STATISTICA, anno LXII, n. 2, 2002

LA GEOMETRIA DELLE FAMIGLIE DI DISTRIBUZIONI DI PROBABILITÀ ESPONENZIALI

Angela De Sanctis

1. INTRODUZIONE

La Statistica parametrica studia famiglie parametrizzate di distribuzioni di probabilità. I parametri non hanno la funzione esclusiva di indicizzare le distribuzioni di probabilità. Infatti, anche gran parte della Inferenza statistica viene a dipendere dalla parametrizzazione prescelta. Questo è dovuto soprattutto al fatto che, nella stima dei parametri e nella successiva verifica di ipotesi, un ruolo fondamentale è svolto dal Calcolo differenziale. Il problema di come il Calcolo differenziale dipende dalle coordinate è l'argomento centrale di quella branca della Geometria differenziale, chiamata Calcolo sulle varietà (Koboyashi e Nomizu, 1963; O'Neill, 1983).

Una delle famiglie parametrizzate di distribuzioni di probabilità più studiata e applicata nella Statistica parametrica è certamente la famiglia esponenziale, per la cui definizione si utilizza un particolare sistema di coordinate. Il punto è che non si può escludere che una famiglia parametrizzata sia esponenziale solo perché non appare nella forma esponenziale, ma deve essere provato che tale famiglia non può essere parametrizzata nella forma esponenziale. Questo lavoro si propone innanzi tutto di dare una definizione intrinseca di famiglia esponenziale, cioè invariante per riparametrizzazioni. Per fare ciò cerchiamo il significato geometrico della famiglia esponenziale. Vedremo che per tale famiglia la geometria è quella affine, il che spiega il profondo significato dei parametri canonici nella parametrizzazione esponenziale. Essi sono coordinate affini cioè le coordinate più adatte a descrivere la geometria affine.

Ma la Geometria differenziale ci permette di andare oltre. Infatti, sotto opportune condizioni di regolarità, i modelli statistici parametrici hanno una naturale struttura di Varietà Riemanniana con metrica data dalla metrica di informazione di Fisher, introdotta da Rao nel 1945 (come letteratura classica, si vedano: Fisher, 1925; Rao, 1945; Efron, 1975). In questo modo classici invarianti geometrici, come ad esempio la curvatura gaussiana, consentono di definire indici statistici e dedurre quindi proprietà statistiche difficilmente immaginabili altrimenti.

Il presente lavoro, sulla base degli studi effettuati negli ultimi venti anni da parte di ricercatori come Amari, Barndorff-Nielsen, Kass, Lauritzen, ecc. (come riferimenti bibliografici, si vedano: Barndorff-Nielsen, 1978 e 1988; Amari, 1985;

Amari et al., 1987; Burbea, 1986; Burbea e Rao, 1984; Murray e Rice, 1993), fornisce tre caratterizzazioni geometriche differenziali della famiglia esponenziale. Infine, essendo quella affine la geometria più semplice, in quanto a curvatura zero, lo studio da un punto di vista geometrico della famiglia esponenziale risulta propedeutico per affrontare quello di modelli statistici parametrici con geometria più complessa.

2. SPAZI AFFINI E FAMIGLIE ESPONENZIALI

Ricordiamo la definizione di spazio affine.

Un insieme X viene detto spazio affine se esiste uno spazio vettoriale V ad ogni vettore v del quale resta associata una trasformazione $+v$ di X in X, detta traslazione per v, che si indica col simbolo $+v(p) := p+v, \quad \forall p \in X$ e che soddisfa le seguenti proprietà:

i) $(p+v)+w = p+(v+w) \qquad \forall p \in X \qquad v,w \in V$

ii) $\forall p,q \in X$ esiste un unico $v \in V$ tale che $q = p+v$

Come esempio possiamo citare quello in cui $X = R^+, V = R$ e $x+v := \exp(vx), \forall x \in R^+, v \in R$.

Fissato un punto O in X, detto origine, scegliamo una base ordinata $v^1,\dots,v^r$ in V. Sappiamo che per ogni $v \in V$ esistono unici $\theta^1,\dots,\theta^r$ in R tali che $v = \theta^1 v^1 + \dots + \theta^r v^r$. I coefficienti θ^i dipendono da v cioè sono funzioni da V in R e possiamo riguardarli anche come coordinate dei punti di X componendoli con la biiezione $X \to V$ determinata identificando ogni punto $p \in X$ con l'unico $v \in V$ tale che $O = p+v$. Questo processo equivale a fissare un sistema di assi per l'origine in X. $\theta = (\theta^1,\dots,\theta^r)$ è detto un sistema di coordinate affini.

Se θ e Φ sono sistemi di coordinate affini sullo spazio affine X allora esistono una matrice X_j^i e un vettore di coordinate $(u^1,\dots,u^r)$ tali che

$$\theta^i(p) = \sum_{j=1}^{r} X_j^i \Phi^j(p) + u^i \tag{1}$$

infatti $(u^1,\dots,u^r)$ sono le θ-coordinate dell'origine del Φ-sistema e X_j^i è la matrice del cambio di base che esprime la base che produce Φ in termini di quella che produce θ. Due sistemi di coordinate legati secondo (1) si dicono legati in modo affine. Nella teoria degli spazi vettoriali si dimostra anche il viceversa della proposizione precedente cioè se X è un insieme dotato di una collezione di sistemi di coordinate a due a due legati in modo affine allora X è uno spazio affine. La Statistica parametrica studia famiglie di misure di probabilità del tipo

$$P = \left\{ p(\theta) : \theta \in \Omega \subset R^d \right\}.$$

Fra queste la più nota è la famiglia esponenziale che può essere parametrizzata nella forma:

$$p(\theta) = \exp\left(\theta^1 X^1 + \ldots + \theta^r X^r - K(\theta)\right) d\mu,$$

dove $X^1,\ldots,X^r$ sono variabili aleatorie e μ una misura su qualche spazio campionario. I parametri θ^i , i=(1,…,n), vengono detti "parametri canonici". La famiglia normale è una particolare famiglia esponenziale dato che può essere parametrizzata nella forma:

$$p\left(\theta^1, \theta^2\right) = \exp\left(x^2\theta^1 + x\theta^2 - K(\theta)\right) dx$$

con la scelta $\theta^1 = \dfrac{-1}{2\sigma^2}$, $\theta^2 = \dfrac{\mu}{\sigma^2}$, $K(\theta) = \dfrac{1}{2} ln\left(\dfrac{-\pi}{\theta^1}\right) - \dfrac{\left(\theta^2\right)^2}{4\theta^1}$

Il parametro $\theta = (\theta^1, \theta^2)$ è il parametro canonico e giace nel semipiano aperto di R^2 definito da $\theta^1 < 0$. Lo spazio campionario è R con la misura di Lebesgue *dx*. Si dimostra che due insiemi θ e Φ di parametri canonici sono legati dalla relazione:

$$\theta^i = \sum_{j=1}^{r} X_j^i \Phi^j + \xi^i .$$

per qualche X_j^i matrice e vettore ξ. Poiché i parametri canonici formano una collezione di sistemi di coordinate legati in modo affine, si deduce che le famiglie esponenziali sono spazi affini e i parametri canonici sono coordinate affini. Questa può essere assunta come definizione di famiglia esponenziale e, ovviamente, è indipendente dalla scelta dei parametri.

3. LA GEOMETRIA DEGLI SPAZI AFFINI

Vogliamo determinare ora una caratterizzazione geometrica delle famiglie esponenziali, utilizzando le proprietà note degli spazi affini.

Sia Ω uno spazio di misura. Nell'insieme delle misure positive su Ω, si consideri la relazione di assoluta continuità reciproca. È facile vedere che tale relazione è di equivalenza. Indichiamo con M una classe dello spazio quoziente. M è uno spazio affine rispetto allo spazio vettoriale R_Ω delle variabili aleatorie f su Ω con la traslazione per f definita da $d\mu + f := e^f d\mu$. $d\mu$ corrisponde ad una scelta dell'origine in M. Si chiama "log-verosimiglianza" l'applicazione $\ell : M \to R_\Omega$

definita da $\ell\left(pd\mu\right) := \ln p$. Le misure di probabilità costituiscono un sottoinsieme P di M che non è un sottospazio affine in quanto la traslazione non conserva la proprietà che la massa totale è unitaria. In M definiamo la relazione rispetto a cui due misure sono in relazione se si ottengono l'una dall'altra per un fattore di scala. Tale relazione è di equivalenza e lo spazio quoziente continua ad essere uno spazio affine rispetto allo spazio vettoriale quoziente $\frac{R_\Omega}{R}$ in cui una classe è del tipo $f + R,\quad f \in R_\Omega$. Le misure finite di M danno luogo nello spazio quoziente ad un sottoinsieme identificabile con P, poiché dividendo una misura non negativa finita per la sua massa totale si ottiene una misura di probabilità. Una famiglia di misure di probabilità $P = \left\{p\left(\theta\right) : \theta \in \Omega \subset \mathrm{R}^{\mathrm{d}}\right\}$ può essere vista come una superficie in P e quindi in M. È noto che il piano tangente a una superficie in un punto è un sottospazio affine nello spazio affine ambiente. In questo caso ha la forma $\left\{p + v = \exp(v)\,p \, : v \in V\right\}$ dove V è un sottospazio di variabili aleatorie su Ω, che coincide con lo spazio tangente a P in p che si indica con $T_p P$. Si può dimostrare il seguente:

Lemma 1. L'insieme $\left\{\ell_i := \frac{\partial l}{\partial \theta^i}\left(p\right),\quad i = 1,\ldots,n\right\}$ è una base di V.

Posto $\tilde{P} = \left\{\exp(\lambda)\,p \;:\; \lambda \in \mathrm{R} \quad p \in P\right\}$, un sistema di coordinate per $\tilde{P}$ è dato da

$$\begin{cases} \theta^0\left(\exp\left(\lambda\right)p\right) = \lambda \\ \theta^i\left(\exp\left(\lambda\right)p\right) = \theta^i\left(p\right) \end{cases}$$

se θ^i è un sistema di coordinate per P. È facile verificare che P è un sottospazio affine di P generato da un sottospazio vettoriale V di $\frac{R_\Omega}{R}$ se e solo se $\tilde{P}$ è un sottospazio affine di M generato dal sottospazio vettoriale di R_Ω definito da $\tilde{V} = \left\{f \in R_\Omega : \quad f + R \in V\right\}$. È possibile estendere la log-verosimiglianza a $\tilde{P}$ mediante $\tilde{\ell}\left(\exp(\lambda)\,p\right) = \ell\left(p\right) + \lambda$. Ne segue che una base di $\tilde{V}$ è data da $\left\{\tilde{\ell}_0 = 1\,, \tilde{\ell}_i = \ell_i, \qquad i = 1,\ldots,n\right\}$. In Geometria differenziale si dimostra che una varietà differenziale è uno spazio affine se e solo se il piano tangente in ogni punto coincide con la varietà stessa cioè le derivate parziali di una base dello spazio tangente in ogni punto continuano a giacere in esso.

Teorema 1. $\tilde{P}$ è uno spazio affine se e solo se $\ell_{ij} := \frac{\partial^2 \ell}{\partial \theta^i \partial \theta^j}\,,\quad i,j = 1,\ldots,n$, giacciono nello spazio generato da $\ell_i, i = 1,\ldots,n$, e 1.

Dimostrazione. L'osservazione precedente, nella simbologia adottata, si traduce dicendo che $\tilde{P}$ è uno spazio affine se e solo se $\tilde{\ell}_{ij} := \dfrac{\partial^2 \tilde{\ell}}{\partial\theta^i \partial\theta^j} \in \tilde{V}$, $i, j = 1,\ldots,n$, cioè esistono unici coefficienti γ_{ij}^k che consentono di esprimere $\tilde{\ell}_{ij}$ come combinazione lineare della base di $\tilde{V}$ fissata: $\tilde{\ell}_{ij}(\theta) = \sum_{k=0}^{r} \gamma_{ij}^k(\theta) \tilde{\ell}_k(\theta)$. Poiché $\tilde{\ell}_{0i}(\theta) = 0$, si deduce la tesi.

Se $p \in P$, si definisce matrice di informazione di Fisher la matrice $g_{ij}(p) := \mathrm{E}_P(\ell_i, \ell_j)$ dove E_P indica il valore atteso rispetto alla misura *p*.

$g_{ij}(p)$ definisce un prodotto scalare su $T_p P$ e quindi, nel linguaggio della geometria riemanniana, g_{ij} rappresenta i coefficienti della prima forma quadratica fondamentale di una metrica riemanniana su *P*.

La definizione può essere estesa ovviamente a $\tilde{P}$. La caratterizzazione, fornita dal teorema 1, prende in considerazione esclusivamente il rapporto fra la varietà $\tilde{P}$ e il suo spazio tangente, espresso dai coefficienti della prima forma fondamentale, in questo senso è intrinseca cioè ignora l'immersione della varietà nell'ambiente. Guardiamo ora $\tilde{P}$ nella varietà ambiente *M*.

$g_{ij}(p)$ si può intendere come la restrizione a $T_p \tilde{P}$ del prodotto scalare $\langle f, g \rangle_p = \mathrm{E}_p(fg)$ sullo spazio vettoriale R_Ω, supponendo per comodità che tutte le variabili aleatorie su Ω siano di quadrato integrabili. Indichiamo con N_p lo spazio normale a $T_p \tilde{P}$ rispetto al suddetto prodotto scalare , quindi R_Ω si decompone nella somma diretta $R_\Omega = N_P \oplus T_P \tilde{P}$.

Vale il seguente :

Lemma 2. Se $f \in R_\Omega$ la sua componente normale in N_p è $\pi_p(f) = f - \sum_{m,n} g^{mn} \mathrm{E}_P(f\ell_m)\ell_n - \mathrm{E}_p(f)$ dove g^{mn} è l'inversa della matrice d'informazione di Fisher.

Dal teorema 1 segue facilmente

Teorema 2. $\tilde{P}$ è uno spazio affine se e solo se $\alpha_{ij} := \pi_p(\ell_{ij}) = 0$ per ogni $i, j = 1,\ldots,n$.

Tale condizione esprime il fatto che è nulla la componente delle derivate seconde perpendicolare allo spazio tangente cioè $\tilde{P}$ è "piatta" in *M*.

In Geometria differenziale la curvatura gaussiana è un indice sintetico della metrica della varietà riemanniana.

In particolare i coefficienti della seconda forma fondamentale, espressi da α_{ij}, sono identicamente nulli se e solo se la curvatura è identicamente nulla. Vale allora:

Teorema 3. $\tilde{P}$ è uno spazio affine se e solo se si annulla la quantità scalare $\gamma(p) = \sum_{i,j,k,\ell} g^{ij}(p) g^{k\ell}(p) \mathrm{E}_P\left(\alpha_{ik}(p)\alpha_{j\ell}(p)\right)$

Dimostrazione. Poiché l'inversa della matrice di Fisher $g^{ij}(p)$ è anch'essa definita positiva, γ si annulla se e solo se si annullano tutti gli α_{ij}.

Dipartimento di Scienze
Università degli Studi "G. d'Annunzio" di Chieti

ANGELA DE SANCTIS

RIFERIMENTI BIBLIOGRAFICI

S.I. AMARI (1985), *Differential-geometric Methods in Statistics*, "Lecture Notes in Statistics", Vol. 28, Springer-Verlag, Berlin.

S.I. AMARI, O.E. BARNDORFF-NIELSEN, R.E. KASS, S.L. LAURITZEN, C.R. RAO (1987), *Differential Geometry in Statistical Inference*, "Lecture Notes Monograph Series", Vol. 10, Institute of Mathematical Statistics, Hayward California.

O.E. BARNDORFF-NIELSEN (1978), *Information and exponential families in statistical theory*, Wiley, New York.

O.E. BARNDORFF-NIELSEN (1988), *Parametric Statistical Models and Likelihood,* "Lecture Notes in Statistics" 50*,* Springer-Verlag, Berlin.

J. BURBEA (1986), *Informative geometry of probability space*, "Experimental Mathematics", 4.

J. BURBEA, J. DEL CASTILLO (1992), *Geodesic submanifolds of statistical models with location parameters,* "Computational Statistics and Data Analysis", Vol. 14, North-Holland, Amsterdam.

J. BURBEA, C.R. RAO (1984), *Differential metrics in probability spaces.* "Probability and Mathematical Statistics", 3.

D.R. COX, D.V. HINKLEY (1974), *Theoretical Statistics*, Chapman and Hall, London.

A. DE SANCTIS (1996), *Geodesic submanifolds of a family of statistical models,* "Statistics & Probability Letters" 30.

B. EFRON (1975), *Defining the curvature of a statistical problem* (*with applications to second order efficiency*) (*with discussion*), "Annals of Statistics" 3.

R.A. FISHER (1925), *Theory of statistical estimation*, "Proceedings of Cambridge Philosophical Society" 22.

S. KOBOYASHI, K. NOMIZU (1963), *Foundations of Differential Geometry,* Interscience, Wiley, New York.

M.K. MURRAY, J.W. RICE (1993), *Differential Geometry and Statistics*, Monographs on Statistics and Applied Probability, 48, Chapman and Hall, London.

B. O'NEILL (1983), *Semi-Riemannian Geometry (with Applications to Relativity)*, Academic Press, New York.

C.R. RAO (1945), *Information and accuracy attainable in the estimation of statistical parameters*, "Bullettin of Calcutta Mathematical Society" 37.

RIASSUNTO

La geometria delle famiglie di distribuzioni di probabilità esponenziali

Sono state dimostrate quattro caratterizzazioni della famiglia esponenziale: la prima è una caratterizzazione algebrica come spazio affine, la seconda è una caratterizzazione geometrica intrinseca (teorema 1), la terza è una caratterizzazione geometrica come varietà immersa (teorema 2), la quarta è una caratterizzazione scalare (teorema 3).

SUMMARY

The geometry of exponential families of probability distributions

We proved four characterisations of the exponential family: the first is an algebrical characterisation as affine space, the second is a geometrical intrinsic characterisation (theorem 1), thirdly there is a geometrical characterisation as immersed manifold (theorem 2) and finally there is a scalar characterisation (theorem 3).