

Tentamentsskrivning i **Matematisk Statistik MVE091**

Tid: den 28 oktober, 2022 em

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. Berra har två lådor med strumpor. I den första lådan finns det 2 gröna, 2 gula och en blå strumpa. I den andra lådan finns det 3 blåa och 2 gula strumpor. Berra väljer först en låda slumpmässigt (dvs med samma sannolikhet) och väljer därefter två strumpor slumpmässigt ur den valda lådan.

(a) Vad är sannolikheten att Berra får en gul och en blå strumpa? (4p)

(b) Givet att Berra fick en gul och en blå strumpa, vad är sannolikheten att Berra valde låda nummer 1? (2p)

Lösning: Låt L_k vara händelsen att Berra valde låda nummer k där $k = 1, 2$. Vi låter sedan BG vara händelsen att Berra får en gul och en blå strumpa.

(a) Vi har att

$$\mathbb{P}(BG|L_1) = \frac{\binom{2}{1}\binom{1}{1}}{\binom{5}{2}} = \frac{2}{10} = \frac{1}{5},$$

(för om han valde lådan nr 1 så finns det två gula att välja mellan och en blå) och att

$$\mathbb{P}(BG|L_2) = \frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} = \frac{6}{10} = \frac{3}{5}.$$

Vi ser därför att

$$\mathbb{P}(BG) = \mathbb{P}(BG|L_1)\mathbb{P}(L_1) + \mathbb{P}(BG|L_2)\mathbb{P}(L_2) = \frac{1}{5} \cdot \frac{1}{2} + \frac{3}{5} \cdot \frac{1}{2} = \frac{2}{5}.$$

- (b) Vi söker alltså $\mathbb{P}(L_1|BG)$ och använder Bayes sats för att se att

$$\mathbb{P}(L_1|BG) = \mathbb{P}(BG|L_1) \frac{\mathbb{P}(L_1)}{\mathbb{P}(BG)} = \frac{1}{5} \frac{\frac{1}{2}}{\frac{2}{5}} = \frac{1}{4}.$$

2. Låt X ha täthetsfunktion

$$f(x) = Ax^4 \text{ för } 0 \leq x \leq 1.$$

- (a) Bestäm A så att detta verkligen är en täthetsfunktion. (2p)
 (b) Beräkna väntevärdet och standardavvikelsen av X . (3p)
 (c) Beräkna $\mathbb{P}(0.7 \leq X \leq 0.9)$ (2p)

Lösning:

- (a) Vi har att A måste uppfylla villkoret

$$1 = \int_{-\infty}^{\infty} f(x)dx = \int_0^1 Ax^4dx = \frac{A}{5}$$

så att $A = 5$.

- (b) Väntevärdet blir

$$\begin{aligned} \mathbb{E}[X] &= \int_{-\infty}^{\infty} xf(x)dx \\ &= \int_0^1 x5x^4dx = 5 \int_0^1 x^5dx = 5 \left[\frac{x^6}{6} \right]_0^1 = \frac{5}{6}. \end{aligned}$$

Vidare är

$$\begin{aligned} \mathbb{E}[X^2] &= \int_{-\infty}^{\infty} x^2f(x)dx \\ &= \int_0^1 x^25x^4dx = 5 \int_0^1 x^6dx = 5 \left[\frac{x^7}{7} \right]_0^1 = \frac{5}{7}, \end{aligned}$$

så att standardavvikelsen blir

$$\sqrt{\text{Var}(X)} = \sqrt{\mathbb{E}[X^2] - \mathbb{E}[X]^2} = \sqrt{\frac{5}{7} - \frac{25}{36}} = \sqrt{\frac{5}{252}} (\approx 0.1409)$$

- (c) Vi har att

$$\mathbb{P}(0.7 \leq X \leq 0.9) = \int_{0.7}^{0.9} 5x^4dx = [x^5]_{0.7}^{0.9} = 0.9^5 - 0.7^5 (\approx 0.4224).$$

3. En flodhäst på ett zoo äter antingen 0, 1 eller 2 paket flodhästmat per dygn. Sannolikheten att den äter 0 paket är 0.01, sannolikheten att den äter 1 paket är 0.1 och sannolikheten att den äter 2 paket är 0.89. Vi antar att hur mycket flodhästen äter på olika dagar är oberoende av varandra.

- (a) Beräkna sannolikheten att flodhästen äter fler än 190 paket mat under 100 dygn. (4p)
- (b) Beräkna sannolikheten att flodhästen äter fler än 190 paket mat under 100 dygn *och* fler än 190 paket under de därpå följande 100 dygnen. (2p)

Lösning: Låt X_k vara antalet paket som flodhästen äter under dygn nummer k där $k = 1, 2, 3, \dots$

- (a) Vi söker här $\mathbb{P}\left(\sum_{k=1}^{100} X_k \geq 190\right)$ och måste använda normalapproximation. Vi har att

$$\sum_{k=1}^{100} X_k \approx N\left(\mathbb{E}\left[\sum_{k=1}^{100} X_k\right], \text{Var}\left(\sum_{k=1}^{100} X_k\right)\right).$$

Vidare är

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

och

$$\text{Var}\left(\sum_{k=1}^{100} X_k\right) = \sum_{k=1}^{100} \text{Var}(X_k) = 100\text{Var}(X_1)$$

där vi använde oberoendeantagandet i sista likheten. Dessutom är

$$\begin{aligned}\mathbb{E}[X_1] &= 0\mathbb{P}(X_1 = 0) + 1\mathbb{P}(X_1 = 1) + 2\mathbb{P}(X_1 = 2) \\ &= 0 \cdot 0.01 + 1 \cdot 0.1 + 2 \cdot 0.89 = 1.88,\end{aligned}$$

och

$$\begin{aligned}\mathbb{E}[X_1^2] &= 0^2\mathbb{P}(X_1 = 0) + 1^2\mathbb{P}(X_1 = 1) + 2^2\mathbb{P}(X_1 = 2) \\ &= 0 \cdot 0.01 + 1 \cdot 0.1 + 4 \cdot 0.89 = 3.66,\end{aligned}$$

så att $\text{Var}(X_1) = \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 = 0.1256$. Därför blir

$$\sum_{k=1}^{100} X_k \approx N(188, 12.56).$$

Till sist ser vi då att

$$\begin{aligned}\mathbb{P}\left(\sum_{k=1}^{100} X_k \geq 190\right) &= \mathbb{P}\left(\frac{\sum_{k=1}^{100} X_k - 188}{\sqrt{12.56}} \geq \frac{190 - 188}{\sqrt{12.56}}\right) \\ &\approx \mathbb{P}(Z \geq 0.56) = 1 - \mathbb{P}(Z \leq 0.56) \approx 1 - 0.71 = 0.29.\end{aligned}$$

- (b) De två händelserna är oberoende och vi har därför att den sökta sannolikheten blir

$$\begin{aligned} & \mathbb{P}\left(\sum_{k=1}^{100} X_k \geq 190, \sum_{k=101}^{200} X_k \geq 190\right) \\ &= \mathbb{P}\left(\sum_{k=1}^{100} X_k \geq 190\right) \mathbb{P}\left(\sum_{k=101}^{200} X_k \geq 190\right) \\ &= \mathbb{P}\left(\sum_{k=1}^{100} X_k \geq 190\right)^2 \approx 0.29^2 \approx 0.084. \end{aligned}$$

4. Låt (X, Y) vara två kontinuerliga slumpvariabler med gemensam täthetsfunktion

$$f_{XY}(x, y) = 6(x - y) \text{ för } 0 \leq y \leq 1 \text{ och } y \leq x \leq 1.$$

- (a) Bestäm marginaltäthetsfunktionen för Y . (2p)
 (b) Bestäm den betingade täthetsfunktionen för X givet $Y = y$. (1p)
 (c) Beräkna $\mathbb{E}[X|Y = y]$ för alla värden på y . (3p)

Lösning:

- (a) Vi har att

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x, y) dx = \int_y^1 6(x - y) dx \\ &= 6 \left[\frac{x^2}{2} - xy \right]_y^1 = 6 \left(\frac{1}{2} - y - \frac{y^2}{2} + y^2 \right) \\ &= 3 - 6y + 3y^2 = 3(y - 1)^2 \text{ för } 0 \leq y \leq 1. \end{aligned}$$

- (b) Enligt definition så är denna

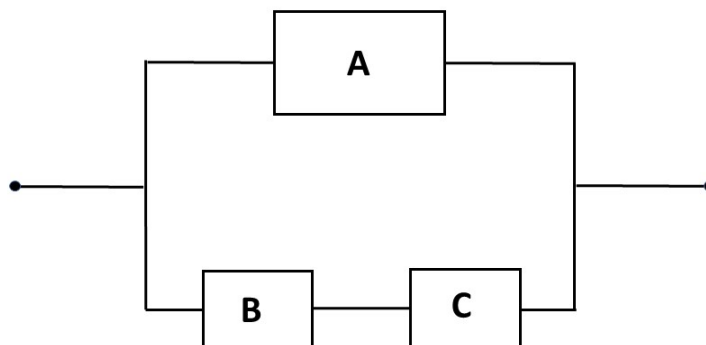
$$f_{X|Y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)} = \frac{6(x - y)}{3(y - 1)^2} = \frac{2(x - y)}{(y - 1)^2},$$

för $0 \leq y \leq 1$ och $y \leq x \leq 1$.

- (c) Enligt definition så har vi att

$$\begin{aligned} \mathbb{E}[X|Y = y] &= \int_{-\infty}^{\infty} x f_{X|Y}(x) dx \\ &= \int_y^1 x \frac{2(x - y)}{(y - 1)^2} dx = \frac{2}{(y - 1)^2} \int_y^1 (x^2 - xy) dx \\ &= \frac{2}{(y - 1)^2} \left[\frac{x^3}{3} - \frac{x^2 y}{2} \right]_y^1 = \frac{2}{(y - 1)^2} \left(\frac{1}{3} - \frac{y}{2} - \frac{y^3}{3} + \frac{y^3}{2} \right) \\ &= \frac{2}{(y - 1)^2} \frac{2 - 3y + y^3}{6} = \frac{y^3 - 3y + 2}{3(y - 1)^2} = \frac{y + 2}{3} \text{ för } 0 \leq y \leq 1. \end{aligned}$$

5. Betrakta ett elektriskt system enligt Figur 1. Systemet fungerar om antingen komponent A fungerar, eller om både B och C fungerar. Vi låter F_A vara händelsen att komponent A fungerar efter ett år och vi låter $p_A = \mathbb{P}(F_A)$ vara motsvarande sannolikhet. På liknande sätt definierar vi F_B, F_C och p_B, p_C .



Figur 1: Ett elektriskt system.

- (a) Under antagandet att alla händelser är oberoende av varandra, vad är sannolikheten (uttryckt i p_A, p_B och p_C) att systemet fungerar efter ett år? (3p)
- (b) Vad blir sannolikheten i (a) om vi vet att händelsen F_A^c inträffar (dvs att komponent A ej fungerar efter ett år)? (3p)

Lösning:

- (a) För att systemet skall fungera måste F_A inträffa och/eller $F_B \cap F_C$. Sannolikheten att hela systemet fungerar är därför

$$\begin{aligned}
 \mathbb{P}(F_A \cup (F_B \cap F_C)) &= \mathbb{P}(F_A) + \mathbb{P}(F_B \cap F_C) - \mathbb{P}(F_A \cap (F_B \cap F_C)) \\
 &= \mathbb{P}(F_A) + \mathbb{P}(F_B)\mathbb{P}(F_C) - \mathbb{P}(F_A)\mathbb{P}(F_B)\mathbb{P}(F_C) \\
 &= p_A + p_B p_C - p_A p_B p_C,
 \end{aligned}$$

där vi använde oberoende i andra likheten.

- (b) Om vi vet att komponent A är sönder så krävs det att både B och C fungerar för att systemet skall fungera. Alternativt ser vi att

$$\mathbb{P}(F_A \cup (F_B \cap F_C) | F_A^c) = \mathbb{P}(F_B \cap F_C | F_A^c) = \mathbb{P}(F_B \cap F_C).$$

Om alla händelserna är oberoende så gäller att det seriekopplade sub-systemet (komponenter B, C) fungerar med sannolikhet

$$\mathbb{P}(F_B \cap F_C) = \mathbb{P}(F_B)\mathbb{P}(F_C) = p_B p_C.$$

6. Låt $X_1, X_2, \dots, X_n$ vara i.i.d. (dvs oberoende och likafördelade) där fördelningen är $\text{Gamma}(\alpha, \beta)$.

(a) Bestäm likelihooden för $X_1, \dots, X_n$. (1p)

(b) Bestäm log-likelihooden för $X_1, \dots, X_n$ och expandera detta uttryck (dvs skriv inte bara log av svaret i (a)). (2p)

(c) Antag att $\alpha = 3$. Hitta maximum likelihood skattaren (MLE:n) för β . Använd att $\Gamma(3) = 2$. (4p)

Lösning:

(a) Enligt föreläsningen så ges likelihooden $L(\alpha, \beta)$ av den gemensamma täthetsfunktionen som blir

$$L(\alpha, \beta) = \prod_{k=1}^n f(x_k) = \prod_{k=1}^n \frac{1}{\Gamma(\alpha)\beta^\alpha} x_k^{\alpha-1} e^{-x_k/\beta}.$$

(b) Log-likelihooden $l(\alpha, \beta)$ fås genom att logaritmera $L(\alpha, \beta)$. Vi ser att

$$\begin{aligned} l(\alpha, \beta) &= \log \left(\prod_{k=1}^n \frac{1}{\Gamma(\alpha)\beta^\alpha} x_k^{\alpha-1} e^{-x_k/\beta} \right) \\ &= -n \log(\Gamma(\alpha)\beta^\alpha) + (\alpha - 1) \sum_{k=1}^n \log(x_k) - \frac{1}{\beta} \sum_{k=1}^n x_k. \end{aligned}$$

(c) Om $\alpha = 3$ så ser vi att

$$\begin{aligned} l(3, \beta) &= -n \log(2\beta^3) + 2 \sum_{k=1}^n \log(x_k) - \frac{1}{\beta} \sum_{k=1}^n x_k \\ &= -n \log(2) - 3n \log(\beta) + 2 \sum_{k=1}^n \log(x_k) - \frac{1}{\beta} \sum_{k=1}^n x_k. \end{aligned}$$

Vi deriverar denna med avseende på β och får att

$$l'(3, \beta) = -\frac{3n}{\beta} + \frac{1}{\beta^2} \sum_{k=1}^n x_k$$

och att

$$l''(3, \beta) = \frac{3n}{\beta^2} - \frac{2}{\beta^3} \sum_{k=1}^n x_k.$$

Om vi sätter $0 = l'(3, \beta)$ så får vi ekvationen

$$3n = \frac{1}{\beta} \sum_{k=1}^n x_k \text{ så att } \beta = \frac{1}{3n} \sum_{k=1}^n x_k = \frac{\bar{x}}{3}.$$

För att se om detta motsvarar ett maximum så undersöker vi $l''(3, \bar{x}/3)$. Vi ser att

$$l''(3, \bar{x}/3) = \frac{3n}{(\bar{x}/3)^2} - \frac{2}{(\bar{x}/3)^3} \sum_{k=1}^n x_k = \frac{27n}{(\bar{x})^2} - \frac{54}{(\bar{x})^3} n\bar{x} = -\frac{27n}{(\bar{x})^2} < 0,$$

så vi har hittat ett maximum. Vår maximum likelihood skattare blir därför

$$\hat{\beta} = \frac{\bar{x}}{3}.$$

7. SkumRask tillverkar isolerskum som skall användas i byggen. De hävdar att halten asbest i deras isolerskum inte är mer än 4%. Konsumentverket misstänker att de ljugar och köper därför in 30 varuprover med isolerskum från SkumRask. Konsumentverket låter därefter testa de 30 varuproverna och får att samplingsmedelvärdet blir $\bar{x} = 5.2$ med samplingsvariansen $s^2(x) = 2.2$. Konsumentverket antar att data kommer från en normalfördelning.
- (a) Sätt upp ett lämpligt hypotestest för att testa Konsumentverkets misstankar. (2p)
 - (b) Om man väljer signifikansnivån $\alpha = 0.05$, vad blir då motsvarande förkastningsregion (rejection region, RR)? (3p)
 - (c) Utför testet i (a) på signifikansnivån $\alpha = 0.05$. Vilken slutsats drar du? Vilken typ av fel riskerar du att göra? (2p)

Lösning:

- (a) Om $X_1, \dots, X_n$ får representera halterna asbest i de 30 varuproverna (innan mätning) så har vi enligt antagandet att $X_k \sim N(\mu, \sigma^2)$ där μ och σ^2 är okända. För att testa ifall halten asbest överstiger 4% så använder vi följande hypotestest:

$$\begin{aligned} H_0 : \quad & \mu = 4 \\ H_1 : \quad & \mu > 4. \end{aligned}$$

- (b) Punktskattaren för μ är $\bar{X}$ och vi vet att

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{30}\right) = N\left(4, \frac{\sigma^2}{30}\right)$$

där likheten är sann om H_0 är sann. Vi ser därför att

$$\frac{\bar{X} - 4}{\sigma/\sqrt{30}} \sim N(0, 1) \text{ och att } \frac{\bar{X} - 4}{s(X)/\sqrt{30}} \sim t(n-1) = t(29)$$

om H_0 är sann. Vi väljer därför test-statistikan

$$T(X) = \frac{\bar{X} - 4}{s(X)/\sqrt{30}}$$

och förkastar H_0 ifall $T(X)$ "blir för stor". Förkastningsregionen blir då på formen $RR = [c, \infty)$ där c bestäms av att

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

och t-tabellen ger att $t_{0.05}(29) \approx 1.699$. Vi har därför att $c = 1.699$ och vår förkastningsregion blir därför

$$RR = [1.699, \infty).$$

(c) Med insatta data får vi att

$$T(x) = \frac{\bar{x} - 4}{s(x)/\sqrt{30}} = \frac{5.2 - 4}{\sqrt{2.2}/\sqrt{30}} \approx 4.43$$

och då $T(x) \in RR$ så förkastar vi H_0 på signifikansnivån $\alpha = 0.05$. Vi riskerar då att göra ett typ I fel, i detta fall att vi anklagar SkumRask för att ljuga om asbesthalten trots att de inte gör det.

Alternativt: Man kan lösa denna utan att använda svaret i (b) genom att beräkna p-värdet. I detta fall så blir

$$\begin{aligned} \text{p-värde} &= \mathbb{P}(T(X) \geq T(x)) \\ &= \mathbb{P}(T(X) \geq 4.43) \leq \mathbb{P}(T(X) \geq 3.659) = 0.0005, \end{aligned}$$

(där olikheten kommer sig av att t-tabellen med 29 frihetsgrader bara sträcker sig upp till punkten 3.659). Då p-värde $< 0.05 = \alpha$ drar vi såklart samma slutsats och riskerar att göra samma fel.

8. Ett företag skall välja en underleverentör som skall leverera en speciell komponent. Det är viktigt att komponenterna inte är defekta och man köper därför in två provserier från vardera underleverentör. Efter att man testat alla inköpta komponenter fick man följande resultat.

Underleverentör nr:	1	2
Antal inköpta komponenter:	800	900
Varav defekta:	32	41

- (a) Skatta proportionerna defekta komponenter från de två underleverentörerna. (1p)
- (b) Ange ett 95% konfidensintervall för skillnaden mellan proportionerna. (4p)

Lösning:

- (a) Låt p_1, p_2 beteckna de två proportionerna. Våra punktskattare är

$$\hat{p}_1 = \frac{32}{800} (\approx 0.04) \text{ och } \hat{p}_2 = \frac{41}{900} (\approx 0.0456).$$

(b) Här använder vi att enligt föreläsningarna så är

$$\hat{p}_k \approx N\left(p_k, \frac{p_k(1-p_k)}{n_k}\right) \text{ för } k = 1, 2,$$

och därmed blir

$$\begin{aligned} \hat{p}_1 - \hat{p}_2 &\approx N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right) \\ &\approx N\left(p_1 - p_2, \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}\right). \end{aligned}$$

Vi väljer därför referensvariabeln

$$R = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{S_{\hat{p}_1 - \hat{p}_2}},$$

där

$$S_{\hat{p}_1 - \hat{p}_2}^2 = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}.$$

Ett $(1-\alpha)100\%$ K.I. ges därför av

$$\begin{aligned} 1-\alpha &= \mathbb{P}(-z_{\alpha/2} \leq R \leq z_{\alpha/2}) \\ &= \mathbb{P}(-z_{\alpha/2} S_{\hat{p}_1 - \hat{p}_2} \leq \hat{p}_1 - \hat{p}_2 - (p_1 - p_2) \leq z_{\alpha/2} S_{\hat{p}_1 - \hat{p}_2}) \\ &= \mathbb{P}(\hat{p}_1 - \hat{p}_2 - z_{\alpha/2} S_{\hat{p}_1 - \hat{p}_2} \leq p_1 - p_2 \leq \hat{p}_1 - \hat{p}_2 + z_{\alpha/2} S_{\hat{p}_1 - \hat{p}_2}). \end{aligned}$$

Då $z_{0.025} \approx 1.96$ så får vi att ett 95% numeriskt K.I. för $p_1 - p_2$ ges av

$$I_{p_1 - p_2} = \hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} S_{\hat{p}_1 - \hat{p}_2} \approx [-0.0248, 0.0137].$$