Tentamentsskrivning i **Matematisk Statistik TMA321/MVE395/MVE302**

Tid: den 12 Oktober, 2024, 08:30-12:30 (-13:30)
Examinator och jour: Erik Broman, mob. 073 7320791,
Hjälpmedel: Typgodkänd miniräknare, 4 A4-sidor egenhändigt skrivna anteckningar (2 ark fram och bak eller 4 ark på en sida) samt utdelade tabeller.

Tentamen består av 6 (7) frågor om sammanlagt 40 (50) poäng. Preliminära betygsgränser är satta till:
F/Kf:
betyg 3: 16 till 23.5 poäng
betyg 4: 24 till 31.5 poäng
betyg 5: 32 eller fler poäng.
TM:
betyg 3: 20 till 29.5 poäng
betyg 4: 30 till 39.5 poäng
betyg 5: 40 eller fler poäng.

OBS! Alla lösningar skall vara väl redovisade, motiverade och fullständiga. Svar skall ges på enklast möjliga form. Talen är ej ordnade efter svårighetsgrad.

1. Elin har 3 lådor med pärlor. I den första har hon 3 rosa och 2 lila, i den andra har hon 2 rosa, 1 lila och 2 blå och i den tredje har hon 3 lila och 3 blå. Elin väljer en låda slumpmässigt. Efter att hon valt låda plockar hon upp 2 pärlor (slumpmässigt) ur den valda lådan.

   (a) Vad är sannolikheten att hon väljer två pärlor av olika färg? (3p)

   (b) Om Elin fick en lila och en blå pärla, vad är sannolikheten att hon valde låda nummer 2? (3p)

   Lösning:

   (a) Låt L_1, L_2 respektive L_3 vara händelserna att Elin valde låda nummer 1,2 respektive 3. Vi väljer att räkna ut sannolikheten för händelsen S att Elin väljer två pärlor av samma färg. Vi har då att

$$\begin{aligned}\mathbb{P}(S) &= \mathbb{P}(S|L_1)\mathbb{P}(L_1) + \mathbb{P}(S|L_2)\mathbb{P}(L_2) + \mathbb{P}(S|L_3)\mathbb{P}(L_3) \\ &= \frac{1}{3}\left(\frac{\binom{3}{2}+\binom{2}{2}}{\binom{5}{2}} + \frac{\binom{2}{2}+\binom{2}{2}}{\binom{5}{2}} + \frac{\binom{3}{2}+\binom{3}{2}}{\binom{6}{2}}\right) \\ &= \frac{1}{3}\left(\frac{4}{10}+\frac{2}{10}+\frac{6}{15}\right) = \frac{1}{3}.\end{aligned}$$

   Vi får sedan att $\mathbb{P}(S^c) = 1 - \mathbb{P}(S) = \frac{2}{3}$.

(b) Låt BL beteckna händelsen att Elin får en blå och en lila pärla. Vi söker här $\mathbb{P}(L_2|BL)$ och vi använde Bayes sats för att se att

$$\mathbb{P}(L_2|BL) = \mathbb{P}(BL|L_2)\frac{\mathbb{P}(L_2)}{\mathbb{P}(BL)}.$$

Uppenbarligen är $\mathbb{P}(L_2) = 1/3$ och

$$\mathbb{P}(BL|L_2) = \frac{\binom{1}{1}\binom{2}{1}}{\binom{5}{2}} = \frac{2}{10} = \frac{1}{5}.$$

Vidare är

$$\begin{aligned}\mathbb{P}(BL) &= \mathbb{P}(BL|L_1)\mathbb{P}(L_1) + \mathbb{P}(BL|L_2)\mathbb{P}(L_2) + \mathbb{P}(BL|L_3)\mathbb{P}(L_3)\\ &= \frac{1}{3}\left(0 + \frac{\binom{1}{1}\binom{2}{1}}{\binom{5}{2}} + \frac{\binom{3}{1}\binom{3}{1}}{\binom{6}{2}}\right) = \frac{1}{3}\left(0 + \frac{2}{10} + \frac{9}{15}\right) = \frac{4}{15},\end{aligned}$$

så att

$$\mathbb{P}(L_2|BL) = \mathbb{P}(BL|L_2)\frac{\mathbb{P}(L_2)}{\mathbb{P}(BL)} = \frac{1}{5}\frac{\frac{1}{3}}{\frac{4}{15}} = \frac{1}{4}.$$

2. Låt (X, Y) vara likformigt fördelad på området

$$A = \{(x, y) \in \mathbb{R}^2 : 0 \le x \le 1, -1 \le y \le \sqrt{1 - x^2}\}.$$

(a) Bestäm den gemensamma täthetsfunktionen för (X, Y). (1p)

(b) Hitta marginaltäthetsfunktionerna för X respektive Y. (3p)

(c) Bestäm $\mathbb{E}[X|Y = y]$ för alla värden på y. (3p)

Lösning:

(a) Likformighet innebär att

$$f_{X,Y}(x, y) = \frac{1}{|A|} = \frac{1}{1 + \pi/4} = \frac{4}{4 + \pi} \text{ för } (x, y) \in A,$$

där $|A|$ betecknar arean av A.

(b) Vi har att

$$f_X(x) = \int_{-1}^{\sqrt{1-x^2}} f_{X,Y}(x, y)dy = (1 + \sqrt{1 - x^2})\frac{4}{4 + \pi}$$

och att

$$f_Y(y) = \int_0^1 f_{X,Y}(x, y)dx = \frac{4}{4 + \pi} \text{ om } -1 \le y \le 0,$$

medans

$$f_Y(y) = \int_0^{\sqrt{1-y^2}} f_{X,Y}(x, y)dx = \sqrt{1 - y^2}\frac{4}{4 + \pi} \text{ om } 0 \le y \le 1.$$

(c) Den betingade täthetsfunktionen för X givet Y ges av

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{\frac{4}{4+\pi}}{\frac{4}{4+\pi}} = 1 & \text{för } 0 \leq x \leq 1 \\ & \text{om } -1 \leq y \leq 0 \\ \frac{\frac{4}{4+\pi}}{\sqrt{1-y^2}\frac{4}{4+\pi}} = \frac{1}{\sqrt{1-y^2}} & \text{för } 0 \leq x \leq \sqrt{1-y^2} \\ & \text{om } 0 \leq y \leq 1 \\ 0 & \text{annars.} \end{cases}$$

Därför blir

$$\mathbb{E}[X|Y=y] = \begin{cases} \int_0^1 x dx = \frac{1}{2} & \text{för } -1 \leq y \leq 0 \\ \int_0^{\sqrt{1-y^2}} \frac{x}{\sqrt{1-y^2}} dx = \frac{\sqrt{1-y^2}}{2} & \text{för } 0 \leq y \leq 1 \\ 0 & \text{annars.} \end{cases}$$

3. Under orienteringstävlingar brukar Jöns-Harald stå i skogen och spana efter vilsna orienterare. Han bevakar alltid en specifik glänta och antalet vilsna orienterare som dyker upp där följer en Poisssonfördelning med parameter $\lambda = 2$. Antalet vilsna orienterare som Jöns-Harald observerar under olika tävlingar anses vara oberoende.

(a) Beräkna sannolikheten att antalet vilsna orienterare som springer genom Jöns-Haralds glänta är minst 3 under en given tävling. (2p)

(b) Beräkna sannolikheten att det under två tävlingar sammanlagt springer förbi exakt 2 vilsna orienterare. (3p)

(c) Antag att Jöns-Harald står där vecka in och vecka ut och stirrar. Om han bevakar sammanlagt 100 tävlingar, beräkna sannolikheten att han får syn på minst 220 vilsna orienterare. (3p)

Lösning:

(a) Låt X_1 vara antalet vilsna orienterare som Jöns-Harald ser. Vi söker då $\mathbb{P}(X_1 \geq 3)$ och vi har att

$$\begin{aligned} \mathbb{P}(X_1 \geq 3) &= 1 - \mathbb{P}(X_1 \leq 2) = 1 - \frac{2^0}{0!}e^{-2} - \frac{2^1}{1!}e^{-2} - \frac{2^2}{2!}e^{-2} \\ &= 1 - e^{-2}(1+2+2) = 1 - 5e^{-2} (\approx 0.3233). \end{aligned}$$

(b) Låt X_k vara antalet vilsna orienterare som Jöns-Harald ser under tävling nummer k. Man kan räkna ut denna uppgift på flera sätt, det enklaste är att observera att $X_1 + X_2 \sim$Poi$(2+2)$. Vi beskriver här dock den längre metoden. Vi har att

$$\begin{aligned}
&\mathbb{P}(X_1 + X_2 = 2) \\
&= \mathbb{P}(X_1 = 0, X_2 = 2) + \mathbb{P}(X_1 = 1, X_2 = 1) + \mathbb{P}(X_1 = 2, X_2 = 0) \\
&= 2\mathbb{P}(X_1 = 0)\mathbb{P}(X_2 = 2) + \mathbb{P}(X_1 = 1)^2 = 2\frac{2^0}{0!}e^{-2}\frac{2^2}{2!}e^{-2} + \left(\frac{2^1}{1!}e^{-2}\right)^2 \\
&= 4e^{-4} + 4e^{-4} = 8e^{-4}\left(= \frac{4^2}{2!}e^{-4} \approx 0.1465\right).
\end{aligned}$$

(c) Vi får här använda normalapproximation. Vi har att

$$S = \sum_{k=1}^{100} X_k \approx N(\mathbb{E}[S], \mathrm{Var}(S))$$

där

$$\mathbb{E}[S] = \sum_{k=1}^{100} \mathbb{E}[X_k] = 200\mathbb{E}[X_1] = 200$$

och

$$\mathrm{Var}(S) = \sum_{k=1}^{100} \mathrm{Var}[X_k] = 200\mathrm{Var}(X_1) = 200.$$

Därför är

$$S \approx N(200, 200).$$

Vi söker nu

$$\begin{aligned}
\mathbb{P}(S \geq 220) &= \mathbb{P}\left(\frac{S - 200}{\sqrt{200}} \geq \frac{220 - 200}{\sqrt{200}}\right) \\
&\approx \mathbb{P}(Z \geq 1.4142) = 1 - \mathbb{P}(Z \leq 1.4142) \approx 0.079,
\end{aligned}$$

där vi använde tabell i den sista approximationen.

4. **Enbart TM:** Låt $X \sim N(0, 1)$.

(a) Beräkna den momentgenererande funktionen för X. **OBS, enbart svar ger 0 poäng.** (2p)

(b) Använd ditt svar i (a) för att beräkna $\mathbb{E}[X^n]$ för alla $n \in \mathbb{N}$. (3p)

Lösning:

(a) Vi har att

$$\begin{aligned}
M_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx} e^{-x^2/2} dx = \left\{\frac{x^2}{2} - tx = \frac{(x-t)^2}{2} - \frac{t^2}{2}\right\} \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{t^2/2} e^{-(x-t)^2/2} dx = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-s^2/2} ds \\
&= e^{t^2/2},
\end{aligned}$$

där vi i sista likeheten använde att integranden är täthetsfunktionen för en $N(0,1)$-fördelad slumpvariabel.

(b) Vi använder här att

$$M_X(t) = \mathbb{E}[e^{Xt}] = \mathbb{E}\left[\frac{(Xt)^n}{n!}\right] = \sum_{n=0}^{\infty} \mathbb{E}[X^n]\frac{t^n}{n!}.$$

Genom att Taylorutveckla uttrycket $M_X(t) = e^{t^2/2}$ bör vi kunna avläsa vad alla moment blir. Vi har att

$$M_X(t) = e^{t^2/2} = \sum_{k=0}^{\infty}\left(\frac{t^2}{2}\right)^k \frac{1}{k!} = \sum_{k=0}^{\infty} \frac{t^{2k}}{2^k k!} = \sum_{k=0}^{\infty} \frac{t^{2k}}{(2k)!}\frac{(2k)!}{2^k k!}$$

så vi ser att koefficienterna för udda k blir 0 vilket innebär att

$$\mathbb{E}[X^{2n+1}] = 0 \text{ för } n = 0, 1, \ldots$$

Avläsning av koeffeicienterna för jämna värden på k ger sedan att

$$\mathbb{E}[X^{2n}] = \frac{(2n)!}{2^n n!} \text{ för } n = 0, 1, \ldots$$

5. Låt X ha täthetsfunktion

$$f_X(x) = (\alpha - 1)x^{-\alpha} \text{ för } x \geq 1 \text{ där } \alpha > 1.$$

(a) Hitta momentskattaren (MME:n) för α. Reflektera över ditt svar. (3p)

(b) Hitta maximum likelihood skattaren (MLE:n) för α. (3p)

(c) Vad blir värdena på dina skattningar om data blev $x_1 = 2.34, x_2 = 1.28, x_3 = 3.48$ och $x_4 = 1.78$. (1p)

Lösning:

(a) Vi har att

$$\mathbb{E}[X] = (\alpha - 1)\int_1^{\infty} x f_X(x)dx = (\alpha - 1)\int_1^{\infty} x^{1-\alpha}dx = \frac{\alpha - 1}{\alpha - 2}$$

där den sista likheten kräver att $\alpha > 2$. Momentmetoden säger oss då att vi skall sätta

$$\overline{X} = \frac{\hat{\alpha} - 1}{\hat{\alpha} - 2} \Rightarrow (\hat{\alpha} - 2)\overline{X} = \hat{\alpha} - 1 \Rightarrow \hat{\alpha}(\overline{X} - 1) = 2\overline{X} - 1$$

så att

$$\hat{\alpha} = \frac{2\overline{X} - 1}{\overline{X} - 1}.$$

Detta skapar dock en mycket märklig situation då vi måste veta att $\alpha > 2$ för att momentmetoden skall fungera. Utan apriori-kunskap om detta bör metoden undvikas.

(b) Vi börjar med att hitta likelihooden

$$\begin{aligned} L(\alpha) &= \prod_{k=1}^{n} f(X_k|\alpha) \\ &= \prod_{k=1}^{n} (\alpha-1)X_k^{-\alpha} = (\alpha-1)^n \left(\prod_{k=1}^{n} X_k\right)^{-\alpha}. \end{aligned}$$

Vi logaritmerar och får

$$l(\alpha) = n\log(\alpha-1) - \alpha\sum_{k=1}^{n}\log(X_k).$$

Vi ser att

$$l'(\alpha) = \frac{n}{\alpha-1} - \sum_{k=1}^{n}\log(X_k) \text{ så att } l''(\alpha) = -\frac{n}{(\alpha-1)^2} < 0.$$

Om $l'(\hat{\alpha}) = 0$ har vi hittat ett maximum. Vi får då att

$$0 = \frac{n}{\hat{\alpha}-1} - \sum_{k=1}^{n}\log(X_k) \Rightarrow \hat{\alpha} = 1 + \frac{n}{\sum_{k=1}^{n}\log(X_k)},$$

som alltså är vår MLE.

(c) Med dessa mätvärde har vi för MME:n att

$$\hat{\alpha}(x) = \frac{2\overline{x}-1}{\overline{x}-1} \approx 2.82$$

medans för MLE:n blir

$$\hat{\alpha}(x) = 1 + \frac{4}{\sum_{k=1}^{4}\log(x_k)} \approx 2.37.$$

6. Vid en undersökning vill man jämföra effekten hos två olika typer av ångturbiner, Typ A och Typ B. Man genomför därför en serie experiment med lite olika ångtryck och mäter vilken effekt de två turbinerna producerar. Resultaten blev som följer:

Experiment nummer:	1	2	3	4	5	6	7	8
Effekt Turbin A (kW):	5.5	8.8	3.7	7.9	8.3	3.7	8.6	10.3
Effekt Turbin B (kW):	6.0	9.6	4.3	7.9	8.7	3.5	9.2	10.0

(a) Hitta ett konfidensintervall för skillnaden i effekt med konfidensgrad 0.95. Motivera dina räkningar och dina val. (4p)

(b) Turbintillverkaren Turbo hävdar att den nya turbinsorten (B) är bättre än den gamla (A). Stödjer data detta påstående på signifikansnivå 99%? (3p)

Lösning:

(a) Då ingen ytterligare information anges och vi skall undersöka skillnaden bör vi välja ett tvåsidigt konfidensintervall. Det viktiga är att förstå från texten att vi har att göra med situationen med parade data. Vi börjar därför med att räkna ut skillnaderna $z_k = y_k - y_k$ och får att

Experiment nummer:	1	2	3	4	5	6	7	8
z_k :	0.5	0.8	0.6	0	0.4	-0.2	0.6	-0.3

Det står inget om vilken fördelning data kan tänkas komma från, så för att komma vidare antar vi att de kommer från en normalfördelning och att de är oberoende. Vi bildar då referensvariabeln

$$R = \frac{\overline{Z} - \Delta}{s/\sqrt{n}} \approx t(7)$$

(andra konstruktioner kan vara acceptabla beroende på vilka grundantaganden man gör) och observerar att enlit tabell så blir $t_{0.025}(7) \approx 2.36$. Vi ser att

$$0.95 \approx \mathbb{P}(-1.96 \leq R \leq 1.96) = \mathbb{P}\left(-2.36 \leq \frac{\overline{Z} - \Delta}{s/\sqrt{n}} \leq 2.36\right)$$
$$= \mathbb{P}\left(\overline{Z} - 2.36\frac{s}{\sqrt{n}} \leq \Delta \leq \overline{Z} + 2.36\frac{s}{\sqrt{n}}\right).$$

Ett 95 % numeriskt K.I. för Δ blir därför

$$I_\Delta = \overline{Z}(x, y) \pm 2.36\frac{s}{\sqrt{8}}.$$

Här blir

$$\overline{Z}(x, y) = 0.3 \text{ och } s^2 = \frac{1}{7}\sum_{k=1}^{8}(Z_k - \overline{Z})^2 \approx 0.1686$$

så att

$$I_\Delta \approx [-0.043, 0.693].$$

(b) Här är hypoteserna vi skall undersöka

$$H_0 : \Delta = 0$$
$$H_1 : \Delta > 0.$$

Skillnaden mot uppgift (a) är att vi skall ta fram ett konfidensintervall som är ensidigt och på 99% nivån. Liknande räkningar som ovan ger då att (med $t_{0.01}(7) \approx 3$),

$$0.99 = \mathbb{P}(Z \leq 3) = \mathbb{P}\left(\frac{\overline{Z} - \Delta}{s/\sqrt{n}} \leq 3\right) = \mathbb{P}\left(\overline{Z} - 3\frac{s}{\sqrt{n}} \leq \Delta\right),$$

och ett 99% K.I blir då

$$I_\Delta = \left[\overline{Z} - 3\frac{s}{\sqrt{8}}, \infty\right) \approx [-0.136, \infty).$$

Då $0 \in I_\Delta$ kan vi inte förkasta H_0 på den föreskrivna signifikansnivån.

7. Alice mäter väntetider $T_1, \ldots, T_{20}$ som anses vara oberoende och likafördelade och komma från en exponentialfördelning med parameter λ. Alice vill undersöka följande hypoteser:

$$H_0 : \lambda = 1$$
$$H_1 : \lambda > 1.$$

Data gav medelvärdet $\bar{t} = 0.83$ på väntetiderna.

(a) Bestäm en lämplig test-statistika och ange fördelningen för denna. (2.5p)

(b) Utgå ifrån (a) och bestäm tillhörande kritiska region om signifikansnivån är 0.1. Vilken slutsats drar du om testet? (2.5p)

(c) (**Enbart TM:**) Vad blir p-värdet av testet? (2p)

(d) (**Enbart TM:**) Vilken styrka har testet om $\lambda = 2$? (3p)

Lösning:

(a) Då vi har att $\mathbb{E}[T_1] = \frac{1}{\lambda}$ och att $\mathrm{Var}(T_1) = \frac{1}{\lambda^2}$ så blir

$$\overline{T} \approx N\left(\mathbb{E}[T_1], \frac{\mathrm{Var}(T_1)}{n}\right) = N\left(\frac{1}{\lambda}, \frac{1}{20\lambda^2}\right).$$

Vi observerar att under $H_0 : \lambda = 1$ så blir

$$\overline{T} \approx N\left(1, \frac{1}{20}\right),$$

och en lämplig test-statistika T blir därför

$$\tilde{T} = \frac{\overline{T} - 1}{\sqrt{1/20}} = \sqrt{20}(\overline{T} - 1) \approx N(0, 1).$$

(b) Om $\lambda > 1$ så blir väntetiderna i snitt mindre än om $\lambda = 1$. Vi bör därför förkasta H_0 till förmån för H_1 ifall $\overline{T}$ blir liten, dvs om test-statistikan $\tilde{T}$ blir "stor" och negativ. Vi ser att det kritiska området blir på formen $(-\infty, c]$ där vi skall bestämma c utifrån villkoret

$$0.1 = \mathbb{P}(\tilde{T} \leq c)$$

så att $c \approx -1.28$ enligt tabell. I Alices fall är $\bar{t} = 0.83$ och värdet på vår test-statistika blir

$$\tilde{T}(t_1, \ldots, t_{20}) = \sqrt{20}(\bar{t} - 1) = \sqrt{20}(0.83 - 1) \approx -0.76.$$

Vi förkastar därför inte H_0 på signifikansnivån 0.1.

(c) p-värdet bestäms av sannolikheten att test-statistikan antar ett minst lika extremt värde som det som observerades. Då vi observerade värdet $\tilde{T}(t) \approx -0.76$ och då testet är ensidigt får vi att (igen med hjälp av tabell)

$$\begin{aligned} p - \text{värde} &= \mathbb{P}(\tilde{T}(t) \leq -0.76) = 1 - \mathbb{P}(\tilde{T}(t) \geq -0.76) \\ &= 1 - \mathbb{P}(\tilde{T}(t) \leq 0.76) \approx 1 - 0.776 = 0.224. \end{aligned}$$

(d) Styrkan är sannolikheten att testet förkastas givet att H_1 är sann (vilket det kan vara på flera olika sätt, t.ex. $\lambda = 1.1, 2, 987$ etc). I vårt fall har vi då att

$$\text{styrka} = \mathbb{P}(\tilde{T}(t) \leq -0.76 | \lambda = 2).$$

Om $\lambda = 2$ får vi så klart att

$$\overline{T} \approx \left(\frac{1}{\lambda}, \frac{1}{20\lambda^2}\right) = \left(\frac{1}{2}, \frac{1}{80}\right).$$

Därmed blir (under antagandet att $\lambda = 2$)

$$\begin{aligned} &\mathbb{P}(\tilde{T} \leq -1.28) = \mathbb{P}(\sqrt{20}(\overline{T} - 1) \leq -1.28) \\ &= \mathbb{P}\left(\overline{T} \leq 1 - \frac{1.28}{\sqrt{20}}\right) = \mathbb{P}\left(\frac{\overline{T} - 1/2}{\sqrt{1/80}} \leq \frac{1 - \frac{1.28}{\sqrt{20}} - 1/2}{\sqrt{1/80}}\right) \\ &\approx \mathbb{P}(Z \leq 1.912) \approx 0.972. \end{aligned}$$