Statistica A – Esercitazione 2

Esercizio 1. Data una verosimiglianza $(L_\theta)_\theta$ coniugata rispetto ad una famiglia di distribuzioni $\mathcal{P}$, dimostrare che anche la famiglia di distribuzioni

$$\mathcal{P}^* := \{\alpha\pi_1 + (1-\alpha)\pi_2 : \pi_1, \pi_2 \in \mathcal{P}\}$$

è coniugata per la verosimiglianza ed inoltre che la posterior è la mistura delle posterior.

Esercizio 2. Sia $(X_i)_{i\geq 1}$ una successione bernoulliana scambiabile e sia $S_n = \sum_{i=1}^n X_i$. Dimostrare che

1. per ogni $k = 0, 1, \ldots, n$ si ha
$$\mathbb{P}\left(S_n = k\right) = \binom{n}{k} \mathbb{P}\left(X_1 = \cdots = X_k = 1, X_{k+1} = \cdots = X_n = 0\right);$$
2. fissato $m < n$ si ha
$$\mathbb{P}\left(\boldsymbol{X}_m = \boldsymbol{x}_m | S_n = s\right) = \frac{\binom{n-m}{s - \sum_{i=1}^m x_i}}{\binom{n}{s}}$$
se $\max\{0, m-n+s\} \sum_{i=1}^m x_i \leq \min\{s, m\}$ e 0 altrimenti.

Esercizio 3. Sia X_n una successione aleatoria scambiabile e A un insieme misurabile. Si dimostri che la successione $Y_n = \mathbb{I}_A(X_n)$ è scambiabile e se ne determini la rapresentazione della legge con il teorema di de Finetti.

Esercizio 4. Si ripete n volte il seguente esperimento: si lancia una moneta fino a che non esce testa, dopo di che si interrompe l'esperimento. Si indicano con X_i il numero di lanci effettuato fino a che non esce testa e con p la probabilità di ottenere testa in un lancio. Calcolare:

1. la distribuzione di X_i;
2. la distribuzione di $Y_n := X_1 + \cdots + X_n$;
3. la funzione generatrice dei momenti di Y_n;
4. $\mathbb{E}\left[Y_n\right]$ e $\mathrm{Var}\left(Y_n\right)$ [che sono rispettivamente $n(1-p)/p$ e $n(1-p)/p^2$].

Esercizio 5. Si assegna il seguente modello bayesiano: le osservazioni X_i condizionatamente al parametro incognito $\Theta = \theta$ sono i.i.d. con distribuzione geometrica di parametro θ e la distribuzione a priori per il parametro è una $Beta(\alpha, \beta)$.

1. Si costruisca il modello associato ad n osservazioni (Si specifichi spazio campionario, dei parametri, etc.).
2. Si calcoli la distribuzione predittiva a priori.
3. Si calcoli la distribuzione a posteriori e quella predittiva a posteriori.
4. Si confrontino i risultati con quelli ottenuti per il modello beta–binomiale e si stabilisca quali connessioni esistono tra i due modelli.

Esercizio 6. Sia $\mathcal{U}$ un'urna contenente N palline marcate con 0 e 1, rispettivamente con numerosità N_0 e $N_1 := N - N_0$. Si compiono n estrazioni dall'urna senza reimmissione e si indicano i risultati con X_i, $i = 1, \ldots, n$ in cui $n \leq \min\{N_0, N_1\}$,

1. si calcoli la distribuzione di $(X_1, \ldots, X_m), m \leq n$;
2. si calcoli la distribuzione di $\left(X_{n-m+1}, \ldots, X_n\right), m \leq n$;
3. si calcoli la distribuzione di $(X_{n-m+1}, \ldots, X_n)$ condizionatamente a $(X_1, \ldots, X_{n-m})$;

4. (*) stabilire se la successione $(X_i)_i$ è finitamente scambiabile e si calcoli la distribuzione di $\sum_{i=1}^n X_i$.
5. (*) supposto che la numerosità dell'urna N tenda ad infinito e che $N_1/N \to p \in [0,1]$, calcolare la distribuzione di probabilità limite del numero di successi nelle n prove.

Esercizio 7. Sia $\mathcal{U}$ un'urna contenente N marcate con 0 e 1, rispettivamente con numerosità N_0 e $N_1 := N - N_0$. Si compiono n estrazioni come segue:

(A) si compie un'estrazione dall'urna e si registra il risultato X_i;

(B) si reimmette nell'urna la pallina estratta più r palline con lo stesso numero.

Supposto di compiere n osservazioni in queste condizioni:

1. si calcoli la distribuzione di $(X_1, \ldots, X_m), m \leq n$;
2. si calcoli la distribuzione di $(X_{n-m+1}, \ldots, X_n), m \leq n$;
3. si calcoli la distribuzione di $(X_{n-m+1}, \ldots, X_n)$ condizionata a $(X_1, \ldots, X_{n-m})$;
4. si calcoli la distribuzione di $\sum_{i=1}^n X_i$.;
5. si ripetano i punti precedenti nel caso in cui ad ogni estrazione, supponendo che si reintroducano r_i palline, se si estrae una pallina di tipo i.

Esercizio 8. Un'urna $\mathcal{U}$ contiene N palline marcate con numeri che vanno da 0 a $d-1$ e di cui indichiamo le rispettive numerosità n_j, $j = 0, \ldots, d-1$. Si compiono n estrazioni dall'urna con reimmissione e si indicano i risultati delle estrazioni con X_i, $i = 1, \ldots, n$.

1. Si calcoli la distribuzione di X_1 e quella congiunta i $X_1, \ldots, X_n$.
2. Sia $\boldsymbol{Y}_n$ il vettore aleatorio il cui j–esimo elemento è $Y_j := \sum_{i=1}^n [X_i = j]$, in cui per ogni evento A $[A] = 1$ se A è vero e 0 altrimenti. Si determini il supporto e si calcoli la distribuzione di $\boldsymbol{Y}_n$.

Esercizio 9. Un'urna $\mathcal{U}$ contiene palline marcate con numeri che vanno da 0 a $d-1$, le proporzioni di palline di tipo j è Θ_j, $j = 0, \ldots, d-1$ ($\Theta_j \geq 0$, $\sum_{i=1}^d \Theta_j = 1$) sono incognite. Siamo interessati a studiare la proporzione delle palline nell'urna e a tal fine compiamo n estrazioni con reimmissione dall'urna e indicando i risultati con X_i, $i = 1, \ldots, n$ in cui ogni $X_i \in \{0, \ldots, d-1\}$.

Sia $\Theta = (\Theta_1, \ldots, \Theta_{d-1})$ il vettore dei parametri incogniti, assegnamo come distribuzione a priori la distribuzione di Dirichlet, $Dir\,(\alpha_1, \alpha_2, \ldots, \alpha_d)$, cioè per ogni $\boldsymbol{\theta} \in \mathbb{R}^{d-1}$

$$\pi(\boldsymbol{\theta}) = \left[\frac{\Gamma\left(\sum_{i=1}^d \alpha_i\right)}{\prod_{i=1}^d \Gamma(\alpha_i)} \left(1 - \sum_{i=1}^{d-1} \theta_i\right)^{\alpha_d - 1} \prod_{i=1}^{d-1} \theta^{\alpha_i - 1} \right] \mathbb{I}_{S_{d-1}}(\boldsymbol{\theta})$$

in cui $S_{d-1} = \left\{ \boldsymbol{y} \in \mathbb{R}^{d-1} : y_i \geq 0, i = 1, \ldots, d-1; \sum_{i=1}^{d-1} y_i \leq 1 \right\}$.

1. Si costruisca il modello associato ad n osservazioni (Si specifichi spazio campionario, dei parametri, etc.).
2. Si calcoli la distribuzione predittiva a priori.
3. Si calcoli la distribuzione a posteriori e quella predittiva a posteriori.

Esercizio 10. Sia assegnato il seguente modello

$$X_i|\Theta \overset{i.i.d.}{\sim} f_\theta \text{ e } \Theta \sim f_\Theta, i \geq 1$$

Data la successione di osservazioni $(X_n)_{n\geq 1}$ (o $(x_i)_{i\geq 1}$), indichiamo con $\boldsymbol{X}_n = (X_1, X_2, \ldots, X_n)$ (rispettivamente $\boldsymbol{x}_n = (x_1, x_2, \ldots, x_n)$). Dimostrare che

1. $f_{\Theta|\boldsymbol{X}_{n+1}}(\theta|\boldsymbol{x}_{n+1}) \propto f_{X_{n+1}|\Theta}(x_{n+1}|\theta)\pi_{\Theta|\boldsymbol{x}_n}$;
2. $f_{X_{n+1}|\boldsymbol{X}_n}(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n) \propto \int_\Theta f_{X_1|\Theta}(x_1|\theta) f_{\Theta|\boldsymbol{X}_n}(\theta|\boldsymbol{x}_n)\, \mathrm{d}\theta$

Esercizio 11 (*)**.** Siano $Z = (Z_i)_{i\geq 1}$ una successione scambiabile e $\rho_{ij} = \frac{Cov(Z_i,Z_j)}{\sqrt{\mathrm{Var}(Z_i)\mathrm{Var}(Z_j)}}$ il coefficiente di correlazione di Z_i e Z_j. Dimostrare che

1. per ogni $m \in \mathbb{R}$ e per ogni $\sigma \in \mathbb{R}_+$ il vettore $[\sigma (Z_i - m)]_{i=1}^N N$ è N–scambiabile;
2. $|\rho_{ij}| \leq 1$ e che se le variabili sono scambiabili, allora $\rho_{ij} = \rho$;
3. $\mathrm{Var}\left(\sum_{i=1}^n Z_i\right) = N + N(N-1)\rho$ e dedurre che se la successione è N–scambiabile allora $\rho \geq -1/(N-1)$, mentre se è infinitamente scambiabile $\rho \geq 0$.

Esercizio 12 (*)**.** Data una successione i.i.d. standardizzata $(\xi_i)_{i\geq 0}$ (i.e. $\mathbb{E}(\xi_1) = 0$ e $\mathbb{E}[\xi_1^2] = 1$) si definiscano le seguenti successioni:

$$Z_i = \xi + c\sum_{i=1}^{N} \xi_i \text{ e } \tilde{Z}_i = c\xi_0 + \xi_i$$

in cui c è una costante.

1. Si dimostri che, per ogni N e per ogni c fissati il vettore $\boldsymbol{Z}_N = (Z_1, \ldots, Z_N)$ è N–scambiabile;
2. fissato $\rho_0 \in [-1/(N-1), 1]$ dimostrare che esiste un vettore scambiabile con tale struttura di correlazione;
3. Si dimostri che, per ogni N e per ogni c la successione $\tilde{Z}_i$ è infinitamente scambiabile e se ne calcoli la funzione di correlazione;
4. fissato $\rho_0 \in [0, 1]$ dimostrare che esiste un vettore scambiabile con tale struttura di correlazione.
5. fissato

Esercizio 13. Siano $X \sim Gauss(0, 1)$ e $V \sim \chi^2_{(n)}$ due variabili aleatorie indipendenti.

1. Si dimostri che la variabile aleatoria $T := X/\sqrt{V/n}$ ha distribuzione simmetrica rispetto all'origine e se ne calcoli la distribuzione. In questo caso diciamo che T ha distribuzione T di Student con n gradi di libertà e scriviamo $T \sim \mathcal{T}_{(n)}$;
2. Fissati $\alpha \in \mathbb{R}$ e $\beta > 0$ si calcoli la distribuzione di $\tilde{T} := \alpha + \beta T$. In questo caso diremo che $\tilde{T}$ ha distribuzione T di Student con parametro di non centralità α e di scala β e scriviamo $\tilde{T} \sim \mathcal{T}_{n,\alpha,\beta}$.

Esercizio 14. Siano $X_i|\Sigma = \sigma^2 \overset{i.i.d.}{\sim} Gauss(\mu, \sigma^2)$ e $\Sigma \sim InvGamma(\nu_0/2, \nu_1/2)$ in cui $\mu \in \mathbb{R}$ è un parametro noto.

1. Dimostrare che $(X_1 - \mu)/\nu_1$ ha una distribuzione $\mathcal{T}_{(\nu_0)}$.
2. Dimostrare che il modello è coniugato e sfruttare il risultato dell'Esercizio 10 per calcolare la distribuzione di $\Sigma|\boldsymbol{X}_n$.

Esercizio 15. Dato un modello statistico bayesiano $\mathcal{M}$ con spazio dei parametri $\mathbb{R}$ e distribuzione a priori π_Θ continua e strettamente monotona, dimostrare che lo stesso modello $\mathcal{M}$ con parametro $\Phi := F_\Theta(\Theta)$ ha distribuzione iniziale uniforme su $(0, 1)$.

Esercizio 16. Si ottenga la distribuzione di Jeffreys, si stabilisca se è propria e si calcoli la distribuzione a posteriori per i seguenti modelli:

1. modello gaussiano con media nota e varianza incognita;
2. modello bernoulliano;
3. modello di Poisson.

Si confrontino i risultati con quelli ottenuti nei relativi modelli coniugati.

Esercizio 17. Data una verosimiglianza $f_{X|\Theta}(\cdot, |\theta)$, $\theta \in \Theta$, dimostrare che se $\boldsymbol{\theta} \mapsto f_{X|\boldsymbol{\Theta}}(\boldsymbol{x}|\boldsymbol{\theta})$ è continua per ogni $\boldsymbol{x}$ è una verosimiglianza, allora la classe $\mathcal{P} \subset M_1(\Theta)$ delle distribuzioni continue è coniugata.

Esercizio 18. Sia $f_{\boldsymbol{X}_n|\boldsymbol{\Theta}}(\boldsymbol{x}_n|\cdot)$ la verosimiglianza associata ad un campione $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ i.i.d. condizionatamente a Θ. Dimostrare che, se $\mathcal{P} \subset M_1(\Theta)$ è coniugato rispetto $f_{X_1|\Theta}$, allora lo è anche rispetto a $f_{\boldsymbol{X}_n|\Theta}$.

Esercizio 19. Sia $\mathcal{F} \subset M_1\left(X^{(n)}\right)$ in cui $X^{(n)}$ è lo spazio campionario relativo alle prime n osservazioni $(X_1, \ldots, X_n)$. Diciamo che il modello è *coniugato predittivamente* se fissato un sistema di predittive a priori in $\mathcal{F}$, allora anche le predittive a posteriori appartengono allo stesso insieme $\mathcal{F}$.

Dimostrare che se un modello è coniugato, allora è anche coniugato predittivamente.