Matematiska Vetenskaper 2023–10–07 EM
Chalmers Tekniska Högskola Examinator: Tony Johansson (1069)

MVE302 – Sannolikhet och Statistik
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. (5p) Två kontinuerliga slumpvariabler X, Y har simultan täthetsfunktion

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

Tips. För heltal $k \geq 0$ gäller att $\int_0^\infty x^k e^{-x} dx = k!$.

(a) Hitta $f_X(x)$ för alla $x > 0$.

(b) Hitta $f_{Y|X}(y \mid x)$ för alla $0 < x < y$.

(c) Visa att X och $Y - X$ är oberoende och likfördelade.

(d) Beräkna $\mathbb{E}[Y]$ och $\mathrm{Var}(Y)$, förslagsvis med hjälp av **(d)**.

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

$$\lim_{n\to\infty} \mathbb{P}\left(an - b\sqrt{n} \leq S_n \leq an + b\sqrt{n}\right) = 0.99.$$

3. (5p) För att avgöra om ett mynt är rättvist, det vill säga ger 50% chans till klave vid singling, gör Boel ett hypotestest. Hon låter p beteckna myntets sannolikhet för klave, och ställer upp nollhypotes $H_0 : p = 0.5$ och alternativhypotes $H_1 : p \neq 0.5$. Hon singlar slanten $n = 400$ gånger, och låter X beteckna antalet klave.

För testet väljer Boel signifikansnivå $\alpha = 0.05$, och en region $\mathcal{R}$ sådan att $X \in \mathcal{R}$ leder till att hon förkastar nollhypotesen. Ange $\mathcal{R}$, med hjälp av lämpliga approximationer.

4. (5p) Antag att $X \sim \Gamma(2, \theta)$, så att X har följande täthetsfunktion med parameter $\theta > 0$:

$$f_\theta(x) = \theta^2 x e^{-\theta x}, \quad x > 0.$$

Hitta en maximum likelihood–skattning av θ.

5. (5p) Låt N vara en Poissonfördelad slumpvariabel med intensitet $\lambda > 0$, och låt $X_1, X_2, \ldots$ vara oberoende slumpvariabler, likformigt fördelade på $[0, 1]$, som även är oberoende av N. Definiera en slumpvariabel

$$T = \begin{cases} \min\{X_1, \ldots, X_N\} & \text{om } N \geq 1, \\ 1 & \text{om } N = 0. \end{cases}$$

Beräkna $F(t) = \mathbb{P}(T \leq t)$ för alla $0 < t < 1$.

Tips. För alla $x \in \mathbb{R}$ gäller att $\sum_{n=1}^{\infty} x^n/n! = e^x - 1$.

6. (5p) En kontinuerlig populationsvariabel X är exponentialfördelad med okänd intensitet $\theta > 0$. Alltså har X täthetsfunktion

$$f_X(x) = \theta e^{-\theta x}, \quad x > 0.$$

Ett stickprov av storlek $n = 4$ tas, och vi noterar teststatistikan

$$\min\{x_1, x_2, x_3, x_4\} = 0.1.$$

Ange, baserat på detta, ett 95% konfidensintervall för θ på formen $[a, \infty)$.

7. (**MVE302**, 3+2p) Anta att vi känner till att en populationsvariabel X är normalfördelad med varians 20 och okänt väntevärde μ. Vi vill testa nollhypotesen $H_0 : \mu = 0$ mot alternativhypotesen $H_1 : \mu > 0$, med ett stickprov av storlek $n = 5$. Som teststatistika används stickprovsmedelvärdet $\overline{X}$. Vi kommer fram till att H_0 ska förkastas på signifikansnivå α om $\overline{X} > 2.5$.

(a) Vad är signifikansnivån α?

(b) För vilket värde på μ gäller att testets styrka är 90%?

8. (**MVE302**, 3+2p) En populationsvariabel X är Bernoullifördelad, $X \sim \mathrm{Bin}(1, \lambda)$, och vi är intresserade av parametern $\lambda \in [0, 1]$. Vi väljer en Bayesiansk approach, och introducerar en slumpvariabel Λ och gör observationer $x_1, \ldots, x_n$.

(a) (3p) Som priorfördelning väljer vi Betafördelningen $\mathcal{B}(4, 3)$. Vad är posteriorfördelningen för Λ?

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

Tips. En stokastisk variabel $Y \sim \mathcal{B}(a, b)$ har täthetsfunktion

$$f_Y(y) = \frac{1}{B(a,b)} y^{a-1}(1-y)^{b-1}, \quad 0 < y < 1,$$

där den normaliserande konstanten $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$ ges av

$$B(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!} \quad \text{om } a \text{ och } b \text{ är heltal.}$$

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. (5p) Två kontinuerliga slumpvariabler X, Y har simultan täthetsfunktion

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

Tips. För heltal $k \geq 0$ gäller att $\int_0^\infty x^k e^{-x} dx = k!$.

(a) Hitta $f_X(x)$ för alla $x > 0$.

(b) Hitta $f_{Y|X}(y \mid x)$ för alla $0 < x < y$.

(c) Visa att X och $Y - X$ är oberoende och likfördelade.

(d) Beräkna $\mathbb{E}[Y]$ och $\mathrm{Var}(Y)$, förslagsvis med hjälp av **(d)**.

Lösning.

(a) För varje $x > 0$ integrerar vi över $y \in (x, \infty)$ för att få fram marginalfördelningen:

$$\begin{aligned} f_X(x) &= \int_x^\infty f(x, y) dy \\ &= \int_x^\infty e^{-y} dy \\ &= e^{-x}. \end{aligned}$$

(b) Den betingade täthetsfunktionen ges, för $0 < x < y$, av

$$\begin{aligned} f_{Y|X}(y \mid x) &= \frac{f(x, y)}{f_X(x)} \\ &= \frac{e^{-y}}{e^{-x}} \\ &= e^{-(y-x)}. \end{aligned}$$

(c) Låt $U = Y - X$. Då gäller att

$$\begin{aligned} \mathbb{P}(U > u \mid X = x) &= \mathbb{P}(Y - X > u \mid X = x) \\ &= \mathbb{P}(Y > u + x \mid X = x) \\ &= \int_{u+x}^\infty e^{-(y-x)} dy \\ &= e^{-u}, \quad u > 0. \end{aligned}$$

Eftersom denna är oberoende av x, så måste X och U vara oberoende. Det gäller alltså att

$$\mathbb{P}(U > u) = \mathbb{P}(U > u \mid X = x) = e^{-u}.$$

Täthetsfunktionen för U är då

$$f_U(u) = -\frac{d}{du}\mathbb{P}(U > u) = e^{-u},$$

samma som för X.

(d) Vi skriver $Y = X + (Y - X)$. Variablerna X och $Y - X$ är oberoende och likfördelade (exponentialfördelade med intensitet 1). Vi har alltså

$$\begin{aligned}\mathbb{E}[Y] &= \mathbb{E}[X] + \mathbb{E}[Y - X] \\ &= 2\mathbb{E}[X] \\ &= 2\int_0^\infty xe^{-x}dx \\ &= 2,\end{aligned}$$

samt

$$\begin{aligned}\mathrm{Var}(Y) &= \mathrm{Var}(X) + \mathrm{Var}(X - Y) \\ &= 2\mathrm{Var}(X) \\ &= 2(\mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2) \\ &= 2(\mathbb{E}\left[X^2\right] - 1).\end{aligned}$$

Vi behöver beräkna

$$\mathbb{E}\left[X^2\right] = \int_0^\infty x^2e^{-x}dx = 2.$$

Vi får då $\mathrm{Var}(Y) = 2(2-1) = 2$.

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

$$\lim_{n\to\infty} \mathbb{P}\left(an - b\sqrt{n} \le S_n \le an + b\sqrt{n}\right) = 0.99.$$

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

$$\mu = \mathbb{E}[X_1] = \frac{1+2+3+4+5+6}{6} = \frac{7}{2} = 3.5,$$

$$\sigma^2 = \mathrm{Var}(X_1) = \frac{1}{6}\sum_{k=1}^6 (k-3.5)^2 = \frac{35}{12} \approx 2.92.$$

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

$$2\Phi\left(\frac{b}{\sigma}\right) - 1 = 0.99,$$

vilket ger

$$\Phi\left(\frac{b}{\sigma}\right) = 0.995 \implies \frac{b}{\sigma} = \Phi^{-1}(0.995) \implies b = \sigma\Phi^{-1}(0.995).$$

Från tabellen ser vi att $\Phi^{-1}(0.995) \approx 2.5$ (egentligen är svaret ≈ 2.578), så

$$b = \sqrt{\frac{35}{12}}\Phi^{-1}(0.995) \approx 1.71 \cdot 2.5 = 4.275.$$

Svar. $a = 3.5$ och $b = 4.275$.

3. (5p) För att avgöra om ett mynt är rättvist, det vill säga ger 50% chans till klave vid singling, gör Boel ett hypotestest. Hon låter p beteckna myntets sannolikhet för klave, och ställer upp nollhypotes $H_0 : p = 0.5$ och alternativhypotes $H_1 : p \neq 0.5$. Hon singlar slanten $n = 400$ gånger, och låter X beteckna antalet klave.

För testet väljer Boel signifikansnivå $\alpha = 0.05$, och en region $\mathcal{R}$ sådan att $X \in \mathcal{R}$ leder till att hon förkastar nollhypotesen. Ange $\mathcal{R}$, med hjälp av lämpliga approximationer.

Lösning. Det gäller att $X \sim \text{Bin}(400, p)$. Under H_0 är specifikt

$$X \overset{H_0}{\sim} \text{Bin}(400, 1/2).$$

Vi vill hitta en lämplig mängd $\mathcal{R} \subseteq \{0, 1, \ldots, 400\}$ sådan att

$$\mathbb{P}_{H_0}(X \in \mathcal{R}) \approx \alpha = 0.05.$$

Eftersom alternativhypotesen är tvåsidig, är det rimligt att vi väljer en region $\mathcal{R}$ av formen

$$\mathcal{R} = \{0, 1, \ldots, 200 - r\} \cup \{200 + r, 32 + r + 1, \ldots, 400\}$$

för något heltal r. På grund av symmetri räcker det att bestämma r så att

$$\mathbb{P}_{H_0}(X \geq 200 + r) \approx \frac{\alpha}{2} = 0.025.$$

Det återstår att hitta r.

Normalapproximation. Under H_0 kan vi approximera $X \sim \text{Bin}(400, 1/2)$ med en normalfördelad variabel $Y \sim N(\mu, \sigma^2)$ med

$$\mu = \mathbb{E}[X] = 400 \cdot \frac{1}{2} = 200, \quad \sigma^2 = \text{Var}(X) = 400 \cdot \frac{1}{2}\left(1 - \frac{1}{2}\right) = 100.$$

Approximationen (med heltalskorrigering för stilpoäng) är att

$$\mathbb{P}(X \geq x) \approx \mathbb{P}\left(Y \geq x + \frac{1}{2}\right).$$

Vi vill då lösa

$$\begin{aligned} 0.025 &\approx \mathbb{P}_{H_0}(X \geq 200 + r) \\ &\approx \mathbb{P}(Y \geq 200.5 + r) \\ &= \mathbb{P}\left(\frac{Y - 200}{\sqrt{100}} \leq \frac{200.5 + r - 200}{\sqrt{100}}\right) \\ &= 1 - \Phi\left(\frac{r + 0.5}{10}\right). \end{aligned}$$

Vi får

$$r \approx 10\Phi^{-1}(0.975) - 0.5 \approx 10 \cdot 1.96 - 0.5 \approx 19.$$

Vi förkastar H_0 om $X \geq 200 + 19 = 219$ eller $X \leq 200 - 19 = 181$.

4. (5p) Antag att $X \sim \Gamma(2, \theta)$, så att X har följande täthetsfunktion med parameter $\theta > 0$:

$$f_\theta(x) = \theta^2 x e^{-\theta x}, \quad x > 0.$$

Hitta en maximum likelihood–skattning av θ.

Lösning. För ett stickprovsutfall $x_1, \ldots, x_n$ ges loglikelihoodfunktionen av

$$\begin{aligned}\ell(\theta) &= \sum_{i=1}^{n} \ln\left(\theta^2 x_i e^{-\theta x}\right) \\ &= n\ln(\theta^2) + \sum_{i=1}^{n} \ln(x_i) - \theta \sum_{i=1}^{n} x_i \\ &= 2n\ln\theta + \sum_{i=1}^{n} \ln(x_i) - \theta \sum_{i=1}^{n} x_i.\end{aligned}$$

Denna ska maximeras med avseende på θ, så vi deriverar:

$$\ell'(\theta) = \frac{2n}{\theta} - \sum_{i=1}^{n} x_i.$$

Derivatan är noll med negativ andraderivata vid

$$\widehat{\theta} = \frac{2n}{\sum_{i=1}^{n} x_i} = \frac{2}{\overline{x}}.$$

Detta är vår ML–skattning.

5. (5p) Låt N vara en Poissonfördelad slumpvariabel med intensitet $\lambda > 0$, och låt $X_1, X_2, \ldots$ vara oberoende slumpvariabler, likformigt fördelade på $[0, 1]$, som även är oberoende av N. Definiera en slumpvariabel

$$T = \begin{cases} \min\{X_1, \ldots, X_N\} & \text{om } N \geq 1, \\ 1 & \text{om } N = 0. \end{cases}$$

Beräkna $F(t) = \mathbb{P}(T \leq t)$ för alla $0 < t < 1$.

Tips. För alla $x \in \mathbb{R}$ gäller att $\sum_{n=1}^{\infty} x^n/n! = e^x - 1$.

Lösning. Totala sannolikhetslagen ger att

$$\mathbb{P}(T > t) = \sum_{n=0}^{\infty} \mathbb{P}(T > t \mid N = n)\mathbb{P}(N = n).$$

För varje $0 < t < 1$ gäller att

$$\mathbb{P}(T > t \mid N = 0) = 1,$$

så vi kan skriva

$$\mathbb{P}(T > t) = \mathbb{P}(N = 0) + \sum_{n=1}^{\infty} \mathbb{P}(T > t \mid N = n)\mathbb{P}(N = n).$$

Eftersom $N \sim \mathrm{Po}(\lambda)$ gäller att

$$\mathbb{P}(N = n) = \frac{e^{-\lambda}\lambda^n}{n!}.$$

För $n \geq 1$ gäller (eftersom alla X_k är oberoende likformiga)

$$\begin{aligned}
\mathbb{P}(T > t \mid N = n) &= \mathbb{P}(\min\{X_1, \ldots, X_n\} > t) \\
&= \prod_{k=1}^{n} \mathbb{P}(X_k > t) \\
&= (1-t)^n.
\end{aligned}$$

Det följer att

$$\begin{aligned}
\mathbb{P}(T > t) = \mathbb{P}(N = 0) + \sum_{n=1}^{\infty} \mathbb{P}(T > t \mid N = n)\mathbb{P}(N = n) \\
&= e^{-\lambda} + \sum_{n=1}^{\infty} (1-t)^n \frac{e^{-\lambda}\lambda^n}{n!} \\
&= e^{-\lambda} + e^{-\lambda} \sum_{n=1}^{\infty} \frac{((1-t)\lambda)^n}{n!} \\
&= e^{-\lambda} + e^{-\lambda}\left(e^{(1-t)\lambda} - 1\right) \\
&= e^{-\lambda} + e^{-t\lambda} - e^{-\lambda} \\
&= e^{-t\lambda}.
\end{aligned}$$

6. (5p) En kontinuerlig populationsvariabel X är exponentialfördelad med okänd intensitet $\theta > 0$. Alltså har X täthetsfunktion

$$f_X(x) = \theta e^{-\theta x}, \quad x > 0.$$

Ett stickprov av storlek $n = 4$ tas, och vi noterar teststatistikan

$$\min\{x_1, x_2, x_3, x_4\} = 0.1.$$

Ange, baserat på detta, ett 95% konfidensintervall för θ på formen $[a, \infty)$.

Lösning. Vår teststatistika är alltså

$$M = \min\{X_1, X_2, X_3, X_4\}.$$

Målet blir att hitta en funktion $a(M)$ sådan att, med ett slumpmässigt stickprov $X_1, \ldots, X_4$,

$$\mathbb{P}(\theta > a(M)) = 0.95. \tag{1}$$

För att kunna hitta en sådan funktion behöver vi förstå fördelningen hos M. För $x > 0$ gäller att

$$\begin{aligned}\mathbb{P}(M > x) &= \mathbb{P}(X_i > x, \text{ alla } i) \\ &= \mathbb{P}(X_1 > x)^4 \\ &= e^{-4\theta x}.\end{aligned}$$

Genom att sätta

$$x = -\frac{1}{4\theta}\ln(0.95)$$

ser vi att

$$\mathbb{P}\left(M > -\frac{1}{4\theta}\ln(0.95)\right) = 0.95.$$

Genom att flytta runt termer får vi

$$\mathbb{P}\left(\theta > -\frac{1}{4M}\ln(0.95)\right) = 0.95.$$

Alltså är vår funktion

$$a(M) = -\frac{1}{4M}\ln(0.95).$$

Med det givna värdet $M = 3.1$, får vi

$$a(0.1) = -\frac{1}{4\cdot 0.1}\ln(0.95) = 0.128.$$

Svaret blir då $[0.128, \ \infty)$.

7. (**MVE302**, 3+2p) Anta att vi känner till att en populationsvariabel X är normalfördelad med varians 20 och okänt väntevärde μ. Vi vill testa nollhypotesen $H_0 : \mu = 0$ mot alternativhypotesen $H_1 : \mu > 0$, med ett stickprov av storlek $n = 5$. Som teststatistika används stickprovsmedelvärdet $\overline{X}$. Vi kommer fram till att H_0 ska förkastas på signifikansnivå α om $\overline{X} > 2.5$.

(a) Vad är signifikansnivån α?

(b) För vilket värde på μ gäller att testets styrka är 90%?

Lösning.

(a) Vi ska ha

$$\alpha = \mathbb{P}_{H_0}(\overline{X} > 2.5).$$

Under H_0 gäller att $\overline{X} \sim N(0, 20/5) = N(0, 4)$. För beräkningens skull noterar vi att $\overline{X}/2$ är standardnormalfördelad. Alltså är, enligt tabell,

$$\alpha = \mathbb{P}_{H_0}(\overline{X} > 2.5) = \mathbb{P}_{H_0}\left(\frac{\overline{X}}{2} > 1.25\right) = 1 - \Phi(1.25) = 0.11.$$

(b) Styrkan för $\mu > 0$ ges av

$$\beta(\mu) = \mathbb{P}_\mu(\overline{X} > 2.5).$$

Vi har $\overline{X} \sim N(\mu, 4)$, alltså $(\overline{X} - \mu)/2 \sim N(0,1)$, och

$$\beta(\mu) = \mathbb{P}_\mu\left(\frac{\overline{X} - \mu}{2} > \frac{2.5 - \mu}{2}\right) = 1 - \Phi\left(\frac{2.5 - \mu}{2}\right) = \Phi\left(\frac{\mu - 2.5}{2}\right).$$

Enligt tabell är detta 0.9 när $(\mu - 2.5)/2 \approx 1.25$, alltså $\mu \approx 5$.

8. (**MVE302**, 3+2p) En populationsvariabel X är Bernoullifördelad, $X \sim \text{Bin}(1, \lambda)$, och vi är intresserade av parametern $\lambda \in [0,1]$. Vi väljer en Bayesiansk approach, och introducerar en slumpvariabel Λ och gör observationer $x_1, \dots, x_n$.

(a) (3p) Som priorfördelning väljer vi Betafördelningen $\mathcal{B}(4,3)$. Vad är posteriorfördelningen för Λ?

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

Tips. En stokastisk variabel $Y \sim \mathcal{B}(a,b)$ har täthetsfunktion

$$f_Y(y) = \frac{1}{B(a,b)} y^{a-1}(1-y)^{b-1}, \quad 0 < y < 1,$$

där den normaliserande konstanten $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$ ges av

$$B(a,b) = \frac{(a-1)!(b-1)!}{(a+b-1)!} \quad \text{om } a \text{ och } b \text{ är heltal.}$$

Lösning.

(a) Vår prior är

$$f_\Lambda(\lambda) \propto \lambda^3(1-\lambda)^2.$$

Eftersom $X \sim \text{Bin}(1, \lambda)$, gäller att

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

Vi har då posterior

$$\begin{aligned} f_{\Lambda|X}(\lambda \mid x_1, \dots, x_n) &\propto \lambda^3(1-\lambda)^2 \lambda^{\sum x_k} \lambda^{n - \sum x_k} \\ &= \lambda^{3 + \sum x_k}(1-\lambda)^{2 + n - \sum x_k}. \end{aligned}$$

Detta är $B(4 + \sum x_k,\ 3 + n - \sum x_k)$.

(b) Vi behöver ta reda på vad Betafördelningen har för väntevärde. Detta går att slå upp i tabell, men vi kan också beräkna integralen. Generellt gäller att om $X \sim \mathcal{B}(a,b)$, så är (se tipset)

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

Givet en posteriorfördelning $\Lambda \sim \mathcal{B}(4+\sum x_k, 3+n-\sum x_k)$, får vi

$$\hat{\lambda}_{\text{pme}} = \mathbb{E}[\Lambda] = \frac{4+\sum x_k}{(4+\sum x_k)+(3+n-\sum x_k)} = \frac{4+\sum x_k}{7+n}.$$