Matematiska Vetenskaper VT 2023
Chalmers Tekniska Högskola Examinator: Tony Johansson

MVE302/MVE395/TMA321
Exempeltenta

Betygsgränser

Kurs	Frågor	3	4	5	Total
MVE302	Åtta	16	24	32	40
TMA321	Sex	12	18	24	30
MVE395	Sex	12	18	24	30

Hjälpmedel. Valfri miniräknare; Mathematics Handbook (Beta); 2 blad (totalt 4 sidor) handskrivna anteckningar (utskrivna anteckningar från skrivplatta godkänns); tabeller finns bifogade till tesen.

1. (5p) Paret (X, Y) av slumpvariabler har täthetsfunktion

$$f(x, y) = c(2x + y - xy) \quad \text{för } 0 < x < 1,\ 0 < y < 1.$$

- **(a)** Bestäm konstanten c.
- **(b)** Bestäm f_X och f_Y.
- **(c)** Bestäm $\mathbb{E}[X]$ och $\mathbb{E}[Y]$.
- **(d)** Bestäm $\mathrm{Cov}(X, Y)$.
- **(e)** Är X och Y oberoende? Motivera ditt svar.

2. (5p) I en påse finns fyra tärningar, med 6, 10, 12 samt 20 sidor. Varje tärning är rättvis (med samma sannolikhet för varje sida), och sidorna på den k–sidiga tärningen är numrerade från 1 till k, för alla $k \in \{6, 10, 12, 20\}$.

En tärning dras på måfå från påsen, kastas, och numret som syns på tärningen är 7. Givet denna information, vad är sannolikheten att det var den 12–sidiga tärningen som drogs?

3. (5p) En fördelning med parameter $\theta \in (0, \infty)$ beskrivs av täthetsfunktionen

$$f_\theta(x) = \theta x^{-(\theta+1)}, x > 1.$$

Hitta ML-skattningen $\hat{\theta}$ av θ för ett stickprovsutfall $x_1, \ldots, x_n$.

4. (4p) För en populationsvariabel $X \sim N(\mu, \sigma^2)$ med okänt σ^2 dras följande stickprov:

$$12.2,\ 13.4,\ 15.1,\ 15.3,\ 17.0.$$

Ange ett 95% konfidensintervall för μ.

5. (7p) Betrakta följande data.

x	0	1	2	3	4
y	5.0	2.4	2.5	2.1	0.0

(a) (3p) Hitta en linjär regressionsmodell $y = a + bx$ för datan.

(b) (3p) Ange ett symmetriskt 90% konfidensintervall för b.

(c) (1p) Baserat på **(b)**, vad kan vi säga om hypotesen $H_0 : b = 0$ kontra $H_A : b \neq 0$ på signifikansnivå 0.1?

6. (4p) Låt $Y_1, Y_2, \ldots$ vara stokastiska variabler med $Y_k \sim \text{Bin}(1, 1/k)$ för alla k. Definiera

$$X_n = \sum_{k=1}^{n} \frac{Y_k}{k}.$$

Visa att om a_n är en sekvens med $a_n \to +\infty$, så är

$$\lim_{n\to\infty} \mathbb{P}(X_n > a_n) = 0.$$

Tips. Serien $\sum_{k=1}^{\infty} k^{-2}$ är konvergent.

7. (MVE302, 5p) För datamängden i Uppgift 4, ange övre 95% konfidensintervall för σ^2. Alltså, hitta ett tal s_c sådant att

$$\sigma^2 \leq s_c \qquad (95\%)$$

8. (MVE302, 5p) Generalisering av CGS. Låt $X_1, X_2, \ldots,$ vara oberoende slumpvariabler, inte nödvändigtvis med samma fördelning, med $\mathbb{E}[X_k] = 0$ för alla k, och sådana att

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \text{Var}(X_k)$$

existerar med $\sigma^2 > 0$. Antag också att den momentgenererande funktionen $M_k(t) = \mathbb{E}\left[e^{tX_k}\right]$ existerar för alla k och t.

Låt

$$Y_n = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^{n} X_k$$

och visa att den momentgenererande funktionen för Y_n konvergerar till den för $N(0,1)$;

$$\lim_{n\to\infty} M_{Y_n}(t) = e^{t^2/2} \quad \text{för alla } t \in \mathbb{R}.$$

Tips. För $t \approx 0$ är

$$1 + t = e^{t + O(t^2)}.$$

Tabeller

Om det värde du söker saknas i en tabell, använd det närmast tillgängliga.

Fördelningsfunktionen $\Phi(z) = F_Z(z)$ för $Z \sim N(0, 1)$

z	0	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.5	3
$\Phi(z)$	0.5	0.6	0.69	0.77	0.84	0.89	0.93	0.96	0.977	0.994	0.999

Kvantilfunktionen $F^{-1}_{t_{\mathrm{df}}}(q)$ för t-fördelningen

	$q = 0.9$	$q = 0.95$	$q = 0.975$	$q = 0.99$
df = 1	3.08	6.31	12.71	31.82
df = 2	1.89	2.92	4.30	6.96
df = 3	1.64	2.35	3.18	4.54
df = 4	1.53	2.13	2.78	3.75
df = 5	1.48	2.02	2.57	3.36
df = 7	1.41	1.89	2.36	3.00
df = 10	1.37	1.81	2.23	2.76
df = 15	1.34	1.75	2.13	2.60
df = 20	1.33	1.72	2.09	2.53
df = 30	1.31	1.70	2.04	2.46
df = 50	1.30	1.68	2.01	2.40
df = 100	1.29	1.66	1.98	2.36
df = ∞	1.28	1.64	1.96	2.33

Kvantilfunktionen $F^{-1}_{\chi^2_{\mathrm{df}}}(q)$ för χ^2-fördelningen

	$q = 0.025$	$q = 0.05$	$q = 0.1$	$q = 0.9$	$q = 0.95$	$q = 0.975$
df = 1	0.001	0.004	.02	2.71	3.84	5.02
df = 2	0.05	0.10	0.21	4.61	5.99	7.38
df = 3	0.22	0.35	0.58	6.25	7.81	9.35
df = 4	0.48	0.71	1.06	7.78	9.49	11.1
df = 5	0.83	1.15	1.61	9.24	11.1	12.8
df = 7	1.69	2.17	2.83	12.02	14.1	16.0
df = 10	3.25	3.94	4.87	16.0	18.3	20.5
df = 15	6.26	7.26	8.55	22.3	25.0	27.5
df = 20	9.59	10.9	12.4	28.4	31.4	34.2

Lösningar

1. **(a)** Konstanten c bestäms av att funktionens integral ska vara 1:

$$\begin{aligned}
1 = \iint f(x,y)dxdy &= c\int_0^1 \left(\int_0^1 2x + y - xydx\right) dy \\
&= c\int_0^1 1 + y - \frac{y}{2}dy \\
&= c\int_0^1 1 + \frac{y}{2}dy \\
&= c\left(1 + \frac{1}{4}\right) = \frac{5}{4}c.
\end{aligned}$$

Alltså är $c = 4/5$.

(b) Marginalfördelningarna fås genom:

$$\begin{aligned}
f_X(x) = \int_0^1 f(x,y)dy &= \frac{4}{5}\int_0^1 2x + y - xydy \\
&= \frac{4}{5}\left(2x + \frac{1}{2} - \frac{1}{2}x\right) \\
&= \frac{2}{5} + \frac{6}{5}x.
\end{aligned}$$

Vi har faktiskt redan hittat

$$f_Y(y) = c\left(1 + \frac{y}{2}\right) = \frac{4}{5} + \frac{2}{5}y$$

i del **(a)**.

Tips. Dubbelkolla att $\int_0^1 f_X(x)dx = 1$ och $\int_0^1 f_Y(y)dy = 1$.

(c) Vi har

$$\begin{aligned}
\mathbb{E}[X] &= \int_0^1 x f_X(x)dx \\
&= \int_0^1 \frac{2}{5}x + \frac{6}{5}x^2 dx \\
&= \frac{1}{5} + \frac{2}{5} = \frac{3}{5},
\end{aligned}$$

och

$$\begin{aligned}
\mathbb{E}[Y] &= \int_0^1 y f_Y(y)dy \\
&= \int_0^1 \frac{4}{5}y + \frac{2}{5}y^2 dy \\
&= \frac{2}{5} + \frac{2}{15} = \frac{8}{15}.
\end{aligned}$$

(d) För att beräkna kovariansen $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ behöver vi

$$\begin{aligned}\mathbb{E}[XY] = \iint xyf(x,y)dxdy &= \int_0^1 y\left(\int_0^1 x\frac{4}{5}(2x+y-xy)dx\right)dy \\ &= \frac{4}{5}\int_0^1 y\left(\frac{2}{3}+\frac{y}{2}-\frac{y}{3}\right)dy \\ &= \frac{4}{5}\int_0^1 \frac{2}{3}y+\frac{1}{6}y^2dy \\ &= \frac{4}{5}\left(\frac{1}{3}+\frac{1}{18}\right) = \frac{14}{45}.\end{aligned}$$

Då är alltså

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\ &= \frac{14}{45} - \frac{3}{5}\times\frac{8}{15} \\ &= \frac{14}{45} - \frac{24}{75} \\ &= \frac{14\cdot 5 - 3\cdot 24}{225} \\ &= -\frac{2}{225}.\end{aligned}$$

(e) Eftersom $\mathrm{Cov}(X,Y) \neq 0$ kan vi direkt utesluta oberoende.

2. Låt X vara antalet sidor på tärningen som dras, så

$$\mathbb{P}(X = x) = \frac{1}{4}, \quad x = 6, 10, 12, 20.$$

Låt Y vara värden som visas på tärningen som kastas. Vi inser att

$$\mathbb{P}(Y = y \mid X = x) = \frac{1}{x}, \quad y = 1, 2, \ldots, x.$$

Med hjälp av Bayes formel kan vi skriva den önskade betingade sannolikheten som

$$\begin{aligned}\mathbb{P}(X = 12 \mid Y = 7) &= \frac{\mathbb{P}(X = 12 \text{ och } Y = 7)}{\mathbb{P}(Y = 7)} \\ &= \frac{\mathbb{P}(Y = 7 \mid X = 12)\mathbb{P}(X = 12)}{\mathbb{P}(Y = 7)}.\end{aligned}$$

I täljaren använder vi totala sannolikhetslagen:

$$\begin{aligned}\mathbb{P}(Y=7) &= \mathbb{P}(Y=7 \mid X=6)\mathbb{P}(X=6)\\ &\quad + \mathbb{P}(Y=7 \mid X=10)\mathbb{P}(X=10)\\ &\quad + \mathbb{P}(Y=7 \mid X=12)\mathbb{P}(X=12)\\ &\quad + \mathbb{P}(Y=7 \mid X=20)\mathbb{P}(X=20)\\ &= \frac{1}{4}\left(0 + \frac{1}{10} + \frac{1}{12} + \frac{1}{20}\right)\\ &= \frac{14}{4\cdot 60}.\end{aligned}$$

Då får vi alltså

$$\begin{aligned}\mathbb{P}(X=12 \mid Y=7) &= \frac{\mathbb{P}(Y=7 \mid X=12)\mathbb{P}(X=12)}{\frac{14}{4\cdot 60}}\\ &= \frac{\frac{1}{12}\cdot\frac{1}{4}}{\frac{14}{4\cdot 60}} = \frac{5}{14}.\end{aligned}$$

3. Loglikelihoodfunktionen ges av

$$\begin{aligned}\ell(\theta) &= \sum_{k=1}^{n} \ln(\theta x_k^{-(\theta+1)})\\ &= n\ln\theta - (\theta+1)\sum_{k=1}^{n}\ln(x_k)\\ &= n\ln\theta - (\theta+1)\ln\left(\prod_{k=1}^{n} x_k\right).\end{aligned}$$

Låt $\Pi = \prod_{k=1}^{n} x_k$. För att maximera $\ell(\theta)$ hittas derivatan:

$$\ell'(\theta) = \frac{n}{\theta} - \ln\Pi.$$

Denna är noll då $\theta = n/\ln\Pi$. Vi har också $\ell''(\theta) = -n/\theta^2$, så ℓ är en konkav funktion och därmed är derivatans nollpunkt ett globalt maximum. Så ML-skattningen är

$$\hat{\theta} = \frac{n}{\ln\Pi} = \frac{n}{\sum_{k=1}^{n}\ln x_k}.$$

4. Kalla datan x_1, x_2, x_3, x_4, x_5 (med $n = 5$). Vi beräknar

$$\begin{aligned}\Sigma_1 &= \sum_{k=1}^{5} x_k = 73,\\ \Sigma_2 &= 1079.5.\end{aligned}$$

Detta ger oss

$$\overline{x} = \frac{1}{5}\Sigma_1 = 14.6,$$
$$s^2 = \frac{1}{4}\left(\Sigma_2 - \frac{1}{5}\Sigma_1^2\right) = 3.425.$$

Eftersom σ^2 är okänt kommer vårt konfidensintervall vara baserat på att

$$\frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}.$$

Vi tar fram 0.975-kvantilen för fördelningen t_4 (från tabell eller miniräknare):

$$F_{t_4}^{-1}(0.975) = 2.78.$$

Då får vi ett 95% konfidensintervall:

$$\mu \in 14.6 \pm 2.78\frac{\sqrt{3.425}}{\sqrt{5}} \quad (95\%),$$

eller

$$12.3 \leq \mu \leq 16.9 \quad (95\%).$$

Kommentar. Uppgiften ber inte explicit om ett symmetriskt intervall, så det går bra att till exempel svara med

$$\mu \leq 16.36 \quad (95\%),$$

där

$$16.36 = 14.6 + F_{t_4}^{-1}(0.95)\sqrt{\frac{3.425}{5}}.$$

5. (a) Jag brukar börja med att beräkna de "rena" summorna

$$\Sigma_x = \sum_{k=1}^{5} x_k = 10,$$
$$\Sigma_y = \sum_{k=1}^{5} y_k = 12,$$
$$\Sigma_{xx} = \sum_{k=1}^{5} x_k^2 = 30,$$
$$\Sigma_{yy} = \sum_{k=1}^{5} y_k^2 = 41.42,$$
$$\Sigma_{xy} = \sum_{k=1}^{5} x_k y_k = 13.7.$$

Vi kan nu beräkna de "centrerade" summorna:

$$\begin{aligned}
S_{xx} &= \Sigma_{xx} - \frac{1}{n}\Sigma_x^2 = 30 - \frac{1}{5}\cdot 10^2 = 10, \\
S_{yy} &= \Sigma_{yy} - \frac{1}{n}\Sigma_y^2 = 41.42 - \frac{12^2}{5} = 12.62, \\
S_{xy} &= \Sigma_{xy} - \frac{1}{n}\Sigma_x\Sigma_y = 13.7 - \frac{10\cdot 12}{5} = -10.3.
\end{aligned}$$

Vi får då följande skattningar av parametrarna a, b;

$$\begin{aligned}
\hat{b} &= \frac{S_{xy}}{S_{xx}} = \frac{-10.3}{10} = -1.03, \\
\hat{a} &= \overline{y} - \hat{b}\overline{x} = \frac{\Sigma_y}{5} + 1.03\frac{\Sigma_x}{5} = 4.46.
\end{aligned}$$

Vår linjära modell är

$$y = 4.46 - 1.03x.$$

(b) Eftersom σ^2 är okänd bygger konfidensintervallet på att

$$\frac{\hat{b} - b}{s/\sqrt{S_{xx}}} \sim t_{n-2},$$

där s fås genom följande skattning av σ^2:

$$\begin{aligned}
s^2 &= \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) \\
&= \frac{1}{3}\left(12.62 - \frac{(-10.3)^2}{10}\right) \\
&= 0.67.
\end{aligned}$$

Ett 90% konfidensintervall motsvaras av $\alpha = 0.1$, och vi får

$$b \in \hat{b} \pm F_{t_3}^{-1}(0.95)\sqrt{\frac{0.67}{10}} \quad (90\%).$$

Vi har $F_{t_3}^{-1}(0.95) = 2.35$, så

$$b \in -1.03 \pm 2.35 \cdot \sqrt{0.067} \quad (90\%),$$

eller

$$-1.63 \leq b \leq -0.42 \quad (90\%).$$

(c) Efterom konfidensintervallet inte innehåller talet 0, så förkastar vi $H_0 : b = 0$ till förmån för $H_A : b \neq 0$ på signifikansnivå $\alpha = 0.1$.

6. Eftersom väntevärde är linjär operation, har vi

$$
\begin{aligned}
\mathbb{E}[X_n] &= \sum_{k=1}^{n} \frac{\mathbb{E}[Y_k]}{k} \\
&= \sum_{k=1}^{n} \frac{1}{k^2} \\
&< \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}.
\end{aligned}
$$

(Det exakta värdet är irrelevant, vi behöver bara veta att serien är konvergent.)

För alla $a_n > 0$ ger Markovs olikhet att

$$\mathbb{P}(X_n \geq a_n) \leq \frac{\mathbb{E}[X_n]}{a_n} < \frac{\pi^2}{6a_n}.$$

Särskilt gäller att om $a_n \to +\infty$, så finns något N sådant att $a_n > 0$ för alla $n \geq N$, och

$$0 \leq \mathbb{P}(X_n \geq a_n) < \frac{\pi^2}{6a_n} \quad \text{för alla } n \geq N.$$

Eftersom $\pi^2/6a_n \to 0$, så ger instängningsprincipen att

$$\lim_{n\to\infty} \mathbb{P}(X_n \geq a_n) = 0.$$

7. Vi utnyttjar det faktum att datan kommer från en normalfördelning, och då gäller att

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

Här är $n = 5$, och vi har att

$$\mathbb{P}\left(\frac{4S^2}{\sigma^2} \geq F^{-1}_{\chi^2_4}(0.05)\right) = 0.95.$$

Vi har $F^{-1}_{\chi^2_4}(0.05) = 0.71$, så

$$\mathbb{P}\left(\sigma^2 \leq \frac{4S^2}{0.71}\right) = 0.95.$$

Med vår datamängd har vi $s^2 = 3.425$, så vi får

$$\sigma^2 \leq \frac{4 \cdot 3.425}{0.71} = 19.3 \quad (95\%).$$

Alltså $s_c = 19.3$.

8. Låt $\sigma_k^2 = \mathrm{Var}\,(X_k)$, och låt $t \neq 0$ vara ett fixt tal (oberoende av n). Vi har

$$\begin{aligned} M_{Y_n}(t) &= \mathbb{E}\left[\exp\left\{\frac{t}{\sigma\sqrt{n}}\sum_{k=1}^{n} X_k\right\}\right] \\ &= \prod_{k=1}^{n} \mathbb{E}\left[\exp\left\{\frac{t}{\sigma\sqrt{n}} X_k\right\}\right] \\ &= \prod_{k=1}^{n} M_k\left(\frac{t}{\sigma\sqrt{n}}\right), \end{aligned}$$

där M_k är den momentgenererande funktionen för X_k. För små s gäller att

$$M_k(s) = \mathbb{E}\left[X_k^0\right] + \mathbb{E}\left[X_k^1\right] s + \mathbb{E}\left[X_k^2\right] \frac{s^2}{2} + O(s^2),$$

och särskilt gäller då att

$$\begin{aligned} M_k\left(\frac{t}{\sigma\sqrt{n}}\right) &= 1 + \frac{t^2\sigma_k^2}{2\sigma^2 n} + O(n^{-3/2}) \\ &= \exp\left\{\frac{t^2\sigma_k^2}{2\sigma^2 n} + O(n^{-3/2})\right\}. \end{aligned}$$

Då är

$$\begin{aligned} M_{Y_n}(t) &= \prod_{k=1}^{n} M_k\left(\frac{t}{\sigma\sqrt{n}}\right) \\ &= \prod_{k=1}^{n} \exp\left\{\frac{t^2\sigma_k^2}{2\sigma^2 n} + O(n^{-3/2})\right\} \\ &= \exp\left\{\left(\sum_{k=1}^{n} \frac{t^2\sigma_k^2}{2\sigma^2 n}\right) + O(n^{-1/2})\right\} \\ &= \exp\left\{\frac{t^2}{2\sigma^2 n}\sum_{k=1}^{n} \sigma_k^2 + O(n^{-1/2})\right\}. \end{aligned}$$

Men, definitionen av σ^2 ger då att

$$\lim_{n\to\infty} M_{Y_n}(t) = e^{t^2/2}.$$