

Tentamentsskrivning i **Matematisk Statistik MVE091/MVE090**

Tid: den 28e augusti 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:

$$\sum_{k=1}^n k = \frac{n(n+1)}{2}, \quad \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6}, \quad \sum_{k=0}^n x^k = \frac{1-x^{n+1}}{1-x}$$

$$\lim_{n \rightarrow \infty} (1 + a_n)^{b_n} = e^c \text{ om } \lim_{n \rightarrow \infty} a_n b_n = c \text{ och } \lim_{n \rightarrow \infty} a_n = 0.$$

1. Merre har tre lådor med osorterade skruv av längderna 60mm, 70mm och 80mm. Antalet skruvar av varje sort kan vi se i följande tabell:

	60mm	70mm	80mm
Låda nummer 1:	2	2	3
Låda nummer 2:	1	4	2
Låda nummer 3:	3	3	1.

Tabellen avläses så att det finns 2 st 60mm skruv i Låda nummer 1, 2 st 70mm skruv i Låda nummer 1 etc.

- Antag att Merre väljer en låda slumpmässigt och sedan tar han tre skruvar (återigen slumpmässigt) ur den valda lådan. Vad är sannolikheten att han får en skruv vardera av de tre längderna? (2.5p)
- Antag återigen att Merre väljer en låda slumpmässigt men nu tar han två skruvar (återigen slumpmässigt) ur den valda lådan. Vad är sannolikheten att dessa två skruvar har samma längd? (2.5p)
- Antag återigen att han väljer en låda slumpmässigt, men att Merre sedan bara tar en enda skruv (vald slumpmässigt) från den valda lådan. Givet att denna skruv har längd 70mm, vad är sannolikheten att han valde låda nummer 2? (2p)

Lösning:

- (a) Vi låter $\mathcal{O}$ beteckna händelsen att Merre får tre skruvar av olika längd och vi låter L_k vara händelsen att Merre valde låda nummer k där $k = 1, 2, 3$. Vi ser då att

$$\begin{aligned}\mathbb{P}(\mathcal{O}) &= \sum_{k=1}^3 \mathbb{P}(\mathcal{O}|L_k)\mathbb{P}(L_k) \\ &= \frac{1}{3} \left(\frac{\binom{2}{1}\binom{2}{1}\binom{3}{1}}{\binom{7}{3}} + \frac{\binom{1}{1}\binom{4}{1}\binom{2}{1}}{\binom{7}{3}} + \frac{\binom{3}{1}\binom{3}{1}\binom{1}{1}}{\binom{7}{3}} \right) \\ &= \frac{12 + 8 + 9}{3\binom{7}{3}} = \frac{29}{105} (\approx 0.276).\end{aligned}$$

- (b) Låt $\mathcal{I}$ beteckna händelsen att Merre väljer skruvar av identisk längd. Vi har att

$$\begin{aligned}\mathbb{P}(\mathcal{I}) &= \sum_{k=1}^3 \mathbb{P}(\mathcal{I}|L_k)\mathbb{P}(L_k) \\ &= \frac{1}{3} \left(\frac{\binom{2}{2} + \binom{2}{2} + \binom{3}{2}}{\binom{7}{2}} + \frac{\binom{4}{2} + \binom{2}{2}}{\binom{7}{2}} + \frac{\binom{3}{2} + \binom{3}{2}}{\binom{7}{2}} \right) \\ &= \frac{5 + 7 + 6}{3\binom{7}{2}} = \frac{18}{63} = \frac{2}{7} (\approx 0.286).\end{aligned}$$

- (c) Låt $\mathcal{S}$ beteckna händelsen att Merre väljer en skruv av längd 70mm. Vi söker

$$\begin{aligned}\mathbb{P}(L_2|\mathcal{S}) &= \mathbb{P}(\mathcal{S}|L_2) \frac{\mathbb{P}(L_2)}{\mathbb{P}(\mathcal{S})} \\ &= \frac{4}{7} \frac{\mathbb{P}(L_2)}{\sum_{k=1}^3 \mathbb{P}(\mathcal{S}|L_k)\mathbb{P}(L_k)} = \frac{4}{7} \frac{1}{\frac{2}{7} + \frac{4}{7} + \frac{3}{7}} = \frac{4}{9}.\end{aligned}$$

2. En växtgrossist (Dino) planterar chilifrön. Om ett frö gror, så blir det en planta som Dino kan sälja för 35 kronor. Antag att grobarheten för dessa frön är 75%. Det vill säga, sannolikheten att ett frö gror är 0.75. Antag också att fröna gror oberoende av varandra.

- (a) Dino köper 400 frön för 200 kronor och hon sår alla dessa i krukor. Antag vidare att kostnaden (el, värme, näring etc men *inte* inräknat kostnaden för fröet) för en planta som gror är 23 kronor, och att en planta som inte gror kostar Dino 9 kronor. Beräkna den förväntade vinsten. (3p)
- (b) Beräkna sannolikheten att Dinos vinst överstiger 4500 SEK så att hon nästa säsong kan köpa ett större växthus. (3p)

Lösning:

- (a) Låt X =antalet frön som gror, och observera att det enligt uppgiftens formulering gäller att $X \sim \text{Bin}(400, 0.75)$. Om V är vinsten så har vi att

$$V = -200 + (35 - 23)X - 9(400 - X) = -3800 + 21X$$

(startkostnad, vinst för en lyckad planta, kostnad för en misslyckad planta). Vi ser att

$$\mathbb{E}[V] = -3800 + 21\mathbb{E}[X] = -3800 + 21 \cdot 400 \cdot 0.75 = 2500,$$

där vi använde våra kunskaper om binomialfördelningen.

- (b) Vi använder normalapproximation. Enligt CGS har vi att

$$X \stackrel{d}{\approx} N(\mathbb{E}[X], \text{Var}(X)) = N(400 \cdot 0.75, 400 \cdot 0.75 \cdot 0.25) = N(300, 75).$$

Vi ser därför att

$$\begin{aligned} V &\stackrel{d}{\approx} N(\mathbb{E}[V], \text{Var}(V)) = N(2500, \text{Var}(-3800 + 21X)) \\ &= N(2500, 21^2 \text{Var}(X)) = N(2500, 441 \cdot 75) = N(2500, 33075). \end{aligned}$$

Vi får därför att

$$\mathbb{P}(V \geq 4500) = \mathbb{P}\left(\frac{V - 2500}{\sqrt{33075}} \geq \frac{4500 - 2500}{\sqrt{33075}}\right) \approx \mathbb{P}(Z \geq 10.99),$$

så denna sannolikhet är i princip obefintlig.

3. Betrakta den tvådimensionella kontinuerliga slumpvariabeln (X, Y) med gemensam täthetsfunktion

$$f(x, y) = e^{-x} \text{ om } 0 \leq y \leq x < \infty.$$

- (a) Beräkna marginaltäthetsfunktionerna för X respektive Y och verifiera att $f(x, y)$ verkligen är en täthetsfunktion. (3p)
- (b) Beräkna $\text{Cov}(X, Y)$, dvs kovariansen mellan X och Y . (3p)
- (c) Låt $Z = 1$ om $X \leq 1$ och låt $Z = 0$ annars. Antag vidare att Z_1, Z_2, Z_3 är oberoende slumpvariabler som alla har samma fördelning som Z . Beräkna $\mathbb{P}(Z_1 + Z_2 + Z_3 = 1)$. (3p)

Lösning:

- (a) Vi har att

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = \int_0^x e^{-x} dy = xe^{-x} \text{ för } 0 \leq x < \infty$$

och att

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y) dx \\ &= \int_y^{\infty} e^{-x} dx = [-e^{-x}]_y^{\infty} = e^{-y} \text{ för } 0 \leq y < \infty. \end{aligned}$$

Uppenbarligen är $f(x, y) \geq 0$ och vidare är

$$\begin{aligned} &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} f_Y(y) dy = \int_0^{\infty} e^{-y} dy = [-e^{-y}]_0^{\infty} = 1. \end{aligned}$$

(b) Vi har att

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

där

$$\mathbb{E}[Y] = \int_0^{\infty} ye^{-y} dy \stackrel{P.I.}{=} [-ye^{-y}]_0^{\infty} + \int_0^{\infty} e^{-y} dy = 1$$

och

$$\mathbb{E}[X] = \int_0^{\infty} x \cdot xe^{-x} dx \stackrel{P.I.}{=} [-x^2e^{-x}]_0^{\infty} + \int_0^{\infty} 2xe^{-x} dx = 2.$$

Vidare är

$$\begin{aligned} \mathbb{E}[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y) dx dy = \int_0^{\infty} \int_0^x xye^{-x} dy dx \\ &= \int_0^{\infty} \frac{x^3}{2} e^{-x} dx \stackrel{P.I.}{=} \left[-\frac{x^3}{2} e^{-x} \right]_0^{\infty} + \frac{3}{2} \int_0^{\infty} x^2 e^{-x} dx = 0 + \frac{3}{2} \cdot 2 = 3 \end{aligned}$$

så att

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 3 - 2 = 1.$$

(c) Vi har att

$$\begin{aligned} \mathbb{P}(Z = 1) &= \mathbb{P}(X \leq 1) = \int_0^1 f_X(x) dx = \int_0^1 xe^{-x} dx \\ &\stackrel{P.I.}{=} [-xe^{-x}]_0^1 + \int_0^1 e^{-x} dx = -e^{-1} + [-e^{-x}]_0^1 = 1 - 2e^{-1} (\approx 0.2642). \end{aligned}$$

Vi får därför att $Z_1 + Z_2 + Z_3 \sim \text{Bin}(3, 1 - 2e^{-1})$ och den sökta sannolikheten blir då

$$\begin{aligned} &\mathbb{P}(Z_1 + Z_2 + Z_3 = 1) \\ &= \binom{3}{1} (1 - 2e^{-1})^1 (1 - (1 - 2e^{-1}))^2 = 3 (1 - 2e^{-1}) 4e^{-2} (\approx 0.429). \end{aligned}$$

4. Låt X ha täthetsfunktion

$$f_X(x) = \cos(x) \text{ för } 0 \leq x \leq \pi/2$$

(a) Bestäm den momengenererande funktionen för X . (3p)

(b) Använd svaret i (a) för att beräkna $\mathbb{E}[X^3]$. (3p)

Lösning:

(a) Vi har att

$$\begin{aligned} I &= \mathbb{E}[e^{Xt}] = \int_0^{\pi/2} e^{xt} \cos(x) dx \\ &\stackrel{P.I.}{=} [\sin(x)e^{xt}]_0^{\pi/2} - t \int_0^{\pi/2} e^{xt} \sin(x) dx \\ &\stackrel{P.I.}{=} e^{\pi t/2} - t \left([-\cos(x)e^{xt}]_0^{\pi/2} + t \int_0^{\pi/2} e^{xt} \cos(x) dx \right) \\ &= e^{\pi t/2} - t(1 + tI) = e^{\pi t/2} - t - t^2 I \end{aligned}$$

så att

$$I(1 + t^2) = e^{\pi t/2} - t \Rightarrow I = \frac{e^{\pi t/2} - t}{1 + t^2}.$$

(b) Vi använder att

$$\mathbb{E}[e^{Xt}] = \sum_{k=0}^{\infty} \mathbb{E}[X^k] \frac{t^k}{k!},$$

så för att hitta $\mathbb{E}[X^3]$ kan vi leta upp koefficienten framför t^3 . Vi har att

$$\mathbb{E}[e^{Xt}] = \frac{e^{\pi t/2} - t}{1 + t^2} = \left[\sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{\pi t}{2} \right)^k - t \right] \times \left[\sum_{n=0}^{\infty} (-t^2)^n \right].$$

Vi ser att $k = 1, n = 1$ och $k = 3, n = 0$ är de enda termerna i dubbelsumman som bidrar till koefficienten framför t^3 . Detta ger oss

$$\left(\frac{\pi t}{2} - t \right) (-t^2) + \frac{1}{3!} \left(\frac{\pi t}{2} \right)^3 = \frac{t^3}{3!} \left(6 \left(1 - \frac{\pi}{2} \right) + \left(\frac{\pi}{2} \right)^3 \right)$$

så att

$$\mathbb{E}[X^3] = 6 \left(1 - \frac{\pi}{2} \right) + \left(\frac{\pi}{2} \right)^3.$$

5. Betrakta en diskret slumpvariabel X med sannolikhetsfunktion

$$p_X(k) = (1 - e^{-b})^2 k e^{-b(k-1)} \text{ för } k = 1, 2, \dots$$

där $b > 0$ är en parameter.

- (a) Bestäm maximum likelihood skattaren för b . (4p)
- (b) Antag att fyra oberoende försök ger stickprovet $x_1 = 4, x_2 = 6, x_3 = 2, x_4 = 5$. Använd dessa data för att uppskatta sannolikheten att nästa datapunkt blir minst 3. (2p)

Lösning:

- (a) Vi har att

$$L(b) = \prod_{l=1}^n (1 - e^{-b})^2 X_l e^{-b(X_l - 1)}$$

så att

$$l(b) = \log(L(b)) = 2n \log(1 - e^{-b}) + \sum_{l=1}^n \log X_l - b \sum_{l=1}^n (X_l - 1).$$

För att hitta maximat deriverar vi och får

$$l'(b) = 2n \frac{e^{-b}}{1 - e^{-b}} - \sum_{l=1}^n (X_l - 1).$$

Villkoret $l'(b) = 0$ ger oss då att

$$\frac{e^{-b}}{1 - e^{-b}} = \frac{1}{2n} \sum_{l=1}^n (X_l - 1).$$

Vi använder sedan att $\frac{x}{1-x} = c$ har lösningen $x = \frac{c}{1+c}$ för att se att

$$e^{-b} = \frac{\frac{1}{2n} \sum_{l=1}^n (X_l - 1)}{1 + \frac{1}{2n} \sum_{l=1}^n (X_l - 1)}$$

så att

$$\begin{aligned} b &= \log \left(1 + \frac{1}{2n} \sum_{l=1}^n (X_l - 1) \right) - \log \left(\frac{1}{2n} \sum_{l=1}^n (X_l - 1) \right) \\ &= \log \left(1 + \frac{1}{2} (\bar{X} - 1) \right) - \log \left(\frac{1}{2} (\bar{X} - 1) \right) = \log \left(\frac{\bar{X} + 1}{\bar{X} - 1} \right). \end{aligned}$$

För att säkerställa att detta är ett maximum så betraktar vi

$$l''(b) = 2n \frac{-e^{-b}(1 - e^{-b}) - e^{-b}(e^{-b})}{(1 - e^{-b})^2} = 2n \frac{-e^{-b}}{(1 - e^{-b})^2} < 0$$

då $b > 0$.

(b) Låt $\hat{b}$ beteckna svaret från (a). Vi har att

$$\begin{aligned}\hat{b} &= \log \left(1 + \frac{1}{8} \sum_{l=1}^n (x_l - 1) \right) - \log \left(\frac{1}{8} \sum_{l=1}^n (x_l - 1) \right) \\ &= \log(1 + 13/8) - \log(13/8) \approx 0.4796.\end{aligned}$$

Vi söker sedan sannolikheten

$$\begin{aligned}\mathbb{P}(X \geq 3) &= 1 - \mathbb{P}(X = 2) - \mathbb{P}(X = 1) \\ &= 1 - \left(1 - e^{-\hat{b}}\right)^2 1e^{-\hat{b}(1-1)} - \left(1 - e^{-\hat{b}}\right)^2 2e^{-\hat{b}(2-1)} \\ &= 1 - \left(1 - e^{-\hat{b}}\right)^2 \left(1 + 2e^{-\hat{b}}\right) \approx 0.6752.\end{aligned}$$

6. Ett företag vill hitta fyndigheter av den sällsynta jordartsmetallen Yttrium. Företaget kommer dock inte ihåg huruvida man kan hitta Yttrium i Ytterby utanför Stockholm (Y-Sthlm), eller i Ytterby utanför Göteborg (Y-Gbg). De skickar därför ut två gäng geologer för att ta prover. Gänget som åker till Y-Sthlm tar 5 prover och gänget som åker till Y-Gbg tar också 5 prover. Resultatet blev som följer:

Prov nr (Y-Sthlm):	1	2	3	4	5
Halt (ppm):	64	120	97	53	87
Prov nr (Y-Gbg):	1	2	3	4	5
Halt (ppm):	31	26	43	18	37

Geologerna antar att data är oberoende och kommer från normalfördelningar med samma varians σ^2 i båda proven.

- (a) Bestäm ett 99% dubbelsidigt konfidensintervall för skillnaden i halten Yttrium i marken mellan Y-Sthlm och Y-Gbg. (4p)
- (b) Testa huruvida halten i Y-Sthlm är större än halten i Y-Gbg på signifikansnivån $\alpha = 0.01$. Gör detta genom att använda lämpligt konfidensintervall och dualiteten mellan konfidensintervall och test. (3p)

Lösning:

- (a) Vi kallar halten Yttrium i Y-Sthlm respektive Y-Gbg för μ_S respektive μ_G . Vi ombeds bilda ett konfidensintervall för $\mu_S - \mu_G$. Om $X = (X_1, \dots, X_5)$ betecknar stickprovet från Y-Sthlm och $Y = (Y_1, \dots, Y_5)$ betecknar stickprovet från Y-Gbg så säger uppgiftsformuleringen att

$$\bar{X} \sim N\left(\mu_S, \frac{\sigma^2}{5}\right) \text{ och } \bar{Y} \sim N\left(\mu_G, \frac{\sigma^2}{5}\right).$$

Vi har därför att

$$\bar{X} - \bar{Y} \sim N\left(\mu_S - \mu_G, \frac{2\sigma^2}{5}\right)$$

så att

$$R = \frac{\bar{X} - \bar{Y} - (\mu_S - \mu_G)}{\sqrt{2s_p^2/5}} \sim t(5 + 5 - 2) = t(8)$$

där

$$s_p^2 = \frac{\sum_{k=1}^5 (X_k - \bar{X})^2 + \sum_{k=1}^5 (Y_k - \bar{Y})^2}{8}.$$

Ett 99% konfidensintervall ges av (med $t_{0.005}(8) \approx 3.355$ och insatta data)

$$\begin{aligned} I_{\mu_S - \mu_G} &= \bar{x} - \bar{y} \pm t_{0.005}(8)s_p\sqrt{2/5} \\ &\approx 84.2 - 31 \pm 3.355 \cdot 20.03 \cdot \sqrt{2/5} \approx [10.7, 95.7]. \end{aligned}$$

(b) Vi vill testa

$$\begin{aligned} H_0 : \mu_S &= \mu_G \\ H_1 : \mu_S &> \mu_G. \end{aligned}$$

genom att använda dualitetsprincipen. Vi skapar därför ett ensidig intervall på formen $[c, \infty)$ för $\mu_S - \mu_G$, och om 0 inte tillhör detta intervall så förkastar vi H_0 . Anledningen är förstås att om $0 \notin [c, \infty)$ så har vi gott stöd för att säga att skillnaden är positiv. Räkningarna blir mycket lika de i uppgift (a) och vi får att

$$\begin{aligned} I_{\mu_S - \mu_G} &= [\bar{x} - \bar{y} - t_{0.01}(8)s_p\sqrt{2/5}, \infty) \\ &\approx [84.2 - 31 - 2.896 \cdot 20.03 \cdot \sqrt{2/5}, \infty) \approx [16.52, \infty). \end{aligned}$$

Då 0 inte tillhör detta intervall kan vi förkasta H_0 till förmån för H_1 .

7. Antalet jordbävningar (över gränsvärdet 2 på Richterskalan) i Kalifornien under ett år anses vara Poissonfördelat med parameter $\lambda > 0$. Historiskt sett har λ ansetts vara 23.4, men The Rock misstänker att det nu har ökat (vilket bådär illa för framtiden). The Rock samlar in data under ett år och registrerar 28 jordbävningar.

- (a) Använd The Rocks nyinsamlade data för att testa på signifikansnivån 5% huruvida The Rock har rätt i sin misstanke att antalet jordbävningar per år har ökat. Ställ tydligt upp dina hypoteser och beräkna det kritiska området. (5p)
- (b) Beräkna p-värde av ditt test. (2p)
- (c) Om λ i själva verket är 30, vad är då styrkan för testet? (2p)

Lösning:

- (a) Vi skall alltså testa

$$H_0 : \lambda = 23.4$$

$$H_1 : \lambda > 23.4.$$

Om X =antalet jordbävningar under ett år, så är alltså $X \sim \text{Poi}(\lambda)$. Vi har vidare att då λ_0 är stort så kan vi normalapproximera och se att $X \approx N(\mathbb{E}[X], \text{Var}(X)) = N(\lambda, \lambda)$ så att

$$T = \frac{X - \lambda_0}{\sqrt{\lambda_0}} \stackrel{H_0}{\approx} N(0, 1)$$

där så klart $\lambda_0 = 23.4$. Vi förkastar H_0 till förmån för H_1 om värdet på test-statistikan blir för stor. Ett förkastningsområde för T blir därför $[1.645, \infty)$ då $z_{0.05} \approx 1.645$. I vårt fall får vi att

$$T(x) = \frac{x - \lambda_0}{\sqrt{\lambda_0}} \approx 0.951,$$

så att $T(x)$ inte tillhör det kritiska området. Vi skall därför inte förkasta H_0 till förmån för H_1 .

- (b) p-värdet ges i detta fall av

$$\begin{aligned} \mathbb{P}(T(X) \geq T(x) | H_0) &= \mathbb{P}(T(X) \geq 0.951) \\ &= 1 - \mathbb{P}(T(X) \leq 0.951) \approx 1 - 0.829 \approx 0.17, \end{aligned}$$

med hjälp av tabell.

- (c) Styrkan ges av

$$\begin{aligned} \mathbb{P}\left(\frac{X - \lambda_0}{\sqrt{\lambda_0}} \geq 1.645 \mid \lambda = 30\right) &= \mathbb{P}(X \geq \lambda_0 + 1.645\sqrt{\lambda_0}) \\ &= \mathbb{P}\left(\frac{X - 30}{\sqrt{30}} \geq \frac{\lambda_0 + 1.645\sqrt{\lambda_0} - 30}{\sqrt{30}}\right) \\ &\approx \mathbb{P}(Z \geq 0.248) \approx 1 - 0.597 \approx 0.4. \end{aligned}$$