Teoria statistica delle decisioni - I verifica intermedia – a.a. 2018-2019 - 31 ottobre 2018

Cognome, nome e n. di matricola: SOLUZIONI

8 1. Siano

- $X_1, \ldots, X_n | \theta \sim \text{Ber}(\theta)$ i.i.d., $\theta \in [0, 1]$;
- $\Theta \sim \text{Beta}(\alpha, \beta), \quad \alpha, \beta > 0$;
- $\mathbf{z}_n = (x_1, \ldots, x_n)$ un campione osservato.

2 (a) Determinare la funzione di densità $\pi_\Psi(\psi)$, dove $\Psi = \frac{1}{\Theta}$.

1 (b) Determinare la distribuzione a posteriori di Θ e di Ψ.

2 (c) Determinare $\mathbb{M}(\Psi|\mathbf{z}_n)$ (moda della densità a posteriori di Ψ).

2 (d) Determinare $\mathbb{E}[\Psi|\mathbf{z}_n]$ (valore atteso a posteriori di Ψ) [sugg.: integrare rispetto a θ].

1 (e) Calcolare i valori numerici delle due stime ottenute per $n = 5, \sum_{i=1}^n x_i = 3, \alpha = 2, \beta = 3$.

(a) $\pi_\psi(\psi) = \pi_\Theta(g^{-1}(\psi)) \left| \frac{d}{d\psi} g^{-1}(\psi) \right|$ $\quad g^{-1}(\psi) = \frac{1}{\psi}$; $\left| \frac{d}{d\psi} g^{-1}(\psi) \right| = \frac{1}{\psi^2}$

$\Rightarrow$
$$\pi_\psi(\psi) = \frac{1}{B(\alpha,\beta)} \left(\frac{1}{\psi}\right)^{\alpha-1} \left(1 - \frac{1}{\psi}\right)^{\beta-1} \frac{1}{\psi^2}$$

$\boxed{\text{n.b.}}$ $\theta \in [0,1] \Rightarrow \boxed{\psi \geq 1}$

$$\boxed{\pi_\psi(\psi) = \frac{1}{B(\alpha,\beta)} \left(\frac{1}{\psi}\right)^{\alpha+\beta} (\psi - 1)^{\beta-1} \mathbb{1}_{\psi \geq 1}}$$

(b) $\Theta | \mathbf{z}_n \sim \text{Beta}(\bar\alpha, \bar\beta)$ $\quad \begin{cases} \bar\alpha = \alpha + \sum x_i \\ \bar\beta = \beta + n - \sum x_i \end{cases}$

$$\boxed{\pi_\psi(\psi|\mathbf{z}_n) = \frac{1}{B(\bar\alpha,\bar\beta)} \left(\frac{1}{\psi}\right)^{\bar\alpha+\bar\beta} (\psi-1)^{\bar\beta-1} \mathbb{1}_{\psi \geq 1}}$$

(c) $\mathbb{M}(\psi|\mathbf{z}) = \underset{\psi \in [1,+\infty)}{\arg\max} \; \ln \pi_\psi(\psi|\mathbf{z}_n)$

$$\ln \pi_\psi(\psi|\mathbf{z}_n) = c + -(\bar\alpha+\bar\beta)\ln\psi + (\bar\beta-1)\ln(\psi-1)$$

$$\frac{d}{d\psi} \;'' \quad = -\frac{\bar\alpha+\bar\beta}{\psi} + \frac{\bar\beta-1}{\psi-1} = 0$$

$$-(\bar\alpha+\bar\beta)(\psi-1) + (\bar\beta-1)\psi = 0$$

$$+\bar\alpha + \bar\beta - \bar\alpha\psi - \cancel{\bar\beta\psi} + \cancel{\bar\beta\psi} - \psi = 0 \Rightarrow \boxed{\tilde\psi = \frac{\bar\alpha+\bar\beta}{\bar\alpha+1}} = \frac{\alpha+\beta+n}{\alpha+\sum x_i+1}$$

(d) $\mathbb{E}\left[\psi \mid \underline{x}_n\right] = \mathbb{E}\left[\frac{1}{\theta} \mid \underline{x}_n\right] = \int_0^1 \frac{1}{\theta}\, \pi_\theta(\theta \mid \underline{x}_n)\, d\theta$

$$= \frac{1}{B(\bar{\alpha},\bar{\beta})} \int_0^1 \frac{1}{\theta}\, \theta^{\bar{\alpha}-1} (1-\theta)^{\bar{\beta}-1} d\theta$$

$$= \frac{1}{B(\bar{\alpha},\bar{\beta})} \int_0^1 \theta^{(\bar{\alpha}-1)-1} (1-\theta)^{\bar{\beta}-1} d\theta = \frac{B(\bar{\alpha}-1,\bar{\beta})}{B(\bar{\alpha},\bar{\beta})}$$

$$= \frac{\dfrac{\Gamma(\bar{\alpha}-1)\cancel{\Gamma(\bar{\beta})}}{\Gamma(\bar{\alpha}+\bar{\beta}-1)}}{\dfrac{\Gamma(\bar{\alpha})\cancel{\Gamma(\bar{\beta})}}{\Gamma(\bar{\alpha}+\bar{\beta})}} = \frac{(\bar{\alpha}+\bar{\beta}-1)\cancel{\Gamma(\bar{\alpha}+\bar{\beta}-1)}\cancel{\Gamma(\bar{\alpha}-1)}}{\cancel{\Gamma(\bar{\alpha}+\bar{\beta}-1)}(\bar{\alpha}-1)\cancel{\Gamma(\bar{\alpha}-1)}} = \boxed{\frac{\bar{\alpha}+\bar{\beta}-1}{\bar{\alpha}-1}}$$

$$\begin{cases} \bar{\alpha} = \alpha + \sum x_i = 2+3 = 5 \\ \bar{\beta} = \beta + \underbrace{n - \sum x_i}_{2} = 3+2 = 5 \end{cases}$$

(e) $\mathbb{M}(\psi \mid \underline{x}_n) = \frac{10}{6}$

$\mathbb{E}(\psi \mid \underline{x}_n) = \frac{9}{4}$

8 2. Siano

- $X_1, \ldots, X_n|\theta$ i.i.d.;
- $p_\theta(x) = \theta x^{\theta-1}, x \in [0,1],\ \theta > 0$.

Determinare:

2 (a) $\pi^J(\theta)$ (distribuzione a priori di Jeffreys);
2 (b) $\pi^J(\theta|\mathbf{z}_n)$ (a meno della costante di normalizzazione);
2 (c) $\tilde\theta(\mathbf{z}_n)$, $I_B(\tilde\theta)$ e approssimazione normale della distribuzione a posteriori;
1 (d) espressione esplicita dell'insieme $\tilde C$, HPD di livello $1-\gamma$ per θ basato sull'approssimazione normale;
1 (e) valori numerici di $\tilde\theta(\mathbf{z}_n)$, $I_B(\tilde\theta)$ e $\tilde C$ che si ottengono in corrispondenza di un campione osservato con $n = 10$ e $\sum_{i=1}^n \ln x_i = -3$ e $\gamma = 0.05$.

(a) $\pi^J(\theta) = \sqrt{I_n(\theta)}$ $\quad I_n(\theta) = -n\,\mathbb{E}\left[\frac{d^2}{d\theta^2}\ln p_\theta(x)\right]$

$\ln p_\theta(x) = \ln\theta + (\theta-1)\ln x$

$\frac{d}{d\theta}$ " $= \frac{1}{\theta} + \ln x$

$\frac{d^2}{d\theta^2}$ " $= -\frac{1}{\theta^2}$ $\Rightarrow$ $\boxed{\pi^J_\theta(\theta) \propto \frac{1}{\theta}}$

(b) $\pi(\theta|\mathbf{z}_n) = c \cdot \theta^n \left(\prod^n x_i\right)^\theta \cdot \frac{1}{\theta}$

$= c \cdot \theta^{n-1} \exp\{-t\theta\}$ $\quad \boxed{t = -\ln\prod_{i=1}^n x_i}$

$\Rightarrow$ $\boxed{\begin{array}{l}\theta|\mathbf{z}_n \sim Ga(n, t) \\ \text{con } t = -\sum^n \ln x_i\end{array}}$

(c) $\tilde\theta = \frac{\alpha-1}{\beta}$ fun $\boxed{\tilde\theta = \frac{n-1}{t}}$

CALCOLO di $I_B = -\frac{d^2}{d\theta^2}\ln\pi(\theta|\mathbf{z}_n)\Big|_{\theta=\tilde\theta}$

$\ln\pi(\theta|\mathbf{z}_n) = c' + (n-1)\ln\theta - t\theta$

$\frac{d}{d\theta}$ " $= 0 + \frac{n-1}{\theta} - t \Rightarrow \tilde\theta = \frac{n-1}{t}$

$\frac{d}{d\theta}$ " $= -\frac{n-1}{\theta^2}$

$\Rightarrow$

$I_B(\tilde\theta) = \frac{n-1}{\tilde\theta^2} \Rightarrow$

$\boxed{I_B(\tilde\theta) = \frac{t^2}{n-1}}$

$$\Rightarrow \boxed{\Theta \mid x_n \overset{\cdot}{\sim} N\left(\frac{n-1}{t}, \frac{n-1}{t^2}\right)}$$

(d) $$\tilde{q} = \frac{n-1}{t} \pm z_{1-\frac{\alpha}{2}} \frac{\sqrt{n-1}}{t}$$

$$\left(\begin{array}{l} \text{da} \\ \tilde{\theta} \pm z_{1-\frac{\alpha}{2}} \sqrt{I_B^{-1}(\tilde{\theta})} \end{array} \right)$$

(e) $$\tilde{q} = \frac{9}{3} \pm 2 \cdot \frac{\sqrt{9}}{3}$$

$$= 3 \pm 2 = [1, 5]$$

7 3. Siano:

- $X_1, \ldots, X_n | \theta \sim \text{Pois}(\theta)$ i.i.d., $\theta > 0$;
- $H_0 : \theta = \theta_0$ vs. $H_1 : \theta = \theta_1$;
- $\theta_0 > \theta_1 > 0$;
- $\mathbf{z}_n$ un campione osservato.

1 (a) Determinare l'espressione di $B_{01}(\mathbf{z}_n)$.

2 (b) Verificare che, preso un valore $k > 0$, si ha

$$B_{01}(\mathbf{z}_n) > k \quad \Leftrightarrow \quad \overline{x}_n > \xi(k),$$

dove $\xi(k)$ è un valore che dipende da k.

1 (c) Determinare l'espressione di $\xi(k)$ (funzione di k, n, θ_0, θ_1).

2 (d) Supponendo ora che $\Theta \sim \text{Ga}(\alpha, \beta)$, calcolare il valore atteso predittivo a priori di

$$Y = \ln B_{01}(\mathbf{Z}_n).$$

[Sugg.: utilizzare la distribuzione predittiva di $\sum_{i=1}^n X_i$].

1 (e) Determinare il valore numerico di $\mathbb{E}[Y]$ in corrispondenza dei seguenti valori: $\alpha = \beta = 2$, $n = 10$, $\theta_0 = \text{e}$, $\theta_1 = 1$ (e indica il numero di Nepero).

(a) $B_{01}(z_n) = B_{01}(t)$ con $t = \sum x_i$ $\qquad \sum x_i | \theta \sim \text{Poiss}(n\theta)$

$$\Rightarrow \boxed{B_{01}(t) = \left(\frac{\theta_0}{\theta_1}\right)^t \exp\{-n(\theta_0 - \theta_1)\}}$$

(b) $B_{01}(t) \geq k \Leftrightarrow \ln B_{01}(t) \geq \ln k \Leftrightarrow t \underbrace{\ln \underbrace{\frac{\theta_0}{\theta_1}}_{>1}}_{>0} \geq \ln k + n(\theta_0 - \theta_1)$

$$\Leftrightarrow t \geq \frac{\ln k + n(\theta_0 - \theta_1)}{\ln \theta_0/\theta_1}$$

(c)
$$\Leftrightarrow \boxed{\bar{x}_n \geq \frac{1}{n} \cdot \frac{\ln k + n(\theta_0 - \theta_1)}{\ln \theta_0/\theta_1} = \xi(k)}$$

(d) $Y = \ln B_{01}(t) = t \ln\left(\frac{\theta_0}{\theta_1}\right) - n(\theta_0 - \theta_1)$

$\mathbb{E}[Y] = \mathbb{E}[T] \ln\left(\frac{\theta_0}{\theta_1}\right) - n(\theta_0 - \theta_1)$

$\mathbb{E}[\sum X_i] = n\,\mathbb{E}(X_i)$

$\mathbb{E}(X_i) = \mathbb{E}_\Theta(\mathbb{E}(X_i|\Theta))$

$= \mathbb{E}_\Theta(\Theta) = \frac{\alpha}{\beta}$

$$= n \frac{\alpha}{\beta} \ln\left(\frac{\theta_0}{\theta_1}\right) - n(\theta_0 - \theta_1)$$

$$= \boxed{n\left[\frac{\alpha}{\beta} \ln\left(\frac{\theta_0}{\theta_1}\right) - (\theta_0 - \theta_1)\right]}$$

(e) $\mathbb{E}[Y] = 10\left[\frac{2}{2} \ln\left(\frac{e}{1}\right) - (e - 1)\right] = 10(1 - e + 1) = 10(2 - e)$

7 4. I seguenti valori numerici indicano la concentrazione di acido lattico (misurata in una opportuna unità di misura) di alcuni campioni di confezioni di formaggio di una determinata marca:

$$1.5,\ 1,\ 0.5,\ 2,\ 2,\ 1,\ 2.5,\ 1,\ 0.5,\ 1.$$

I ricercatori ipotizzano che i valori campionari osservati siano realizzazioni di v.a. che tendono a concentrarsi intorno al valore vero incognito di concentrazione di acido lattico con una varianza pari a 1/9.

In uno studio precedente, basato su soli tre campioni di formaggio, la concentrazione media di acido lattico era risultata pari a 2.

Siamo interessati a determinare la distribuzione di probabilità del valore vero di concentrazione di acido lattico in una confezione di formaggio.

3 (a) Formalizzare quanto sopra in termini di un problema di inferenza bayesiana [sugg.: associare le quantità in gioco a variabili, parametro, dimensione campionaria, iperparametri...].

(b) Scegliere le distribuzioni di probabilità per variabili e parametro incognito.

4 (c) Determinare una stima puntuale della concentrazione di acido lattico in una confezione, utilizzando tutte le informazioni disponibili [sugg.: i valori numerici possono essere espressi in frazioni].

(d) Determinare una stima intervallare al 95% della concentrazione di acido lattico sulla base di tutte le informazioni disponibili [sugg.: i valori numerici possono essere espressi in frazioni].

(e) Sappiamo che, in una confezione formaggio, la concentrazione di acido lattico non debba superare il valore 2. Calcolare la probabilità che ciò avvenga, sulla base dell'informazione disponibile [sugg.: i valori numerici possono essere espressi in frazioni].

(a) X_i = Concentr di acido lattico in una confezione di formaggio

θ = Concentr. media incognita di acido lattico nella "popolazione" delle confezioni di formaggio.

$n = 10 \qquad \sigma^2 = 1/9$

(b) $X_i|\theta \sim N(\theta, \sigma^2)$, $\Theta \sim N\left(\mu_0, \frac{\sigma^2}{n_0}\right) \Rightarrow \begin{cases} \bar{X}_n|\theta \sim N\left(\theta, \frac{\sigma^2}{n}\right) \\ \Theta \sim N\left(\mu_0, \frac{\sigma^2}{n_0}\right) \end{cases}$

Con $\begin{cases} \mu_0 = 2 \\ n_0 = 3 \end{cases} \Rightarrow \begin{cases} \bar{X}_n|\theta \sim N\left(\theta, \frac{1}{90}\right) \\ \Theta \sim N\left(2, \frac{1}{27}\right) \end{cases}$

(c) $$\mathbb{E}(\Theta|\bar{x}_n) = \frac{\mu_0 n_0 + \bar{x}_n n}{n_0 + n} = \frac{(2)\cdot(3) + (1.3)10}{3 + 10} = \boxed{\frac{19}{13}}$$

(d) $C = \frac{n_0\mu_0 + n\bar{x}_n}{n_0 + n} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n_0+n}}$

$$\frac{\sigma}{\sqrt{n_0+n}} = \frac{1}{3\sqrt{3+10}} = \frac{1}{3\sqrt{13}}$$

$$\Rightarrow \boxed{C = \frac{19}{13} \pm 1.96 \frac{1}{3\sqrt{13}}}$$

(e) $P[\Theta > 2 \mid x_n] = 1 - \Phi\left(\frac{2 - 19/13}{1/3\sqrt{13}}\right) = 1 - \Phi\left(3\sqrt{13}\left(2 - \frac{19}{13}\right)\right)$

$$= 1 - \Phi\left(3\sqrt{13}\ \frac{7}{13}\right) = 1 - \Phi\left(\frac{21\sqrt{13}}{13}\right)$$

$=$