikhetsfunktionen för U_1, U_2 samt marginalsannolikhetsfunktionen för U_2. (3p)

   (b) Bestäm $\mathbb{E}[U_2|U_1]$ och $\mathbb{E}[U_2]$. (3p)

   (c) Mer generellt låter vi fördelningen för U_m betingat på att $U_{m-1} = k$ vara $U_m \sim U(\{1,\ldots,2k\})$. Bestäm $\mathbb{E}[U_m]$ för alla värden på m. (2p)

   Lösning:

   (a) Vi har att

   $$\mathbb{P}(U_2 = l, U_1 = k) = \mathbb{P}(U_2 = l|U_1 = k)\mathbb{P}(U_1 = k) = \frac{1}{2k}\frac{1}{n} = \frac{1}{2nk},$$

   för $1 \le k \le n$ och $1 \le l \le 2k$. Vidare blir då

   $$\begin{aligned}\mathbb{P}(U_2 = l) &= \sum_{k=1}^{n} \mathbb{P}(U_2 = l, U_1 = k)\\ &= \sum_{k=\lceil l/2\rceil}^{n} \mathbb{P}(U_2 = l, U_1 = k) = \sum_{k=\lceil l/2\rceil}^{n} \frac{1}{2nk}.\end{aligned}$$

   Här betyder $\lceil x\rceil$ det minsta heltal större än eller likamed x. Om $l = 5$ får vi t.ex. att k måste vara minst 3.

   (b) Vi har att $\mathbb{E}[U_2|U_1 = k] = \frac{2k+1}{2}$ så att

   $$\mathbb{E}[U_2|U_1] = \frac{2U_1 + 1}{2}.$$

   Vidare blir då

   $$\mathbb{E}[U_2] = \mathbb{E}[\mathbb{E}[U_2|U_1]] = \mathbb{E}\left[\frac{2U_1+1}{2}\right] = \mathbb{E}[U_1] + \frac{1}{2} = \frac{n+1}{2} + \frac{1}{2} = \frac{n+2}{2}.$$

(c) Från svaret i uppgift (*b*) ser vi att $\mathbb{E}[U_m|U_{m-1}] = \frac{2U_{m-1}+1}{2}$ så att

$$\mathbb{E}[U_3] = \mathbb{E}[\mathbb{E}[U_3|U_2]] = \frac{2\mathbb{E}[U_2]+1}{2} = \frac{n+3}{2}.$$

Vi använder induktionshypotesen att $\mathbb{E}[U_m] = \frac{n+m}{2}$ och induktionssteget blir då

$$\mathbb{E}[U_m] = \mathbb{E}[\mathbb{E}[U_m|U_{m-1}]] = \frac{2\frac{n+m-1}{2}+1}{2} = \frac{n+m}{2},$$

och vi kan då dra slutsatsen att

$$\mathbb{E}[U_m] = \frac{n+m}{2}$$

för varje värde på m.

2. Låt $X_1, \ldots, X_n$ vara i.i.d. $N(\mu, \sigma^2)$ där μ är känt. Carlita vill testa hypoteserna

$$H_0 : \sigma^2 = \sigma_0^2$$
$$H_1 : \sigma^2 = \sigma_1^2$$

(a) Bilda lämplig likelihood ratio, och härled ur denna enklaste möjliga test. (3p)

(b) Hitta en lämplig förkastningsregion (RR) om signifikansnivån skall vara 2.5% och $n = 17$. (3p)

Lösning:

(a) Likelihooden blir

$$f(X|\sigma^2) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \prod_{k=1}^{n} \exp\left(-\frac{(X_k-\mu)^2}{2\sigma^2}\right),$$

så att den sökta likelihoodkvoten blir

$$\lambda(X) = \frac{f(X|\sigma_0^2)}{f(X|\sigma_1^2)} = \left(\frac{\sigma_0}{\sigma_1}\right)^n \prod_{k=1}^{n} \exp\left(\frac{(X_k-\mu)^2}{2\sigma_1^2} - \frac{(X_k-\mu)^2}{2\sigma_0^2}\right).$$

Enligt paradigmen skall H_0 förkastas till förmån för H_1 ifall kvoten $\lambda(X)$ är tillräckligt liten. Vi har två fall.
Fall 1 ($\sigma_1 > \sigma_0$): Kvoten är liten ifall

$$\sum_{k=1}^{n} \frac{(X_k-\mu)^2}{2\sigma_1^2} - \frac{(X_k-\mu)^2}{2\sigma_0^2} = \left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2}\right) \sum_{k=1}^{n} (X_k-\mu)^2$$

är mycket negativ. Detta innebär att vi förkastar H_0 ifall

$$\sum_{k=1}^{n} (X_k-\mu)^2$$

är stor.

Fall 2 ($\sigma_1 < \sigma_0$): Kvoten är nu liten ifall

$$= \left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2}\right) \sum_{k=1}^{n} (X_k - \mu)^2$$

är lite positiv (den kan ju inte vara negativ). Detta innebär att vi förkastar H_0 ifall

$$\sum_{k=1}^{n} (X_k - \mu)^2$$

är liten.

(b) Under H_0 är

$$\sum_{k=1}^{17} \frac{(X_k - \mu)^2}{\sigma_0} \sim \chi^2_{17}.$$

Tabellen innehåller inga siffror för $n = 17$ så vi tar halvvägs mellan värdet för $n = 16$ och $n = 18$.

Fall 1 ($\sigma_1 > \sigma_0$): Vår förkastningsregion blir

$$\left\{ \sum_{k=1}^{17} \frac{(X_k - \mu)^2}{\sigma_0} \geq 30 \right\}$$

(värdet mitt emellan är $(28.85 + 31.53)/2 = 30.19$ men tabellen är inte linjär i n).

Fall 2 ($\sigma_1 < \sigma_0$): Vår förkastningsregion blir

$$\left\{ \sum_{k=1}^{17} \frac{(X_k - \mu)^2}{\sigma_0} \leq 7.5 \right\}$$

(värdet mitt emellan är $(6.91 + 8.23)/2 = 7.57$ men tabellen är inte linjär i n).

3. Låt (X, Y) ha gemensam täthetsfunktion

$$f(x, y) = 2y \text{ för } 0 \leq x, y \leq 1,$$

och låt $W = XY$.

(a) Bestäm den momentgenererande funktionen för W. (2p)

(b) Använd ditt svar i (a) för att bestämma $\mathbb{E}[W^n]$ för varje värde på n. (3p)

Lösning:

(a) Vi har att

$$M_W(t) = \mathbb{E}[e^{Wt}] = \int_0^1 \int_0^1 2ye^{xyt}dxdy = \int_0^1 2y \left[\frac{e^{xyt}}{yt}\right]_0^1 dy$$
$$= \int_0^1 2\left(\frac{e^{yt}-1}{t}\right)dy = 2\left[\frac{e^{yt}}{t^2} - \frac{y}{t}\right]_0^1 = 2\frac{e^t - 1 - t}{t^2}.$$

(b) Vi har att

$$M_W(t) = 2\frac{\sum_{k=0}^{\infty}\frac{t^k}{k!} - 1 - t}{t^2} = \sum_{k=2}^{\infty}\frac{t^{k-2}}{k!} = \sum_{k=0}^{\infty}\frac{t^k}{(k+2)!} = \sum_{k=0}^{\infty}\frac{t^k}{k!}\frac{k!}{(k+2)!},$$

så vi ser att

$$\mathbb{E}[W^k] = \frac{k!}{(k+2)!} = \frac{1}{(k+1)(k+2)}.$$

för $k = 0, 1, \ldots$

4. En slumpvariabel har fördelningen

$$f(x) = \sqrt{\frac{2}{\pi}}a^{3/2}x^2e^{-ax^2/2} \text{ för } x \geq 0,$$

där $a > 0$ är okänd.

(a) Hitta momentskattaren (MME) för a. (3p)

(b) Hitta maximum likelihhodskattaren (MLE) för a. (3p)

Lösning:

(a) Vi har att

$$\mathbb{E}[X] = \sqrt{\frac{2a}{\pi}}\int_0^{\infty} ax^3e^{-ax^2/2}dx$$
$$= \sqrt{\frac{2a}{\pi}}\left[-x^2e^{-ax^2/2}\right]_0^{\infty} + \sqrt{\frac{2a}{\pi}}\int_0^{\infty} 2xe^{-ax^2/2}dx$$
$$= 0 + \sqrt{\frac{2a}{\pi}}\left[-\frac{2}{a}e^{-ax^2/2}\right]_0^{\infty} = \sqrt{\frac{8}{\pi a}}.$$

Momentmetoden säger då att vi skall sätta

$$\bar{X} = \sqrt{\frac{8}{\pi\hat{a}}} \text{ så att } \hat{a} = \frac{8}{\pi\bar{X}^2}.$$

(b) Likelihooden är

$$L(a) = \prod_{k=1}^{n} \sqrt{\frac{2}{\pi}} a^{3/2} X_k^2 e^{-aX_k^2/2} = \left(\frac{2}{\pi}\right)^{n/2} \left(X_k^2\right)^n a^{3n/2} \exp\left(-\frac{a}{2}\sum_{k=1}^{n} X_k^2\right),$$

så att

$$l(a) = \log\left(\left(\frac{2}{\pi}\right)^{n/2} \left(X_k^2\right)^n\right) + \frac{3n}{2}\log(a) - \frac{a}{2}\sum_{k=1}^{n} X_k^2$$

och därmed blir

$$l'(a) = \frac{3n}{2a} - \frac{1}{2}\sum_{k=1}^{n} X_k^2.$$

Om vi sätter $l'(a) = 0$ ser vi att

$$\hat{a} = \frac{3n}{\sum_{k=1}^{n} X_k^2} = \frac{3}{\overline{X^2}}.$$

Vi får inte glömma att kolla att vi har hittat ett maximum. Vi har att $l''(a) = -\frac{3n}{2a^2} < 0$ så vi har hittat ett maximum.

5. Bengt-Arne undersöker cellförändringar i cellprov. För ett givet cellprov antar Bengt-Arne att antalet celler med förändringar är Poissonfördelat med parameter $\lambda > 0$ (som är okänt). Bengt-Arne undersöker ett antal cellprov och får följande dataserie

#cellprov:	13	30	34	17	16	8	3
#förändringar:	0	1	2	3	4	5	6.

Data skall tolkas som att Bengt-Arne fick 13 cellprov med 0 cellförändringar etc.

(a) Hjälp Bengt-Arne med att skatta $\lambda > 0$. (2p)

(b) Hitta ett tvåsidigt 95% K.I. för λ åt Bengt-Arne. (2p)

(c) Hitta ett ensidigt 95% K.I. för λ åt Bengt-Arne som är på formen $[c, \infty)$. (2p)

Lösning:

(a) Vi skattar λ med hjälp av $\bar{X}$ (ty detta är en väntevärdesriktig skattare) och får

$$\hat{\lambda} = \frac{0*13 + 1*30 + 2*34 + 3*17 + 4*16 + 5*8 + 6*3}{13 + 30 + 34 + 17 + 16 + 8 + 3} = \frac{271}{121} \approx 2.24.$$

(b) Vi använder centrala gränsvärdessatsen för att approximera

$$\bar{X} \approx N(\mathbb{E}[\bar{X}], \mathrm{Var}(\bar{X})) = N\left(\lambda, \frac{\lambda}{n}\right) \approx N\left(\lambda, \frac{\hat{\lambda}}{n}\right).$$

Därmed har vi att

$$\frac{\bar{X} - \lambda}{\sqrt{\hat{\lambda}/n}} \approx N(0, 1),$$

så att ett 95% konfidensintervall ges av

$$0.95 = \mathbb{P}(-1.96 \leq Z \leq 1.96) \approx \mathbb{P}\left(-1.96 \leq \frac{\bar{X} - \lambda}{\sqrt{\hat{\lambda}/n}} \leq 1.96\right)$$
$$= \mathbb{P}\left(\bar{X} - 1.96\sqrt{\hat{\lambda}/n} \leq \lambda \leq \bar{X} + 1.96\sqrt{\hat{\lambda}/n}\right).$$

Med våra data får vi då att

$$I_\lambda = [\bar{x} - 1.96\sqrt{\bar{x}/121}, \bar{x} + 1.96\sqrt{\bar{x}/121}] \approx [1.973, 2.506].$$

(c) I detta fall gör vi nästan samma som i uppgift (b):'

$$0.95 = \mathbb{P}(-1.645 \leq Z)$$
$$\approx \mathbb{P}\left(-1.645 \leq \frac{\bar{X} - \lambda}{\sqrt{\hat{\lambda}/n}}\right) = \mathbb{P}\left(\bar{X} - 1.645\sqrt{\hat{\lambda}/n} \leq \lambda\right).$$

Med våra data får vi då att

$$I_\lambda = [\bar{x} - 1.645\sqrt{\bar{x}/121}, \infty) \approx [2.016, \infty).$$

6. Märta-Lisa funderar på att köpa ett hus i centrala Älmhultaryd. Hon gör en enkel marknadsundersökning där hon tittar på ett antal hus som tidigare varit ute för försäljning. För varje sålt hus noterar hon slutpris och antalet kvadratmeter. Märta-Lisa fick följande dataserie

m^2:	83	98	107	119	143	164	179	210	245
Pris:	160	170	190	210	230	200	240	220	240

Här anges priset i $kSEK$ (dvs i tusental kronor). Märta-Lisa ansätter en linjär regressionsmodell $y = \beta_0 + \beta_1 x$ där y är priset och x är antalet kvadratmeter. Data sammanfattas med att $S_{xx} \approx 23714$, $S_{yy} \approx 6800$ och $S_{xy} \approx 10323$.

Hjälp Märta-Lisa med att lösa följande uppgifter:

(a) Skatta β_0 och β_1. (2p)

(b) Skapa ett 95% konfidensintervall för β_1 och testa huruvida $\beta_1 = 0$ på 95%-nivån. (2p)

(c) Ange förklaringsgraden och beräkna residualerna. Kommentera ditt resultat. (2p)

(d) Fundera kring Märta-Lisas val av modell och analysen ovan, är den rimlig? Vilka är bristerna? (1p)

Lösning

(a) Vi har att

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \approx 0.435 \text{ och } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \approx 141.$$

(b) Vi betraktar hypteserna $H_0 : \beta_1 = 0$ och $H_1 : \beta_1 \neq 0$. Vi använder att

$$\frac{\hat{\beta}_1 - \beta_1}{s_r / \sqrt{S_{xx}}} \sim t(7)$$

och får med hjälp av tabell att $t_{0.025}(5) \approx 2.365$ så att

$$\begin{aligned} 0.95 &= \mathbb{P}\left(-2.365 \leq \frac{\hat{\beta}_1 - \beta_1}{s_r/\sqrt{S_{xx}}} \leq 2.365 \right) \\ &= \mathbb{P}\left(\hat{\beta}_1 - 2.365 \frac{s_r}{\sqrt{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + 2.365 \frac{s_r}{\sqrt{S_{xx}}} \right). \end{aligned}$$

Vi har att

$$s_r^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}} \right) \approx 329.4$$

så att $s_r = 18.1$. Ett 95% numeriskt K.I. för β_1 blir då

$$I_{\beta_1} = \hat{\beta}_1 \pm 2.365 \frac{s_r}{\sqrt{S_{xx}}} \approx [0.157, 0.714].$$

Då $0 \notin I_{\beta_1}$ förkastar vi H_0 på 95% nivån.

(c) Förklaringsgraden blir

$$R^2 = \frac{S_{xy}^2}{S_{xx} S_{yy}} \approx 0.661$$

vilket inte är speciellt bra. Residualerna är $e_k = y_k - (\hat{\beta}_0 + \hat{\beta}_1 x_k)$ blir

e_1	e_2	e_3	e_4	e_5	e_6	e_7		
-17.6	-14.1	2	16.7	26.3	-12.9	20.6	-12.9	-8.1

Inga direkta mönster kan ses, så vi kan inte dra slutsatsen att ett specifikt annat samband skulle passa bättre. Man bör dock vara något försiktig då antalet mätpunkter är så lågt.

(d) Punkt (b) indikerar att priset beror på antalet kvadratmeter och detta är såklart fullt rimligt. Problemet med Märta-Lisas ansats är att modellen inte tar hänsyn till skicket på huset, huruvida det finns garage, balkong etc etc. Om man inte tar med alla dessa faktorer i modellen kommer den inte bli speciellt bra.

7. En klubb består av 15 stalinister och 20 trotskyister. En arbetarkommitté skall bildas bestående av exakt 7 medlemmar.

(a) Vad är sannolikheten att kommittén innehåller minst en trotskyist men även fler stalinister än trotskyister, om varje medlem väljs oberoende och med samma sannolikhet? (2p)

(b) Stalinisterna vill genomföra en utrensning av oönskade element (dvs trotskyisterna). För att arbetarkomittén skall kunna besluta om en sådan måste de ha minst fem röster. Vad är sannolikheten att komittén kan besluta om en laglig utrensning givet att kommittén är som i punkt a? Vi antar här att alla stalinister röstar för utrensning men alla trotskyister röstar emot. (2p)

(c) Efter en lyckad utrensning består nu klubben av enbart 15 stalinister. Snart splittras dessa upp i en revolutionär och en småborgerlig gren. På hur många sätt kan medlemmarna delas upp så att de båda grenarna innehåller minst 6 medlemmar? (2p)

Lösning:

(a) Låt A beteckna den sökta händelsen. Antalet kommittéer som uppfyller kravet är

$$\binom{15}{4}\binom{20}{3} + \binom{15}{5}\binom{20}{2} + \binom{15}{6}\binom{20}{1},$$

och det sammanlagda antaler möjliga kommittéer är $\binom{35}{7}$ så med hjälp av divisionsregeln ser vi att den sökta sannolikheten blir

$$\mathbb{P}(A) = \frac{\binom{15}{4}\binom{20}{3} + \binom{15}{5}\binom{20}{2} + \binom{15}{6}\binom{20}{1}}{\binom{35}{7}} \approx 0.331$$

(b) Låt B vara händelsen att en utrensning kan genomföras. Vi söker då

$$\begin{aligned}
\mathbb{P}(B|A) &= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B)}{\mathbb{P}(A)} \\
&= \frac{\frac{\binom{15}{5}\binom{20}{2}+\binom{15}{6}\binom{20}{1}}{\binom{35}{7}}}{\frac{\binom{15}{4}\binom{20}{3}+\binom{15}{5}\binom{20}{2}+\binom{15}{6}\binom{20}{1}}{\binom{35}{7}}} \\
&= \frac{\binom{15}{5}\binom{20}{2} + \binom{15}{6}\binom{20}{1}}{\binom{15}{4}\binom{20}{3} + \binom{15}{5}\binom{20}{2} + \binom{15}{6}\binom{20}{1}} \approx 0.301.
\end{aligned}$$

(c) Antalet sätt att välja exakt k medlemmar till den revolutionära grenen är $\binom{15}{k}$. Det sökta antalet är därför

$$\binom{15}{6} + \binom{15}{7} + \binom{15}{8} + \binom{15}{9}.$$

8. Låt $T \sim \text{Exp}(\lambda)$ och $a > 0$.

(a) Bestäm täthetsfunktionen för $Y = (T - a)^2$. (3p)

(b) Låt Z vara en slumpvariabel vars fördelningsfunktion $F_Z(t)$ definieras av att

$$F_Z(t) = \mathbb{P}((T-a)^2 \leq t | T \geq a)$$

for alla $t \in \mathbb{R}$. Hitta täthetsfunktionen för Z. (3p)

Lösning:

(a) Vi har att

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}((T-a)^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq T - a \leq \sqrt{y}) \\ &= \mathbb{P}(a - \sqrt{y} \leq T \leq a + \sqrt{y}) = \int_{\max(0, a-\sqrt{y})}^{a+\sqrt{y}} \lambda e^{-\lambda x} dx \\ &= \left[-e^{-\lambda x}\right]_{\max(0,a-\sqrt{y})}^{a+\sqrt{y}} = e^{-\lambda \max(0, \sqrt{y}-a)} - e^{-\lambda(a+\sqrt{y})}. \end{aligned}$$

Vi ser då att täthetsfunktionen blir

$$f_Y(y) = F_Y'(y) = \begin{cases} \frac{\lambda}{2\sqrt{y}} e^{-\lambda(a+\sqrt{y})} & \text{om } a \leq y^2 \\ \frac{\lambda}{2\sqrt{y}} e^{-\lambda(a+\sqrt{y})} + \frac{\lambda}{2\sqrt{y}} e^{\lambda(a-\sqrt{y})} & \text{om } a > y^2 \end{cases}$$

(b) Vi börjar med att bestämma ett uttryck för fördelningsfunktionen.

$$\begin{aligned} F_Z(t) &= \mathbb{P}((T-a)^2 \leq t | T \geq a) = \frac{\mathbb{P}((T-a)^2 \leq t, T \geq a)}{\mathbb{P}(T \geq a)} \\ &= \frac{\mathbb{P}(T - a \leq \sqrt{t}, T \geq a)}{\mathbb{P}(T \geq a)} = \frac{\mathbb{P}(a \leq T \leq a + \sqrt{t})}{\mathbb{P}(T \geq a)} \\ &= \frac{\int_a^{a+\sqrt{t}} \lambda e^{-\lambda s} ds}{\int_a^{\infty} \lambda e^{-\lambda s} ds} = \frac{e^{-\lambda a} - e^{-\lambda(a+\sqrt{t})}}{e^{-\lambda a}} = 1 - e^{-\lambda\sqrt{t}} \end{aligned}$$

om $t \geq 0$ och annars är $F_Z(t) = 0$. Vi får därför att

$$f_Z(t) = \frac{1}{\sqrt{t}} e^{-\lambda\sqrt{t}} \text{ för } t \geq 0.$$