Teoria statistica delle decisioni – a.a. 24-25 - Esonero 1 - 06-12-2024 [**SI CORREGGONO SOLO QUESTI FOGLI**]

Cognome, nome e n. di matricola: ______________________________

Problema A. Sia S_Δ la porzione del primo quadrante cartesiano limitato inferiormente dalla retta $y = 2$, superiormente dalla retta $y = 4$ e lateralmente dagli archi della parabola γ di equazione $y = (x-2)^2$, ovvero l'insieme convesso chiuso (frontiera inclusa) di vertici A, C, D e E con: A, C punti di intersezione di γ con $y = 2$ (con A avente ascissa inferiore a C) e D, E punti di intersezione di γ con $y = 4$ (con E avente ascissa inferiore a D).

1. (a) Determinare le coordinate di A, C, D e E
   (b) Rappresentare graficamente l'insieme S_Δ.

> **Soluzione**
>
> (a) $A = (2-\sqrt{2}, 2);\ C = (2+\sqrt{2});\ D = (4,4);\ E = (0,4)$
> A e C si trovano come soluzioni di $\begin{cases} y = (x-2)^2 \\ y = 2 \end{cases}$; E e D si trovano come soluzioni di $\begin{cases} y = (x-2)^2 \\ y = 4 \end{cases}$.
>
> (b) La parabola ha vertice in $V = (2, 0)$. Il grafico della parabola e quindi S_Δ si ottengono facilmente analiticamente. Verifica con `R` usando il seguente codice:
>
> ```
> y=function(x){(x-2)^2}
> curve(y(x),from = 0, to=4,xlab="x", ylab="y",ylim=c(0,5))
> abline(h=2)
> abline(h=4)
> abline(0,1,lty=2)
> text(x=0.5,y=1.8,labels=expression(A))
> text(x=3.5,y=1.8,labels=expression(C))
> text(x=4,y=4.5,labels=expression(D))
> text(x=0,y=4.5,labels=expression(E))
> text(x=2,y=1.6,labels=expression(P))
> ```

2. (a) Determinare gli insiemi S_{Δ^+}, S_{B^+} e S_B e rappresentarli graficamente.
   (b) Determinare $S_{\Delta_m^\star}$, insieme delle decisioni ottime rispetto al criterio del minimax, e rappresentarlo graficamente.

> **Soluzione**
>
> (a) $S_{\Delta^+} = S_{B^+} =$ arco AE estremi inclusi; $S_B =$ arco AE unito al segmento $\overline{AC}$ estremi inclusi
>
> (b) La bisettrice $y = x$ interseca la frontiera inferiore di S_Δ nel punto $P = (2, 2)$. Pertanto $S_{\Delta_m^\star} = \overline{AP}$ (il solo ottimo minimax ammissibile è quindi il punto A).

3. (a) Determinare il valore del criterio del valore atteso K_{va} nel punto-decisione A di S_Δ in corrispondenza della generica distribuzione $(p, 1-p)$, con $p \in (0, 1)$ assegnata ai due stati di natura.
   (b) Calcolare il valore di $p \in (0, 1)$ tale che il valore del criterio K_{va} in A sia pari a 1.

> **Soluzione**
>
> (a) Il valore di K_{va} in $A = (2-\sqrt{2}, 2)$ per la distribuzione proposta è $(2-\sqrt{2})p + 2(1-p) = 2-\sqrt{2}p$.
>
> (b) K_{va} in A vale 1 $\iff 2-\sqrt{2}p = 1$ ovvero se $p = \frac{\sqrt{2}}{2}$.

Problema B. Siano $X_1, \ldots, X_n$ le risposte aleatorie a un trattamento medico somministrato a n pazienti. Per ciascun paziente le risposte sono dicotomiche: il trattamento può avere o non avere successo. Sia θ la *probabilità di successo* nella popolazione da cui provengono i dati, ovvero la probabilità che, nella prova i-esima, il trattamento funzioni, $i = 1, \ldots, n$.
Si dispone di informazione pre-sperimentale e sperimentale sulla *probabilità di successo* oggetto di interesse.
`Informazione sperimentale`: su 10 osservazioni sperimentali, le risposte positive al trattamento sono 8.
`Informazione pre-sperimentale`: basandoci su dati storici è lecito supporre che la *probabilità di successo* abbia distribuzione con valore atteso pari a 1/2 e con *prior sample size* pari a $n_0 = 5$.

1. (a) Formalizzare opportunamente il problema per mezzo di un modello bayesiano, ovvero: indicare quali sono le distribuzioni di $X_i|\theta$ e di Θ da utilizzare e i valori da associare agli iperparametri della densità a priori.
(b) Scrivere le espressioni generiche (formule) e specifiche (valori numerici) della distribuzione a posteriori di Θ.

> **Soluzione**
>
> (a) $X_i|\theta \sim \text{Ber}(\theta)$ e $\Theta \sim \text{Beta}(\alpha, \beta)$. Dalle informazioni sperimentali sappiamo che $s_n = \sum_{i=1}^n x_i = 8$ e $n = 10$. Ricordando che nel modello considerato la *prior sample size* è $n_0 = \alpha + \beta$ e che $\mathbb{E}[\Theta] = \frac{\alpha}{\alpha+\beta}$,
> dal sistema $\begin{cases} \frac{\alpha}{\alpha+\beta} = \frac{1}{2} \\ \alpha + \beta = 5 \end{cases}$ otteniamo $\alpha = \beta = \frac{5}{2}$. Pertanto
> $\overline{\alpha} = \alpha + s_n = \frac{5}{2} + 8 = \frac{21}{2}$ e
> $\overline{\beta} = \beta + n - s_n = \frac{9}{2}$.
>
> (b)

2. (a) Fornire una stima puntuale $\widehat{\theta}_B$ per θ basata sulla distribuzione a posteriori di tale parametro [formule generali e valori numerici in frazioni].
(b) Determinare l'insieme di stima C per θ che si ottiene con la formula $\widehat{\theta}_B \pm 2\sqrt{\frac{\widehat{\theta}_B(1-\widehat{\theta}_B)}{n}}$ [formule generali e valori numerici in frazioni].

> **Soluzione**
>
> (a) $\widehat{\theta}_B = \mathbb{E}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}}{\overline{\alpha}+\overline{\beta}} = \frac{21}{30} = \frac{7}{10} = 0.7$
>
> (b) $C = \frac{\overline{\alpha}}{\overline{\alpha}+\overline{\beta}} \pm 2\sqrt{\frac{\overline{\alpha}\overline{\beta}}{(\overline{\alpha}+\overline{\beta})^2}} = 0.7 \pm 2\sqrt{\frac{0.7\times 0.3}{10}} = \ldots$

3. (a) Determinare la distribuzione a posteriori di Θ che si ottiene ponendo $n_0 = 0$ (caso non informativo).
(b) Per questo caso non informativo, determinare stima puntuale e intervallare di θ. [formule generali e valori numerici in frazioni].

> **Soluzione**
>
> (a) $n_0 = \alpha + \beta = 0 \implies \alpha = \beta = 0$. Pertanto
> $\overline{\alpha} = s_n = 8$
> e
> $\overline{\beta} = n - s_n = 2$.
> Abbiamo quindi che $\Theta|\mathbf{z}_n \sim \text{Beta}(s_n, n - s_n) = \text{Beta}(8, 2)$.
>
> (b) $\widehat{\theta}_B = \frac{s_n}{n} = \overline{x}_n = \frac{8}{10} = 0.8$ e $C = \overline{x}_n \pm 2\sqrt{\frac{\overline{x}_n(1-\overline{x}_n)}{n}} = 0.8 \pm 2\sqrt{\frac{0.8\times 0.2}{10}} = \ldots$

Problema C. Siano $X_1, \ldots, X_n|\theta \sim \text{Pois}(\theta)$ iid con ciascuna X_i avente funzione di massa di probabilità $p_\theta^X(x_i) = \frac{e^{-\theta}\theta^{x_i}}{x_i!}$, $x = 0, 1, 2, \ldots$ e $\theta \in \mathbb{R}^+$.

1. (a) Determinare l'espressione di $p_\theta(\mathbf{z}_n)$.
(b) Per il confronto tra $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$, determinare l'espressione di $\lambda_{01}(\mathbf{z}_n)$ in funzione di $t_n = \sum_{i=1}^n X_i$ (realizzazione di $T_n = \sum_{i=1}^n X_i$).

> **Soluzione**
>
> (a) $p_\theta(\mathbf{z}_n) = \prod_{i=1}^n p_\theta^X(x_i) = e^{-n\theta}\frac{\theta^{t_n}}{\prod_{i=1}^n x_i!}$, con $t_n = T_n(\mathbf{z}_n) = \sum_{i=1}^n x_i$.
>
> (b) $\lambda_{01}(\mathbf{z}_n) = \frac{p_{\theta_0}(\mathbf{z}_n)}{p_{\theta_1}(\mathbf{z}_n)} = \left(\frac{\theta_0}{\theta_1}\right)^{t_n} e^{-(\theta_0-\theta_1)}$

2. (a) Verificare che $\ln \lambda_{01}(\mathbf{Z}_n)$ può essere scritto nella forma $c_1 T_n + c_2$ e specificare a cosa corrispondono c_1 e c_2.
(b) Determinare le espressione di $\mathbb{E}_{\theta_0}[\ln \lambda_{01}(\mathbf{Z}_n)]$ e $\mathbb{V}_{\theta_0}[\ln \lambda_{01}(\mathbf{Z}_n)]$. [**Sugg.**: usare valore atteso e varianza di $X_i|\theta_0$].

> **Soluzione**
>
> (a) $\ln \lambda_{01}(\mathbf{Z}_n) = T_n \ln(\theta_0/\theta_1) - n(\theta_0 - \theta_1) = c_1 T_n + c_2$ con $c_1 = \ln(\theta_0/\theta_1)$ e $c_2 = -n(\theta_0 - \theta_1)$.
>
> (b) Per le proprietà di valore atteso e varianza di somme di v.a. iid (in questo caso Poisson), si ha
> $\mathbb{E}_{\theta_0}[\ln \lambda_{01}(\mathbf{Z}_n)] = c_1 \mathbb{E}_{\theta_0}[\sum_{i=1}^n X_i] + c_2 = n c_1 \theta_0 + c_2$
> $\mathbb{V}_{\theta_0}[\ln \lambda_{01}(\mathbf{Z}_n)] = c_1^2 \mathbb{V}_{\theta_0}[\sum_{i=1}^n X_i] = n c_1^2 \theta_0$

3. (a) Determinare l'espressione di λ_{01} che si ottiene ponendo $n = 1$, $\theta_0 = 1$ e $\theta_1 = 2$.
(b) Verificare che $\lambda_{01}(x) > 1 \iff x < k$ e determinare il valore di k.

> **Soluzione**
>
> (a) Per $n = 1$ abbiamo che $T_n(\mathbf{z}_n) = x$ (una sola osservazione). Pertanto $\lambda_{01}(x) = \left(\frac{1}{2}\right)^x e$
>
> (b) $\lambda_{01}(x) > 1 \iff \left(\frac{1}{2}\right)^x e > 1 \iff \ldots \iff x < \frac{1}{\ln 2}$

Problema D. Siano $X_1, \ldots, X_n|\theta \sim \text{EN}(\theta)$ un campione iid con ciascuna X_i avente funzione di densità $p_\theta^X(x) = \theta e^{-x\theta}$, $x \geq 0$ e $\theta > 0$. Sia $\pi(\theta) = c \times \frac{1}{\theta^2}$, $\quad c > 0$ e $\mathbf{z}_n$ un generico campione osservato. Determinare quanto segue.

1. (a) Funzione di verosimiglianza $\ell(\theta; \mathbf{z}_n)$ e stima di massima verosimiglianza $d_{mv}(\mathbf{z}_n)$.
   (b) Distribuzione a posteriori di Θ e espressione di $\mathbb{E}[\Theta|\mathbf{z}_n]$.

> **Soluzione**
>
> (a) $\ell(\theta; \mathbf{z}_n) = \theta^n \, e^{-s_n\theta}$, $\theta > 0$. Si verifica che $d_{mv} = \frac{1}{\bar{x}_n}$.
>
> (b) $\pi(\theta|\mathbf{z}_n) = c \times \frac{1}{\theta^2} \times \theta^n \, e^{-s_n\theta} = c \times \theta^{(n-1)-1} e^{-s_n\theta}$, $\theta > 0$.
> Quindi $\Theta|\mathbf{z}_n \sim \text{Ga}(\overline{\alpha}, \texttt{rate} = \overline{\beta})$ con $\overline{\alpha} = n - 1$ e $\overline{\beta} = s_n$. Pertanto $\mathbb{E}[\Theta|\mathbf{z}_n] = \frac{\overline{\alpha}}{\overline{\beta}} = \frac{n-1}{s_n}$.

2. (a) Espressione di $\mathbb{E}\left[\Theta^{-1}|\mathbf{z}_n\right]$.
   (b) Espressione di $\mathbb{E}[\Theta^k|\mathbf{z}_n]$, $\quad k > 0$.

> **Soluzione**
>
> (a) $\Theta|\mathbf{z}_n \sim \text{Ga}(\overline{\alpha}, \texttt{rate} = \overline{\beta}) \Longrightarrow \Theta^{-1}|\mathbf{z}_n \sim \text{InvGa}(\overline{\alpha}, \texttt{rate} = \overline{\beta})$ e quindi $\mathbb{E}[\Theta^{-1}|\mathbf{z}_n] = \frac{\overline{\beta}}{\overline{\alpha}-1} = \frac{s_n}{n-2}$ (ricorda che $\overline{\alpha} = n - 1$ e $\overline{\beta} = s_n$). Se non si ricorda questo risultato, il valore atteso si calcola ricordando l'integrale gamma, in base al quale, per $a, b > 0$,
>
> $$\int_0^\infty \theta^{a-1} e^{-b\theta} d\theta = \frac{\Gamma(a)}{b^a}.$$
>
> Si ha quindi che $\mathbb{E}[\Theta^{-1}|\mathbf{z}_n] = \int_0^\infty \frac{1}{\theta} \pi(\theta|\mathbf{z}_n) d\theta = \ldots = \frac{\overline{\beta}}{\overline{\alpha}-1} = \frac{s_n}{n-2}$.
>
> (b) $\mathbb{E}[\Theta^k|\mathbf{z}_n] = \int_0^\infty \theta^k \times \pi(\theta|\mathbf{z}_n) d\theta = \ldots = \int_0^\infty \theta^k \times \frac{\overline{\beta}^{\overline{\alpha}}}{\Gamma\overline{\alpha}} \theta^{\overline{\alpha}-1} e^{-\overline{\beta}\theta} d\theta = \ldots = \frac{\overline{\beta}^{\overline{\alpha}}}{\Gamma(\overline{\alpha})} \frac{\Gamma(\overline{\alpha}+k)}{\overline{\beta}^{\overline{\alpha}+k}} = \frac{\Gamma(n+k-1)}{\Gamma(n-1) s_n^k}$ (anche in questo caso si sfrutta l'integrale gamma).

3. (a) Espressione di $m(y|\mathbf{z}_n)$ (densità marginale a posteriori di una osservazione futura Y).
   (b) Espressione di $C_h = (y \in \mathbb{R}^+ : \ m(y|\mathbf{z}_n) \geq h)$, $h \geq 0$, insieme HD marginale a posteriori per Y <u>nel caso in cui $n = 2$</u>.

> **Soluzione**
>
> (a) $m(y|\mathbf{z}_n) = \int_0^\infty p_\theta(y) \, \pi(\theta|\mathbf{z}_n) d\theta = \theta e^{-y\theta} \times \frac{\overline{\beta}^{\overline{\alpha}}}{\Gamma(\overline{\alpha})} \theta^{\overline{\alpha}-1} e^{-\overline{\beta}\theta} d\theta = \ldots$ (integrale gamma) $\ldots \frac{\overline{\alpha}\overline{\beta}^{\overline{\alpha}}}{(\overline{\beta}+y)^{\overline{\alpha}+1}}$.
> Per $n = 2$ si trova $m(y|\mathbf{z}_n) = \frac{s_n}{(s_n+y)^2}$, $y \geq 0$.
> Notare che si tratta di una funzione di densità definita in $[0, \infty)$ che in $y = 0$ vale 1 e poi decresce a zero al crescere di y.
>
> (b) Per quanto appena osservato, l'insieme HD C_h deve essere del tipo $[0, K_h]$, con $K_h > 0$ da determinare. Poichè $m(y|\mathbf{z}_n) \geq h \iff \frac{s_n}{(s_n+y)^2} \iff (s_n + y)^2 \geq \frac{s_n}{h}$, risolvendo la disequazione rispetto a y e ricordando il vincolo $y \geq 0$, troviamo che $C_h = [0, K_h]$ con $K_h = \sqrt{\frac{s_n}{h}} - s_n$.