

Tentamentsskrivning i **Matematisk Statistik MVE091**

Tid: den 13 januari, 2025, 08:30-13:30

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 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, motiverade och fullständiga. Svar skall ges på enklast möjliga form. Talen är ej ordnade efter svårighetsgrad.

1. Tre fåglar (fågel F_1 , F_2 och F_3) kommer flygande och sätter sig oberoende och helt slumpmässigt på telefontråden mellan stolpe S_1 och stolpe S_2 . Det är 40 meter mellan stolparna. Låt X_k beteckna avståndet mellan fågel F_k och S_1 , för $k = 1, 2, 3$. Vi antar att X_1, X_2 och X_3 är oberoende och att de är kontinuerligt likformigt fördelade på intervallet $[0, 40]$.
 - (a) Vad är sannolikheten att exakt två av fåglarna sätter sig inom 10 meter från stolpe S_1 ? (2p)
 - (b) Vad är väntevärdet av antalet fåglar som sätter sig inom 10 meter från stolpe S_1 ? (2p)
 - (c) Vad är den betingade sannolikheten att fågel F_1 sätter sig inom 10 meter från stolpe S_1 om du vet att den sätter sig inom 35 meter från stolpe S_2 ? (2p)

Lösning:

- (a) Sannolikheten att en given fågel F_k sätter sig inom 10 meter från stolpe 1 är

$$\mathbb{P}(X_k \leq 10) = \frac{10}{40} = \frac{1}{4}$$

då $X_k \sim U([0, 40])$. Svaret blir därför

$$\begin{aligned} & 3\mathbb{P}(X_1 \leq 10, X_2 \leq 10, X_3 > 10) \\ &= 3\mathbb{P}(X_1 \leq 10)^2(1 - \mathbb{P}(X_1 \leq 10)) = 3\frac{1}{4^2} \left(1 - \frac{1}{4}\right) = \frac{3^2}{4^3} \end{aligned}$$

där vi använde oberoende i första likheten.

- (b) Låt Y = antalet fåglar som sätter sig inom 10 meter från S_1 . Vi ser att $Y \sim \text{Bin}(3, 1/4)$ så att

$$\mathbb{E}[Y] = 3 \frac{1}{4} = \frac{3}{4}.$$

- (c) Vi söker här

$$\begin{aligned} \mathbb{P}(X_1 \leq 10 | X_1 \geq 5) \\ = \frac{\mathbb{P}(X_1 \leq 10, X_1 \geq 5)}{\mathbb{P}(X_1 \geq 5)} = \frac{\mathbb{P}(5 \leq X_1 \leq 10)}{\mathbb{P}(X_1 \geq 5)} = \frac{5/40}{35/40} = \frac{5}{35} = \frac{1}{7}. \end{aligned}$$

2. Låt X ha täthetsfunktion

$$\mathbb{P}(X = k) = \frac{C}{k} \text{ för } k \in \{1, 2, 4\}.$$

- (a) Hitta C så att detta blir en täthetsfunktion. (2p)
 (b) Beräkna väntevärdet av X utan att använda dig av den momentgenererande funktionen. (2p)
 (c) Bestäm den momentgenererande funktionen för X . (2p)
 (d) Använd den momentgenererande funktionen du räknat fram i (c) för att återigen beräkna väntevärdet av X . (2p)

Lösning:

- (a) Vi har att

$$1 = \sum_{k=1}^5 \frac{C}{k} = C \left(1 + \frac{1}{2} + \frac{1}{4} \right) = C \frac{7}{4}$$

så att

$$C = \frac{4}{7}.$$

- (b) Vi har att

$$\mathbb{E}[X] = \sum_{k \in \{1, 2, 4\}} k \frac{C}{k} = 3C = \frac{12}{7}.$$

- (c) Här ser vi att

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \sum_{k \in \{1, 2, 4\}} e^{kt} \frac{C}{k} \\ &= \frac{4}{7} \left(e^t + \frac{e^{2t}}{2} + \frac{e^{4t}}{4} \right) = \frac{4e^t}{7} + \frac{2e^{2t}}{7} + \frac{e^{4t}}{7}. \end{aligned}$$

(d) Vi deriverar uttrycket i (c) och får

$$M'_X(t) = \frac{4e^t}{7} + \frac{4e^{2t}}{7} + \frac{4e^{4t}}{7}$$

så att

$$\mathbb{E}[X] = M'_X(0) = \frac{4}{7} + \frac{4}{7} + \frac{4}{7} = \frac{12}{7}.$$

3. Låt (X, Y) vara en tvådimensionell kontinuerlig slumpvariabel med täthetsfunktion

$$f_{X,Y}(x, y) = 2 \text{ för } x, y \geq 0 \text{ och } x + y \leq 1.$$

- (a) Beräkna $\mathbb{P}(X \leq Y)$. (2p)
 (b) Beräkna $\mathbb{P}(Y \leq (X - 1)^2)$. (2p)
 (c) Låt $Z = \max(X, Y)$. Beräkna $\mathbb{P}(Z \leq 3/4)$. (2p)

Lösning: Vi börjar med att observera att (X, Y) är likformigt fördelat på området $A = \{(x, y) : 0 \leq x, y \text{ och } x + y \leq 1\}$.

- (a) Man kan här använda symmetri för att direkt se att $\mathbb{P}(X \leq Y) = 1/2$ (se bild 1a).
 (b) Återigen blir det relativt enkelt om vi ritar upp en bild (se bild 1b). Utifrån denna ser vi att

$$\begin{aligned} \mathbb{P}(Y \leq (X - 1)^2) &= \int \int_{x, y \geq 0, x+y \leq 1} I(y \leq (x - 1)^2) f_{X,Y}(x, y) dy dx \\ &= \int_0^1 (x - 1)^2 2 dx = 2 \int_0^1 x^2 dx = 2 \left[\frac{x^3}{3} \right]_0^1 = \frac{2}{3}. \end{aligned}$$

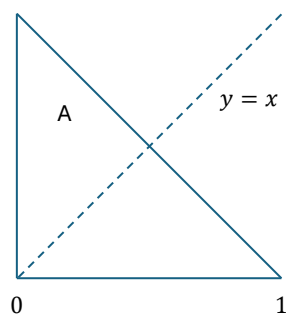
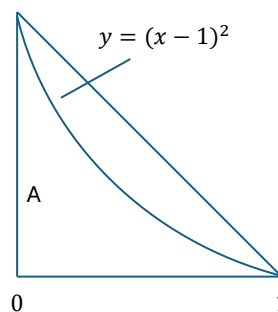
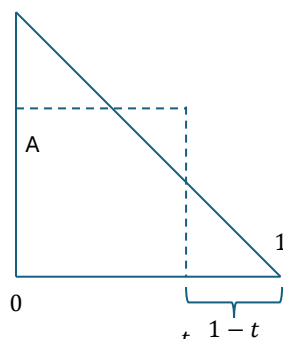
- (c) Vi har att $Z \in [0, 1]$ och för $t \in [0, 1]$ så gäller att

$$\mathbb{P}(Z \leq t) = \mathbb{P}(\max(X, Y) \leq t) = \mathbb{P}(X \leq t, Y \leq t).$$

Betrakta figur 1c. Den sökta sannolikheten blir integralen av täthetsfunktionen över arean av A minus arean av de två små triangelarna i figuren (när $t = 3/4$). Vi ser att

$$\begin{aligned} \mathbb{P}(Z \leq 3/4) &= \mathbb{P}(X \leq 3/4, Y \leq 3/4) \\ &= 2 \left(\frac{1}{2} - \left(1 - \frac{3}{4}\right)^2 \right) = 2 \left(\frac{1}{2} - \frac{1}{16} \right) = \frac{7}{8}. \end{aligned}$$

4. Betrakta följande seriekopplade system med två komponenter. Systemet fungerar om båda komponenterna fungerar. Vi antar att tiderna (T_1 respektive T_2) till att de två komponenterna fallerar (dvs går sönder) båda är Weibull-fördelade med parametrar α_1, β_1 respektive α_2, β_2 . Vi antar dessutom att dessa falleringstider är oberoende av varandra.

(a) Området A och linjen $y = x$.(b) Området A och kurvan $y = (x-1)^2$.(c) Området A och begränsningarna att $x, y \leq t$.

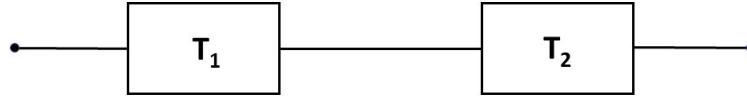
(a) Bestäm överlevnadsfunktionen för systemet. (2.5p)

(b) Bestäm feltätheten hos systemet. (2.5p)

Lösning:

(a) Låt T beteckna tiden det tar till att systemet fallerar. Vi har då att $T = \min(T_1, T_2)$ så att överlevnadsfunktionen för systemet blir

$$\begin{aligned} R_T(t) &= \mathbb{P}(T \geq t) = \mathbb{P}(\min(T_1, T_2) \geq t) = \mathbb{P}(T_1 \geq t) \mathbb{P}(T_2 \geq t) \\ &= \int_t^\infty \alpha_1 \beta_1 s^{\beta_1-1} e^{-\alpha_1 s^{\beta_1}} ds \int_t^\infty \alpha_2 \beta_2 s^{\beta_2-1} e^{-\alpha_2 s^{\beta_2}} ds \\ &= \left[-e^{-\alpha_1 s^{\beta_1}} \right]_t^\infty \left[-e^{-\alpha_2 s^{\beta_2}} \right]_t^\infty = e^{-\alpha_1 t^{\beta_1}} e^{-\alpha_2 t^{\beta_2}}. \end{aligned}$$



Figur 2: Ett system med två komponenter.

(b) Feltätheten ges av

$$\begin{aligned}
 f_T(t) &= F'_T(t) = \frac{d}{dt}(1 - R_T(t)) = -R'_T(t) \\
 &= -\left(-\alpha_1\beta_1 t^{\beta_1-1} e^{-\alpha_1 t^{\beta_1}} e^{-\alpha_2 t^{\beta_2}} - \alpha_2\beta_2 t^{\beta_2-1} e^{-\alpha_1 t^{\beta_1}} e^{-\alpha_2 t^{\beta_2}}\right) \\
 &= (\alpha_1\beta_1 t^{\beta_1-1} + \alpha_2\beta_2 t^{\beta_2-1})R_T(t)
 \end{aligned}$$

5. Låt X vara en slumpvariabel med täthetsfunktion

$$f_X(x) = \frac{2}{\theta} \left(1 - \frac{x}{\theta}\right) \text{ för } 0 \leq x \leq \theta, \quad (1)$$

där $\theta > 0$ är en parameter vars värde är okänt.

- (a) Bestäm momentskattaren (dvs MME:n) $\hat{\theta}$ av θ . (3p)
- (b) Är din skattare väntevärdesriktig (VVR)? (1p)
- (c) Följande data kom från ett stickprov med oberoende slumpvariabler dragna från täthetsfunktionen (1):

1.2 4.3 3.6 0.7 4.5

Vad blir värdet på din skattning av θ ?

(1p)

(d) Använd Chebyshevs olikhet för att ge en uppskattning av

$$\mathbb{P}(|\hat{\theta} - \theta| \geq 1/10).$$

Du kan här anta att stickprovet består av 1000 oberoende slumpvariabler där alla har täthetsfunktion som i (1). (4p)

Lösning:

(a) Vi beräknar väntevärdet av X och får

$$\begin{aligned}
 \mathbb{E}[X] &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\theta} x \frac{2}{\theta} \left(1 - \frac{x}{\theta}\right) dx \\
 &= \frac{2}{\theta} \left[\frac{x^2}{2} - \frac{x^3}{3\theta} \right]_0^{\theta} = \frac{2}{\theta} \left(\frac{\theta^2}{2} - \frac{\theta^3}{3\theta} \right) = \theta - \frac{2\theta}{3} = \frac{\theta}{3}.
 \end{aligned}$$

Momentmetoden säger att vi skall sätta

$$\bar{X} = \frac{\hat{\theta}}{3} \text{ så att } \hat{\theta} = 3\bar{X}.$$

Vår MME är alltså $\hat{\theta} = 3\bar{X}$.

(b) Vi har att

$$\mathbb{E}[\hat{\theta}] = 3\mathbb{E}[\bar{X}] = 3\mathbb{E}\left[\frac{1}{n} \sum_{k=1}^n X_k\right] = \frac{3}{n} \sum_{k=1}^n \mathbb{E}[X_k] = \frac{3}{n} \sum_{k=1}^n 3\theta = \theta,$$

så att skattaren $\hat{\theta} = 3\bar{X}$ från (a) är VVR.

(c) Data ger att $\bar{x} \approx 2.86$ så att $\hat{\theta}(x) = 3\bar{x} \approx 8.58$.

(d) Chebyshevs olikhet säger att

$$\mathbb{P}(|Y - \mathbb{E}[Y]| \geq a) \leq \frac{\text{Var}(Y)}{a^2}.$$

I vårt fall är $Y = \hat{\theta} = 3\bar{X}$ så att

$$\begin{aligned} \mathbb{P}(|\hat{\theta} - \theta| \geq 1/10) &= \mathbb{P}(|3\bar{X} - \theta| \geq 1/10) \\ &\leq \frac{\text{Var}(3\bar{X})}{1/10^2} = 100 \cdot 9 \cdot \text{Var}(\bar{X}) = 900 \text{Var}\left(\frac{1}{1000} \sum_{k=1}^{1000} X_k\right) \\ &= \frac{900}{1000^2} \sum_{k=1}^{1000} \text{Var}(X_k) = \frac{9}{10} \text{Var}(X_1), \end{aligned}$$

och det återstår att beräkna $\text{Var}(X_1)$. Vi har att $\text{Var}(X_1) = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2$ där $\mathbb{E}[X_1]$ beräknades i uppgift (a). Vi ser att

$$\begin{aligned} \mathbb{E}[X_1^2] &= \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^{\theta} x^2 \frac{2}{\theta} \left(1 - \frac{x}{\theta}\right) dx \\ &= \frac{2}{\theta} \left[\frac{x^3}{3} - \frac{x^4}{4\theta} \right]_0^{\theta} = \frac{2}{\theta} \left(\frac{\theta^3}{3} - \frac{\theta^4}{4\theta} \right) = \frac{2\theta^2}{3} - \frac{\theta^2}{2} = \frac{\theta^2}{6}, \end{aligned}$$

så att

$$\text{Var}(X_1) = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = \frac{\theta^2}{6} - \frac{\theta^2}{9} = \frac{\theta^2}{18}.$$

Därmed blir

$$\mathbb{P}(|\hat{\theta} - \theta| \geq 1/10) \leq \frac{9}{10} \text{Var}(X_1) = \frac{9}{10} \cdot \frac{\theta^2}{18} = \frac{\theta^2}{20}.$$

6. Guldletargänet GoldDiggaz vaskar guld i Klondyke. De gräver upp slam från gamla bäckfåror och denna slam vaskar de i sin maskin GoldVasquez.

Vaskningen är aldrig perfekt, så en del av guldet hamnar i slagghögen och kommer inte till nytta. GoldDiggaz funderar på att investera i en ny maskin (GoldGramps) som enligt försäljaren ger en större avkastning.

För att undersöka detta grävde GoldDiggaz fram 8 prover, och varje prov delades in i två lika stora delar. Sedan vaskades de två delarna med GoldVasquez respektive GoldGramps med följande resultat:

Prov nr:	1	2	3	4	5	6	7	8
Avkastning GV (g):	10.3	14.2	3.5	17.7	8.8	9.7	19.5	1.3
Avkastning GG (g):	10.3	16.4	4.5	21.4	8.1	11.3	22.1	5.0

Den första raden innehåller resultatet för GoldVasquez medans den andra innehåller resultatet för GoldGramps.

- Formulera hypoteser för att testa huruvida GoldGramps ger ett bättre resultat än GoldVasquez. (1p)
- Ange en lämplig test-statistika och en förkastningsregion motsvarande signifikansnivån $\alpha = 0.05$. Vad kan du dra för slutsats med de angivna data? (5p)
- Antag att GoldGramps ger 2 extra gram i avkastning i genomsnitt. Bestäm styrkan av ditt test under detta antagande. (3p)

Lösning: Vi har uppenbarligen att göra med en situation där vi har parade data. Vi bildar därför skillnaderna $Z_k = Y_k - X_k$ där X_k är mängden guld för GoldVasquez i prov nr k och Y_k är mängden guld för GoldGramps i prov nr k . Vi har då att $Z_1, \dots, Z_8$ är i.i.d. $N(\Delta, \sigma^2)$ där vi vill undersöka Δ och där σ^2 är okänd.

- De misstänker ju att GoldGramps är bättre än GoldVasquez vilket skulle innebära att $\Delta > 0$. Hypoteserna blir därför

$$\begin{aligned} H_0 : \quad \Delta &= 0 \\ H_1 : \quad \Delta &> 0 \end{aligned}$$

- Vi har att

$$\bar{Z} \approx N\left(\Delta, \frac{\sigma^2}{8}\right) \stackrel{H_0}{=} N\left(0, \frac{\sigma^2}{8}\right),$$

så vi använder test-statistikan

$$T = \frac{\bar{Z}}{s/\sqrt{8}} \sim t(7)$$

där som vanligt

$$s^2 = \frac{1}{7} \sum_{k=1}^8 (Z_k - \bar{Z})^2$$

används för att skatta σ^2 . Om H_1 gäller så blir T typiskt större än 0 så vi förkastar H_0 om T hamnar i en förkastningsregion på formen $[c, \infty)$. Här ges c av sambandet

$$\alpha = 0.05 = \mathbb{P}(T \in [c, \infty)) = \mathbb{P}(T \geq c)$$

där $T \sim t(7)$. Tabell ger att $c \approx 1.895$. Vi kan nu beräkna skillnaderna från data och ser att

Prov nr:	1	2	3	4	5	6	7	8
Skillnad Δ (g):	0	2.2	1.0	3.7	-0.7	1.6	2.6	3.7

Insatta data ger att $s^2 \approx 2.597$ och test-statistikan blir

$$T(z) = \frac{\bar{z}}{s(z)/\sqrt{8}} \approx 3.09.$$

Detta tillhör förkastningsregionen så vi förkastar H_0 på signifikansnivån $\alpha = 0.05$.

(c) Om H_1 är korrekt med $\Delta = 2$ så har vi istället att

$$\tilde{T} = \frac{\bar{Z} - 2}{s/\sqrt{8}} \sim t(7).$$

Styrkan ges av $\mathbb{P}(H_0 \text{ förkastas} | H_1 \text{ sann med } \Delta = 2)$ vilket i detta fall blir

$$\begin{aligned} \mathbb{P}(T \geq 1.895) &= \mathbb{P}\left(\frac{\bar{Z}}{s/\sqrt{8}} \geq 1.895\right) = \mathbb{P}\left(\bar{Z} \geq 1.895 \frac{s}{\sqrt{8}}\right) \\ &= \mathbb{P}\left(\bar{Z} - 2 \geq 1.895 \frac{s}{\sqrt{8}} - 2\right) = \mathbb{P}\left(\frac{\bar{Z} - 2}{s/\sqrt{8}} \geq 1.895 - \frac{2}{s/\sqrt{8}}\right). \end{aligned}$$

Vi har att

$$1.895 - \frac{2}{s/\sqrt{8}} \approx -1.615$$

så att styrkan blir

$$\mathbb{P}(\tilde{T} \geq -1.615) = \mathbb{P}(\tilde{T} \leq 1.615) \approx 0.92$$

där vi approximerat från tabell med 7 frihetsgrader.

7. Vid tillverkning av datorchipp med hög prestanda blir en viss andel defekta. Tillverkaren ChIP4U tar ett stickprov om $n = 500$ chipp och finner att 23 av dessa är defekta.

- (a) Hitta en punktskattning och ett 99% konfidensintervall för proportionen defekta chipp som ChIP4U's tillverkningsprocess resulterar i. (5p)

- (b) ChIP4U vill ha ett mycket smalt konfidensintervall för proportionen defekta chipp. Ungefär hur stort stickprov måste de ha för att deras 99% konfidensintervall skall ha bredd 0.001? (2p)

Lösning:

- (a) Vi använder punktskattaren $\hat{p} = \frac{X}{n}$ där $X \sim \text{Bin}(n, p)$ och där $n = 500$. Med insatta data får vi

$$\hat{p}(x) = \frac{x}{n} = \frac{23}{500} (\approx 0.046).$$

Vi har att dessutom att

$$\begin{aligned} \hat{p}(X) &\approx N(\mathbb{E}[\hat{p}(X)], \text{Var}(\hat{p}(X))) \\ &= N\left(\frac{\mathbb{E}[X]}{n}, \frac{\text{Var}(X)}{n^2}\right) = N\left(p, \frac{p(1-p)}{n}\right) \approx N\left(p, \frac{\hat{p}(1-\hat{p})}{n}\right) \end{aligned}$$

så en lämplig referensvariabel blir

$$R = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx N(0, 1).$$

Vi har från tabell att $z_{0.005} \approx 2.576$ så att

$$\begin{aligned} 0.99 &= \mathbb{P}(-z_{0.005} \leq Z \leq z_{0.005}) \approx \mathbb{P}\left(-z_{0.005} \leq \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq z_{0.005}\right) \\ &= \mathbb{P}\left(\hat{p} - z_{0.005} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{0.005} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \end{aligned}$$

så ett 99% konfidensintervall för p blir

$$I_p = \hat{p} \pm z_{0.005} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \frac{23}{500} \pm 2.576 \sqrt{\frac{\frac{23}{500} \frac{477}{500}}{500}} \approx [0.022, 0.07].$$

- (b) Bredden på konfidensintervallet är

$$2z_{0.005} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Om vi sätter detta till 0.001 så får vi att

$$0.001 = 2z_{0.005} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \Rightarrow n = \hat{p}(1-\hat{p}) \left(\frac{2z_{0.005}}{0.001}\right)^2 \approx 1164817.$$