Tentamentsskrivning i **Matematisk Statistik TMA321**

Tid: den 28 maj (fm), 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. Antag att det i urna A finns 3 röda och 4 gröna bollar, och att det i urna B finns 2 röda och 5 gröna bollar. Först väljer man en boll slumpmässigt från urna A och lägger den i urna B. Sedan väljer man en boll slumpmässigt från urna B (som nu innehåller 8 bollar), och lägger den i urna A. I det tredje och sista steget drar vi slumpässigt en boll från urna A. Vi antar hela tiden att bollarna i urnorna är väl blandade.

   (a) Vad är sannolikheten att den andra bollen som dras (dvs den från urna B) är grön? (3p)

   (b) Vad är sannolikheten att bollen vi drar i det sista steget är grön? (3p)

   Lösning:

   (a) Låt G_i vara händelsen att en grön boll dras i drag nummer i. Vi har då att

   $$\mathbb{P}(G_2) = \mathbb{P}(G_2|G_1)\mathbb{P}(G_1) + \mathbb{P}(G_2|G_1^c)\mathbb{P}(G_1^c) = \frac{6}{8} \times \frac{4}{7} + \frac{5}{8} \times \frac{3}{7} = \frac{39}{56} \approx 0.6964$$

   (b) Det finns 4 fall som gör att vi drar en grön i sista och tredje steget. Vi betecknar fallen med A_1, A_2, A_3 och A_4. Fallet A_1 är följande sekvens av händelser: Först flyttas en grön, sedan en grön, och en grön dras i tredje steget. A_2: grön - röd-grön. A_3: röd - grön - grön. A_4: röd - röd -grön. Sannolikheten för A_1 kan beräknas enligt följande:

   $$P(A_1) = \frac{4}{7} \times \frac{6}{8} \times \frac{4}{7} = \frac{96}{392}.$$

   Den första faktorn kommer sig av att vi har 4 gröna bollar av 7 bollar när vi väljer i första steget. Den andra faktorn kommer från att vi

har 6 gröna av 8 i andra steget, ifall det flyttades en grön i första. Det tredje faktorn kommer från att vi har 4 gröna bollar av 7 i tredje steget ifall det flyttades gröna bollar i de två första stegen.
På liknande sätt fås att

$$P(A_2) = \frac{4}{7} \times \frac{2}{8} \times \frac{3}{7} = \frac{24}{392},$$

$$P(A_3) = \frac{3}{7} \times \frac{5}{8} \times \frac{5}{7} = \frac{75}{392},$$

$$P(A_4) = \frac{3}{7} \times \frac{3}{8} \times \frac{4}{7} = \frac{36}{392}.$$

Eftersom A_1, A_2, A_3, A_4 samtliga är disjunkta, fås

$$\begin{aligned} &P(\text{grön dras i tredje steget}) \\ &\quad = P(A_1) + P(A_2) + P(A_3) + P(A_4) = \frac{231}{392} = \frac{33}{56} \approx 0.5893. \end{aligned}$$

2. En djurpark skall köpa in ett parti om 13 fullvuxna flodhästar. Deras leverantör plockar flodhästarna från lämplig flod och väger dom. Vi kan anta att flodhästarnas vikter är oberoende och normalfördelade med parametrar μ, σ^2. De uppmätta vikterna blev:

flodhäst nr:	1	2	3	4	5	6	7
vikt:	2220	1857	1877	2099	2407	1946	2174
flodhäst nr:	8	9	10	11	12	13	
vikt:	2055	2323	1882	2107	2211	2320	

(a) Hitta ett 95% K.I. för μ om $\sigma = 200$ är känd. (2p)

(b) Hitta ett 95% K.I. för μ om σ är okänd. (2p)

(b) Hitta ett 99% K.I. för σ^2. (2p)

Lösning:

(a) Vi har att

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

så att

$$0.95 = \mathbb{P}(-1.96 \leq R \leq 1.96) = \mathbb{P}\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu\bar{X} + 1.96\frac{\sigma}{\sqrt{n}} \leq\right),$$

så att ett 95% numeriskt K.I. ges av

$$I_\mu = \bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} \approx 2113.7 \pm 1.96\frac{200}{\sqrt{13}} \approx [2005, 2222.4]$$

(b) Vi får skatta σ^2 med

$$s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2 \approx 33686$$

så att $s \approx 183.54$. Dessutom är $t_{0.025}(12) \approx 2.179$ så att ett numeriskt 95% K.I. ges av

$$I_\mu = \bar{x} \pm 2.179\frac{s}{\sqrt{n}} \approx 2113.7 \pm 2.179\frac{183.54}{\sqrt{13}} \approx [2002.8, 2224.6].$$

(c) Vi får använda att

$$(n-1)\frac{s^2}{\sigma^2} \sim \chi(n-1)^2$$

Tabell ger att $\chi^2_{0.005}(12) \approx 28.3$ medans $\chi^2_{0.995}(12) \approx 3.074$. Därför blir

$$0.95 \approx \mathbb{P}\left(3.074 \leq (n-1)\frac{s^2}{\sigma^2} \leq 28.3\right) = \mathbb{P}\left((n-1)\frac{s^2}{28.3} \leq \sigma^2 \leq (n-1)\frac{s^2}{3.074}\right)$$

så att ett 99% numeriskt K.I. för σ^2 blir

$$I_{\sigma^2} = \left[12\frac{33686}{28.3}, 12\frac{33686}{3.074}\right] \approx [14284, 131500].$$

3. Låt X vara en slumpvariabel med täthetsfunktion

$$f(x) = cxe^{\alpha x} \text{ för } 0 \leq x \leq 2,$$

där parametern $\alpha > 0$.

(a) Bestäm konstanten c. (2p)

(b) Bestäm täthetsfunktionen för $Y = \sqrt{X}$. (2p)

(c) Beräkna $\mathbb{E}[Y^2]$. (2p)

Lösning:

(a) Vi får att

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty} f(x)dx = \int_0^2 cxe^{\alpha x}dx \\
&= c\left[x\frac{e^{\alpha x}}{\alpha}\right]_0^2 - \frac{c}{\alpha}\int_0^2 e^{\alpha x}dx = \frac{c}{\alpha}2e^{2\alpha} - \frac{c}{\alpha^2}\left(e^{2\alpha} - 1\right) \\
&= c\left(\frac{1 + (2\alpha - 1)e^{2\alpha}}{\alpha^2}\right),
\end{aligned}$$

så att $c = \frac{\alpha^2}{1 + e^{2\alpha}(2\alpha - 1)}$.

(b) Vi har att för $0 \le y \le \sqrt{2}$

$$
\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(\sqrt{X} \le y) = \mathbb{P}(X \le y^2) \\
&= \int_0^{y^2} cxe^{\alpha x} dx = c\left[x\frac{e^{\alpha x}}{\alpha}\right]_0^{y^2} - \frac{c}{\alpha}\int_0^{y^2} e^{\alpha x} dx \\
&= \frac{c}{\alpha} y^2 e^{y^2\alpha} - \frac{c}{\alpha^2}\left(e^{y^2\alpha} - 1\right) \\
&= c\left(\frac{1 + (y^2\alpha - 1)e^{y^2\alpha}}{\alpha^2}\right),
\end{aligned}
$$

så att

$$
\begin{aligned}
f_Y(y) &= \frac{c}{\alpha^2}\left(2y\alpha e^{y^2\alpha} + (y^2\alpha - 1)2y\alpha e^{y^2\alpha}\right) \\
&= 2cy^3 e^{\alpha y^2} = \frac{2\alpha^2 y^3 e^{\alpha y^2}}{1 + e^{2\alpha}(2\alpha - 1)}
\end{aligned}
$$

för $0 \le y \le \sqrt{2}$.

(c) Uppenbarligen är $\mathbb{E}[Y^2] = \mathbb{E}[X]$ så vi räknar ut den senare. Vi har att

$$
\begin{aligned}
\mathbb{E}[X] &= c\int_0^2 x^2 e^{\alpha x} = c\left[x^2\frac{e^{\alpha x}}{\alpha}\right]_0^2 - \frac{2c}{\alpha}\int_0^2 xe^{\alpha x} dx \\
&= \frac{4c}{\alpha}e^{2\alpha} - \frac{2c}{\alpha}\left(\frac{1 + (2\alpha - 1)e^{2\alpha}}{\alpha^2}\right) = \frac{4c}{\alpha}e^{2\alpha} - \frac{2c}{\alpha}\frac{1}{c} \\
&= \frac{4e^{2\alpha}}{\alpha}\frac{\alpha^2}{1 + (2\alpha - 1)e^{2\alpha}} - \frac{2}{\alpha}
\end{aligned}
$$

4. Låt X ~Poi(λ) (dvs Poisson fördelad med parameter λ), och betingat på att $X = k$, låt $Y \sim U(\{1, 2, \dots, k+1\})$ (dvs likformigt fördelad på $1, \dots, k+1$.).

   (a) Hitta den gemensamma sannolikhetsfunktionen för (X, Y). (2p)

   (b) Hitta marginalsannolikhetsfunktionen för Y. (2p)

   (c) Hitta den betingade sannolikhetsfunktionen för X givet Y. (2p)

   (d) Hitta $\mathbb{E}[X + 1 | Y = 1]$. (2p)

Lösning:

(a) Vi har att

$$
\begin{aligned}
&\mathbb{P}(X = k, Y = l) \\
&\qquad = \mathbb{P}(Y = l | X = k)\mathbb{P}(X = k) = \frac{1}{k+1}\frac{\lambda^k}{k!}e^{-\lambda} = \frac{\lambda^k}{(k+1)!}e^{-\lambda},
\end{aligned}
$$

där $k \in \mathbb{N}$ och $l \in \{1, \dots, k+1\}$.

(b) Använd LTP för att se att

$$\begin{aligned}\mathbb{P}(Y=l) &= \sum_{n=0}^{\infty} \mathbb{P}(X=n, Y=l) \\ &= \sum_{n=l-1}^{\infty} \mathbb{P}(X=n, Y=l) = \sum_{n=l-1}^{\infty} \frac{\lambda^n}{(n+1)!} e^{-\lambda} = \frac{1}{\lambda} \sum_{n=l}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda}.\end{aligned}$$

(c) Vi söker

$$\begin{aligned}\mathbb{P}(X=k|Y=l) &= \frac{\mathbb{P}(X=k, Y=l)}{\mathbb{P}(Y=l)} \\ &= \frac{\lambda^k}{(k+1)!} e^{-\lambda} \frac{1}{\frac{1}{\lambda}\sum_{n=l}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda}} \\ &= \frac{\lambda^{k+1}}{(k+1)! \sum_{n=l}^{\infty} \frac{\lambda^n}{n!}} = \left(\sum_{n=l}^{\infty} \frac{(k+1)!}{n!} \lambda^{n-k-1} \right)^{-1},\end{aligned}$$

där $l \in \{1, 2, \ldots\}$ och $k \in \{l-1, l, \ldots\}$.

(d) Vi ser att

$$\begin{aligned}\mathbb{E}[X+1|Y=1] &= \sum_{k=0}^{\infty} (k+1) \mathbb{P}(X=k|Y=1) \\ &= \sum_{k=0}^{\infty} (k+1) \frac{\lambda^{k+1}}{(k+1)! \sum_{n=1}^{\infty} \frac{\lambda^n}{n!}} = \frac{\lambda}{e^{\lambda}-1} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \frac{\lambda e^{\lambda}}{e^{\lambda}-1}.\end{aligned}$$

5. Antag att $X_1, X_2, \ldots, X_n$ är oberoende och likafördelade med fördelning $U[0, \theta]$ där θ är okänt.

   (a) Hitta maximum likelihood skattaren (MLE) för θ. (3p)

   (b) Är din skattare från (a) väntevärdesriktig? Det räcker att du löser denna deluppgift i fallet att $n = 2$. (3p)

Lösning:

(a) Vi har att

$$\begin{aligned}f(X_1, \ldots, X_n|\theta) &= \prod_{k=0}^{n} f(X_k|\theta) = \prod_{k=0}^{n} \frac{I(0 \le X_k \le \theta)}{\theta} \\ &= \frac{I(0 \le \min(X_1, \ldots, X_n) \le \max(X_1, \ldots, X_n) \le \theta)}{\theta^n}.\end{aligned}$$

Detta skall maximeras med avseende på θ, vilket innebär att vi måste ta $\hat{\theta} = \max(X_1, \ldots, X_n)$.

(b) Vi har att

$$\begin{aligned}\mathbb{E}[\hat{\theta}] &= \mathbb{E}[\max(X_1, X_2)] = \frac{1}{\theta^2}\int_0^\theta \int_0^\theta \max(x_1, x_2)dx_1dx_2 \\ &= \frac{1}{\theta^2}\int_0^\theta \left(\int_0^{x_2} x_2dx_1 + \int_{x_2}^\theta x_1dx_1\right)dx_2 = \frac{1}{\theta^2}\int_0^\theta x_2^2 + \frac{\theta^2 - x_2^2}{2}dx_2 \\ &= \frac{1}{2\theta^2}\left[\theta^2 x_2 + \frac{x_2^3}{3}\right]_0^\theta = \frac{2}{3}\theta.\end{aligned}$$

Således är skattaren inte väntevärdesriktig i detta fall.

6. Låt $X_1, X_2, \ldots$ vara en sekvens av oberoende och likafördelade slumpvariabler. Antag att fördelningen är sådan att

$$\mathbb{P}(X_1 = 1) = \frac{1}{2},\ \mathbb{P}(X_1 = 2) = \frac{1}{4} \text{ och } \mathbb{P}(X_1 = 3) = \frac{1}{4}$$

(a) Bestäm den momentgenererande funktionen för X_1. (2p)

(b) Använd ditt svar i (a) för att bestämma $\mathbb{E}[X_1^n]$ för varje värde på n. (2p)

(c) Beräkna $\mathbb{E}[(X_1+X_2)^n]$ för varje värde på n. Svara på enklast möjliga form. (2p)

Lösning:

(a) Vi har att

$$M_{X_1}(t) = \mathbb{E}[e^{X_1t}] = \frac{1}{2}e^t + \frac{1}{4}e^{2t} + \frac{1}{4}e^{3t} = \frac{2e^t + e^{2t} + e^{3t}}{4}.$$

(b) Taylorutveckling av exponentialfunktionerna ger att

$$\begin{aligned}M_{X_1}(t) &= \frac{1}{4}\left(2\sum_{k=0}^\infty \frac{t^k}{k!} + \sum_{k=0}^\infty \frac{(2t)^k}{k!} + \sum_{k=0}^\infty \frac{(3t)^k}{k!}\right) \\ &= \sum_{k=0}^\infty \frac{t^k}{k!}\left(\frac{1}{2} + \frac{2^k}{4} + \frac{3^k}{4}\right)\end{aligned}$$

Vi kan därför läsa av att

$$\mathbb{E}[X_1^k] = \frac{1}{2} + \frac{2^k}{4} + \frac{3^k}{4}.$$

(c) Vi har att

$$\begin{aligned}M_{X_1+X_2}(t) &= M_{X_1}(t)^2 = \frac{1}{16}(2e^t + e^{2t} + e^{3t})^2 \\ &= \frac{4e^{2t} + 4e^{3t} + 5e^{4t} + 2e^{5t} + e^{6t}}{16} \\ &= \sum_{k=0}^\infty \frac{t^k}{k!}\frac{4 * 2^k + 4 * 3^k + 5 * 4^k + 2 * 5^k + 6^k}{16}\end{aligned}$$

så att

$$\mathbb{E}[(X_1 + X_2)^k] = \frac{4 * 2^k + 4 * 3^k + 5 * 4^k + 2 * 5^k + 6^k}{16}$$

för $k = 0, 1, \ldots$

7. Alexandra brukar göra sin egen cement. Hon tar då lika delar kalk och stenmineral, som hon sedan bränner under hög temperatur. Efter bränning kan materialet användas för diverse byggnadsarbeten. En viktig egenskap är då torktiden efter att den brända blandningen blandas med vatten. Alexandra testar 8 olika bränntemperaturer, och undersöker sedan motsvarande torktider (alla yttre faktorer såsom utomhustemperatur, mängd tillsatt vatten etc kan anses vara samma för alla experiment). Följande data uppmättes

Temp (C):	2100	2200	2300	2400	2500	2600	2700	2800
Tid (min):	620	594	533	524	475	451	446	448

Alexandra ansatta en linjär regressionsmodell $y = \beta_0 + \beta_1 x$ där x är temperaturen (i C) och y torktiden (i minuter). Data kan sammanfattas med att $S_{xx} = 420000$, $S_{yy} = 32512$ och att $S_{xy} = -111950$. Hjälp Alexandra med följande uppgifter:

(a) Skatta β_0, β_1. (2p)

(b) Skapa ett 99% konfidensintervall för β_1 och testa huruvida $\beta_1 = -0.3$ på 99%-nivån. (2p)

(c) Ange förklaringsgraden och beräkna residualerna. Verkar Alexandras modell rimlig? (2p)

Lösning

(a) Vi har att

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \approx -0.267 \text{ och } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \approx 1164.4.$$

(b) Vi använder att

$$\frac{\hat{\beta}_1 - \beta_1}{s_r/\sqrt{S_{xx}}} \sim t(6)$$

och får med hjälp av tabell att $t_{0.005}(6) \approx 3.707$ så att

$$\begin{aligned} 0.99 &= \mathbb{P}\left(-3.707 \leq \frac{\hat{\beta}_1 - \beta_1}{s_r/\sqrt{S_{xx}}} \leq 3.707\right) \\ &= \mathbb{P}\left(\hat{\beta}_1 - 3.707\frac{s_r}{\sqrt{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + 3.707\frac{s_r}{\sqrt{S_{xx}}}\right). \end{aligned}$$

Vi har att

$$s_r^2 = \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) \approx 445.31$$

så att $s_r = 21.1$. Ett 99% numeriskt K.I. för β_1 blir då

$$I_{\beta_1} = \hat{\beta_1} \pm 3.707\frac{s_r}{\sqrt{S_{xx}}} \approx [-0.387, -0.146].$$

Då $-0.3 \in I_{\beta_1}$ förkastar vi ej H_0 på 99% nivån.

(c) Förklaringsgraden blir

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} \approx 0.918$$

vilket är bra. Residualerna blir

e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8
15.3	16	-18.4	-0.7	-23	-20.4	1.3	29.9.

Det faktum att residualerna i mitten alla är negativa medans de på kanterna alla är positiva är kanske lite mystiskt. Med tanke på hur hög förklaringsgraden var, och hur små residualerna är till sin storlek, bör man dock inte dra alltför stora växlar på detta.

Det är fysikaliskt självklart att ett linjärt samband inte kan vara giltigt för alla temperaturer (man skulle få negativa torktider). Det är dock inte heller fallet att man kan förkasta en linjär modell för det aktuella temperaturspannet enbart baserat på residualerna.

8. Gamma-fördelningen har täthetsfunktion

$$\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} \text{ för } x > 0.$$

Denna fördelning har väntevärde α/β och varians α/β^2.

Låt $X_1, \ldots, X_n$ vara i.i.d. Gamma-fördelade enligt ovan angivna täthetsfunktion. Betrakta hypoteserna

$$H_0 : \beta = 20 \text{ och } H_1 : \beta = 25.$$

(a) Sätt upp likelihood ration, och härled ur denna enklaste möjliga test (för hypoteserna H_0, H_1) med tillhörande förkastningsområde. (3p)

(b) Om $n = 57$, bestäm förkastningsområde (RR) motsvarande signifikansnivå 1% för ditt härledda test. Du kan här anta att parametern $\alpha = 4$. (3p)

Lösning:

(a) Vi har att

$$\begin{aligned}\text{lik}(\beta) &= f(X_1,\ldots,X_n|\beta) = \prod_{k=1}^{n} f(X_k|\beta) \\ &= \prod_{k=1}^{n} \frac{\beta^\alpha}{\Gamma(\alpha)} X_k^{\alpha-1} e^{-\beta X_k} = \frac{\beta^{n\alpha}}{\Gamma(\alpha)^n}\left(\prod_{k=1}^{n} X_k\right)^{\alpha-1} e^{-\beta(X_1+\cdots+X_n)},\end{aligned}$$

så den sökta LR kvoten blir

$$\frac{\text{lik}(\beta_0)}{\text{lik}(\beta_1)} = \left(\frac{\beta_0}{\beta_1}\right)^{n\alpha} e^{-(\beta_0-\beta_1)(X_1+\cdots+X_n)}.$$

Vi förkastar H_0 om denna kvot är liten, vilket är ekvivalent med att vi förkastar om $X_1+\cdots+X_n$ är litet (Observera att $-(\beta_0-\beta_1) = 5 > 0$). Vårt RR blir nu $\{X_1+\cdots+X_n \le c\}$ där konstanten c bestäms av signifikansnivån.

(b) Vi har att

$$X_1+\cdots+X_n \approx N(n\alpha/\beta, n\alpha/\beta^2),$$

så att om $\alpha = 4$

$$R = \frac{\bar{X}-\alpha/\beta}{\sqrt{\alpha}/(\beta\sqrt{n})} = \frac{\bar{X}-4/\beta}{2/(\beta\sqrt{n})} = \frac{\sqrt{n}}{2}(\beta\bar{X}-4) \approx N(0,1).$$

Därför blir under H_0

$$\begin{aligned}0.01 &= \mathbb{P}(Z \le -2.33) \approx \mathbb{P}\left(\frac{\sqrt{n}}{2}(\beta_0\bar{X}-4) \le -2.33\right) = \mathbb{P}\left(\bar{X} \le \frac{1}{\beta_0}\left(-\frac{4.66}{\sqrt{n}}+4\right)\right) \\ &= \mathbb{P}\left(X_1+\cdots+X_n \le \frac{1}{\beta_0}\left(-4.66\sqrt{n}+4n\right)\right) \approx \mathbb{P}\left(X_1+\cdots+X_n \le 9.64\right).\end{aligned}$$

Vårt förkastningsområde blir därför

$$RR = [0, 9.64],$$

med teststatistika $X_1+\cdots+X_n$ eller alternativt

$$RR = [0, 0.1691]$$

med teststatistika $\bar{X}$.