Matematiska Vetenskaper 2023–05–31 FM

Chalmers Tekniska Högskola Examinator: Tony Johansson

MVE302, MVE395, TMA321
Tentamen

Detta är en sammanslagning av tentamen för de tre kurserna.
Notera att uppgifter märkta **MVE302** ej ingår på tentamen
för MVE395 och TMA321.

TMA321 & MVE395: 12–17p ger trea, 18–23p ger fyra, 24–30p ger femma.

MVE302: 16–23p ger trea, 24–31p ger fyra, 33–40p ger femma.

Tillåtna hjälpmedel:

- valfri miniräknare,
- Mathematics Handbook (Beta),
- 2 blad (totalt 4 sidor) handskrivna anteckningar (utskrivna anteckningar från skrivplatta godkänns),
- tabeller bifogade till tesen.

---

1. (6p) Två kontinuerliga stokastiska variabler X, Y har simultanfördelning

$$f(x,y) = \frac{1}{x}e^{-x}, \quad 0 < y < x.$$

Bestäm korrelationskoefficienten

$$\rho(X,Y) = \frac{\mathrm{Cov}\,(X,Y)}{\sqrt{\mathrm{Var}\,(X)\,\mathrm{Var}\,(Y)}}.$$

2. (5p) En vanlig sexsidig tärning kastas n gånger, och S_n betecknar summan av utfallen av de n kasten. Välj konstanterna a, b så att

$$\lim_{n\to\infty} \mathbb{P}\left(an - b\sqrt{n} \leq S_n \leq an + b\sqrt{n}\right) = 0.95.$$

3. (5p) Följande data kommer från en normalfördelning med okänt väntevärde μ och varians σ^2:

$$1.01,\ 1.22,\ 1.31,\ 1.66,\ 1.87,\ 2.15,\ 2.75,\ 3.21.$$

Ange ett 95% konfidensintervall för μ, och ta ställning till $H_0 : \mu = 1$ mot $H_A : \mu \neq 1$ på signifikansnivå 0.05.

4. Låt $\theta > 0$. För en populationsvariabel $X \sim \mathcal{U}[0, \theta]$ vill vi testa $H_0 : \theta = 3$ mot $H_A : \theta > 3$, baserat på teststatistikan

$$M = \max\{X_1, \dots, X_n\}.$$

(a) (4p) Utforma ett sådant test på signifikansnivå $\alpha = 0.05$.

(b) (**MVE302**, 3p) Bestäm styrkan $\beta(\theta)$ för testet med $n = 4$, som en funktion av $\theta > 3$.

5. Johanna är ett stort fan av Eurovision Song Contest, och undersöker sambandet mellan juryns poäng (X) och publikens poäng (Y) till ett bidrag. Baserat på data (x_k, y_k) från $n = 51$ bidrag från 2021 och 2022 beräknar hon följande summor:

$$\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} y_k = 4582,$$

$$S_{xx} = \sum_{k=1}^{n} (x_k - \overline{x})^2 = 327000,$$
$$S_{xy} = \sum_{k=1}^{n} (x_k - \overline{x})(y_k - \overline{y}) = 215000,$$
$$S_{yy} = \sum_{k=1}^{n} (y_k - \overline{y})^2 = 493000.$$

(a) (1p) Skatta koefficienterna a, b, σ^2 i en linjär regressionsmodell

$$Y = a + bX + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

(b) (3p) Sveriges bidrag 2023 fick 340 poäng från juryn, och Johanna vill känna sig säker på att det räcker för vinst innan publikens poäng tillkännages. Hon beräknar därför ett nedre begränsat 90% prediktionsintervall för publikens poäng för samma bidrag. Ange detta.

6. Betrakta följande process, där $n \geq 2$ är ett heltal. Välj X_1 likformigt från $\{1, \dots, n\}$. Givet X_1, välj X_2 likformigt från $\{1, \dots, X_1\}$. Generellt; för alla $k \geq 2$ väljs X_k likformigt från $\{1, \dots, X_{k-1}\}$. Låt

$$T_n = \min\{k : X_k = 1\}.$$

(a) (1p) Hitta sannolikhetsfunktionen för T_2.

(b) (3p) Hitta sannolikhetsfunktionen för T_3.

(c) (2p) Beräkna $\mathbb{E}[T_4]$.

7. (**MVE302**) För en populationsvariabel $X \sim \Gamma(2, \lambda)$ är vi intresserade av parametern λ. Vi väljer en Bayesiansk approach, och introducerar en slumpvariabel Λ och gör observationer $x_1, \ldots, x_n$.

(a) (3p) Givet priorfördelning $\Gamma(\alpha, \beta)$ med $\alpha \in \mathbb{N}$ och $\beta \in \mathbb{R}_{>0}$, vad är posteriorfördelningen för Λ?

(b) (2p) Skatta λ via ett posterior mean estimate $\hat{\lambda}_{\text{pme}}$.

(c) (2p) Visa att skattningen $\hat{\lambda}_{\text{pme}}$ är konsistent.

Tips. En stokastisk variabel $X \sim \Gamma(\alpha, \lambda)$ med $\alpha \in \mathbb{N}$ har täthetsfunktion

$$f_X(x) = \frac{\lambda^\alpha x^{\alpha-1}}{(\alpha-1)!} e^{-\lambda x}, \quad x > 0.$$

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. (6p) Två kontinuerliga stokastiska variabler X, Y har simultanfördelning

$$f(x,y) = \frac{1}{x}e^{-x}, \quad 0 < y < x.$$

Bestäm korrelationskoefficienten

$$\rho(X,Y) = \frac{\mathrm{Cov}\,(X,Y)}{\sqrt{\mathrm{Var}\,(X)\,\mathrm{Var}\,(Y)}}.$$

Lösning. Vi behöver beräkna ett gäng väntevärden; först och främst skriver vi om

$$\begin{aligned}
\mathrm{Cov}\,(X,Y) &= \mathbb{E}\,[XY] - \mathbb{E}\,[X]\,\mathbb{E}\,[Y]\,, \\
\mathrm{Var}\,(X) &= \mathbb{E}\left[X^2\right] - \mathbb{E}\,[X]^2\,, \\
\mathrm{Var}\,(Y) &= \mathbb{E}\left[Y^2\right] - \mathbb{E}\,[Y]^2\,.
\end{aligned}$$

Det underlättar om vi kan hitta marginal- eller betingade fördelningar. Vi har

$$f_X(x) = \int_0^x \frac{1}{x}e^{-x}dy = e^{-x},$$

medan $f_Y(y)$ inte är lätt att få något uttryck för. Istället noterar vi att

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)} = \frac{1}{x}, \quad 0 < y < x.$$

Kommentar. Nu kan vi inse (för att underlätta intuitionen) att $X \sim \exp(1)$, och Y har den betingade fördelningen $\mathcal{U}[0,X]$. Jag kommer fortsätta lösningen under antagandet att vi *inte* har insett detta.

I väntevärdesberäkningarna nedan kommer vi upprepade gånger använda

$$\int_0^\infty e^{-x}dx = 1, \quad \int_0^\infty xe^{-x}dx = 1, \quad \int_0^\infty x^2e^{-x}dx = 2.$$

(Dessa värden kan vi beräkna med hjälp av partiell integration, alternativt genom att inse att det är värden av Gammafunktionen.)

Vi beräknar väntevärden (med hjälp av väntevärdesvarianten av totala sannolikhetslagen (TSL) i vissa fall):

$$\begin{aligned}
\mathbb{E}\,[X] &= \int_0^\infty xe^{-x}dx = 1, \\
\mathbb{E}\,[Y] &\overset{\text{TSL}}{=} \int_0^\infty \mathbb{E}\,[Y \mid X = x]\,f_X(x)dx = \int_0^\infty \left(\int_0^x \frac{y}{x}dy\right)e^{-x}dx \\
&= \int_0^\infty \frac{x}{2}e^{-x}dx = \frac{1}{2}.
\end{aligned}$$

Vidare har vi (med $A = \{(x,y) : 0 < y < x\}$)

$$\begin{aligned}
\mathbb{E}\left[XY\right] &= \iint_A xyf(x,y)dxdy = \int_0^\infty \int_y^\infty ye^{-x}dxdy \\
&= \int_0^\infty ye^{-y}dy = 1, \\
\mathbb{E}\left[X^2\right] &= \int_0^\infty x^2e^{-x}dx \\
&= 2, \\
\mathbb{E}\left[Y^2\right] &\overset{\text{TSL}}{=} \int_0^\infty \mathbb{E}\left[Y^2 \mid X = x\right] f_X(x)dx = \int_0^\infty \left(\int_0^x \frac{y^2}{x}dy\right) e^{-x}dx \\
&= \int_0^\infty \frac{x^2}{3}e^{-x}dx = \frac{2}{3}.
\end{aligned}$$

Då har vi

$$\begin{aligned}
\operatorname{Cov}(X,Y) &= \mathbb{E}\left[XY\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right] = \frac{1}{2}, \\
\operatorname{Var}(X) &= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2 = 1, \\
\operatorname{Var}(Y) &= \mathbb{E}\left[Y^2\right] - \mathbb{E}\left[Y\right]^2 = \frac{2}{3} - \frac{1}{4} = \frac{5}{12}.
\end{aligned}$$

Alltså är

$$\rho(X,Y) = \frac{\frac{1}{2}}{\sqrt{1 \cdot \frac{5}{12}}} = \sqrt{\frac{3}{5}}.$$

2. (5p) En vanlig sexsidig tärning kastas n gånger, och S_n betecknar summan av utfallen av de n kasten. Välj konstanterna a, b så att

$$\lim_{n\to\infty} \mathbb{P}\left(an - b\sqrt{n} \leq S_n \leq an + b\sqrt{n}\right) = 0.95.$$

Lösning. Vi använder centrala gränsvärdessatsen. Låt $X_k \in \{1,2,3,4,5,6\}$ vara utfallet av kast k, så att

$$S_n = \sum_{k=1}^n X_k.$$

Vi har

$$\begin{aligned}
\mu = \mathbb{E}\left[X_1\right] &= \frac{1+2+3+4+5+6}{6} = \frac{7}{2} = 3.5, \\
\sigma^2 = \operatorname{Var}(X_1) &= \frac{1}{6}\sum_{k=1}^6 (k-3.5)^2 = \frac{35}{12} \approx 2.92.
\end{aligned}$$

Notera att Stora Talens Lag säger att om $a \neq \mu$, så är gränsvärdet i fråga 0; om $a < \mu$ kan vi låta $\varepsilon > 0$ vara så att $a = \mu - 2\varepsilon$, och för stora n är

$$\begin{aligned}\mathbb{P}(S_n \leq an + b\sqrt{n}) &= \mathbb{P}\left(\frac{S_n}{n} \leq a + \frac{b}{\sqrt{n}}\right) \\ &\leq \mathbb{P}\left(\frac{S_n}{n} \leq \mu - \varepsilon\right) \\ &\xrightarrow{\text{STL}} 0.\end{aligned}$$

Så vi måste ha $a \geq \mu$, och på samma sätt visas att $a \leq \mu$, alltså $a = \mu$.
Låt oss välja b.

Vi har $\sigma \approx \sqrt{2.92} \approx 1.71$. Centrala gränsvärdessatsen säger att

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - \mu n}{\sigma\sqrt{n}} \leq x\right) = \Phi(x)$$

för alla $x \in \mathbb{R}$. Vi har då

$$\begin{aligned}\lim_{n\to\infty} \mathbb{P}\left(\mu n - b\sqrt{n} \leq S_n \leq \mu n + b\sqrt{n}\right) &= \Phi\left(\frac{b}{\sigma}\right) - \Phi\left(\frac{-b}{\sigma}\right) \\ \text{(symmetri hos } \Phi) &= 2\Phi\left(\frac{b}{\sigma}\right) - 1.\end{aligned}$$

Vi behöver välja b så att

$$2\Phi\left(\frac{b}{\sigma}\right) - 1 = 0.95,$$

vilket ger

$$\Phi\left(\frac{b}{\sigma}\right) = 0.975 \Longrightarrow \frac{b}{\sigma} = \Phi^{-1}(0.975) \Longrightarrow b = \sigma\Phi^{-1}(0.975).$$

Från tabellen ser vi att $\Phi^{-1}(0.975) \approx 2$ (egentligen är svaret ≈ 1.96), så

$$b = \sqrt{\frac{32}{15}}\Phi^{-1}(0.975) \approx 2 \cdot 1.71 = 3.42.$$

Svar. $a = 3.5$ och $b = 3.42$.

3. (5p) Följande data kommer från en normalfördelning med okänt väntevärde μ och varians σ^2:

$$1.01,\ 1.22,\ 1.31,\ 1.66,\ 1.87,\ 2.15,\ 2.75,\ 3.21.$$

Ange ett 95% konfidensintervall för μ, och ta ställning till $H_0 : \mu = 1$ mot $H_A : \mu \neq 1$ på signifikansnivå 0.05.

Lösning. Frågan är, som många påpekade, tvetydig – känner vi till variansen eller ej? Båda tolkningar är okej (även om min intention är att variansen ska vara okänd), så jag presenterar båda lösningarna.

Lösning A: okänd σ^2.

Vi beräknar stickprovsmedelvärde och -standardavvikelse:

$$\overline{x} = 1.9, \quad s = 0.77.$$

Ett 95% konfidensintervall för μ får vi via formeln

$$\mu \in \overline{x} \pm F_{t_{n-1}}^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{s}{\sqrt{n}} \quad (1 - \alpha).$$

Konfidensgrad 95% motsvarar $\alpha = 0.05$. Kvantilen $F_{t_7}^{-1}(0.975) = 2.36$ hämtas från tabell, så vi får

$$\mu \in 1.9 \pm 2.36 \cdot \frac{0.77}{\sqrt{8}} = 1.9 \pm 0.64 \quad (95\%).$$

Konfidensintervallet innehåller inte talet 1, så vi förkastar H_0 på signifikansnivå $\alpha = 0.05$.

Lösning B: känd σ^2.

Vi beräknar stickprovsmedelvärde

$$\overline{x} = 1.9.$$

Ett 95% konfidensintervall för μ får vi via formeln

$$\mu \in \overline{x} \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}} \quad (1 - \alpha).$$

Konfidensgrad 95% motsvarar $\alpha = 0.05$, och vi har kvantilen $\Phi^{-1}(0.975) \approx 2$ (från tabell). Så vi får (med $n = 8$)

$$\mu \in 1.9 \pm \frac{2\sigma}{\sqrt{8}} = 1.9 \pm \frac{\sigma}{\sqrt{2}}.$$

Hur vi ställer oss till hypotesen $H_0 : \mu = 1$ mot $H_A : \mu \neq 1$ beror på huruvida intervallet innehåller 1. Detta beror på σ, och gränsen går vid $\sigma = 0.9\sqrt{2}$.

Om $\sigma > 0.9\sqrt{2}$ behålls H_0. Om $\sigma < 0.9\sqrt{2}$ förkastas H_0.

Kommentar till båda lösningar. Det går också bra att välja ett ensidigt konfidensintervall för μ (eftersom uppgiften inte specificerar), men endast det symmetriska intervallet i lösningen kan användas för att utföra hypotestestet.

4. Låt $\theta > 0$. För en populationsvariabel $X \sim \mathcal{U}[0, \theta]$ vill vi testa $H_0 : \theta = 3$ mot $H_A : \theta > 3$, baserat på teststatistikan

$$M = \max\{X_1, \ldots, X_n\}.$$

(a) (4p) Utforma ett sådant test på signifikansnivå $\alpha = 0.05$.

(b) (**MVE302**, 3p) Bestäm styrkan $\beta(\theta)$ för testet med $n = 4$, som en funktion av $\theta > 3$.

Lösning.

(a) Vi börjar med att undersöka teststatistikans fördelning. För $0 < x < \theta$ gäller att

$$\begin{aligned} \mathbb{P}(M \leq x) &= \mathbb{P}(X_k \leq x \ \forall k) \\ \text{(oberoende)} &= \mathbb{P}(X_1 \leq x)^n \\ &= \left(\frac{x}{\theta}\right)^n. \end{aligned}$$

Då är

$$\mathbb{P}(M > x) = 1 - \left(\frac{x}{\theta}\right)^n.$$

Förkastningsregionen för $H_A : \theta > 3$ ska ha formen $\mathcal{R}_\alpha = (m, \infty)$ för något $m \in \mathbb{R}$, eftersom ju större M är desto mer tror vi på $\theta > 3$. Vi väljer m så att

$$\begin{aligned} 0.05 &= \mathbb{P}_{H_0}(M > m) \\ &= 1 - \left(\frac{m}{3}\right)^n. \end{aligned}$$

Vi löser ut m och får

$$m = 3 \cdot 0.95^{1/n}.$$

Alltså är $\mathcal{R}_\alpha = (3 \cdot 0.95^{1/n},\ \infty)$.

(b) För $n = 4$ får vi $m = 3 \cdot 0.95^{1/4} = 2.96$. Styrkan är, för $\theta > \theta_0$,

$$\beta(\theta) = \mathbb{P}_\theta(M \in \mathcal{R}_\alpha) = \mathbb{P}_\theta(M > 2.96) = 1 - \left(\frac{2.96}{\theta}\right)^4.$$

5. Johanna är ett stort fan av Eurovision Song Contest, och undersöker sambandet mellan juryns poäng (X) och publikens poäng (Y) till ett bidrag. Baserat på data (x_k, y_k) från $n = 51$ bidrag från 2021 och 2022 beräknar hon följande summor:

$$\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} y_k = 4582,$$

$$
\begin{aligned}
S_{xx} &= \sum_{k=1}^{n}(x_k - \overline{x})^2 = 327000,\\
S_{xy} &= \sum_{k=1}^{n}(x_k - \overline{x})(y_k - \overline{y}) = 215000,\\
S_{yy} &= \sum_{k=1}^{n}(y_k - \overline{y})^2 = 493000.
\end{aligned}
$$

(a) (1p) Skatta koefficienterna a, b, σ^2 i en linjär regressionsmodell

$$Y = a + bX + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2).$$

(b) (3p) Sveriges bidrag 2023 fick 340 poäng från juryn, och Johanna vill känna sig säker på att det räcker för vinst innan publikens poäng tillkännages. Hon beräknar därför ett nedre begränsat 90% prediktionsintervall för publikens poäng för samma bidrag. Ange detta.

Lösning.

(a) Vi använder formler

$$
\begin{aligned}
\hat{b} &= \frac{S_{xy}}{S_x x} = 0.65,\\
\hat{a} &= \overline{y} - \hat{b}\overline{x} = \frac{4582}{51}(1 - \hat{b}) = 30.65,\\
s^2 &= \frac{1}{n-2}\left(S_{yy} - \frac{S_{xy}^2}{S_{xx}}\right) = 7160.
\end{aligned}
$$

(b) Ett nedre prediktionsintervall baserat på observation $x = 340$ har nedre gräns

$$\hat{a} + \hat{b}x - F_{t_{n-2}}^{-1}(0.9)s\sqrt{1 + \frac{1}{n} + \frac{(x - \overline{x})^2}{S_{xx}}} = 133.4.$$

Intervallet är då $(133.4, \infty)$.

(Johanna kunde alltså inte känna sig särskilt säker på de 287 poäng som behövdes för vinst. Jag kan också påpeka att modellen som togs fram inte är särskilt bra, och att en kvadratisk modell $Y^2 = a + bX^2$ är bättre lämpad.)

6. Betrakta följande process, där $n \geq 2$ är ett heltal. Välj X_1 likformigt från $\{1, \ldots, n\}$. Givet X_1, välj X_2 likformigt från $\{1, \ldots, X_1\}$. Generellt; för alla $k \geq 2$ väljs X_k likformigt från $\{1, \ldots, X_{k-1}\}$. Låt

$$T_n = \min\{k : X_k = 1\}.$$

(a) (1p) Hitta sannolikhetsfunktionen för T_2.

(b) (3p) Hitta sannolikhetsfunktionen för T_3.

(c) (2p) Beräkna $\mathbb{E}[T_4]$.

Lösning.

(a) Låt t vara ett positivt heltal. Händelsen $T_2 = t$ innebär att $X_1 = X_2 = \cdots = X_{t-1} = 2$, medan $X_t = 1$. Detta har sannolikhet 2^{-t}, så

$$\mathbb{P}(T_2 = t) = 2^{-t}, \quad t = 1, 2, \ldots.$$

Alltså är $T_2 \sim \text{Geom}(1/2)$.

(b) När $n = 3$, så kan vi även definiera

$$K = \min\{k : X_k < 3\}.$$

Då är $K \sim \text{Geom}(2/3)$, enligt samma argument som i **(a)**. Då har vi

$$\begin{aligned}
\mathbb{P}(T_3 = t) &= \sum_{k=1}^{t} \mathbb{P}(T_3 = t \mid K = k)\mathbb{P}(K = k) \\
&= \sum_{k=1}^{t} \mathbb{P}(T_3 = t \mid K = k)\frac{2}{3}\left(1 - \frac{2}{3}\right)^{k-1} \\
&= 2\sum_{k=1}^{t} \frac{1}{3^k}\mathbb{P}(T_3 = t \mid K = k).
\end{aligned}$$

Notera att

$$\mathbb{P}(T_3 = k \mid K = k) = \frac{1}{2},$$

eftersom X_K kan vara 1 eller 2 med sannolikhet $1/2$. Vi har generellt, för $t \geq k$,

$$\mathbb{P}(T_3 = t \mid K = k) = \left(\frac{1}{2}\right)^{t-k+1}.$$

Det gäller då för alla positiva heltal t att

$$\begin{aligned}
\mathbb{P}(T_3 = t) &= 2\sum_{k=1}^{t} \frac{1}{3^k}\left(\frac{1}{2}\right)^{t-k+1} \\
&= 2^{-t}\sum_{k=1}^{t} \left(\frac{2}{3}\right)^{k} \\
&= 2^{-t}\frac{2}{3}\frac{1 - \left(\frac{2}{3}\right)^t}{1 - \frac{2}{3}} \\
&= 2(2^{-t} - 3^{-t})
\end{aligned}$$

(c) Anta att vi väljer något godtyckligt $n \geq 2$, och definiera

$$K_n = \min\{k : X_k < n\}.$$

Då har vi $K_n \sim \text{Geom}(1 - 1/n)$, så

$$\mathbb{E}[K_n] = \frac{1}{1 - \frac{1}{n}} = \frac{n}{n-1}.$$

Låt $N = X_{K_n}$ vara det första värdet lägre än n. Då är N likformigt fördelat i $\{1, \ldots, n-1\}$. Vi har

$$\begin{aligned}
\mathbb{E}[T_n] &= \mathbb{E}[K_n] + \mathbb{E}[T_n - K_n] \\
&= \frac{n}{n-1} + \sum_{j=1}^{n-1} \mathbb{E}[T_n - K_n \mid N = j] \, \mathbb{P}(N = j) \\
&= \frac{n}{n-1} + \frac{1}{n-1} \sum_{j=1}^{n-1} \mathbb{E}[T_n - K_n \mid N = j].
\end{aligned}$$

Vi noterar att om $N = 1$ så är $T_n = K_n$. För övriga värden har vi

$$\mathbb{E}[T_n - K_n \mid N = j] = \mathbb{E}[T_j].$$

Då gäller rekursionen

$$\mathbb{E}[T_n] = \frac{1}{n-1} \left(n + \sum_{j=2}^{n-1} \mathbb{E}[T_j] \right).$$

Från **(a)** har vi $\mathbb{E}[T_2] = 2$, och med hjälp av rekursionen får vi

$$\begin{aligned}
\mathbb{E}[T_3] &= \frac{1}{2}(3+2) = \frac{5}{2} = 3 - \frac{1}{2}, \\
\mathbb{E}[T_4] &= \frac{1}{3}\left(4 + 2 + \frac{5}{2}\right) = \frac{17}{6}.
\end{aligned}$$

7. (**MVE302**) För en populationsvariabel $X \sim \Gamma(2, \lambda)$ är vi intresserade av parametern λ. Vi väljer en Bayesiansk approach, och introducerar en slumpvariabel Λ och gör observationer $x_1, \ldots, x_n$.

(a) (3p) Givet priorfördelning $\Gamma(\alpha, \beta)$ med $\alpha \in \mathbb{N}$ och $\beta \in \mathbb{R}_{>0}$, vad är posteriorfördelningen för Λ?

(b) (2p) Skatta λ via ett posterior mean estimate $\hat{\lambda}_{\text{pme}}$.

(c) (2p) Visa att skattningen $\hat{\lambda}_{\text{pme}}$ är konsistent.

Tips. En stokastisk variabel $X \sim \Gamma(\alpha, \lambda)$ med $\alpha \in \mathbb{N}$ har täthetsfunktion

$$f_X(x) = \frac{\lambda^\alpha x^{\alpha-1}}{(\alpha-1)!} e^{-\lambda x}, \quad x > 0.$$

Lösning.

(a) Vår prior är

$$f_\Lambda(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1}}{(\alpha-1)!} e^{-\beta\lambda} \propto \lambda^{\alpha-1} e^{-\beta\lambda}.$$

Vi har

$$\begin{aligned} f_{X|\Lambda}(x_1, \ldots, x_n \mid \lambda) &= \prod_{k=1}^{n} \lambda^2 x_k e^{-\lambda x_k} \\ &\propto \lambda^{2n} e^{-\lambda \sum x_k}. \end{aligned}$$

Vi har då posterior

$$\begin{aligned} f_{\Lambda|X}(\lambda \mid x_1, \ldots, x_n) &\propto \lambda^{2n} e^{-\lambda \sum x_k} \lambda^{\alpha-1} e^{-\beta\lambda} \\ &= \lambda^{\alpha+2n-1} e^{-\lambda(\beta + \sum x_k)}. \end{aligned}$$

Detta är $\Gamma\left(\alpha + 2n, \beta + \sum_{k=1}^{n} x_k\right)$.

(b) Vi behöver ta reda på vad Gammafördelningen har för väntevärde. Detta går att slå upp i tabell, eller genom att inse att det (eftersom första parametern är ett heltal) är summan av oberoende exponentialfördelade variabler. Annars kan vi också göra såhär: om $X \sim \Gamma(\alpha, \beta)$ där α är ett heltal, så är

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty x \frac{\beta^\alpha x^{\alpha-1}}{(\alpha-1)!} e^{-\beta x} dx \\ &= \frac{\alpha!}{\beta(\alpha-1)!} \int_0^\infty \frac{\beta^{\alpha+1} x^\alpha}{\alpha!} e^{-\beta x} dx \\ &= \frac{\alpha}{\beta}. \end{aligned}$$

Här har vi använt att den sista integralen är av täthetsfunktionen för $\Gamma(\alpha + 1, \beta)$.

Givet en posteriorfördelning $\Lambda \sim \Gamma(\alpha + 2n, \beta + \sum_{k=1}^{n} x_k)$, får vi

$$\hat{\lambda}_{\text{pme}} = \mathbb{E}[\Lambda] = \frac{\alpha + 2n}{\beta + \sum_{k=1}^{n} x_k}.$$

(c) Vi använder stora talens lag för att visa att $\hat{\lambda}_{\text{pme}}$ är konsistent. Låt $X_1, X_2, \ldots$ vara oberoende med fördelning $\Gamma(2, \lambda)$, och låt $S_n = \sum_{k=1}^n S_n$. Definiera $Y_n = \hat{\lambda}_{\text{pme}} - \lambda$. Eftersom $\mu = \mathbb{E}[X] = 2/\lambda$ har vi

$$\begin{aligned} Y_n = \hat{\lambda}_{\text{pme}} - \lambda &= \frac{\alpha + 2n}{\beta + S_n} - \lambda \\ &= \frac{\alpha/n + 2}{\beta/n + S_n/n} - \frac{2}{\mu} \\ &= \frac{\alpha\mu/n + 2(\mu - S_n/n)}{\mu(\beta/n + S_n/n)}. \end{aligned}$$

Nu handlar det om att inse att $S_n/n \to \mu$ via stora talens lag, så $\hat{\lambda}_{\text{pme}} \to \lambda$. Formellt bevisar vi det som följer. Låt $\varepsilon > 0$ vara godtyckligt, och låt $\delta > 0$ (som vi väljer senare), och definiera händelsen

$$\mathcal{E}_n = \left\{ \left| \frac{S_n}{n} - \mu \right| \leq \delta \right\}.$$

Då säger stora talens lag att $\mathbb{P}(\mathcal{E}_n) \to 1$ då $n \to \infty$. Vi har

$$\begin{aligned} \mathbb{P}\left(|Y_n| \geq \varepsilon\right) &= \mathbb{P}(|Y_n| \geq \varepsilon \mid \mathcal{E}_n)\mathbb{P}(\mathcal{E}_n) + \mathbb{P}(|Y_n| \geq \varepsilon \mid \overline{\mathcal{E}_n})\mathbb{P}(\overline{\mathcal{E}_n}) \\ &\leq \mathbb{P}(|Y_n| \geq \varepsilon \mid \mathcal{E}_n) + \mathbb{P}(\overline{\mathcal{E}_n}). \end{aligned}$$

Om $\mathcal{E}_n$ inträffar har vi, för tillräckligt stora n,

$$\begin{aligned} |Y_n| &= \left| \frac{\alpha\mu}{\mu\beta + \mu S_n} + \frac{2(\mu - S_n/n)}{\mu(\beta/n + S_n/n)} \right| \\ &\leq \frac{\alpha\mu}{\mu\beta + \mu(\mu - \delta)n} + \frac{2\delta}{\mu(\beta/n + (\mu - \delta))} \\ &\leq \frac{\varepsilon}{2} + \frac{2\delta}{\mu(\mu - 2\delta)} \\ &< \varepsilon, \end{aligned}$$

om vi väljer δ tillräckligt litet. Detta ger

$$\mathbb{P}(|Y_n| \geq \varepsilon \mid \mathcal{E}_n) = 0, \qquad \text{för stora } n.$$

Så,

$$\lim_{n\to\infty} \mathbb{P}(|Y_n| \geq \varepsilon) \leq \lim_{n\to\infty} \mathbb{P}(\overline{\mathcal{E}_n}) = 0,$$

enligt stora talens lag. Detta visar att $\hat{\lambda}_{\text{pme}}$ är konsistent.