Tentamentsskrivning i **Matematisk Statistik MVE091**

Tid: den 23 oktober, 2023, 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.

1. Ahmed är projektledare på ett vägbygge och skall använda armeringsjärn. Det finns två typer (typ a och typ b) av armeringsjärn tillverkade av olika stål-legeringar. Om man bockar ett armeringsjärn av typ a så knäcks det med sannolikhet 0.05, men om man bockar ett armeringsjärn av typ b så knäcks det med sannolikhet 0.6. Praktikanten Berra råkade blanda armeringsjärnen så att det i en hög ligger 100 stycken av typ a och 10 stycken av typ b.

   (a) Ahmed plockar ett armeringsjärn ur högen slumpmässigt och bockar det. Vad är sannolikheten att det knäcks? (3p)

   (b) Om armeringsjärnet som Ahmed bockar knäcks, vad är då sannolikheten att det var av typ a? (2p)

   (c) Antag istället att Ahmed plockar tre armeringsjärn slumpmässigt ur högen. Vad är sannolikheten att högst ett av armeringsjärnen är av typ b? (2p)

   Lösning:

   (a) Låt A vara händelsen att Ahmed plockar ett armeringsjärn av typ a och låt B vara händelsen att Ahmed plockar ett armeringsjärn av typ b. Vi låter sedan K beteckna händelsen att järnet knäcks vid bockning. Vi söker $\mathbb{P}(K)$ och har att

$$\begin{aligned}\mathbb{P}(K) &= \mathbb{P}(K \cap A) + \mathbb{P}(K \cap B) \\ &= \mathbb{P}(K|A)\mathbb{P}(A) + \mathbb{P}(K|B)\mathbb{P}(B) \\ &= 0.05\frac{100}{110} + 0.6\frac{10}{110} = \frac{5+6}{110} = \frac{1}{10}.\end{aligned}$$

(b) Vi söker nu $\mathbb{P}(A|K)$ och har enligt Bayes sats att

$$\mathbb{P}(A|K) = \mathbb{P}(K|A)\frac{\mathbb{P}(A)}{\mathbb{P}(K)} = 0.05\frac{\frac{100}{110}}{\frac{1}{10}} = \frac{5}{11}.$$

(c) Låt X vara antalet armeringsjärn av typ b som Ahmed plockar upp. Vi söker då

$$\begin{aligned}\mathbb{P}(X \leq 1) &= \mathbb{P}(X = 0) + \mathbb{P}(X = 1) \\ &= \frac{\binom{100}{3}}{\binom{110}{3}} + \frac{\binom{100}{2}\binom{10}{1}}{\binom{110}{3}} = \frac{211200}{215820} = \frac{320}{327} \approx 0.9786.\end{aligned}$$

Här har vi använt att man kan plocka 3 armeringsjärn ur en samling av 110 på $\binom{110}{3}$ sätt, att man kan plocka 3 armeringsjärn ur en samling av 100 på $\binom{100}{3}$ sätt etc.

2. Låt X vara en slumpvariabel med ha täthetsfunktion

$$f(k) = Ck \text{ för } k = 1, 2, 3, 4, 5$$

(a) Bestäm C så att detta verkligen är en täthetsfunktion. (2p)

(b) Låt $Y = e^{-X}$. Bestäm täthetsfunktionen för Y. (2p)

(c) Beräkna $\mathbb{E}[(\log Y)^2]$. (2p)

Lösning:

(a) Vi har att C måste uppfylla villkoret

$$1 = \sum_k f_X(k) = \sum_{k=1}^{5} Ck = C\sum_{k=1}^{5} k = 15C$$

så att $C = 1/15$.

(b) Y kan anta värden i mängden $\{e^{-1}, e^{-2}, e^{-3}, e^{-4}, e^{-5}\}$ och vi har att

$$\mathbb{P}(Y = e^{-k}) = \mathbb{P}(X = k) = \frac{k}{15}$$

för $k = 1, \ldots, 5$.

(c) Vi har att $\log Y = -X$ så att

$$\begin{aligned}\mathbb{E}[(\log Y)^2] = \mathbb{E}[(-X)^2] = \mathbb{E}[X^2] &= \sum_{k=1}^{5} k^2\mathbb{P}(X = k) \\ &= \sum_{k=1}^{5} k^2\frac{k}{15} = \frac{1 + 2^3 + 3^3 + 4^3 + 5^3}{15} = \frac{225}{15} = 15.\end{aligned}$$

3. En spegel som skall användas i ett teleskop får inte ha för många defekter. Defekter uppkommer slumpmässigt och man vet att en viss tillverkningsprocedur i genomsnitt genererar 0.03 defekter per dm^2.

   Företaget MRRor tillverkar speglar med ovan nämnda tillverkningsprocedur. Deras speglar är alla 1 m^2 stora.

   (a) Teleskopstillverkaren TeleSkope köper en spegel av MRRor. För att de skall kunna använda spegeln får den högst ha en enda defekt. Vad är sannolikheten för detta? (3p)

   (b) Om TeleSkope istället köper 1000 speglar, vad är sannolikheten att de kan använda minst 200 av dessa? Vi kan här anta att antalet defekter per spegel är oberoende av varandra för de olika speglarna. (4p)

   Lösning:

   (a) Det första vi måste förstå är att vi har att göra med en Poissonprocess. Detta då vi räknar antal händelser (defekter) i en mängd (som här är ytenhet i enheten dm^2). Låt X vara antalet defekter på en spegel och observera att 1 m^2 är 100 dm^2. Vi har därför att

   $$X \sim \text{Poi}(0.03 * 100) = \text{Poi}(3),$$

   och vi söker att $\mathbb{P}(X \leq 1)$. Vi har att

   $$\mathbb{P}(X \leq 1) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \frac{3^0}{0!}e^{-3} + \frac{3^1}{1!}e^{-3} = 4e^{-3} (\approx 0.1991).$$

   (b) Låt Y vara antalet speglar som kan användas. Under antagandet om oberoende har vi att Y ~Bin$(1000, p)$ där $p = 4e^{-3}$. Vi söker då $\mathbb{P}(Y \geq 200)$ och för att beräkna detta måste vi använda normalapproximation. Centrala gränsvärdessatsen ger att

   $$Y \approx N(\mathbb{E}[Y], \text{Var}(Y)) = N(1000p, 1000p(1-p))$$

   där vi i likheten använder vår kunskap om binomialfördelningen. Vi ser att

   $$\begin{aligned}\mathbb{P}(Y \geq 200) &= \mathbb{P}\left(\frac{Y - 1000p}{\sqrt{1000p(1-p)}} \geq \frac{200 - 1000p}{\sqrt{1000p(1-p)}}\right) \\ &\approx \mathbb{P}(Z \geq 0.0674) = 1 - \mathbb{P}(Z \leq 0.0674) \approx 1 - 0.526 = 0.474.\end{aligned}$$

   där vi satte in värdet på p och sedan har interpolerat i tabell för att få fram den sista approximationen.

4. Betrakta slumpvariablerna X, Y med den gemensamma täthetsfunktionen

   $$f_{X,Y}(x, y) = 2xe^{-y} \text{ för } x \in [0, 1] \text{ och } y \in [0, \infty).$$

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

(b) Är X, Y oberoende? (1p)

(c) Beräkna sannolikheten $\mathbb{P}(X \geq Y)$. (3p)

Lösning

(a) Vi har att

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \\ &= \int_0^{\infty} 2xe^{-y}dy = 2x\left[-e^{-y}\right]_0^{\infty} = 2x \text{ för } 0 \leq x \leq 1, \end{aligned}$$

och att

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \\ &= \int_0^1 2xe^{-y}dx = e^{-y}\left[x^2\right]_0^1 = e^{-y} \text{ för } 0 \leq y < \infty. \end{aligned}$$

(b) Om de är oberoende skall vi ha att $f_{X,Y}(x,y) = f_X(x)f_Y(y)$. Vi ser att

$$f_{X,Y}(x,y) = 2xe^{-y} = (2x)\left(e^{-y}\right) = f_X(x)f_Y(y)$$

för $x \in [0,1]$ och $y \in [0,\infty)$ och annars är allt 0, så de är oberoende.

(c) Vi har att

$$\begin{aligned} \mathbb{P}(X \geq Y) &= \int\int_{x \geq y} f_{X,Y}(x,y)dxdy \\ &= \int_0^1 \int_0^x 2xe^{-y}dydx = \int_0^1 2x\left[-e^{-y}\right]_0^x dx \\ &= \int_0^1 2x\left(1-e^{-x}\right)dx = \int_0^1 2xdx - \int_0^1 2xe^{-x}dx \\ &= [x^2]_0^1 - 2\left[-xe^{-x}\right]_0^1 - 2\int_0^1 e^{-x}dx \\ &= 1 + 2e^{-1} + 2[e^{-x}]_0^1 = 1 + 4e^{-1} - 2 = 4e^{-1} - 1. \end{aligned}$$

5. Låt X vara en slumpvariabel med täthetsfunktion

$$f_X(x) = \theta(\theta+1)x^{\theta-1}(1-x) \text{ för } 0 \leq x \leq 1$$

där $\theta > 0$ är en parameter som vi vill skatta.

(a) Bestäm momentskattaren för parametern θ. (4p)

(b) Antag att man fick följande mätvärden när man simulerade slumpvariabler från fördelningen beskriven av f_X:

data:	x_1	x_2	x_3	x_4	x_5
värde	0.19	0.23	0.09	0.31	0.44

Använd dessa mätvärdena för att hitta en numerisk skattning av θ. (1p)

(c) Använd ditt svar i (b) för att uppskatta sannolikheten att nästa mätvärde (dvs x_6) blir minst 1/2. (2p)

Lösning:

(a) Enligt momentmetoden skall vi börja med att beräkna väntevärdet av X. Vi har att

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

Momentmetoden föreskriver att vi nu skall lösa ut θ ur sambandet

$$\overline{X} = \frac{\hat{\theta}}{\hat{\theta}+2} \Leftrightarrow (\hat{\theta}+2)\overline{X} = \hat{\theta} \Leftrightarrow \hat{\theta}(1-\overline{X}) = 2\overline{X}$$

så att till slut

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

(b) Vi har att $\overline{x} = 0.252$ så att

$$\hat{\theta}(x) = \frac{2\overline{x}}{1-\overline{x}} \approx 0.6738.$$

(c) Vi söker nu

$$\begin{aligned}\mathbb{P}(X \geq 1/2) &= \int_{1/2}^1 \theta(\theta+1)x^{\theta-1}(1-x)dx \\ &= \int_{1/2}^1 \theta(\theta+1)x^{\theta-1}dx - \int_{1/2}^1 \theta(\theta+1)x^\theta dx = [(\theta+1)x^\theta]_{1/2}^1 - [\theta x^{\theta+1}]_{1/2}^1 \\ &= (\theta+1)(1-2^{-\theta}) - \theta(1-2^{-\theta-1}) \approx 0.162,\end{aligned}$$

om vi använder att $\theta = 0.6738$.

6. En kemisk process syftar till att framställa helt ren bensen. Det går dock inte att undvika orenheter, och forskarna vill förstå hur mycket orenheter slutprodukten typiskt innehåller. De gjorde därför 8 försök och fick följande data.

Försök nummer:	1	2	3	4	5	6	7	8
% orenheter:	4.5	2.9	4.4	2.6	2.8	6.0	5.8	5.8.

Forskarna antar att data kommer från en normalfördelning med okända μ och σ^2. De antar också att utfallen av experimenten är oberoende av varandra.

(a) Skatta μ och σ^2. (2p)

(b) Ange ett 95% konfidensintervall för μ. (3p)

(c) Ange ett 99% konfidensintervall för σ^2. (3p)

Lösning:

(a) Vi har att
$$\hat{\mu} = \overline{x} = 4.35$$
och att
$$\hat{\sigma^2} = s^2(x) = \frac{1}{10-1}\sum_{k=1}^{10}(x_k - \overline{x})^2 \approx 2.0743.$$

(b) Vi använder här referensvariabeln
$$R = \frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t(n-1)$$
och ser att
$$\begin{aligned} 0.95 = \mathbb{P}(-t_{0.025} \leq R \leq t_{0.025}) &= \mathbb{P}\left(-t_{0.025} \leq \frac{\overline{X} - \mu}{s/\sqrt{8}} \leq t_{0.025}\right) \\ &= \mathbb{P}\left(-t_{0.025}\frac{s}{\sqrt{8}} \leq \overline{X} - \mu \leq t_{0.025}\frac{s}{\sqrt{8}}\right) \\ &= \mathbb{P}\left(\overline{X} - t_{0.025}\frac{s}{\sqrt{8}} \leq \mu \leq \overline{X} + t_{0.025}\frac{s}{\sqrt{8}}\right). \end{aligned}$$

Ett 95% numeriskt konfidensintervall för μ blir då ($t_{0.025}(7) \approx 2.365$ enligt tabell)
$$I_\mu = \overline{x} \pm t_{0.025}\frac{s}{\sqrt{8}} \approx 4.35 \pm 2.365\frac{\sqrt{2.0743}}{\sqrt{8}} \approx [3.146, 5.554].$$

(c) Här använder vi referensvariabeln

$$R = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

och ser från tabell att $\chi^2_{0.005}(7) \approx 20.3$ och $\chi^2_{0.995}(7) \approx 0.989$. Vi har därför att

$$\begin{aligned} 0.95 &= \mathbb{P}(\chi^2_{0.995} \leq R \leq \chi^2_{0.005}) \\ &= \mathbb{P}\left(\chi^2_{0.995} \leq \frac{(n-1)S^2}{\sigma^2} \leq \chi^2_{0.005}\right) \\ &= \mathbb{P}\left(\frac{(n-1)S^2}{\chi^2_{0.005}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{0.995}}\right). \end{aligned}$$

Ett 99% numeriskt konfidensintervall för σ^2 blir då

$$\begin{aligned} I_{\sigma^2} &= \left[\frac{(n-1)s^2}{\chi^2_{0.005}(7)}, \frac{(n-1)s^2}{\chi^2_{0.995}(7)}\right] \\ &\approx \left[\frac{7 \cdot 2.0743}{20.3}, \frac{7 \cdot 2.0743}{0.989}\right] \approx [0.715, 14.682]. \end{aligned}$$

7. Arabia har två fabriker som tillverkar muminkoppar. Vid tillverkning av porslin sorteras slutprodukterna in i tre kategorier nämligen A, B och C. Kategori A innebär att produkten inte har några synliga fel, kategori B har smärre synliga fel och kategori C har allvarliga synliga fel. Koppar i kategori A säljs till fullpris, kategori B säljs till reducerat pris och kategori C kasseras.

   Ledningen för Arabia vill undersöka huruvida det finns någon skillnad i kvaliteten mellan muminkopparna som tillverkas vid de två fabrikerna. Man valde därför slumpmässigt ut 100 koppar var från de två fabrikerna och fick följande data:

Kategori:	A	B	C
Fabrik nummer 1:	61	19	20
Fabrik nummer 2:	54	18	28

   Tabellen skall läsas så att av de 100 testade kopparna från fabrik nummer 1 så var 61 av kategori A, 19 av kategori B och 20 av kategori C.

   (a) Testa huruvida proportionen koppar av kategori A skiljer sig mellan de två fabrikerna. Formulera ditt test noga och genomför testet på signifikansnivån 0.1. (4p)

   (b) Fabriksledningen misstänker på förhand att proportionen koppar som måste kasseras är högre i den andra fabriken än i den första. Formulera ditt test noga och genomför testet på signifikansnivån 0.1. (4p)

   Lösning:

(a) Låt p_1^A och p_2^A vara proportionen av kategori A i fabrik nummer 1 respektive nummer 2. Vårt hypotestest blir

$$\begin{array}{ll} H_0: & p_1^A = p_2^A \\ H_1: & p_1^A \neq p_2^A. \end{array}$$

Vi använder test-statistikan

$$T(X,Y) = \frac{\hat{p}_1^A - \hat{p}_2^A}{\sqrt{\hat{p}^A(1-\hat{p}^A)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \overset{H_0}{\approx} N(0,1)$$

där som indikerat, nollfördelningen är approximativt $N(0,1)$. Här är

$$\hat{p}^A = \frac{n_1\hat{p}_1^A + n_2\hat{p}_2^A}{n_1+n_2}.$$

Vi vill beräkna p-värdet och börjar med att kolla vad $T(x,y)$ blir. Vi har att

$$\hat{p}_1^A = \frac{61}{100} = 0.61,\ \hat{p}_2^A = \frac{54}{100} = 0.54$$

så att

$$\hat{p}^A = \frac{61+54}{200} = \frac{115}{200} = \frac{23}{40} = 0.575.$$

Detta ger oss att

$$T(x,y) = \frac{0.61-0.54}{\sqrt{0.575(1-0.575)\left(\frac{2}{100}\right)}} \approx 1,$$

så att

$$\begin{aligned} \text{p-värde} &= \mathbb{P}(T(X,Y) \geq |T(x,y)|) \\ &= 2\mathbb{P}(T(X,Y) \geq 1) = 2\mathbb{P}(T(X,Y) \leq -1) \approx 2 \cdot 0.1587 = 0.3174. \end{aligned}$$

Detta värde är långt över signifikansnivån så vi förkastar inte H_0.

(b) Lösningen är mycket lik den ovan, men vi skriver ned den ändå. Låt p_1^C och p_2^C vara proportionen av kategori C i fabrik nummer 1 respektive nummer 2. Vårt hypotestest blir

$$\begin{array}{ll} H_0: & p_1^C = p_2^C \\ H_1: & p_1^C < p_2^C. \end{array}$$

Vi använder test-statistikan

$$T(X,Y) = \frac{\hat{p}_1^C - \hat{p}_2^C}{\sqrt{\hat{p}^C(1-\hat{p}^C)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \overset{H_0}{\approx} N(0,1)$$

där som indikerat, nollfördelningen är approximativt $N(0,1)$. Här är

$$\hat{p}^C = \frac{n_1\hat{p}_1^C + n_2\hat{p}_2^C}{n_1+n_2}.$$

Vi vill beräkna p-värdet och börjar med att kolla vad $T(x,y)$ blir. Vi har att

$$\hat{p}_1^C = \frac{20}{100} = 0.2,\ \hat{p}_2^C = \frac{28}{100} = 0.28$$

så att

$$\hat{p}^C = \frac{20+28}{200} = \frac{48}{200} = \frac{6}{25} = 0.24.$$

Detta ger oss att

$$T(x,y) = \frac{0.2-0.28}{\sqrt{0.24(1-0.24)\left(\frac{2}{100}\right)}} \approx -1.3245,$$

så att

$$\begin{aligned}\text{p-värde} &= \mathbb{P}(T(X,Y) \leq T(x,y)) \\ &= \mathbb{P}(T(X,Y) \leq -1.3245) \approx 0.093.\end{aligned}$$

Detta värde är precis under signifikansnivån så vi förkastar H_0.