

Tentamentsskrivning i **Matematisk Statistik MVE091/MVE090**

Tid: den 15 april, 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 7 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.

Följande formler kan eventuellt vara till nytta:

$$\sum_{i=1}^n i = \frac{n(n+1)}{2} \text{ och } \sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$$

-
1. Betrakta en helt vanlig kortlek. Herr Lööf drar slumpmässigt fem kort ur leken.

- (a) Hur många kombinationer av fem kort finns det? På hur många sätt kan Herr Lööf dra fem kort ur leken som alla är spader? (2p)
- (b) Vad är sannolikheten att Herr Lööf får exakt tre kungar bland de fem korten? (2p)
- (c) Antag att det första kortet som Herr Lööf drog var spader Kung. Vad är nu sannolikheten att han får exakt tre kungar bland de fem korten? (2p)

Lösning:

- (a) Vi har 52 objekt att välja mellan varav 13 är spader. Herr Lööfs hand består av fem kort så det totala antalet händer är därför $\binom{52}{5}$ och antalet händer som består av fem spader är $\binom{13}{5}$.
- (b) Det finns fyra kungar, så antalet sätt man kan välja tre av dem är $\binom{4}{3}$. De två resterande korten kan sedan väljas hur som helst av de återstående 48 korten (kom ihåg att den sista kungen inte får bli vald!). Detta kan ske på $\binom{48}{2}$ sätt. Det finns alltså

$$\binom{4}{3} \binom{48}{2}$$

händer med exakt tre kungar och den sökta sannolikheten blir därför

$$\frac{\binom{4}{3}\binom{48}{2}}{\binom{52}{5}} (\approx 0.0017)$$

- (c) Om det första kortet är en kung så återstår 51 kort varav tre är kungar. Exakt två av dessa skall väljas vilket kan ske på $\binom{3}{2}$ sätt medans de två återstående korten kan väljas på $\binom{48}{2}$ sätt, precis som i föregående uppgift. Den sökta (betingade) sannolikheten blir därför

$$\frac{\binom{3}{2}\binom{48}{2}}{\binom{51}{4}} (\approx 0.0135),$$

där $\binom{51}{4}$ är antalet sätt man kan välja de fyra återstående korten.

2. Låt X vara en slumpvariabel med täthetsfunktion

$$f_X(x) = Cx(1-x) \text{ för } 0 \leq x \leq 1.$$

- (a) Bestäm konstanten C så att detta verkligen blir en täthetsfunktion. (2p)
- (b) Beräkna $\mathbb{P}(X \geq 1/3)$. (2p)
- (c) Beräkna sannolikheten att $X \in [1/4, 3/4]$ givet att $X \geq 1/4$, dvs

$$\mathbb{P}\left(\frac{1}{4} \leq X \leq \frac{3}{4} \mid X \geq \frac{1}{4}\right).$$

(3p)

- (d) Beräkna $\mathbb{E}[X^n]$ för alla heltal $n \geq 1$. (2p)

Lösning:

- (a) Vi har att C måste uppfylla

$$1 = \int_0^1 Cx(1-x)dx = C \left[\frac{x^2}{2} - \frac{x^3}{3} \right]_0^1 = C \left(\frac{1}{2} - \frac{1}{3} \right) = \frac{C}{6}$$

så att $C = 6$.

- (b) Vi har att

$$\begin{aligned} \mathbb{P}(X \geq 1/3) &= \int_{1/3}^1 6x(1-x)dx \\ &= 6 \left[\frac{x^2}{2} - \frac{x^3}{3} \right]_{1/3}^1 = 6 \left(\frac{1}{2} - \frac{1}{3} - \frac{1}{18} + \frac{1}{81} \right) = \frac{20}{27}. \end{aligned}$$

(c) Vi har att

$$\mathbb{P}\left(\frac{1}{4} \leq X \leq \frac{3}{4} \mid X \geq \frac{1}{4}\right) = \frac{\mathbb{P}\left(\frac{1}{4} \leq X \leq \frac{3}{4} \cap X \geq \frac{1}{4}\right)}{\mathbb{P}\left(X \geq \frac{1}{4}\right)} = \frac{\mathbb{P}\left(\frac{1}{4} \leq X \leq \frac{3}{4}\right)}{\mathbb{P}\left(X \geq \frac{1}{4}\right)}.$$

Vi ser att

$$\begin{aligned}\mathbb{P}\left(X \geq \frac{1}{4}\right) &= 1 - \mathbb{P}\left(X \leq \frac{1}{4}\right) = 1 - \int_0^{1/4} 6x(1-x)dx \\ &= 1 - 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^{1/4} = 1 - 6\left(\frac{1}{32} - \frac{1}{192}\right) = 1 - \frac{5}{32} = \frac{27}{32},\end{aligned}$$

och på liknande sätt

$$\begin{aligned}\mathbb{P}\left(\frac{1}{4} \leq X \leq \frac{3}{4}\right) &= 1 - \mathbb{P}\left(X \leq \frac{1}{4}\right) - \mathbb{P}\left(X \geq \frac{3}{4}\right) \\ &= 1 - 2\mathbb{P}\left(X \leq \frac{1}{4}\right) = 1 - 2\frac{5}{32} = \frac{22}{32},\end{aligned}$$

så att

$$\mathbb{P}\left(\frac{1}{4} \leq X \leq \frac{3}{4} \mid X \geq \frac{1}{4}\right) = \frac{\frac{22}{32}}{\frac{27}{32}} = \frac{22}{27} (\approx 0.8148).$$

(d) Vi har att

$$\begin{aligned}\mathbb{E}[X^n] &= \int_0^1 6x^n x(1-x)dx = 6 \int_0^1 x^{n+1} - x^{n+2}dx \\ &= 6\left[\frac{x^{n+2}}{n+2} - \frac{x^{n+3}}{n+3}\right]_0^1 = 6\left(\frac{1}{n+2} - \frac{1}{n+3}\right) = \frac{6}{(n+2)(n+3)}.\end{aligned}$$

3. Låt (X, Y) ha gemensam täthetsfunktion

$$f_{X,Y}(k, n) = \frac{1}{2N} \text{ för } 1 \leq k \leq N \text{ och } n \in \{k-1, k+1\}.$$

(a) Bestäm marginaltäthetsfunktionerna för X respektive Y . (4p)

(b) Beräkna $\text{Cov}(X, Y)$, dvs kovariansen mellan X och Y . (4p)

Lösning:

(a) Vi har att

$$f_X(k) = f_{X,Y}(k, k-1) + f_{X,Y}(k, k+1) = 2\frac{1}{2N} = \frac{1}{N} \text{ för } k = 1, \dots, N$$

så att X är likformigt fördelad på mängden $1, 2, \dots, N$.

Vidare ser vi att Y alltid antar ett värde som är antingen 1 mer än X eller 1 mindre och därmed har vi att $Y \in \{0, 1, \dots, N, N+1\}$. Om

Y ligger “nära kanten”, dvs om $Y \in \{0, 1, N, N+1\}$ så kan X bara ha antagit ett enda värde, men om istället $Y \in \{2, \dots, N-1\}$ så kan X ha antagit ett av två möjliga värden. Därför blir

$$f_Y(0) = \mathbb{P}(Y = 0) = \mathbb{P}(X = 1, Y = 0) = \frac{1}{2N}$$

och på samma sätt får vi att

$$\begin{aligned} f_Y(1) &= \mathbb{P}(Y = 1) = \mathbb{P}(X = 2, Y = 1) = \frac{1}{2N} \\ &= \mathbb{P}(X = N-1, Y = N) = \mathbb{P}(X = N, Y = N+1). \end{aligned}$$

För $n \in \{2, \dots, N-1\}$ ser vi att istället

$$f_Y(n) = \mathbb{P}(X = n-1, Y = n) + \mathbb{P}(X = n+1, Y = n) = 2\frac{1}{2N} = \frac{1}{N}.$$

- (b) Vi använder uttrycket $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ där vi har att $\mathbb{E}[X] = \frac{N+1}{2}$ då X är likformigt fördelad på mängden $\{1, 2, \dots, N\}$. På liknande sätt kan man se att även $\mathbb{E}[Y] = \frac{N+1}{2}$ av symmetri-skäl. Dessutom har vi att

$$\begin{aligned} \mathbb{E}[XY] &= \sum_{k=1}^N k(k+1)\mathbb{P}(X = k, Y = k+1) + k(k-1)\mathbb{P}(X = k, Y = k-1) \\ &= \sum_{k=1}^N (k(k+1) + k(k-1))\frac{1}{2N} = \frac{1}{N} \sum_{k=1}^N k^2 = \frac{1}{N} \frac{N(N+1)(2N+1)}{6} \\ &= \frac{(N+1)(2N+1)}{6} \end{aligned}$$

så att till slut

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \frac{(N+1)(2N+1)}{6} - \left(\frac{N+1}{2}\right)^2 = \frac{N^2 - 1}{12}. \end{aligned}$$

4. Ett mekaniskt system fallerar efter en slumpmässig tid T som anses vara exponentialfördelad med parameter $\beta > 0$.
- (a) Beräkna felintensiteten för T med hjälp av sannolikhetstätheten för T . (OBS, ange inte bara svaret ifall du råkar ha det på din lapp) (2p)
 - (b) Förklara eventuella brister med modellantagandet, dvs att T är exponentialfördelad med parameter $\beta > 0$. (1p)

Lösning:

- (a) felintensiteten ges av

$$\rho_T(t) = \frac{f_T(t)}{R_T(t)} = \frac{f_T(t)}{1 - F_T(t)}$$

där $f_T(t) = \frac{1}{\beta} e^{-t/\beta}$ för $t > 0$. Dessutom är

$$F_T(t) = \int_0^t f_T(s) ds = \int_0^t \frac{1}{\beta} e^{-s/\beta} ds \left[e^{-s/\beta} \right]_0^t = 1 - e^{-t/\beta},$$

så att

$$\rho_T(t) = \frac{f_T(t)}{1 - F_T(t)} = \frac{\frac{1}{\beta} e^{-t/\beta}}{e^{-t/\beta}} = \frac{1}{\beta},$$

för $t > 0$.

- (b) Problemet är att intensiteten är konstant. Därmed reflekterar den inte att systemet slits eller eventuellt har tillverkningsbrister som gör att den kan gå sönder snabbt.

5. Betrakta en slumpvariabel X med täthetsfunktion

$$f_X(x) = \frac{x}{a} \exp\left(-\frac{x^2}{2a}\right) \text{ för } x \geq 0,$$

där $a > 0$ är en parameter.

- (a) Bestäm momentskattaren (MME:n) för a . (4p)
- (b) Bestäm maximum likelihood skattaren (MLE:n) för a . (4p)

Lösning:

- (a) Enligt momentmetoden skall vi beräkna väntevärdet av X . Man kan identifiera att detta är en Weibull-fördelad fördelning med $\alpha = 1/a$ och $\beta = 2$ (notation från föreläsningarna). Enligt föreläsningen är då

$$\mathbb{E}[X] = \alpha^{-1/\beta} \Gamma(1 + 1/\beta) = \sqrt{a} \Gamma(3/2).$$

Alternativt kan man göra som följer. Observera att

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty x \frac{x}{a} \exp\left(-\frac{x^2}{2a}\right) dx \\ &= \left[-x \exp\left(-\frac{x^2}{2a}\right) \right]_0^\infty + \int_0^\infty \exp\left(-\frac{x^2}{2a}\right) dx = \int_0^\infty \exp\left(-\frac{x^2}{2a}\right) dx. \end{aligned}$$

Den sista integralen kan vi räkna ut implicit. Vi har ju att

$$1 = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^\infty \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma} \int_0^\infty \exp\left(-\frac{x^2}{2\sigma^2}\right) dx,$$

så vi ser att

$$\mathbb{E}[X] = \int_0^\infty \exp\left(-\frac{x^2}{2a}\right) dx = \sqrt{a}\sqrt{\frac{\pi}{2}}.$$

Momentmetoden säger då att vi skall ansätta ekvationen

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

blir vår momentskattare. Den första metoden (Weibull) ger

$$\hat{a} = \Gamma(3/2)^{-2}\bar{X}^2,$$

vilket så klart också ger full poäng.

(b) Vi börjar med att beräkna likelihooden:

$$L(a) = \prod_{k=1}^n \frac{X_k}{a} \exp\left(-\frac{X_k^2}{2a}\right) = a^{-n} \left(\prod_{k=1}^n X_k\right) \exp\left(-\sum_{k=1}^n \frac{X_k^2}{2a}\right),$$

och bildar sedan log-likelihooden:

$$l(a) = \log L(a) = -n \log(a) + \sum_{k=1}^n \log(X_k) - \frac{1}{2a} \sum_{k=1}^n X_k^2.$$

Vi ser då att

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

och att

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

Ekvationen $l'(a) = 0$ ger oss lösningen

$$\hat{a} = \frac{1}{2n} \sum_{k=1}^n X_k^2$$

och stoppar man in detta i yttrycket för $l''(a)$ så ser vi att

$$l''(\hat{a}) = \frac{n}{\hat{a}^2} - \frac{1}{\hat{a}^3} \sum_{k=1}^n X_k^2 = \frac{1}{\hat{a}^2} \left(n - \frac{\sum_{k=1}^n X_k^2}{\frac{1}{2n} \sum_{k=1}^n X_k^2} \right) = -\frac{n}{\hat{a}^2} < 0,$$

så vi har hittat ett maximum.

6. En sjö ligger i nära anslutning till ett område som nyligen haft ett utsläpp av det giftiga ämnet PFAS. Området har tidigare drabbats av PFAS utsläpp, och nu undrar man hur detta senaste utsläpp har påverkat halten

PFAS i fiskar som bor i sjön. Vid mätningar som genomfördes innan det senaste utsläppet så var den genomsnittliga koncentrationen PFAS 27.3 ng/kg (nanogram per kilogram).

Man tar ett stickprov (av storlek $n = 8$) som man antar kommer från en normalfördelning med okänt väntevärde μ . Värdet på standardavvikelsen $\sigma = 1.2$ anses vara känt baserat på äldre mätningar från samma sjö. Data blev som följer

Fisk nr:	1	2	3	4	5	6	7	8
Halt PFAS (ng/kg):	29.5	28.8	26.7	28.9	29.3	27.7	28.6	28.3

- Testa på signifikansnivån 0.05 huruvida halten PFAS har ökat. Ange tydligt nollhypotes och alternativ hypotes. Vad blir p -värdet av ditt test? (4p)
- Om den korrekta halten är $\mu = 28.3$ ng/kg, vad blir då styrkan av ditt test? (3p)
- Verkar antagandet att $\sigma = 1.2$ korrekt? Svara genom att hitta ett 95% konfidensintervall för σ och tolka detta. (3p)

Lösning:

- Vi betraktar här hypoteserna

$$\begin{aligned} H_0 : & \mu = 27.3 \\ H_1 & \mu > 27.3. \end{aligned}$$

En lämplig test-statistika blir

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - 27.3}{1.2/\sqrt{8}} \stackrel{H_0}{\sim} N(0, 1).$$

Data ger att $\bar{x} \approx 28.475$ och därmed blir

$$p\text{-värde} = \mathbb{P}\left(T \geq \frac{\bar{x} - 27.3}{1.2/\sqrt{8}}\right) \approx \mathbb{P}(Z \geq 2.77) \approx 1 - 0.9972 = 0.0028.$$

Då detta är långt under $\alpha = 0.05$ så förkastar vi H_0 till förmån för H_1 på denna signifikansnivå.

- För att beräkna styrkan måste vi börja med att hitta förkastningsområdet. Vi har här att $RR = [1.645, \infty)$ då vi förkastar H_0 till förmån för H_1 ifall test-statistikan T antar ett för stort, positivt värde och då $0.05 = \mathbb{P}(T \geq 1.645)$. Styrkan ges sedan av

$$\begin{aligned} \mathbb{P}(T \in RR | \mu = 28.3) &= \mathbb{P}\left(\frac{\bar{X} - 27.3}{1.2/\sqrt{8}} \geq 1.645\right) \\ &= \mathbb{P}\left(\bar{X} \geq 27.3 + 1.645 \frac{1.2}{\sqrt{8}}\right) = \mathbb{P}\left(\frac{\bar{X} - 28.3}{1.2/\sqrt{8}} \geq \frac{27.3 - 28.3}{1.2/\sqrt{8}} + 1.645\right) \\ &\approx \mathbb{P}(Z \geq -0.712) = \mathbb{P}(Z \leq 0.712) \approx 0.76. \end{aligned}$$

- (c) För ett konfidensintervall för σ^2 används referensvariabeln

$$R = \frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2.$$

Ett 95% K.I. för σ^2 ges därför av uttrycket (använd tabell!)

$$0.95 = \mathbb{P}\left(1.69 \leq \frac{(n-1)s^2}{\sigma^2} \leq 16\right) = \mathbb{P}\left(\frac{(n-1)s^2}{16} \leq \sigma^2 \leq \frac{(n-1)s^2}{1.69}\right).$$

Data ger att $s^2 \approx 0.831$ så ett 95% numeriskt K.I. för σ^2 blir därför

$$I_{\sigma^2} = \left[\frac{(n-1)s^2}{16}, \frac{(n-1)s^2}{1.69}\right] \approx [0.3634, 3.4408].$$

Från detta får vi ett 95% numeriskt K.I. för σ som följer

$$I_{\sigma} \approx [\sqrt{0.3634}, \sqrt{3.4408}] \approx [0.61, 1.85].$$

Då $\sigma = 1.2$ tillhör konfidensintervallet så verkar det inte orimligt att $\sigma = 1.2$.

7. Några forskare trodde att härdtiden för ett nytt epoxylim eventuellt var proportionellt mot luftfuktigheten i den omgivande luften. De genomförde därför en liten serie av experiment för att testa detta med följande resultat.

luftfuktighet (i %):	30	40	50	60	70	80
Härdtid (i timmar):	23.4	23.2	20.6	24.4	22.8	21.2

Vid experimenten hölls den omgivande temperaturen konstant, och alla andra externa faktorer var lika. Forskarna antar ett linjärt samband mellan härdtiden och luftfuktigheten på formen $y = a + bx$. Data kan sammanfattas med att

$$S_{xx} = 1750, S_{yy} = 10.24 \text{ och att } S_{xy} = -42.$$

- (a) Skatta parametrarna a, b . (1p)
 (b) Ange ett 95% konfidensintervall för b . (4p)
 (c) Beräkna förklaringsgraden. Verkar modellen rimlig? (1p)

Lösning:

- (a) Vi har att

$$\hat{b} = \frac{S_{xy}}{S_{xx}} \approx -0.024 \text{ och att } \hat{a} = \bar{y} - \hat{b}\bar{x} \approx 23.9.$$

(b) Vi har att

$$\hat{b} \sim N\left(b, \frac{\sigma^2}{S_{xx}}\right)$$

Vi använder därför referensvariabeln

$$R = \frac{\hat{b} - b}{S_r / \sqrt{S_{xx}}} \sim t(4).$$

Här blir

$$S_r^2 = \frac{1}{n-2} \sum_{k=1}^n (y_k - \hat{y})^2 \approx 2.31$$

så att $S_r \approx 1.52$. Ett 95% KI ges därför av (ty $t_{0.975}(4) \approx 2.776$)

$$\begin{aligned} 0.95 &= \mathbb{P}\left(-2.776 \leq \frac{\hat{b} - b}{S_r / \sqrt{S_{xx}}} \leq 2.776\right) \\ &= \mathbb{P}\left(\hat{b} - 2.776 \frac{S_r}{\sqrt{S_{xx}}} \leq b \leq \hat{b} + 2.776 \frac{S_r}{\sqrt{S_{xx}}}\right), \end{aligned}$$

så att ett 95% numeriskt KI blir

$$I_b = \hat{b} \pm 2.776 \frac{S_r}{\sqrt{S_{xx}}} \approx -0.024 \pm 2.776 \frac{1.52}{\sqrt{1750}} \approx [-0.1248, 0.0768].$$

(c) Förklaringsgraden blir

$$R^2 = \frac{S_{xy}^2}{S_{xx} \cdot S_{yy}} \approx 0.098$$

vilket är uselt. Modellantagandet verkar klen underbyggt eller också är beroendet mycket svagt (dvs b är nära 0).