Tentamentsskrivning i **Matematisk Statistik MVE091**

Tid: 5 januari, 2023 em
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. En politisk kommitté skall bildas av en grupp politiker. Gruppen består av 3 personer från partiet A, 4 personer från partiet B och 6 personer från partiet C.

   (a) Antag att kommittéen består av sex politiker och att alla partier måste representeras av två individer. På hur många sätt kan man bilda en sådan kommitté? (3p)

   (b) Antag att kommittéen består av sju politiker och att alla partier måste representeras av minst två individer. På hur många sätt kan man bilda en sådan kommitté? (3p)

   Lösning

   (a) Antal sätt att välja ut två politiker att representera parti A är

   $$\binom{3}{2}$$

   och på liknande sätt för B,C har vi

   $$\binom{4}{2} \text{ respektive } \binom{6}{2}.$$

   Det sammanlagda antalet sätt blir därför

   $$\binom{3}{2}\binom{4}{2}\binom{6}{2} = 3 \cdot 6 \cdot 15 = 270.$$

   (b) Ett sätt att räkna ut detta är att summera över de olika fallen när vi har tre representanter från parti A,B respektive C. Antalet kommittéer med tre represententer från A blir då

   $$\binom{3}{3}\binom{4}{2}\binom{6}{2} = 1 \cdot 6 \cdot 15 = 90,$$

och för B

$$\binom{3}{2}\binom{4}{3}\binom{6}{2} = 3 \cdot 4 \cdot 15 = 180,$$

respektive C

$$\binom{3}{2}\binom{4}{2}\binom{6}{3} = 3 \cdot 6 \cdot 20 = 360.$$

Sammanlagt har vi då

$$90 + 180 + 360 = 630$$

möjliga kommittéer.

2. Låt X ha täthetsfunktion

$$f(x) = \frac{1}{60}x^3 \text{ för } 2 \leq x \leq 4.$$

(a) Beräkna $\mathbb{E}[\sqrt{X}]$. (2p)

(b) Låt $Y = \sqrt{X}$. Beräkna täthetsfunktionen för Y. (3p)

(c) Används ditt svar i (b) för att beräkna $\mathbb{E}[Y]$. (2p)

Lösning:

(a) Vi har att

$$\mathbb{E}[\sqrt{X}] = \int_{-\infty}^{\infty} \sqrt{x} f(x) dx = \int_2^4 \sqrt{x}\frac{x^3}{60} dx$$
$$= \frac{1}{60}\int_2^4 x^{7/2} dx = \frac{1}{60}\left[\frac{2x^{9/2}}{9}\right]_2^4 = \frac{4^{9/2} - 2^{9/2}}{270}$$

(≈ 1.8125).

(b) Vi går via fördelningsfunktionen men noterar först att $Y = \sqrt{X} \in [\sqrt{2}, 4]$. Vi ser därför att för $y \in [\sqrt{2}, 4]$ så har vi att

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\sqrt{X} \leq y) = \mathbb{P}(X \leq y^2)$$
$$= \int_2^{y^2} \frac{x^3}{60} dx = \left[\frac{x^4}{240}\right]_2^{y^2} = \frac{y^8 - 2^4}{240}.$$

Detta innebär i sin tur att

$$f_Y(y) = F_Y'(y) = \frac{8y^7}{240} = \frac{y^7}{30} \text{ för } \sqrt{2} \leq y \leq 2.$$

(c) Vi har nu att

$$\mathbb{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y) dy = \int_{\sqrt{2}}^{2} y\frac{y^7}{30} dy$$
$$= \int_{\sqrt{2}}^{2} \frac{y^8}{30} dy = \left[\frac{y^9}{270}\right]_{\sqrt{2}}^{2} = \frac{2^9 - 2^{9/2}}{270},$$

vilket är samma svar som i (a).

3. Mr Bakos lider av ett märkligt tvångsbeteende. Varje morgon, året om, går Mr Bakos ut med sitt hagelgevär och skjuter lerduvor. Han skjuter tills han fått en träff och sedan går han in och dricker kaffe. Antag att varje skott träffar oberoende av varandra och med sannolikhet 2/3.

   (a) Vad är sannolikheten att Mr Bakos under en given morgon inte behöver skjuta mer än högst 2 skott? (2p)

   (b) Vad är sannolikheten att Mr Bakos under tre på varandra följande mornar inte behöver skjuta mer än högst 2 skott? (2p)

   (c) Beräkna sannolikheten att under minst 320 av årets 365 mornar så behöver Mr Bakos inte skjuta mer än högst 2 skott. (3p)

Lösning:

(a) Låt X =antalet skott Mr Bakos behöver till och med första träffen. Då är X ~Geom(p) där $p = 2/3$ enligt uppgift. Den sökta sannolikheten blir därför

$$\mathbb{P}(X \leq 2) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = \frac{2}{3} + \frac{1}{3} \cdot \frac{2}{3} = \frac{8}{9}.$$

(b) Låt

$$Y_k = \begin{cases} 1 & \text{om Mr Bakos inte behöver fler än två skott dag } k \\ 0 & \text{annars,} \end{cases}$$

där $k = 1, 2, 3$. Vi söker då

$$\mathbb{P}(Y_1 = Y_2 = Y_3 = 1) = \mathbb{P}(Y_1 = 1)^3 = \left(\frac{8}{9}\right)^3,$$

där vi använde att alla träffar är oberoende av varandra.

(c) Låt Y_k vara som i (b) fast för $k = 1, 2, \ldots, 365$. Låt sedan $Y = \sum_{k=1}^{365} Y_k$ och observera att vi söker sannolikheten $\mathbb{P}\left(Y \geq 320\right)$. Vi har att Y ~Bin$(365, 8/9)$ men för att kunna räkna måste vi använda normalapproximation. Vi har då att

$$Y \approx N(\mathbb{E}[Y], \mathrm{Var}(Y)) = N\left(365 \cdot \frac{8}{9}, 365 \cdot \frac{8}{9} \cdot \frac{1}{9}\right) \approx N(324.4, 36.0).$$

Därmed blir

$$\begin{aligned} \mathbb{P}\left(Y \geq 300\right) &= \mathbb{P}\left(\frac{Y - 324.4}{\sqrt{36}} \geq \frac{320 - 324.4}{\sqrt{36}}\right) \\ &\approx \mathbb{P}(Z \geq -0.73) = \mathbb{P}(Z \leq 0.73) \approx 0.7673, \end{aligned}$$

där vi använde symmetri och tabell i de två sista stegen.

4. Låt (X, Y) ha gemensam täthetsfunktion

$$f_{XY}(x,y) = cx^2y \text{ där } 0 \leq x \leq 1 \text{ och } 0 \leq y \leq 1$$

   (a) Bestäm konstanten c så att $f_{XY}(x,y)$ verkligen är en täthetsfunktion. (2p)

   (b) Bestäm marginaltäthetsfunktionen för X. (2p)

   (c) Beräkna sannolikheten $\mathbb{P}(X < Y)$. (2p)

   Lösning:

   (a) Vi använder att $1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy$ och ser att

$$\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)dxdy \\
&= \int_0^1\int_0^1 cx^2ydxdy = c\int_0^1 y\left[\frac{x^3}{3}\right]_0^1 dy = \frac{c}{3}\left[\frac{y^2}{2}\right]_0^1 = \frac{c}{6}
\end{aligned}$$

   så att $c = 6$.

   (b) Vi har att

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy = \int_0^1 cx^2ydy = cx^2\left[\frac{y^2}{2}\right]_0^1 = 3x^2$$

   för $0 \leq x \leq 1$.

   (c) Vi har att

$$\begin{aligned}
\mathbb{P}(X < Y) &= \int\int_{x<y} f_{XY}(x,y)dxdy \\
&= \int_0^1\int_0^y 6x^2ydxdy = 6\int_0^1 y\left[\frac{x^3}{3}\right]_0^y dy = 2\int_0^1 y^4dy = \frac{2}{5}.
\end{aligned}$$

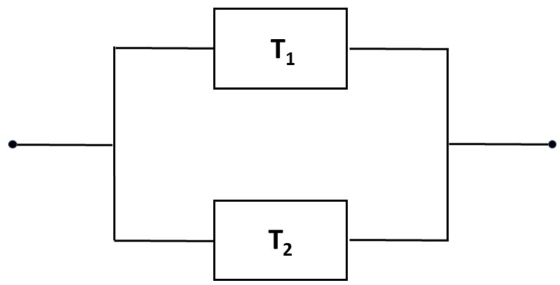
5. Betrakta ett system som i Figur 1 där systemet fungerar om minst en av komponenterna fungerar, och antag att de två komponenterna fungerar oberoende av varandra. Antag också att tiden det tar tills komponenterna slutar fungera är oberoende och exponentialfördelade med parameter β_1 respektive β_2.

   (a) Beräkna överlevnadsfunktionen (reliability function) för hela systemet. (2p)

   (b) Beräkna felintensiteten (hazard rate) för hela systemet. (2p)

   (c) Antag att $\beta_1 = \beta_2 = 1$, vad är felintensiteten då $t = 0$ och när $t \to \infty$? (2p)

   Lösning:



Figur 1: Parallellkopplat system.

(a) Om vi låter $T = \max(T_1, T_2)$ och R_T beteckna överlevnadsfunktionen för hela systemet så gäller att

$$\begin{aligned} R_T(t) &= \mathbb{P}(\max(T_1, T_2) \geq t) \\ &= 1 - \mathbb{P}(\max(T_1, T_2) \leq t) = 1 - \mathbb{P}(T_1 \leq t)\mathbb{P}(T_2 \leq t) \\ &= 1 - \int_0^t \frac{1}{\beta_1} e^{-s/\beta_1} ds \int_0^t \frac{1}{\beta_2} e^{-s/\beta_2} ds \\ &= 1 - \left(1 - e^{-t/\beta_1}\right)\left(1 - e^{-t/\beta_2}\right) \\ &= e^{-t/\beta_1} + e^{-t/\beta_2} - e^{-t(1/\beta_1 + 1/\beta_2)}. \end{aligned}$$

(b) Det gäller att felintensiteten $\rho(t)$ ges av

$$\rho(t) = \frac{f_T(t)}{R_T(t)} = \frac{F_T'(t)}{R_T(t)} = \frac{\frac{d}{dt}(1 - R_T(t))}{R_T(t)}$$

och vidare är

$$\begin{aligned} \frac{d}{dt}(1 - R_T(t)) &= \frac{d}{dt}\left(1 + e^{-t(1/\beta_1 + 1/\beta_2)} - e^{-t/\beta_1} - e^{-t/\beta_2}\right) \\ &= -\left(\frac{1}{\beta_1} + \frac{1}{\beta_2}\right) e^{-t(1/\beta_1 + 1/\beta_2)} + \frac{1}{\beta_1} e^{-t/\beta_1} + \frac{1}{\beta_2} e^{-t/\beta_2} \end{aligned}$$

så att

$$\rho(t) = \frac{-\left(\frac{1}{\beta_1} + \frac{1}{\beta_2}\right) e^{-t(1/\beta_1 + 1/\beta_2)} + \frac{1}{\beta_1} e^{-t/\beta_1} + \frac{1}{\beta_2} e^{-t/\beta_2}}{e^{-t/\beta_1} + e^{-t/\beta_2} - e^{-t(1/\beta_1 + 1/\beta_2)}}$$

(c) Om $\beta_1 = \beta_2 = 1$ så blir

$$\rho(t) = \frac{-2e^{-2t} + 2e^{-t}}{2e^{-t} - e^{-2t}} = \frac{2e^{-2t} - 2e^{-t}}{e^{-2t} - 2e^{-t}} = \frac{2e^{-t} - 2}{e^{-t} - 2} = \frac{2 - 2e^{-t}}{2 - e^{-t}}.$$

Vi ser att $\rho(0) = \frac{2-2}{2-1} = 0$ och att $\lim_{t\to\infty} \rho(t) = \frac{2-0}{2-0} = 1$.

6. Lös följande två uppgifter.

   (a) Låt X vara likformigt fördelat på intervallet $[0, N]$ (dvs $X \sim \mathrm{U}([0, N])$). Bestäm momentskattaren för N. (3p)

   (b) Låt Y vara likformigt fördelat på intervallet $[-N, N]$ (dvs $Y \sim \mathrm{U}([-N, N])$). Bestäm momentskattaren för N. (3p)

   Lösning:

   (a) Enligt momentmetoden skall vi först beräkna $\mathbb{E}[X]$. Vi har att $\mathbb{E}[X] = \frac{N}{2}$ så att enligt momentmetoden skall vi lösa ut $\hat{N}$ ur sambandet $\overline{X} = \frac{\hat{N}}{2}$ och vi får då att vår momentskattare blir

   $$\hat{N} = 2\overline{X}.$$

   (b) Om vi här beräknar $\mathbb{E}[X]$ så ser vi att $\mathbb{E}[X] = 0$. Momentmetoden föreskriver därför att vi istället beräknar $\mathbb{E}[X^2]$. Vi ser att

   $$\mathbb{E}[X^2] = \int_{-N}^{N} x^2 \frac{1}{2N} dx = \frac{1}{2N}\left[\frac{x^3}{3}\right]_{-N}^{N} = \frac{N^3 - (-N)^3}{6N} = \frac{N^2}{3}.$$

   Momentmetoden säger då att vi skall ta $\overline{X^2} = \frac{\hat{N}^2}{3}$ så att

   $$\hat{N} = \sqrt{3\overline{X^2}}$$

   där $\overline{X^2} = \frac{1}{n}\sum_{k=1}^{n} X_k^2$.

7. Företaget Balk'R'Us tillverkar balkar som skall användas i större konstruktioner. Dessa balkar måste hålla för stora påfrestningar och forskningsavdelningen har nyligen hittat en ny legering som de tror kan vara bättre än den gamla. Det testar därför många balkar av vardera typ (existerande legering och ny legering) som de utsätter för extrema belastningar (17 N) och ser hur många av balkarna som brister. De fick följande data:

	Existerande legering	Ny legering
Antal testade balkar	50	40
Antal balkar med brott	13	7

   (a) Punktskatta proportionerna av vardera balktyp som håller för en belastning av 17 N. (1p)

   (b) Skapa ett 99% K.I. (konfidensintervall) för skillnaden mellan de två proportionerna. (4p)

   Lösning:

(a) Vi använder så klart

$$\hat{p}_1 = \frac{37}{50} \approx 0.74 \text{ och } \hat{p}_2 = \frac{33}{40} \approx 0.825.$$

Om man råkat räkna på de som går sönder får man istället:

$$\hat{p}_1 = \frac{13}{50} \approx 0.26 \text{ och } \hat{p}_2 = \frac{7}{40} \approx 0.175.$$

(b) Här använder vi referensvariabeln

$$R = \frac{\hat{p}_2 - \hat{p}_1 - (p_2 - p_1)}{S_{\hat{p}}} \approx N(0,1)$$

där

$$S_{\hat{p}}^2 = \frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2} \text{ och } n_1 = 50,\ n_2 = 40.$$

Vi har vidare att

$$\begin{aligned} 0.99 &\approx \mathbb{P}(-z_{0.005} \leq R \leq z_{0.005}) \\ &\approx \mathbb{P}\left(-2.575 \leq \frac{\hat{p}_2 - \hat{p}_1 - (p_2 - p_1)}{S_{\hat{p}}} \leq 2.575\right) \\ &= \mathbb{P}\left(\hat{p}_2 - \hat{p}_1 - 2.575 \cdot S_{\hat{p}} \leq p_2 - p_1 \leq \hat{p}_2 - \hat{p}_1 + 2.575 \cdot S_{\hat{p}}\right), \end{aligned}$$

så att ett 99% K.I. för $p_2 - p_1$ blir

$$I_{p_2 - p_1} = \hat{p}_2 - \hat{p}_1 \pm 1.96 \cdot S_{\hat{p}} \approx [-0.137, 0.307]$$

där vi använt att $S_{\hat{p}}^2 \approx 0.0075$.

8. Den ryska armén ser ett behov av att kunna konservera mat under en längre period. Speciellt vill man konservera korv i konservburkar. I ett experiment undersöktes därför hur länge konserverade korvar hölls färska som funktion av mängden cyanid som blandades i konserverna. Man fick följande tabell:

Mängd cyanid (ppm):	0	1	2	4	6	10	14	20
Överlevnadstid (dagar):	488	583	514	536	595	829	909	1088

Med överlevnadstid menas här tiden som konserven nasågs vara färsk. Data sammanfattas med att

$$S_{xx} \approx 347 \quad S_{yy} \approx 341600 \quad S_{xy} \approx 10614.$$

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

(b) Genomför ett hypotestest på signifikansnivån $\alpha = 0.01$ hurvida mängden cyanid ökar konservernas hållbarhet. Gör detta genom att beräkna p-värdet (så gott det går) för testet. (4p)

(c) Beräkna alla residualer och plotta dessa. Beräkna också förklaringsgraden. Verkar deras modell rimlig? (2p)

Lösning:

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

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} \approx 30.6 \text{ och } \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x} \approx 474.8.$$

(b) Vi skall alltså testa

$$\begin{aligned} H_0 &: \quad \beta_1 = 0 \\ H_1 &: \quad \beta_1 > 0 \end{aligned}$$

på $\alpha = 0.01$. Vi vet att $\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right)$ så att under H_0 (dvs om $\beta_1 = 0$) är

$$\hat{\beta}_1 \sim N\left(0, \frac{\sigma^2}{S_{xx}}\right).$$

Vi väljer därför test-statistikan

$$T(E) = \frac{\hat{\beta}_1(E)}{S_r(E)/\sqrt{S_{xx}}} \sim t(n-2) = t(6),$$

där $t(6)$ är noll-fördelningen och

$$S_r^2 = \frac{1}{n-2} \sum_{k=1}^{n} (y_k - \hat{\beta}_0 - \hat{\beta}_1 x_k)^2.$$

Om H_1 stämmer så bör $T(E)$ tendera att bli stor och positiv. P-värdet ges därför av

$$p\text{-värde} = \mathbb{P}(T(E) \geq T(\epsilon))$$

och här blir

$$T(\epsilon) = \frac{\hat{\beta}_1(\epsilon)}{S_r(\epsilon)/\sqrt{S_{xx}}} \approx \frac{30.6}{\sqrt{2801}/\sqrt{347}} \approx 10.77$$

där vi använde att data gav att $S_r^2(\epsilon) \approx 2801$. Tabell ger att

$$\begin{aligned} p\text{-värde} &= \mathbb{P}(T(E) \geq T(\epsilon)) \\ &\approx \mathbb{P}(T(E) \geq 10.77) \leq \mathbb{P}(T(E) \geq 5.959) = 0.0005. \end{aligned}$$

Då p-värdet är mindre än α förkastar vi H_0 på signifikansnivån 0.01.

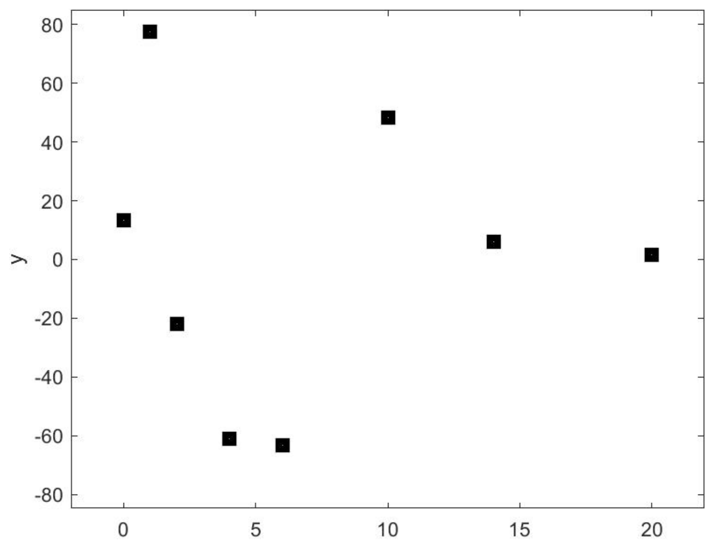
(c) Förklaringsgraden R^2 ges av

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} \approx 0.9504.$$

Residualerna ges av

Mängd cyanid (ppm):	0	1	2	4	6	10	14	20
Överlevnadstid (dagar):	488	583	514	536	595	829	909	1088
Residualer:	13.2	77.6	-22	-61.2	-63.3	48.3	6	1.4

Den höga förklaringsgraden tillsammans med plotten av residualerna gör att vi drar slutsatsen att deras modell är rimlig.



Figur 2: Residualerna.