Tentamentsskrivning i **Matematisk Statistik MVE091**

Tid: 24 augusti, 2023 fm
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 8 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. Kålle kastar vattenballonger på förbipasserande som går på gatan utanför huset där han bor. Varje gång någon passerar kastar Kålle en ballong som träffar med sannolikhet 0.2. Vi antar att träffarna/missarna sker oberoende av varandra. Kålle antas ha 20 ballonger.

   (a) Vad är sannolikheten att Kålle missar alla de fem första kasten? (2p)

   (b) Om Kålle istället kastar två ballonger på varje förbipasserande, vad är då sannolikheten att Kålle träffar den förste förbipasserande med minst en ballong? (2p)

   (c) Vad är sannolikheten att minst en av de fem först förbipasserande blir träffad? Antag som i (*b*) att Kålle kastar två ballonger på varje förbipasserande. (2p)

   Lösning:

   (a) Låt M_k vara händelsen att kast nummer k missar. Vi söker sannolikheten att alla de fem första försöken missar, dvs

   $$\mathbb{P}(M_1 \cap \cdots \cap M_5) = \prod_{k=1}^{5} \mathbb{P}(M_k) = \mathbb{P}(M_1)^5 = (1-0.2)^5 = 0.8^5 \approx 0.3277.$$

   (b) Låt T_1 vara händelsen att första förbipasserande blir träffad. Vi har då att $T_1 = (M_1 \cap M_2)^c$, dvs händelsen att första personen blir träffad är komplementet till att de två första *kasten* missar. Vi ser att

   $$\begin{aligned}\mathbb{P}(T_1) &= \mathbb{P}((M_1 \cap M_2)^c) = 1 - \mathbb{P}(M_1 \cap M_2) \\ &= 1 - \mathbb{P}(M_1)\mathbb{P}(M_2) = 1 - 0.8^2 = 1 - 0.64 = 0.36.\end{aligned}$$

(c) Låt T vara händelsen att minst en av de fem första blir träffade. Vi har då att
$$T^c = M_1 \cap \cdots \cap M_{10},$$
så att
$$\begin{aligned}\mathbb{P}(T) &= 1 - \mathbb{P}(T^c)\\ &= 1 - \mathbb{P}(M_1 \cap \cdots \cap M_{10}) = 1 - \mathbb{P}(M_1)^{10} = 1 - 0.8^{10} \approx 0.8926.\end{aligned}$$

2. Låt X ha täthetsfunktion
$$f(x) = A\cos(x) \text{ för } -\pi/2 \leq x \leq \pi/2.$$

(a) Bestäm värdet på konstanten A. (2p)

(b) Beräkna $\mathbb{E}[|X|]$ (2p)

(c) Beräkna $\mathbb{E}[\sin^3(X)]$. (2p)

Lösning:

(a) Vi har att
$$1 = \int_{-\infty}^{\infty} f(x)dx = A\int_{-\pi/2}^{\pi/2} \cos(x)dx = A[\sin(x)]_{-\pi/2}^{\pi/2} = 2A$$
så att $A = 1/2$.

(b) Vi har att
$$\begin{aligned}\mathbb{E}[|X|] &= \int_{-\infty}^{\infty} |x| f(x)dx = \frac{1}{2}\int_{-\pi/2}^{\pi/2} |x|\cos(x)dx\\ &= \int_0^{\pi/2} x\cos(x)dx = [x\sin(x)]_0^{\pi/2} - \int_0^{\pi/2}\sin(x)dx\\ &= \left(\frac{\pi}{2} - 0\right) - [-\cos(x)]_0^{\pi/2} = \frac{\pi}{2} + 0 - 1 = \frac{\pi}{2} - 1.\end{aligned}$$

(c) Vi har att
$$\mathbb{E}[\sin^3(X)] = \int_{-\infty}^{\infty} \sin^3(x) f(x)dx = \frac{1}{2}\int_{-\pi/2}^{\pi/2} \sin^3(x)\cos(x)dx = 0$$
pga av symmetri. Detta följer av att $g(x) = \sin^3(x)\cos(x)$ så att
$$g(-x) = \sin^3(-x)\cos(-x) = (-1)^3\sin^3(x)\cos(x) = -\sin^3(x)\cos(x)$$
och därmed är $g(x)$ en udda funktion.

3. Bella jobbar på posten och sorterar paket. Paketen väger i genomsnitt 640 gram med variansen 40000 gram.

   Antag att Bella tar emot 100 paket. Beräkna approximativt sannolikheten att den totala vikten överstiger 60 kg. (6p)

   Lösning:

   Låt P_k vara vikten av paket nummer k. Vi söker sannolikheten

$$\mathbb{P}\left(\sum_{k=1}^{100} P_k \geq 60000\right).$$

   För att kunna beräkna denna sannolikhet använder vi normalapproximation (dvs centrala gränsvärdessatsen). Denna säger att

$$S_P = \sum_{k=1}^{100} P_k \approx N\left(100\mathbb{E}[P_k], 100\mathrm{Var}(P_k)\right)$$

   där

$$100\mathbb{E}[P_k] = 100 \cdot 640 = 64000$$

   och

$$100\mathrm{Var}(P_k) = 100 \cdot 40000 = 4000000.$$

   Därför blir

$$\frac{S_P - 100\mathbb{E}[P_k]}{\sqrt{100\mathrm{Var}(P_k)}} = \frac{S_P - 64000}{2000} \approx N(0,1)$$

   så att

$$\mathbb{P}\left(\sum_{k=1}^{100} P_k \geq 60000\right) = \mathbb{P}\left(\frac{S_P - 64000}{2000} \geq \frac{60000 - 64000}{2000}\right) \approx \mathbb{P}(Z \geq -2) \approx 0.0228$$

   där vi använde tabell i sista ”likheten”.

4. Låt (X,Y) ha gemensam fördelning med täthetsfunktion

$$\mathbb{P}(X = k, Y = l) = \begin{cases} \frac{\lambda^k}{k!}e^{-\lambda}p & \text{för } k = 0,1,2,\dots \text{ och } l = 1 \\ \frac{\lambda^k}{k!}e^{-\lambda}(1-p) & \text{för } k = 0,1,2,\dots \text{ och } l = 0, \end{cases}$$

   där $\lambda > 0$ och $p \in (0,1)$.

   (a) Beräkna marginalfördelningarna för X respektive Y. Är X, Y oberoende? (5p)

   (b) Låt $Z = XY$. Beräkna täthetsfunktionen för Z. (3p)

   Lösning:

(a) Vi börjar med att observera att X kan anta värdena $0, 1, 2, \ldots$ medans Y kan anta värdena 0 och 1. För att beräkna marginalfördelningen för X observerar vi att

$$\begin{aligned}\mathbb{P}(X=k) &= \mathbb{P}(X=k, Y=0) + \mathbb{P}(X=k, Y=1)\\ &= \frac{\lambda^k}{k!}e^{-\lambda}p + \frac{\lambda^k}{k!}e^{-\lambda}(1-p) = \frac{\lambda^k}{k!}e^{-\lambda} \text{ för } k = 0, 1, \ldots\end{aligned}$$

Vi ser alltså att $X \sim$Poi(λ). Vidare gäller att

$$\begin{aligned}\mathbb{P}(Y=1) &= \sum_{k=0}^{\infty}\mathbb{P}(X=k, Y=1)\\ &= \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}p = e^{-\lambda}p\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}pe^{\lambda} = p\end{aligned}$$

och på samma sätt blir

$$\mathbb{P}(Y=0) = \sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}(1-p) = 1-p$$

så att $Y \sim$Bern(p). Vi ser vidare att

$$\mathbb{P}(X=k)\mathbb{P}(Y=1) = \frac{\lambda^k}{k!}e^{-\lambda}(1-p) = \mathbb{P}(X=k, Y=1)$$

och på samma sätt ser vi att $\mathbb{P}(X=k)\mathbb{P}(Y=1) = \mathbb{P}(X=k, Y=1)$ så att X, Y är oberoende.

(b) Vi ser att $Z = XY$ kan anta värdena $0, 1, 2, \ldots$. Vidare är

$$\begin{aligned}\mathbb{P}(Z=0) &= \mathbb{P}(Y=0) + \mathbb{P}(X=0, Y=1)\\ &= 1-p+p\frac{\lambda^0}{0!}e^{-\lambda} = 1-p+pe^{-\lambda},\end{aligned}$$

och för $k = 1, 2, \ldots$

$$\mathbb{P}(Z=k) = \mathbb{P}(X=k, Y=1) = p\frac{\lambda^k}{k!}e^{-\lambda}.$$

Täthetsfunktionen är alltså

$$\mathbb{P}(Z=k) = \begin{cases} 1-p+pe^{-\lambda} & \text{för } k=0\\ p\frac{\lambda^k}{k!}e^{-\lambda} & \text{för } k=1,2,\ldots\end{cases}$$

5. Definiera begreppet felintensitet och ge en tolkning av detta (2p)

Lösning:

Felintensiteten $\rho_T(t)$ för systemet som beskrivs av slumpvariabeln T definieras som

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

där som vanligt $F_T(t) = \mathbb{P}(T \leq t)$ och $f_T(t) = F_T'(t)$.

Tolkningen av felintensiteten är att den beskriver den momentana feltakten vid tiden t givet att systemet fungerar vid denna tid.

6. Låt X ha täthetsfunktion

$$f_X(x) = (\theta - 1)x^{-\theta} \text{ för } x \geq 1,$$

där parametern $\theta > 1$.

(a) Hitta momentskattaren för θ. (4p)

(b) Om vi har tre mätvärden $x_1 = 12.4, x_2 = 78.1$ och $x_3 = 3.7$, vad blir då värdet på din skattning? (2p)

Lösning:

(a) Enligt momentmetoden skall vi beräkna $\mathbb{E}[X]$. Vi ser att

$$\begin{aligned}\mathbb{E}[X] &= \int_1^\infty (\theta - 1)x^{1-\theta} dx \\ &= (\theta - 1)\left[\frac{x^{2-\theta}}{2-\theta}\right]_1^\infty = \begin{cases} \frac{\theta-1}{\theta-2} & \text{om } \theta > 2 \\ \infty & \text{om } 1 < \theta \leq 2. \end{cases}\end{aligned}$$

Momentmetoden fungerar alltså bara ifall $\theta > 2$. För $\theta > 2$ får vi att vi skall lösa ekvationen

$$\overline{X} = \frac{\hat{\theta} - 1}{\hat{\theta} - 2} \text{ så att } (\hat{\theta} - 2)\overline{X} = \hat{\theta} - 1 \text{ och } \hat{\theta}(\overline{X} - 1) = 2\overline{X} - 1$$

så att vi till slut har att

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

(b) Med dessa tre värden får vi

$$\bar{x} = \frac{x_1 + x_2 + x_3}{3} = \frac{12.4 + 78.1 + 3.7}{3} = 31.4$$

så att

$$\hat{\theta}(x) = \frac{2 \cdot 31.4 - 1}{31.4 - 1} = \frac{61.8}{30.4} \approx 2.033.$$

7. Olssons restaurang har två sorters vitlökssås. En med extra mycket vitlök (EMV) och en med fantastiskt mycket vitlök (FMV). Herr Cato undrar vad halten vitlök är i såserna och om det egentligen är någon större skillnad. Han köper därför 6 burkar av vardera sort (vid olika tillfällen) och lämnade in dessa till ett vitlökslabb som mätte halten vitlök i vardera 12 burkar. Herr Cato fick därefter följande tabell över halten vitlök i burkarna

Halt EMV:	2.9	3.1	3.7	3.6	2.7	3.6
Halt FMV:	3.5	3.6	3.5	3.3	3.5	3.2

Här tolkas siffrorna i raden “Halt EMV” som halten (i %) vitlök i de 6 burkarna med såsen “extra mycket vitlök”, medans siffrorna i raden “Halt FMV” skall tolkas som halten vitlök i de 6 burkarna med såsen “fantastiskt mycket vitlök”.

Herr Cato antar att de uppmätta halterna kommer från normalfördelningar där varianserna är lika stora.

(a) Ange ett hypotestest för att testa huruvida såsen FMV faktiskt innehåller mer vitlök än EMV. Beräkna en förkastningsregion (RR) för ditt test om du har signifikansnivån $\alpha = 0.05$. (5p)

(b) Vad blir resultatet av ditt test med de data som Herr Cato fick? (3p)

(c) Hitta ett (approximativt) p-värde för ditt test. (1p)

Lösning:

(a) Om X betecknar en slumpvariabel som representerar såsen EMV och Y betecknar en slumpvariabel som representerar såsen FMV så antar vi enligt uppgiftsformuleringen att

$$X_k \sim N(\mu_1, \sigma_1^2) \text{ och att } Y_k \sim N(\mu_2, \sigma_2^2)$$

där alla slumpvariablerna är oberoende och där μ_1, μ_2 och $\sigma_1 = \sigma_2 = \sigma^2$ alla är okända. Vi vill nu testa hypoteserna

$$H_0 : \mu_1 = \mu_2 \text{ mot } H_1 : \mu_1 < \mu_2$$

då det i texten står att det är detta som skall testas. Vi använder test-statistikan

$$T(X,Y) = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{6} + \frac{1}{6}\right)}} \overset{H_0}{=} \frac{\overline{X} - \overline{Y}}{\sqrt{s_p^2\left(\frac{1}{6} + \frac{1}{6}\right)}} \sim t(6+6-2) = t(10)$$

där den andra likheten gäller om H_0 är sann, dvs om $\mu_1 = \mu_2$. Här är

$$s_p^2 = \frac{5s_X^2 + 5s_Y^2}{6+6-2} = \frac{s_X^2}{2} + \frac{s_Y^2}{2}.$$

Om H_1 stämmer så bör $T(X,Y) < 0$, och vi letar därför förkastningsregion på formen $(-\infty, c_\alpha]$ där c_α ges av sambandet

$$\alpha = \mathbb{P}(T(X,Y) \in (-\infty, c_\alpha]) = \mathbb{P}(T(X,Y) \leq c_\alpha).$$

Då $\alpha = 0.05$ ger tabell att

$$0.05 \approx \mathbb{P}(T(X,Y) \leq -1.812),$$

dvs $RR = (-\infty, -1.812]$.

(b) Herr Catos data ger att

$$\overline{x} = \frac{2.9 + 3.1 + 3.7 + 3.6 + 2.7 + 3.6}{6} \approx 3.2667$$

och att

$$\overline{y} = \frac{3.5 + 3.6 + 3.5 + 3.3 + 3.5 + 3.2}{6} \approx 3.4333.$$

Vidare blir

$$s_x^2 = \frac{1}{5}\sum_{k=1}^{6}(x_k - \overline{x})^2 \approx 0.1787$$

och

$$s_y^2 = \frac{1}{5}\sum_{k=1}^{6}(y_k - \overline{y})^2 \approx 0.0227$$

så att

$$s_p^2(x,y) = \frac{s_x^2}{2} + \frac{s_y^2}{2} \approx 0.1007.$$

Därmed får vi att

$$T(x,y) = \frac{\overline{x} - \overline{y}}{\sqrt{s_p^2\left(\frac{1}{6} + \frac{1}{6}\right)}} \approx -0.9098.$$

Vi ser att Herr Catos teststatistika inte tillhör RR, och därmed förkastar vi inte H_0.

(c) p-värdet ges av

$$\mathbb{P}(T(X,Y) \leq t(x,y)) \approx \mathbb{P}(T(X,Y) \leq -0.9098) \approx 1 - 0.82 = 0.18.$$

Tabellen visar inte just siffran 0.9098 utan 0.700 och 1.372 värdet gissas av att 0.9098 är någonstans däremellan.

8. En magiker hävdar att han kan skapa ett fält där tiden går långsammare. Dessutom hävdar magikern att han kan kontrollera styrkan på sitt fält. Ett experiment genomförs där sönderfallstider (i sekunder) av ett radioaktivt

material uppmäts för 7 olika fältstyrkor (enhet mystika). Data blev som följer:

Fältstyrka:	1	2	3	4	5	6	7
Sönderfallstid :	3.71	2.80	2.88	4.49	4.41	4.42	3.67

Data kan sammanfattas av att

$$S_{xx} = 28, \quad S_{yy} \approx 3.097 \text{ och } S_{xy} \approx 4.65$$

(a) Ansätt en linjär regressionsmodell och skatta värdet på de relevanta parametrarna. (1p)

(b) Hitta ett 95% konfidensintervall för lutningen av din linjära modell. (4p)

(c) Bestäm förklaringsgraden för modellen och diskutera vad du kan dra för slutsats om magikerns krafter. (2p)

Lösning:

(a) Vi ansätter $y = \beta_0 + \beta_1 x$ och har att

$$\hat{\beta}_1 = \frac{Sxy}{Sxx} \approx 0.1661 \text{ och } \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \approx 3.1043$$

(b) Enligt föreläsningarna så är

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right),$$

så att

$$R = \frac{\hat{\beta}_1 - \beta_1}{S_r/\sqrt{S_{xx}}} \sim t(n-2)$$

där

$$S_r^2 = \frac{1}{n-2}\sum_{k=1}^{n}(y_k - \hat{y}_k)^2 \approx 0.465$$

Vi får därför att

$$\begin{aligned}
0.95 &= \mathbb{P}(-t_{0.025}(5) \leq R \leq t_{0.025}(5)) \\
&= \mathbb{P}\left(-t_{0.025}(5) \leq \frac{\hat{\beta}_1 - \beta_1}{S_r/\sqrt{S_{xx}}} \sim t(n-2) \leq t_{0.025}(5)\right) \\
&= \mathbb{P}\left(\hat{\beta}_1 - t_{0.025}(5)\frac{S_r}{\sqrt{S_{xx}}} \leq \beta_1 \leq \hat{\beta}_1 + t_{0.025}(5)\frac{S_r}{\sqrt{S_{xx}}}\right).
\end{aligned}$$

Tabell ger att $t_{0.025}(5) \approx 2.571$ så ett 95% numeriskt konfidensintervall för β_1 blir

$$\begin{aligned}
I_{\beta_1} &= \hat{\beta}_1 \pm t_{0.025}(5)\frac{S_r}{\sqrt{S_{xx}}} \\
&\approx 0.1661 \pm 2.571\frac{0.465}{\sqrt{28}} = [-0.0599, 0.3920]
\end{aligned}$$

(c) Förklaringsgraden ges av

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} \approx 0.2493.$$

Detta är mycket lågt och gör att vi bör bortse från hela magiker-grejen. Dessutom står hans påståenden om magiska tidsfält på en ganska svajig teoretisk grund. Om inte detta är nog ser vi också att konfidensintervallet innehåller nollan.