Teoria statistica delle decisioni - I verifica intermedia – a.a. ~~2019-2020~~ 20-21 - 09 novembre 2021 - **PARTE A**

Cognome, nome e n. di matricola: ______________________________

IMPORTANTE: vengono valutate solo le soluzioni scritte su questi fogli

\- - -

1. Sia $\mathbf{Z}_n = (X_1, \ldots, X_n)$ un campione casuale (i.i.d.) con ciascuna X_i avente funzione di massa di probabilità

$$p_\theta^X(x_i) = \binom{r + x_i - 1}{x_i} \theta^r (1-\theta)^{x_i}, \qquad x_i = 0, 1, 2, \ldots,$$

dove r indica un numero intero noto.

$\mathbb{V}_\theta[X] = \frac{\theta r}{(1-\theta)^2}$

$\mathbb{E}_\theta[X] = \frac{\theta r}{1-\theta}$

(a) Determinare la stima di massima verosimiglianza di θ

$\ell(\theta) = c \cdot \theta^{rn} (1-\theta)^{S_n}$

$\ln \ell(\theta) = c + rn \ln \theta + S_n \ln(1-\theta)$

$\frac{d^2}{d\theta^2} \; '' = -\frac{rn}{\theta^2} - \frac{S_n}{(1-\theta)^2} < 0 \quad \forall \theta$ $\Rightarrow$ OK

$\frac{d}{d\theta} \; '' = \frac{rn}{\theta} + \frac{S_n}{1-\theta}(-1) = 0$ per $rn(1-\theta) - S_n\theta = 0$

$rn - rn\theta - S_n\theta = 0$

$\boxed{\frac{rn}{rn + S_n} = \hat{\theta}_{MV}}$

(b) Verificare chela famiglia delle densità Beta(α, β) è coniugata al modello in esame e determinare $\pi(\theta|\mathbf{z}_n)$.

$\pi(\theta|z_n) = c \; \pi(\theta) \cdot \ell(\theta) \Rightarrow \propto \theta^{nr} (1-\theta)^{S_n} \cdot \theta^{\alpha-1} (1-\theta)^{\beta-1}$

$\propto \theta^{(\alpha + nr) - 1} (1-\theta)^{(\beta + S_n) - 1} = \text{Beta}(\theta | \bar{\alpha}, \bar{\beta})$

$\begin{cases} \bar{\alpha} = \alpha + nr \\ \bar{\beta} = \beta + S_n \end{cases}$

(c) Determinare l'espressione di $\mathbb{E}[\Theta|\mathbf{z}_n]$

$\mathbb{E}(\theta|z_n) = \frac{\bar{\alpha}}{\bar{\alpha} + \bar{\beta}} = \frac{\alpha + nr}{\alpha + nr + \beta + S_n}$

(d) Sia $\xi(\mathbf{Z}_n) = \mathbb{E}[\Theta|\mathbf{z}_n]^{-1}$. Determinare l'espressione di $\mathbb{V}_\theta[\xi(\mathbf{Z}_n)]$ (varianza della distribuzione *campionaria* di $\xi(\mathbf{Z}_n)$).

$\xi(Z_n) = \frac{\alpha + nr + \beta + S_n}{\alpha + nr} = \frac{S_n + A}{B}$

$\mathbb{V}_\theta[\xi(Z_n)] = \frac{1}{B^2} \mathbb{V}_\theta[S_n] = \frac{n}{B^2} \mathbb{V}_\theta[X] = \frac{n}{B^2} \frac{\theta r}{(1-\theta)^2}$

2. Sia $\mathbf{Z}_n = (X_1, \ldots, X_n)$ un campione casuale (i.i.d.) con ciascuna X_i avente funzione di densità di probabilità

$$p_\theta^X(x_i) = \theta x_i^{\theta-1}, \qquad x_i \in [0,1], \quad \theta > 0.$$

Verificare che, se $\Theta \sim \text{Gamma}(\alpha, \texttt{rate} = \beta)$, allora $\Theta|\mathbf{z}_n \sim \text{Gamma}(\overline{\alpha}, \texttt{rate} = \overline{\beta})$, con $\overline{\alpha} = n + \alpha$, $\overline{\beta} = \beta - \ln \prod_{i=1}^n x_i$.

$$\ell(\theta) = \theta^n \left(\Pi x_i\right)^\theta \qquad T = \Pi x_i$$

$$\pi(\theta|z_n) = c \cdot \theta^{\alpha-1} e^{-\beta\theta} \cdot \theta^n \exp\{\underbrace{\ln T^\theta}_{= \theta \ln T}\}$$

$$= c \cdot \theta^{(\alpha+n)-1} \exp\{-\theta(\beta - \ln T)\}$$

nucleo di $Ga(\alpha+n, \beta - \ln T)$

3. Sia $X_i|\theta \sim N(\theta, 1)$ i.i.d. e $\Theta \sim N(\mu_0, n_0^{-1})$. Determinare la probabilià predittiva a priori dell'evento $\mathbb{P}[\overline{\mu}_p(\mathbf{Z}_n) > 0]$, dove $\overline{\mu}_p(\mathbf{Z}_n) = \mathbb{E}[\Theta|\mathbf{z}_n]$.

[**N.B.** - Esprimere il risultato in funzione di $\Phi(\cdot)$, funzione di ripartizione della v.a. $N(0,1)$.]

$$X_i|\theta \sim N(\theta, 1)$$

$$\begin{cases} \bar{X}_n|\theta \sim N(\theta, \sigma^2/n) \\ \Theta \sim N(\mu_0, \sigma^2/n_0) \end{cases} \Rightarrow \theta|z_n \sim N(\bar{\mu}_p, \bar{\sigma}_p^2)$$

$$\bar{\mu}_p = \frac{n_0\mu_0 + n\bar{X}_n}{n_0 + n}$$

$$P[\bar{\mu}_p > 0] = P\left[\frac{n_0\mu_0 + n\bar{X}_n}{n_0 + n} > 0\right] =$$

$$= P[n\bar{X}_n > -n_0\mu_0] =$$

$$= P\left[\bar{X}_n > -\frac{n_0}{n}\mu_0\right] = *$$

$$\bar{X}_n \sim N\left(\mu_0, \sigma^2\left(\frac{1}{n} + \frac{1}{n_0}\right)\right) \Rightarrow$$

$$* = 1 - \Phi\left(\frac{-\frac{n_0}{n}\mu_0 - \mu_0}{\sigma\left(\frac{1}{n} + \frac{1}{n_0}\right)^{1/2}}\right)$$

Teoria statistica delle decisioni - I verifica intermedia – a.a. 2019-2020 - 09 novembre 2021 - **PARTE B**

Cognome, nome e n. di matricola: ____________________

1. Si suppone che il numero di casi di una determinata malattia in 10 unità di territorio possano essere assimilati a un campione casuale da una popolazione di Poisson di parametro θ. I valori osservati sono i seguenti: (2, 0, 2, 4, 1, 6, 1, 3, 3, 2). Un esperto propone per Θ una distribuzione a priori Gamma tale con valore atteso uguale a 2 e varianza uguale a 1 (parametrizzazione `rate`).

Determinare:

(a) i valori degli iperparametri α e β di $\pi(\theta)$;

$$\mathbb{E}[\Theta] = \frac{\alpha}{\beta} = 2 \quad e \quad \mathbb{V}[\Theta] = \frac{\alpha}{\beta^2} = 1 \Rightarrow \frac{2}{\beta} = 1 \Rightarrow \begin{cases} \beta = 2 \\ \alpha = 4 \end{cases}$$

$$\boxed{Ga(4,2)}$$

(b) la stima di massima verosimiglianza e una stima puntuale bayesiana di θ;

$$\hat{\theta}_{MV} = \bar{x}_n \Rightarrow \hat{\theta}_{MV} = \frac{24}{10} = \frac{12}{5}$$

$$\begin{cases} \bar{\alpha} = \alpha + s_n = 4+24=28 \\ \bar{\beta} = \beta + n = 2+10=12 \end{cases}$$

$$\hat{\theta}_B = \frac{\bar{\alpha}}{\bar{\beta}} = \frac{\alpha + s_n}{\beta + n} = \frac{4+24}{2+10} = \boxed{\frac{28}{12} = \frac{7}{3}}$$

(c) l'insieme di credibilità per θ che si ottiene utilizzando la formula $\mathbb{E}[\Theta|\mathbf{z}_n] \pm 2\sqrt{\mathbb{V}[\Theta|\mathbf{z}_n]}$;

$$\tilde{C} = \frac{\bar{\alpha}}{\bar{\beta}} \pm 2\sqrt{\frac{\bar{\alpha}}{\bar{\beta}^2}} = \frac{7}{3} \pm 2\sqrt{\frac{28}{12^2}} = \frac{7}{3} \pm \frac{2}{12}\sqrt{28}$$

$$= \frac{7}{3} \pm \frac{1}{6} 2\sqrt{7} = \boxed{\frac{7}{3} \pm \frac{\sqrt{7}}{3}}$$

(d) L'approssimazione normale della distribuzione a posteriori di Θ e l'intervallo di confidenza asintotico di livello 0.95 (per il quale $z_{1-\frac{\alpha}{2}} = 1.96$);

$$\tilde{\theta} = \frac{\bar{\alpha}-1}{\bar{\beta}} \qquad I_B^{-1} = \frac{\bar{\alpha}-1}{\bar{\beta}^2} \Rightarrow \theta|z_n \overset{\cdot}{\sim} N\left(\frac{\bar{\alpha}-1}{\bar{\beta}}, \frac{\bar{\alpha}-1}{\bar{\beta}^2}\right)$$

$$\Rightarrow \frac{27}{12} \pm 1.96 \frac{\sqrt{27}}{12} = \tilde{C} = \boxed{\frac{9}{4} \pm 1.96 \frac{\sqrt{3}}{4}} \overset{\circ}{\sim} N\left(\frac{27}{12}, \frac{27}{12^2}\right)$$

(e) I valori di $\mathbb{E}[Y_m|\mathbf{z}_n]$ e $\mathbb{V}[Y_m|\mathbf{z}_n]$, (valore atteso e varianza predittivi a posteriori) dove Y_m indica la somma di $m = 5$ osservazioni future.

$$\mathbb{E}[Y_m | z_n] = \mathbb{E}_{\theta|z_n}\left[\mathbb{E}_{Y_m|\theta}(Y_m|\theta)\right] = \mathbb{E}_{\theta|z_n}[5\Theta] = 5\,\mathbb{E}(\theta|z_n) = 5\frac{\bar{\alpha}}{\bar{\beta}}$$

$$= 5 \cdot \frac{7}{3} = \frac{35}{3}$$

$$\mathbb{V}[Y_m|z_n] = \mathbb{E}_{\theta|z_n}[m\Theta] + \mathbb{V}_{\theta|z_n}[m\Theta] = m\frac{\bar{\alpha}}{\bar{\beta}} + m^2\frac{\bar{\alpha}}{\bar{\beta}^2} = 5\cdot\frac{7}{3} + 25\cdot\frac{28}{12^2} = \dots$$