Una caratterizzazione delle leggi a priori coniugate per la famiglia esponenziale

L. Di Scala L. La Rocca

Tesina per il corso del Prof. Regazzini, anno 2001

Su $\Re^k$ munito dei boreliani consideriamo una misura σ-finita μ, tale che l'interno dell'involucro convesso del suo supporto non sia vuoto. Sia

$$\Theta = \left\{ \theta \in \Re^k : \int_{\Re^k} e^{<x,\theta>} \, \mu(dx) < \infty \right\}$$

un aperto di $\Re^k$, sul quale è ben definita (e analitica) la log-trasformata di Laplace di μ (cioè il logaritmo dell'integrale in questione) che indicheremo con $M(\theta)$. Si veda Brown [2] come riferimento. Al variare di θ in Θ, le densità di probabilità $p(x, \theta) = e^{<x,\theta>-M(\theta)}$ definiscono un modello statistico su $\Re^k$. Si supponga che tale modello sia identificabile. Si osservi che una condizione sufficiente per l'identificabilità (e quindi per la stretta convessità di $M(\theta)$) è che la dimensione della chiusura dell'involucro convesso del supporto di μ sia piena, ovvero uguale a k. Sia τ una legge su Θ, anch'esso munito dei boreliani. Costruiamo un modello statistico bayesiano, prendendo come spazio di probabilità il prodotto $\Omega = \Re^k \times \Theta$ (di cui indicheremo il generico punto con $\omega = (x, \theta)$), munito dei boreliani e della probabilità $P(dx, d\theta) = p(x, \theta)\mu(dx)\tau(d\theta)$. Ci interesseranno le variabili aleatorie $X(x, \theta) = x$ (osservazione) e $\tilde{\theta}(x, \theta) = \theta$ (parametro). La legge di $\tilde{\theta}$ è individuata da $\tau(d\theta)$, per costruzione, mentre la legge di X è indivduata da $f(x)\,\mu(dx)$ con $f(x) = \int_\Theta p(x, \theta)\,\tau(d\theta)$, come si verifica agevolmente. La legge di $X|\tilde{\theta}$ è individuata da $p(x, \theta)\,\mu(dx)$, per costruzione, mentre la legge di $\tilde{\theta}|X$ è individuata da $\frac{1}{f(x)}\,p(x, \theta)\tau(d\theta)$, come conseguenza del teorema di Bayes.

Diamo per noto che $E[X|\tilde{\theta}] = \nabla M(\tilde{\theta})$ (onde $E[\nabla M(\tilde{\theta})] = E[X]$) e che pertanto $\nabla M(\tilde{\theta})$ può essere interpretato come una riparametrizzazione in termini di valor atteso dell'osservazione. È allora di interesse la variabile aleatoria $E[\nabla M(\tilde{\theta})|X]$, che assume il significato di parametro a posteriori per il valore atteso dell'osservazione (nonché di previsione per una successiva osservzione Y che sia distribuita come X e, condizionalmente a $\tilde{\theta}$, da X indipendente).

Un caso particolarmente semplice si ha quando si assume come legge a priori per il parametro quella individuata da $\tau(d\theta) \propto e^{n_0<x_0,\theta>-n_0 M(\theta)}\,d\theta$ con x_0 interno all'involucro convesso del supporto di μ e $n_0 > 0$. Si tratta di una legge coniugata al modello, perché, se la si adotta, la legge di $\tilde{\theta}|X$ rimane individuata, a meno di un coefficiente di normalizzazione, da $e^{n_1<x_1,\theta>-n_1 M(\theta)}\,d\theta$ con $n_1 = n_0 + 1$ e $x_1(X) = \frac{1}{n_0+1}x_0 + \frac{n_0}{n_0+1}X$, ovvero è dello stesso tipo. La scelta di x_0 e n_0 è giustificata dal seguente teorema.

Teorema 1 (Diaconis e Ylvisaker, 1979) *Se $n_0 > 0$ e x_0 è interno all'involucro convesso del supporto di μ, allora $\tau(\Theta) < +\infty$. Viceversa, se $\tau(\Theta) < +\infty$ e $\Theta = \Re^k$, allora $n_0 > 0$. Se, invece, $\tau(\Theta) < +\infty$ e $n_0 > 0$, allora x_0 è interno all'involucro convesso del supporto di μ.*

Soto le ipotesi del suddetto teorema, vale il seguente risultato.

Teorema 2 (Diaconis e Ylvisaker, 1979) *Sia $\Theta \subset \Re^k$ aperto e $\tilde{\theta}$ abbia legge a priori coniugata individuata da $\tau(d\theta) \propto e^{n_0 <x_0,\theta> - n_0 M(\theta)} d\theta$, con x_0 interno all'involucro convesso del supporto di μ e $n_0 > 0$. Allora si ha*

$$E[\nabla M(\tilde{\theta})] = x_0.$$

DIM.

La tesi può essere riscritta alla seguente maniera:

$$E[\nabla M(\tilde{\theta})] - x_0 = \int_\Theta (\nabla M(\theta) - x_0)\tau(d\theta) = \int_\Theta (\nabla M(\theta) - x_0) f_{n_0,x_0}(\theta) d\theta = 0$$

dove $f_{n_0,x_0}(\theta) = e^{<n_0 x_0,\theta> - n_0 M(\theta)}$. Ovvero, $\int_\Theta \nabla f_{n_0,x_0}(\theta) d\theta = 0$. Si dimostra a seguire che tutte le componenti dell'integrale sono nulle. Se ne consideri la prima: $\int_\Theta \frac{\partial}{\partial\theta_1} f_{n_0,x_0}(\theta) d\theta = 0$. Per il teorema di Fubini, l'integrale si riscrive come

$$\begin{aligned}
&\int \cdots \int \left(\int_{\underline{\theta}_1}^{\overline{\theta}_1} \frac{\partial}{\partial\theta_1} f_{n_0,x_0}(\theta) d\theta_1 \right) d\theta_2 \ldots d\theta_k = \\
= \;&\int \cdots \int \left(\lim_{\theta_1 \to \overline{\theta}_1} f_{n_0,x_0}(\theta) - \lim_{\theta_1 \to \underline{\theta}_1} f_{n_0,x_0}(\theta) \right) d\theta_2 \ldots d\theta_k
\end{aligned}$$

dove $(\underline{\theta}_1, \overline{\theta}_1) = (\underline{\theta}_1(\theta_2, \ldots, \theta_k), \overline{\theta}_1(\theta_2, \ldots, \theta_k))$ è la sezione di Θ sulla prima componente, poiché Θ è un convesso aperto. Si dimostra ora che entrambi i limiti sono 0. Si supponga, dapprima, che $\overline{\theta}_1 = +\infty$. Allora, per la diseguaglianza

$$e^{M(\theta)} \leq \mu(\overline{A})^{-1} e^{<\theta, m(\overline{A})>} \tag{1}$$

con $A \subset \Re^k$ insieme limitato e convesso e $m(\overline{A}) = \int_{\overline{A}} z\mu_A(dz)$, dove $\mu_A(\cdot) = \mu(A \cap \cdot)/\mu(A)$, si ottiene

$$\lim_{\theta_1 \to +\infty} f_{n_0,x_0}(\theta) \leq \lim_{\theta_1 \to +\infty} \mu(\overline{A})^{-n_0} e^{n_0 <\theta, x_0 - m(\overline{A})>}.$$

Tale limite è nullo scegliendo A tale che $(x_0)_1 < (m(\overline{A}))_1$ e tenendo fissi $(\theta_2, \ldots, \theta_k)$. Sia $\overline{\theta}_1 < +\infty$. Si ponga $\overline{\theta} = (\overline{\theta}_1, \theta_2, \ldots, \theta_k)$ e $\theta^\star = (\theta_1^\star, \theta_2, \ldots, \theta_k)$ tale che $\theta^\star \in \Theta$. Quindi

$$\begin{aligned}
&\lim_{\theta_1 \to \overline{\theta}_1} f_{n_0,x_0}(\theta) = \lim_{\theta_1 \to \overline{\theta}_1} e^{<n_0 x_0,\theta>} e^{-n_0 M(\theta)} = \\
= \;&\lim_{\theta_1 \to \overline{\theta}_1} e^{<n_0 x_0,\theta>} \left(\int_{\Re^k} e^{<x,\theta>} \mu(dx) \right)^{-n_0} \\
= \;&\lim_{\theta_1 \to \overline{\theta}_1} e^{<n_0 x_0,\theta>} \left(\int_{x_1 \leq 0} e^{<x,\theta>} \mu(dx) + \int_{x_1 > 0} e^{<x,\theta>} \mu(dx) \right)^{-n_0}
\end{aligned}$$

È ora sufficiente dimostrare, poiché $n_0 > 0$, che la quantità in parentesi è uguale a $+\infty$. Si ha che

$$\int_{x_1 \leq 0} e^{<x,\overline{\theta}>} \mu(dx) \leq \int_{x_1 \leq 0} e^{<x,\theta^\star>} \mu(dx) < +\infty$$

poiché $\theta^\star \in \Theta$ e $\theta_1^\star < \overline{\theta}_1$. Inoltre,

$$\int_{x_1 > 0} e^{<x,\theta>} \mu(dx) \to \int_{x_1 > 0} e^{<x,\overline{\theta}>} \mu(dx) = +\infty$$

per convergenza monotona. Il ragionamento può ripetersi per il secondo limite con il medesimo risultato.

Si dimostra ora che il teorema di Fubini poteva essere correttamente applicato. Essendo il modello identificabile, la matrice Hessiana di $M(\theta)$ è definita positiva. Pertanto, fissati $(\theta_2, \ldots, \theta_k)$, $\frac{\partial}{\partial\theta_1} M(\theta)$ è una funzione strettamente crescente in θ_1. Poiché

$$\frac{\partial}{\partial\theta_1} f_{n_0,x_0}(\theta) = n_0 \left(x_0 - \frac{\partial}{\partial\theta_1} M(\theta) \right) f_{n_0,x_0}(\theta),$$

$\frac{\partial}{\partial\theta_1} f_{n_0,x_0}(\theta)$ è strettamente decrescente in θ_1, cambiando segno al più una volta in $\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)$. Sia ora $\theta_1^\star$ quell'unico $\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)$ che verifichi la condizione $\frac{\partial}{\partial\theta_1} f_{n_0,x_0}(\theta_1^\star, \theta_2, \ldots, \theta_k) = 0$. Avendo già dimostrato che

$$\lim_{\theta_1 \to \overline{\theta}_1} f_{n_0,x_0}(\theta) = \lim_{\theta_1 \to \underline{\theta}_1} f_{n_0,x_0}(\theta) = 0$$

si può concludere che $f_{n_0,x_0}(\theta)$, come funzione di θ_1, ha un unico punto di massimo $\theta^\star = (\theta_1^\star, \theta_2, \ldots, \theta_k)$. Quindi, per il teorema fondamentale del calcolo integrale, si ha

$$\begin{aligned}
\int_{\underline{\theta}_1}^{\overline{\theta}_1} \left| \frac{\partial}{\partial\theta_1} f_{n_0,x_0}(\theta) \right| d\theta_1 &= \int_{\underline{\theta}_1}^{\overline{\theta}_1} \left(\frac{\partial}{\partial\theta_1} f_{n_0,x_0}(\theta) \right)^+ + \left(\frac{\partial}{\partial\theta_1} f_{n_0,x_0}(\theta) \right)^- d\theta_1 \\
&= \int_{\underline{\theta}_1}^{\theta_1^\star} \frac{\partial}{\partial\theta_1} f_{n_0,x_0}(\theta) d\theta_1 - \int_{\theta_1^\star}^{\overline{\theta}_1} \frac{\partial}{\partial\theta_1} f_{n_0,x_0}(\theta) d\theta_1 = \\
&= \lim_{\theta_1 \to \theta_1^\star} f_{n_0,x_0}(\theta) - \lim_{\theta_1 \to \underline{\theta}_1} f_{n_0,x_0}(\theta) + \lim_{\theta_1 \to \theta_1^\star} f_{n_0,x_0}(\theta) - \lim_{\theta_1 \to \overline{\theta}_1} f_{n_0,x_0}(\theta) = \\
&= 2 f_{n_0,x_0}(\theta^\star) = 2 \max_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} f_{n_0,x_0}(\theta).
\end{aligned}$$

Rimane da dimostrare che $2 f_{n_0,x_0}(\theta^\star)$ è integrabile in $d\theta_2 \cdots d\theta_k$. Essendo x_0 interno all'involucro convesso del supporto di μ ed essendo quest'ultimo pieno, $x_0 = \sum_{i=1}^{k+1} \lambda_i m(\overline{A}_{x_i})$, dove $\lambda_i \geq 0 \ \forall i$, $\sum_{i=1}^{k+1} \lambda_i = 1$, gli x_i appartengono al supporto di μ e A_{x_i} è un intorno limitato e convesso di $x_i \ \forall i$. Inoltre, gli x_i non cadono in un iperpiano $(k-1)$-dimensionale di $\Re^k$, come nemmeno gli $m(\overline{A}_{x_i})$. Prendendo $x_0 = 0$ e $i \in \{1, \ldots, k+1\}$, per (1) si ha che

$$\int \cdots \int \left(\int_{\underline{\theta}_1}^{\overline{\theta}_1} \left| \frac{\partial}{\partial\theta_1} f_{n_0,x_0}(\theta) \right| d\theta_1 \right) d\theta_2 \ldots d\theta_k =$$

$$
\begin{aligned}
&= \int \cdots \int 2 \max_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} f_{n_0,x_0}(\theta) d\theta_2 \ldots d\theta_k \leq \\
&\leq 2 \int \cdots \int \max_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} \frac{1}{\mu(\overline{A}_{x_i})} e^{-n_0 < \theta, m(\overline{A}_{x_i}) >} d\theta_2 \ldots d\theta_k \\
&\leq c \int \cdots \int \max_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} e^{-n_0 < \theta, m(\overline{A}_{x_i}) >} d\theta_2 \ldots d\theta_k
\end{aligned}
$$

dove $c = 2 \left[\min_{i=1,\ldots,k+1} \ \mu(\overline{A}_{x_i})\right]^{-1}$. Quindi

$$
\begin{aligned}
& \int \cdots \int \left(\int_{\underline{\theta}_1}^{\overline{\theta}_1} \left| \frac{\partial}{\partial \theta_1} f_{n_0,x_0}(\theta) \right| d\theta_1 \right) d\theta_2 \ldots d\theta_k \\
\leq \ & c \int \cdots \int \max_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} \min_{i=1,\ldots,k+1} e^{-n_0 < \theta, m(\overline{A}_{x_i}) >} d\theta_2 \ldots d\theta_k = \\
= \ & \int \cdots \int \exp \left\{ -n_0 \min_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} \max_{i=1,\ldots,k+1} < \theta, m(\overline{A}_{x_i}) > \right\} d\theta_2 \ldots d\theta_k
\end{aligned}
$$

Sia $0 \neq \theta^{(2)} = (\theta_2, \ldots, \theta_k) =\parallel \theta^{(2)} \parallel \eta$, $\parallel \eta \parallel = 1$; allora

$$
\begin{aligned}
& \min_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} \max_{i=1,\ldots,k+1} < \theta, m(\overline{A}_{x_i}) > = \\
= \ & \min_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} \max_{i=1,\ldots,k+1} \left[\theta_1 m(\overline{A}_{x_i})_1 + \ldots + \theta_k m(\overline{A}_{x_k})_i \right] = \\
= \ & \min_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} \max_{i=1,\ldots,k+1} \parallel \theta^{(2)} \parallel \left[\frac{\theta_1 m(\overline{A}_{x_i})_1}{\parallel \theta^{(2)} \parallel} + < \eta, m(\overline{A}_{x_i})^{(2)} > \right] = \\
= \ & \parallel \theta^{(2)} \parallel \min_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} \max_{i=1,\ldots,k+1} \left[\theta_1 m(\overline{A}_{x_i})_1 + < \eta, m(\overline{A}_{x_i})^{(2)} > \right]
\end{aligned}
$$

dove l'ultima uguaglianza vale poiché sin dal principio si poteva supporre $\theta_1 \in (-\infty, +\infty)$. La tesi segue dalla dimostrazione che

$$
\inf_{\|\eta\|=1} \min_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} \max_{i=1,\ldots,k+1} \left[\theta_1 m(\overline{A}_{x_i})_1 + < \eta, m(\overline{A}_{x_i})^{(2)} > \right] > 0.
$$

Avendo posto, senza ledere la generalità, $x_0 = 0$, si ha

$$
\sum_{i=1}^{k+1} \lambda_i < \theta, m(\overline{A}_{x_i}) >= 0 \ \forall \theta \in \Theta.
$$

Cioè

$$
\max_{i=1,\ldots,k+1} < \theta, m(\overline{A}_{x_i}) > \geq 0 \ \forall \theta \in \Theta.
$$

Pertanto anche

$$
\min_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} \max_{i=1,\ldots,k+1} < \theta, m(\overline{A}_{x_i}) > \geq 0 \ \forall \theta^{(2)}.
$$

Si noti che la funzione

$$
f(\theta_2, \ldots, \theta_k) = f(\theta^{(2)}) = \min_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} \max_{i=1,\ldots,k+1} < \theta, m(\overline{A}_{x_i}) >
$$

è continua in $\theta^{(2)}$. Quindi $\inf_{\|\eta\|=1}\ f(\eta) \geq 0$. Supponendo per assurdo che sia 0 allora esiste $\eta^\star \in \Re^{k-1}$ tale che, per compattezza,

$$\min_{\theta_1 \in (\underline{\theta}_1, \overline{\theta}_1)} \max_{i=1,\ldots,k+1} \left[\theta_1 m(\overline{A}_{x_i})_1 + < \eta^\star, m(\overline{A}_{x_i})^{(2)} >\right] = 0$$

e quindi esisterebbe $(\theta_1, \eta^\star) \in \Theta$ tale che $\max_{i=1,\ldots,k+1} < (\theta_1, \eta^\star), m(\overline{A}_{x_i}) >= 0$. Questo implicherebbe che i punti $\{m(\overline{A}_{x_1}), \ldots, m(\overline{A}_{x_{k+1}})\}$ cadono in un iperpiano di dimensione 1. Ciò è assurdo.

*

Osservazione. Applicando il teorema a posteriori, si ottiene

$$E[\nabla M(\tilde{\theta})|X] = \frac{1}{n_0+1} x_0 + \frac{n_0}{n_0+1} X$$

cioè la previsione è combinazione convessa del valore atteso a priori e dell'osservazione, con pesi dipendenti da n_0, come se n_0 fosse la taglia di un ipotetico campione a priori.

Si vogliono studiare le ipotesi sotto le quali una previsione lineare del tipo $E[\nabla M(\tilde{\theta})|X] = aX + b$ implichi una legge a priori coniugata del tipo visto. A tal fine, si osservi preliminarmente che, prendendo il valore atteso, si ottiene la relazione $(1-a)E[X] = b$. Se $a \neq 1$, allora $E[X] = b/(1-a)$; altrimenti deve essere $b = 0$.

Caso Continuo

Se si fa l'ipotesi che il supporto della dominante μ sia sufficientemente "ricco", allora si ottiene il risultato seguente:

Teorema 3 (Diaconis e Ylvisaker, 1979) *Sia $\Theta \subset \Re^k$ aperto. Sia X un campione di taglia uno da $P_\theta(dx) = e^{<x,\theta>-M(\theta)}\mu(dx)$. Si supponga che il supporto di μ contenga un iper-rettangolo aperto $I_0 \subset \Re^k$. Sia τ la legge a priori di $\tilde{\theta}$, non concentrata in un unico punto. Si supponga, inoltre, che*

$$E[\nabla M(\tilde{\theta})|X] = aX + b \tag{2}$$

dove $a \in \Re$ e $b \in \Re^k$. Allora $a \neq 0$ e τ è assolutamente continua rispetto alla misura di Lebesgue su Θ con $\tau(d\theta) \propto e^{a^{-1}<b,\theta>-a^{-1}(1-a)M(\theta)}d\theta$.

Dim.

Sia $M_i(\theta) = \frac{\partial}{\partial\theta_i}M(\theta)$. Per l'ipotesi (2), $E[M_i^+(\tilde{\theta})|X]$ e $E[M_i^-(\tilde{\theta})|X]$ sono P-q.c. finiti $\forall i = 1, \ldots, k$. Per costruzione, X ha densità strettamente positiva $f(x) = \int_\Theta e^{<x,\theta>-M(\theta)}\tau(d\theta)$ rispetto a μ e quindi $E[M_i^+(\tilde{\theta})|X = x]$ e $E[M_i^-(\tilde{\theta})|X = x]$ sono μ-q.c. finiti $\forall i = 1, \ldots, k$. Per il Teorema di Bayes, si ha μ-q.c. $\forall i = 1, \ldots, k$

$$E[M_i^\pm(\tilde{\theta})|X = x] = \frac{1}{f(x)} \int_\Theta M_i(\theta) e^{<x,\theta>-M(\theta)}\tau(d\theta).$$

Pertanto, l'integrale al secondo membro è μ-q.c. finito $\forall i = 1, \ldots, k$. L'ipotesi (2) può allora essere riscritta come

$$\int_\Theta \nabla M(\theta) e^{<x,\theta>-M(\theta)} \tau(d\theta) = (ax+b)f(x) \qquad \mu - q.c. \tag{3}$$

e liberamente manipolata.

Se $a = 0$, allora (3) diventa

$$\int_\Theta (\nabla M(\theta) - b) e^{<x,\theta>-M(\theta)} \tau(d\theta) = 0 \qquad \mu - q.c.$$

e quindi l'integrale si annulla μ-q.c. su I_0. Se ne deduce che

$$(\nabla M(\theta) - b) = 0 \qquad \tau - q.c.$$

e, dunque, $\nabla M(\theta) \equiv b$ sul supporto di τ, che per ipotesi consta di almeno due punti. Ciò contraddice la stretta convessità della funzione $M(\theta)$. Quindi, $a \neq 0$.

Ripartendo da (3), con $a \neq 0$ e con $z = x + iy$, si definisce la funzione

$$Q(z) = \int_\Theta (\nabla M(\theta) - az - b) e^{<z,\theta>-M(\theta)} \tau(d\theta)$$

e si osserva che è nulla μ-q.c. almeno per $Re(z) \in I_0$. Scegliendo $x_0 \in I_0$ e $z = x_0 + iy$, ponendo $m(\theta) = \frac{1}{a}(\nabla M(\theta) - ax_0 - b)$ e $F(d\theta) = e^{<x_0,\theta>-M(\theta)} \tau(d\theta)$, si ottiene

$$\int_\Theta e^{i<y,\theta>} m(\theta) F(d\theta) = iy \int_\Theta e^{i<y,\theta>} F(d\theta). \tag{4}$$

Al fine di applicare la formula di inversione per funzioni caratteristiche di vettori aleatori, moltiplichiamo ambo i membri dell'equazione per il fattore

$$\left(\frac{1}{2\pi}\right)^k \prod_{j=1}^k \frac{1 - e^{-ih_j y_j}}{iy_j} \frac{e^{-i\alpha_j y_j} - e^{-i\beta_j y_j}}{iy_j}$$

dove $\alpha_j < \beta_j \; \forall j = 1, \ldots, k$ e $h_j \in \Re$. Si inizi considerando la prima componente del membro di destra:

$$\begin{aligned}
& \left(\frac{1}{2\pi}\right)^k iy_1 \prod_{j=1}^k \frac{1 - e^{-ih_j y_j}}{iy_j} \frac{e^{-i\alpha_j y_j} - e^{-i\beta_j y_j}}{iy_j} \int_\Theta e^{i<y,\theta>} F(d\theta) = \\
= \; & \left(\frac{1}{2\pi}\right)^k \frac{(1 - e^{-ih_1 y_1})(e^{-i\alpha_1 y_1} - e^{-i\beta_1 y_1})}{iy_1} \cdot \\
& \cdot \prod_{j=2}^k \frac{1 - e^{-ih_j y_j}}{iy_j} \int_{\alpha_j}^{\beta_j} e^{-iu_j y_j} du_j \int_\Theta e^{i<y,\theta>} F(d\theta) = \\
= \; & \int_{\alpha_2}^{\beta_2} \cdots \int_{\alpha_k}^{\beta_k} \left\{ \frac{1}{(2\pi)^k} \left[\frac{e^{-i\alpha_1 y_1} - e^{-i(\alpha_1 + h_1) y_1}}{iy_1} - \frac{e^{-i\beta_1 y_1} - e^{-i(\beta_1 + h_1) y_1}}{iy_1} \right] \cdot \right. \\
& \left. \cdot \prod_{j=2}^k \frac{e^{-iu_j y_j} - e^{-i(u_j + h_j) y_j}}{iy_j} \int_\Theta e^{i<y,\theta>} F(d\theta) \right\} du_2 \ldots du_k.
\end{aligned}$$

Integrando su $-T \leq y_j \leq T\ \forall j = 1, \ldots, k$, applicando Fubini e il teorema di convergenza dominata, si può portare T a $+\infty$ e ottenere

$$\int_{\alpha_2}^{\beta_2} \cdots \int_{\alpha_k}^{\beta_k} \{\Delta_F[(\alpha_1, u_2, \ldots, u_k), (\alpha_1 + h_1, u_2 + h_2, \ldots, u_k + h_k)] -$$

$$-\Delta_F[(\beta_1, u_2, \ldots, u_k), (\beta_1 + h_1, u_2 + h_2, \ldots, u_k + h_k)]\}\, du_2 \ldots du_k.$$

Invece, posto $G(d\theta) = m_1(\theta)F(d\theta)$, la prima componente del membro di sinistra diventa:

$$\left(\frac{1}{2\pi}\right)^k \prod_{j=1}^k \frac{1 - e^{-ih_j y_j}}{iy_j} \frac{e^{-i\alpha_j y_j} - e^{-i\beta_j y_j}}{iy_j} \int_\Theta e^{i<y,\theta>} G(d\theta) =$$

$$= \int_{\alpha_1}^{\beta_1} \cdots \int_{\alpha_k}^{\beta_k} \left\{ \left(\frac{1}{2\pi}\right)^k \frac{1 - e^{ih_j y_j}}{iy_j} e^{-iu_j y_j} \int_\Theta e^{i<y,\theta>} G(d\theta) \right\} du_1 \ldots du_k$$

Integrando su $-T \leq y_j \leq T\ \forall j = 1, \ldots, k$, applicando Fubini ed il teorema di convergenza dominata, si può portare T a $+\infty$ e ottenere

$$\int_{\alpha_1}^{\beta_1} \cdots \int_{\alpha_k}^{\beta_k} \Delta_G[(u_1, \ldots, u_k), (u_1 + h_1, \ldots, u_k + h_k)]\ du_1 \ldots du_k$$

mediante la formula di inversione per la funzione caratteristica di G.

Per $h_j \to -\infty\ \forall j = 1, \ldots, k$, uguagliando i due membri, il teorema di convergenza dominata fornisce

$$\int_{\alpha_1}^{\beta_1} \cdots \int_{\alpha_k}^{\beta_k} G(u_1, \ldots, u_k)\ du_1 \ldots du_k =$$

$$= \int_{\alpha_2}^{\beta_2} \cdots \int_{\alpha_k}^{\beta_k} [F(\alpha_1, u_2, \ldots, u_k) - F(\beta_1, u_2, \ldots, u_k)]\ du_2 \ldots du_k$$

Questa uguaglianza è vera nei punti α_1 e β_1 di continuità per F. Inoltre, data l'abitrarietà di $(\alpha_2, \beta_2), \ldots, (\alpha_k, \beta_k)$, si deduce che

$$F(\alpha_1, u_2, \ldots, u_k) - F(\beta_1, u_2, \ldots, u_k) = \int_{\alpha_1}^{\beta_1} G(u_1, \ldots, u_k)\ du_1 \tag{5}$$

per $\mathcal{L}$-quasi ogni $u_2, \ldots, u_k$, dove α_1 e β_1 sono fissati e di continuità per F. La suddetta uguaglianza vale per tutti i valori α_1 e β_1 in un insieme numerabile denso di punti di continuità di F, al di fuori di un unico insieme di misura $\mathcal{L}$ nulla. Essendo F continua a destra e valendo il teorema di Lebesgue, l'uguaglianza è vera per ogni α_1 e β_1 e F risulta essere (assolutamente) continua nel primo argomento, per $\mathcal{L}$-quasi ogni $u_2, \ldots, u_k$. Se ne deduce che F (monotona crescente) dipende con continuità anche da $u_2, \ldots, u_k$. Il secondo membro di (5) può riscriversi come

$$\int_{\alpha_1}^{\beta_1} \int_{-\infty}^{u_1} \cdots \int_{-\infty}^{u_k} m_1(\theta)F(d\theta)\ du_1$$

e dipende anch'esso con continuità da $u_2, \ldots, u_k$, in virtù della continuità di F. Pertanto (5) vale in realtà $\forall u_2, \ldots, u_k$ e F è assolutamente continua nel primo argomento, per ogni possibile valore dei restanti. Il ragionamento può ripetersi per le altre componenti di (4), ottenendo così l'assoluta continuità di F in ogni fissato argomento. Allora anche $G_j(u) = \frac{\partial}{\partial u_j} F(u)$ ($\forall j = 1, \ldots, k$) è assolutamente continua in ogni fissato argomento, perché $G_j(du) = m_j(u)F(du)$. Ne segue l'assoluta continuità di F come funzione di k variabili, cioè $F(du) = f(u)\, du$, con $f(u) = \frac{\partial^k}{\partial u_1 \cdots \partial u_k} F(u)$. In virtù della (5), la densità f risulta caratterizzata da

$$\frac{\partial}{\partial u_j} f(u) = \frac{\partial^k}{\partial u_1 \cdots \partial u_k} \frac{\partial}{\partial u_j} F(u) = -m_j(u) f(u) \quad \forall j = 1, \ldots, k$$

che, ricordando l'espressione di m, in forma vettoriale si scrive

$$\nabla f(\theta) = \left(x_0 + a^{-1}b - a^{-1}\nabla M(\theta)\right) f(\theta).$$

Pertanto la densità f dovrà essere della forma

$$f(\theta) \propto e^{<x_0,\theta>+<a^{-1}b,\theta>-a^{-1}M(\theta)}$$

e di conseguenza si avrà

$$\tau(d\theta) = \ e^{-<x_0,\theta>+M(\theta)} F(d\theta) \propto \exp\left[a^{-1} < b, \theta > -a^{-1}(1-a)M(\theta)\right]\, d\theta$$

come volevasi dimostrare.

*

Caso Discreto

Se il supporto della dominante μ non contiene un iper-rettangolo aperto, allora non c'è un risultato unitario che caratterizzi le leggi a priori secondo le quali la previsione è lineare. Vi sono però diversi teoremi, che coprono i casi di maggiore interesse pratico.

Sia, per cominciare, $k = 1$ e si supponga che il supporto di μ sia $\aleph = \{0, 1, 2, \ldots\}$. In questo caso, si ha $\Theta = (-\infty, \theta_0)$ con $\theta_0 \leq +\infty$. Il risultato che segue assume $\theta_0 < +\infty$ e vale, per esempio, nel caso di modello statistico binomiale negativo. In particolare, vale per il modello geometrico. In quest'ultimo, infatti, la dominante μ è la misura che conta e la sua trasformata di Laplace vale $\sum_{x=0}^{\infty} e^{x\theta} = 1/(1-e^{\theta})$ per $e^{\theta} < 1$. Cioè $M(\theta) = -\ln(1-e^{\theta})$ per $\theta < \theta_0 = 0$.

Teorema 4 (Diaconis e Ylvisaker, 1979) *Sia $\Theta = (-\infty, \theta_0)$ con $\theta_0 < +\infty$ e si supponga che $\tilde{\theta}$ abbia legge a priori τ non concentrata in un solo punto. Sia X un campione di taglia uno da $P_\theta(\{x\}) = e^{x\theta - M(\theta)}\mu(\{x\})$. Sotto queste ipotesi, se si assume che*

$$E[\nabla M(\tilde{\theta})|X] = aX + b \tag{6}$$

dove $a, b \in \Re$, allora $a \neq 0$ e τ è assolutamente continua rispetto alla misura di Lebesgue su Θ, con $\tau(d\theta) \propto e^{a^{-1}b\theta - a^{-1}(1-a)M(\theta)} d\theta$.

Dim.

Come nel caso continuo, si verifica che $E[(M')^{\pm}(\tilde{\theta})|X = x] < \infty$ per ogni $x \in \aleph$ e si riscrive l'ipotesi (6) come

$$\int_{-\infty}^{\theta_0} M'(\theta)\, e^{x\theta - M(\theta)}\tau(d\theta) = (ax+b)f(x) \quad x = 0,1,2,\dots \tag{7}$$

dove $f(x) = \int_{-\infty}^{\theta_0} e^{x\theta - M(\theta)}\tau(d\theta)$. Se fosse $a = 0$ la (7) diventerebbe

$$\int_{-\infty}^{\theta_0} (M'(\theta) - b)\, e^{x\theta - M(\theta)}\tau(d\theta) = 0 \quad x = 0,1,2,\dots$$

e con il cambio di variabile $t = e^{\theta}$ si avrebbe

$$\int_{0}^{e^{\theta_0}} (M'(\ln(t)) - b)\, t^x e^{-M(\ln(t))}\sigma(dt) = 0 \quad x = 0,1,2,\dots$$

dove σ è l'immagine di τ attraverso $\theta \longmapsto e^{\theta}$. Ma allora la misura con segno sul compatto $[0, e^{\theta_0}]$ individuata da $(M'(\ln(t)) - b)\, e^{-M(\ln(t))}\sigma(dt)$ avrebbe tutti i momenti nulli e non potrebbe che essere la misura identicamente nulla (unica soluzione del problema omogeneo di Hausdorff). Ne seguirebbe $M'(\theta) = b$ sul supporto di τ, in contraddizione con la stretta convessità di $M(\theta)$, dato che per ipotesi τ non può essere concentrata in un solo punto.

Stabilito che $a \neq 0$, la (7) può riscriversi come

$$\int_{-\infty}^{\theta_0} (M'(\theta) - b)\ e^{x\theta - M(\theta)}\tau(d\theta) = ax \int_{-\infty}^{\theta_0} e^{x\theta - M(\theta)}\tau(d\theta) \quad x \geq 0. \tag{8}$$

Se si osserva che $e^{x\theta} = \int_{-\infty}^{\theta} x e^{xy} dy$ per $x = 1, 2, \dots$, si sostituisce nel primo membro della (8) e si applica il teorema di Fubini, si ottiene

$$\int_{-\infty}^{\theta_0} x e^{xy} \left\{ \int_{y}^{\theta_0} (M'(\theta) - b)\ e^{-M(\theta)}\tau(d\theta) \right\} dy \quad x \geq 1$$

che, utilizzando la (8) con $x = 0$, può essere trasformato in

$$-\int_{-\infty}^{\theta_0} x e^{xy} \left\{ \int_{-\infty}^{y} (M'(\theta) - b)\ e^{-M(\theta)}\tau(d\theta) \right\} dy \quad x \geq 1.$$

Con il primo membro così trasformato, osservando che le variabili mute θ e y possono scambiarsi di ruolo e semplificando $x \neq 0$, la (8) diviene

$$\int_{-\infty}^{\theta_0} e^{x\theta} \left\{ -\int_{-\infty}^{\theta} (M'(y) - b)\ e^{-M(y)}\tau(dy) \right\} d\theta = a \int_{-\infty}^{\theta_0} e^{x\theta} e^{-M(\theta)}\tau(d\theta)$$

valida per ogni $x \geq 1$. Il cambio di variabile $t = e^{\theta}$ produce, analogamente a quanto visto in precedenza, un misura con segno sul compatto $[0, e^{\theta_0}]$ avente uguali a zero i momenti di ordine maggiore o uguale a uno. Tale misura con segno deve necessariamente essere concentrata nell'origine e pertanto non può essere che quella identicamente nulla, visto che $\tilde{\theta}$ non vale mai $-\infty$. Se ne deduce che

$$a\, e^{-M(\theta)}\tau(d\theta) = -\int_{-\infty}^{\theta} (M'(y) - b)\, e^{-M(y)}\tau(dy)\, d\theta$$

e quindi che τ è assolutamente continua rispetto alla misura di Lebesgue su Θ, ammettendo densità

$$g(\theta) = -a^{-1}e^{M(\theta)} \int_{-\infty}^{\theta} (M'(y) - b)\, e^{-M(y)}\tau(dy). \tag{9}$$

Derivando la (9) si ottiene l'equazione differenziale

$$-ag'(\theta) = (M'(\theta) - b)g(\theta) - M'(\theta)ag(\theta)$$

che, una volta risolta, fornisce l'espressione desiderata per $\tau(d\theta)$.

*

Si consideri ora un secondo caso notevole, quello della distribuzione binomiale. Nel tale caso, non basterà supporre la linearità della previsione relativa alla singola osservazione per caratterizzare la legge a priori coniugata. Non basterà neppure la linearità della previsione rispetto alla media di un campione di taglia n. Infatti, in quest'ultima situazione, l'uguaglianza

$$\int_0^1 p^{k+1}(1-p)^{n-k}\pi(dp) = (ak + b)\int_0^1 p^k(1-p)^{n-k}\pi(dp)\ \forall k \in \{0, \dots, n\}$$

permette di determinare solamente i primi $n+1$ momenti della legge a priori π. Sarà, quindi, necessario assumere la linearità rispetto alla media di campioni di taglia arbitrariamente grande. I seguenti due teoremi fanno riferimento a questo specifico caso.

Teorema 5 (Ericson, 1969) *Sia $X = (X_1, \dots, X_n)$ un vettore aleatorio avente distribuzione condizionale rispetto a un parametro θ su cui sia imposta una distribuzione a priori π. Sia $E[X_i|\tilde{\Theta}] = m(\tilde{\Theta})$, $V[X_i|\tilde{\Theta}] < +\infty$ e sia $0 < V[m(\tilde{\Theta})] < +\infty$. Inoltre, si supponga che*

$$E[m(\tilde{\Theta})|X] = \alpha\overline{X} + \beta$$

dove α e β non dipendono da X. Allora

$$E[m(\tilde{\Theta})|X] = \frac{\overline{X}\ V[m(\tilde{\Theta})]\ +\ E[m(\tilde{\Theta})]\ E[V[\overline{X}|\tilde{\Theta}]]}{V[m(\tilde{\Theta})]\ +\ E[V[\overline{X}|\tilde{\Theta}]]}. \tag{10}$$

Teorema 6 (Diaconis e Ylvisaker, 1979) *Sia $(X_1, \dots, X_n)$ un campione di taglia n con densità discreta $\binom{n}{x} p^x(1-p)^{n-x}$ condizionatamente al parametro $\tilde{p} = E[X_i|\tilde{\Theta}]$, dove $\tilde{\Theta}$ è il parametro naturale della famiglia esponenziale. Sia π una distribuzione iniziale su $\tilde{p}$ definita su $\mathcal{B}([0,1])$ tale che π non sia concentrata su un singolo punto. Si supponga che esistono due successioni $(a_n)_{n\geq 1}$ e $(b_n)_{n\geq 1}$ in $\Re$ tali che $\forall n \geq 1$ e $\forall k \in \{0, \dots, n\}$*

$$E[\tilde{p}|S_n = k] = a_n k + b_n. \tag{11}$$

Allora

$$a_n = \frac{a}{1+(n-1)a} \qquad b_n = \frac{b}{1+(n-1)a}$$

con $a > 0$, $b > 0$, $a + b < 1$ *e* $\pi = B(b/a, (1-a-b)/a)$.

DIM.

La condizione (11) si traduce in

$$\int_0^1 p^{k+1}(1-p)^{n-k}\pi(dp) = (a_n k + b_n) \int_0^1 p^{k+}(1-p)^{n-k}\pi(dp).$$

Uguagliando (10) e (11), si ottiene

$$a_n k + b_n = \frac{k\ V[\tilde{p}\ ]\ +\ E[\tilde{p}\ ]E[\tilde{p}\ (1-\tilde{p}\ )]}{nV[\tilde{p}\ ]\ +\ E[\tilde{p}\ (1-\tilde{p}\ )]}.$$

Pertanto:

$$a = \frac{V[\tilde{p}\ ]}{E[\tilde{p}\ ]E[1-\tilde{p}\ ]} \qquad b = \frac{E[\tilde{p}\ (1-\tilde{p}\ )]}{E[1-\tilde{p}\ ]}$$

dove a e b non dipendono da n o k. Considerando (11) con $n = 1$ e sommando su $k = 0, 1$, si ottiene

$$\int_0^1 p\ \pi(dp) = b + a \int_0^1 p\ \pi(dp).$$

Quindi $\int_0^1 p\ \pi(dp) = \frac{b}{1-a}$, determinando così il momento primo di $\tilde{p}$. Si consideri nuovamente (11) con $k = n$. Allora

$$\int_0^1 p^{n+1}\ \pi(dp) = \frac{an+b}{1+a(n-1)} \int_0^1 p^n\ \pi(dp) = \cdots = \frac{b}{1-a}\prod_{j=1}^{n}\frac{aj+b}{1+a(j-1)}$$

determinando così tutti i restanti momenti di $\tilde{p}$. Poiché la distribuzione $B(\alpha, \beta)$ ha momenti

$$E[X^{n+1}] = \frac{\alpha(\alpha+1)\cdots(\alpha+n)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+n)} \qquad n = 0, 1, 2\ldots$$

e poiché π ha supporto $[0, 1]$ compatto, la distribuzione a priori è una Beta di parametri $\alpha = b/a$ e $\beta = (1-a-b)/a$ (quindi $\alpha + \beta = \frac{1}{a} - 1$). La legge τ è dunque la legge coniugata al modello esponenziale in questione.

*

Infine, si può considerare il caso di osservazioni con legge di Poisson condizionatamente al parametro, caso trattato da Johnson [5, 6]: la linearità della previsone rispetto a una singola osservazione implica una legge a priori di tipo gamma, cioè coniugata al modello.

Riferimenti bibliografici

[1] O. Barndorff-Nielsen. *Exponential Families - Exact Theory.* Aarhus Universitet, 1970.

[2] L. Brown. *Fundamentals of Statistical Exponential Families.* Institute of Mathematical Statistics, 1986.

[3] P. Diaconis and D. Ylvisaker. Conjugate priors for exponential families. *The Annals of Statistics*, 7:269–281, 1979.

[4] W. Ericson. A note on the posterior mean of population mean. *Journal of the Royal Statistical Society Seires B*, 31:332–334, 1969.

[5] W. Johnson. Uniqueness of a result in the theory of accident proneness. *Biometrika*, 44:430–531, 1957.

[6] W. Johnson. Note on a uniqueness relation in certain accident proneness models. *Journal of American Statistical Association*, 62:288–289, 1967.

[7] A. Kagan, Y. Linnick, and C. Rao. *Characterization Problems in Mathematical Statistics.* Wiley, 1973.