Butlletí de la Societat Catalana de Matemàtiques
Vol. 11, núm. 1, 1996. Pàgs. 0–0.

Geometria i estadística

José Manuel Corcuera Valverde

1 La geometria en l'anàlisi de dades

És ben conegut i resulta bastant evident el paper que la geometria pot tenir en la representació i l'anàlisi de dades. No resulta tan obvi, en canvi, el que pot tenir la geometria diferencial. No obstant això, importants resultats de l'estadística actual necessiten per a la seva formulació de la geometria diferencial. Anem a veure el paper que desenvolupa la geometria en l'Estadística i com la geometria diferencial s'ha anat introduint en ella.

Imaginem una variable $X = (X_1, X_2, \ldots, X_n)$ de la qual s'observen diferents valors $X^{(1)}, X^{(2)}, \ldots, X^{(s)}$. Cada valor $X^{(i)}$ correspon a un punt de l'espai n–dimensional. Un problema habitual en l'anàlisi de dades consisteix a trobar representacions de menor dimensió, bé per raons de visualització gràfica o bé perquè se suposa que el fenòmen d'estudi pot ser descrit en una dimensió menor. Sorgeix així, de manera natural, la idea de projectar sobre espais de dimensió inferior de manera que es deformi el mínim la representació original. Necessitem una mètrica per a avaluar la deformació.

$d^2(i,j) = \sum_{h=1}^{n} (X_h^{(i)} - X_h^{(j)})^2$ (distància euclidiana)
$d^2(i,j) = (X^{(i)} - X^{(j)})'\Sigma^{-1}(X^{(i)} - X^{(j)})$ (distància de Mahalanobis (1936) [23])
(Σ: matriu de variàncies–covariàncies de X)
$d_r(i,j) = \left(\sum_{h=1}^{n} |X_h^{(i)} - X_h^{(j)}|^r\right)$, $1 \leq r < \infty$ (distàncies de Minkowski).

De vegades, per exemple, quan les variables són dicotòmiques, es parteix d'una matriu de dissimilituds i es busca una representació de s punts $P_1, P_2, \ldots, P_s$ tal que les seves distàncies euclidianes concordin raonablement amb la informació proporcionada per la matriu de dissimilituds (*multidimensional scaling*). Això també es fa quan la distància de partida no és pas euclidiana.

Si tenim una representació euclidiana podem tractar de trobar les $m < n$ coordenades cartesianes més rellevants:

$$d_Y^2(i,j)_m = \sum_{h=1}^{m} (Y_h^{(i)} - Y_h^{(j)})^2, \quad \sum_{i,j=1}^{s} d_Y^2(i,j)_m = \text{ màxima.}$$

Això es coneix com l'*anàlisi de components principals*, que va ser introduïda per Hotteling el 1933 [18].

Un altre problema geomètric que es presenta en l'anàlisi de dades és el de la representació simultània dels s individus i les n variables, $X_1, X_2, \dots, X_n$. Això va donar origen al mètode *biplot* (Gabriel (1971) [16]), que consisteix a descompondre la matriu de dades $D = (X^{(1)}, X^{(2)}, \dots, X^{(s)})'$, $s \times n$ de rang r de la forma

$$D = GH', \quad G \text{ matriu } s \times r \text{ i } H \; n \times r$$

i a utilitzar els s vectors fila de G com a representants dels individus i els n de H per representar les variables. La descomposició anterior utilitza la descomposició en valors singulars de D:

$$D = U\Sigma V'$$

on U i V són matrius unitàries $s \times s$ i $n \times n$, respectivament, i Σ és una *matriu diagonal* $s \times n$.

Per acabar, en els problemes de *classificació* i *discriminació*, les regles basades en distàncies (Matusita (1956) [24]) han fet i fan un paper important.

1.1 Però,..., per què la geometria diferencial?

Els individus, les poblacions i les variables es representen en un espai lineal i els problemes geomètrics que apareixen es resolen amb tècniques de l'àlgebra lineal. No obstant això, si el conjunt de poblacions constitueix un model paramètric $M = \{p(x;\theta)\ \theta \in \Theta \text{ obert de } \mathbb{R}^q\}$ (on $p(x;\theta)$ són funcions de densitat respecte d'alguna mesura de referència) té sentit estudiar la variació de la població $p(x;\theta)$ quan realitzo una modificació infinitesimal del paràmetre ($d\theta$).

El 1945 Rao [29] en el seu article *Information and Accuracy Estimation of Statistical Parameters*, a més de derivar la important desigualtat coneguda amb el nom de desigualtat de Cramér–Rao i que per al cas unidimensional i d'un estimador no esbiaixat de $f(\theta)$, T_n, basat en n observacions independents, ve donat per:

$$V(T_n) \geq \frac{(f'(\theta))^2}{nI(\theta)}, \quad (E(T_n) = f(\theta)), \quad I(\theta) = E\left(\left(\frac{d\log p(x;\theta)}{d\theta}\right)^2\right)$$

a part d'això, diem, introdueix en tres apartats finals: *The population space*, *The distance between two populations* i *Distance in tests of significance and classification*, una distància riemanniana i el seu ús en inferència i anàlisi de dades. Rao procedeix de la manera següent.

Les nostres poblacions vénen caracteritzades per

$$p(x;\theta), \quad \theta = (\theta_1, \theta_2, \dots \theta_q) \quad \text{(paràmetres poblacionals)}$$

llavors, variant θ obtenim un conjunt de poblacions que pot ser representat en un espai de q dimensions (espai de les poblacions). Dues poblacions contigües difereixen (excepte ordres superiors) en:

$$dp = \sum_{i=1}^{q} \frac{\partial p}{\partial \theta_i} d\theta_i$$

i si considerem la discrepància relativa

$$\frac{dp}{p} = \sum_{i=1}^{q} \frac{1}{p} \frac{\partial p}{\partial \theta_i} d\theta_i = \sum_{i=1}^{q} \frac{\partial \log p}{\partial \theta_i} d\theta_i$$

i finalment la seva variància, tindrem la forma quadràtica diferencial, definida positiva:

$$ds^2 = V\left(\frac{dp}{p}\right) = \sum_{i,j=1}^{q} E\left(\frac{\partial \log p}{\partial \theta_i}\frac{\partial \log p}{\partial \theta_j} d\theta_i d\theta_j\right) = \sum_{i,j=1}^{q} g_{ij} d\theta_i d\theta_j$$

amb

$$g_{ij} = E\left(\frac{\partial \log p}{\partial \theta_i}\frac{\partial \log p}{\partial \theta_j}\right).$$

Aquesta forma quadràtica té importants propietats:

- És invariant sota transformacions bijectives (i mesurables) de les dades.
- És invariant sota canvis (suaus) en la parametrització de les poblacions.

Així, g_{ij} constitueix un tensor covariant simètric de segon ordre que proporciona una mètrica riemanniana a l'espai de poblacions.

Per al cas de poblacions normals de mitjana i variàncies desconegudes

$$p(x; m, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-m)^2}{\sigma^2}\right\}$$

(m, σ), $m \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$, tindrem

$$\begin{aligned}
g_{11} &= E\left(\left(\frac{\partial}{\partial m} \log p(x; m, \sigma)\right)^2\right) = E\left(\frac{(x-m)^2}{\sigma^4}\right) = \frac{1}{\sigma^2} \\
g_{12} &= E\left(\frac{\partial}{\partial m} \log p(x; m, \sigma) \frac{\partial}{\partial \sigma} \log p(x; m, \sigma)\right) \\
&= E\left(\frac{x-m}{\sigma^2}\left(-\frac{1}{\sigma} + \frac{1}{\sigma^3}(x-m)^2\right)\right) = 0 \\
g_{22} &= E\left(\frac{1}{\sigma^2} + \frac{1}{\sigma^3}(x-m)^4 - \frac{2}{\sigma^4}(x-m)^2\right) \\
&= \frac{1}{\sigma^2} + \frac{3\sigma^4}{\sigma^6} - 2\frac{\sigma^2}{\sigma^4} = \frac{2}{\sigma^4}
\end{aligned}$$

de manera que

$$ds^2 = \frac{(dm)^2}{\sigma^2} + \frac{2(d\sigma)^2}{\sigma^2},$$

calculant la geodèsica que uneix les dues poblacions (m_1, σ_1), (m_2, σ_2), s'obté la distància entre elles (Atkinson i Mitchell (1981) [2], Burbea i Rao (1982) [9]):

$$s = 2\sqrt{2} \tanh^{-1}\left\{\frac{(m_1 - m_2)^2 + 2(\sigma_1 - \sigma_2)^2}{(m_1 - m_2)^2 + 2(\sigma_1 + \sigma_2)^2}\right\}.$$

En el cas de poblacions normals multivariants:

$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}$$

resulta

$$ds^2 = (d\mu)'\Sigma^{-1}(d\mu) + \frac{1}{2}\operatorname{tr}\left\{(\Sigma^{-1}d\Sigma)^2\right\}.$$

Si Σ és coneguda, la distància resultant és la de Mahalanobis:

$$s = (\mu_1-\mu_2)'\Sigma^{-1}(\mu_1-\mu_2),$$

també està resolt el cas μ coneguda (Atkinson i Mitchell (1981) [2]) però continua obert el problema de quina és la distància en el model general. Els avenços més importants es deuen a Eriksen (1986) [15] i Calvo i Oller (1991) [10].

2 Un altre precedent

El 1946 Jeffreys [19] en el seu article *An invariant form for the prior probability in estimation problems* suggereix utilitzar el volum riemannià:

$$|(g_{ik})|^{\frac{1}{2}} d\theta_1 \cdot d\theta_2 \cdots \theta_q, \quad g_{ik} = E\left(\frac{\partial \log p}{\partial \theta_i}\frac{\partial \log p}{\partial \theta_k}\right).$$

com a mesura *a priori* en l'espai de paràmetres, en el cas de total desconeixement de θ. Notem que resulta ingenu suposar la densitat uniforme en uns paràmetres com a sinònim de desconeixement total:

$$\tilde\theta = g(\theta) \Longrightarrow f_{\tilde\theta}(\tilde\theta) = f_\theta(g^{-1}(\tilde\theta))|J_{g^{-1}}(\tilde\theta)| = K 1_{g(\Theta)}(\tilde\theta)|J_{g^{-1}}(\tilde\theta)|$$

En canvi, la mesura anterior és invariant sota reparametritzacions i, a més a més, si el model presenta certes *simetries* també les presenta la mesura anterior.

Si el model és invariant sota l'acció d'un grup G, de transformacions, el volum riemannià resulta invariant sota les transformacions induïdes en l'espai de paràmetres.

Invariància sota G: $\forall g \in G$ si $X \sim p(x;\theta)$, aleshores $gX \sim p(x;\bar g\theta)$.
Grup induït $\bar G$:

$$\begin{aligned} g \mapsto \bar g : \Theta &\to \Theta \\ \theta &\mapsto \bar g\theta. \end{aligned}$$

Invariància del volum riemannià V: Si $A \subset \Theta$, llavors $V(A) = V(\bar g a)$.

El treball de Jeffreys ha tingut la seva continuació en àmbits bayesians amb les aportacions de Lindley (1961) [21], Hartigan (1964) [17], Villegas (1971) [30] i Bernardo (1975, 1979) [7, 8].

3 Nous impulsos

El 1975 Efron [13] publica *Defining the curvature of a statistical problem (with applications to second order efficiency).* En aquest treball s'analitza el comportament asimptòtic dels estimadors en el context de subfamílies uniparamètriques de famílies exponencials

$$p(x;\theta(u)) = \exp\left\{\sum_{i=1}^{q} \theta_i(u)x_i - \psi(\theta(u))\right\}$$

$\theta_1, \theta_2, \ldots, \theta_q$ són els denominats *paràmetres naturals* i $u \in I$ interval de $\mathbb{R}$.

Escrivim

$$S = \{p(x;\theta),\ \theta \in \Theta \subset \mathbb{R}^q\}, \qquad M = \{p(x;\theta(u)),\ u \in I \subset \mathbb{R}\}.$$

Si considerem estimadors de u que depenguin de la mostra $X_1, X_2, \ldots, X_n$, mitjançant $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i$, tindrem, com veurem, una *foliació* de S associada amb cada estimador, $\hat{u} = \phi(\bar{x})$. L'aplicació

$$\eta : \bar{x} \mapsto \tilde{\theta}, \quad E_{\tilde{\theta}}X = \bar{x}$$

identifica l'espai mostral i S, a més a més, com $\hat{u} = \phi(\eta^{-1}(\tilde{\theta})) = \alpha(\tilde{\theta})$, per cada punt $\theta(u) \in M$ tindrem la subvarietat

$$A(u) = \{\theta \in S,\ \alpha(\theta) = u\}.$$

Es demana que l'estimador sigui consistent: $\theta(u) \in A(u)$. En aquestes condicions $\hat{u}$ és asimptòticament eficient si i només si $A(u)$ és ortogonal (amb la mètrica de Fisher) a M. A més, la variància d'un estimador asimptòticament eficient ve donada per:

$$V(\hat{u}) = \frac{1}{nI(u)} + \frac{1}{n^2}\left\{\gamma_u^2 + 4\frac{\Gamma_u^2}{I(u)} + \Delta_u^{\hat{u}}\right\} + 2\frac{\dot{b}_u}{nI(u)} + o(\frac{1}{n^2}).$$

Amb la denominació d'Efron, γ_u^2 és la *curvatura estadística* i $4\frac{\Gamma_u^2}{I(u)}$ la *naming curvature.* $\Delta_u^{\hat{u}}$ és una quantitat no negativa que s'anul·la per l'estimador del màxim de versemblança. Per acabar, $2\dot{b}_u = \frac{d}{du}(E\hat{u} - u)$.

Amari (1982) [1] va explicar les tres quantitats anteriors en termes geomètrics. Si representem l'espai tangent en θ, S_θ, amb base e_i, $i = 1, 2, \ldots, q$, de la forma

$$e_i \longleftrightarrow \frac{\partial \log p(x;\theta)}{\partial \theta_i} \quad i = 1, 2, \ldots, q,$$

aquesta representació indueix una mètrica en S de forma natural:

$$g_{ij}(\theta) = E\left(\frac{\partial \log p(x;\theta)}{\partial \theta_i}\frac{\partial \log p(x;\theta)}{\partial \theta_j}\right)$$

Com definir el *canvi* de e_i? Si tenim en compte que

$$E\left(\frac{\partial \log p(x;\theta)}{\partial \theta_i}\right) = 0, \text{ i que } E\left(\frac{\partial^2 \log p(x;\theta)}{\partial \theta_i \partial \theta_j}\right) = -g_{ij}(\theta)$$

resulta que

$$l_{ij}(x;\theta) = \frac{\partial^2 \log p(x;\theta)}{\partial \theta_i \partial \theta_j} \notin S_\theta.$$

Podem tractar de modificar la quantitat anterior perquè tingui mitjana zero:

$$l_{ij}^{(\alpha)} = \frac{\partial^2 \log p(x;\theta)}{\partial\theta_i \partial\theta_j} + \frac{1+\alpha}{2} g_{ij}(\theta) + \frac{1-\alpha}{2} \frac{\partial \log p(x;\theta)}{\partial\theta_i} \frac{\partial \log p(x;\theta)}{\partial\theta_j},$$

finalment, si projectem sobre S_θ tindrem les components d'una α-connexió afí:

$$\Gamma_{ijk}^{(\alpha)} = E\left[l_{ij}^{(\alpha)} \frac{\partial \log p(x;\theta)}{\partial\theta_k} \right].$$

Aquestes connexions són no riemannianes, excepte per a $\alpha = 0$, i ja havien estat introduïdes per Čencov (1972) [11] per al cas de distribucions multinomials, demostrant que són les úniques invariants sota transformacions de Markov.

Així, γ_u^2 es pot expressar en termes de la *imbedding curvature* de M en S quant a la 1–connexió, $\Delta_u^{\hat{u}}$ en termes de la *imbedding curvature* de A en S per a la -1–connexió i $4\frac{\Gamma_u^2}{I(u)}$ en termes del símbol de Christoffel en M per a la -1-connexió.

De la mateixa forma interpreta el cas en què M sigui multidimensional (Madsen (1979) [22]). Aquests treballs *culminen* els de Rao (1961 [26], 1962, [28], 1963 [27]).

Un altre problema ha produït un fort desenvolupament de la geometria diferencial en estadística: la descomposició invariant dels ajustaments de Bartlett (1937) [6]. Si w indica l'estadístic *logaritme de la raó de versemblances* per contrastar $\theta = \theta_0$ contra $\theta \neq \theta_0$:

$$w = 2\left\{\log p(x;\hat{\theta}) - \log p(x;\theta_0)\right\} \quad (\hat{\theta} = \text{ MLE})$$

amb gran generalitat w, sota la hipòtesi $\theta = \theta_0$, té aproximadament una distribució χ_q^2. Aleshores és possible ajustar w a un nou estadístic

$$w' = w/(1 + R/q),$$

on R és el denominat ajustament de Bartlett, que segueix de nou una χ_q^2 però amb major grau d'aproximació. $R \approx E_\theta(W) - q$ i s'han proposat diverses aproximacions per R: Lawley (1956) [20], Barndorff-Nielsen i Cox (1984) [5] i Barndorff-Nielsen (1986) [3]. En tots els casos, R és una quantitat escalar suma de termes que no són escalars. McCullagh i Cox (1986) [25] aconsegueixen descompondre l'expressió per R donada per Lawley(1956) [20] en suma de sis escalars que tenen una interpretació geomètrica en termes de la mètrica de Fisher. Per la seva part, Barndorff-Nielsen (1986) [4] obté una expressió per R suma de quatre escalars que s'interpreten amb una nova mètrica, *la informació observada*. Sigui $\hat{\theta}$ l'estimador del màxim de versemblança de θ i a un estadístic auxiliar tal que $(\hat{\theta}, a)$ sigui un estadístic suficient minimal, llavors podem escriure:

$$\log p(x;\theta) = \log h(\hat{\theta}, a; \theta) + C(x)$$

i

$$-\left. \frac{\partial^2 \log h(\hat{\theta}, a; \theta)}{\partial\theta_i \partial\theta_j} \right|_{\theta=\hat{\theta}} = J_{ij}(\hat{\theta}),$$

$J_{ij}(\hat{\theta})$ serà un matriu simètrica definida positiva que es transforma com un tensor covariant de segon ordre i que proporciona una mètrica a la varietat de densitats, tot donant lloc a la denominada geometria observada. La profundització en l'estudi dels objectes geomètrics que apareixen en una i altra geometria han portat Barndorff-Nielsen i un gran grup de col·laboradors a desenvolupar les nocions de *yoke*, *string*, *connection string*, derivada tensorial.

4 L'anàlisi intrínseca

En el període 1991-1994, Oller i Corcuera [12] desenvolupen una teoria de l'estimació puntual que fa compatible les nocions de biaix i error quadràtic mitjà amb la invariància sota reparametritzacions.

Un estimador $\mathcal{U}$ és una aplicació mesurable de l'espai mostral χ en la varietat q-dimensional de les densitats:

$$\mathcal{U} : \chi \to S = \{p(x;\theta), \theta \in \Theta \subset \mathbb{R}^q\}\,.$$

Ara, la mitjana i la variància de $\mathcal{U}$ manquen de significat (S no és pas un espai lineal). Quines mesures de centralització i dispersió alternatives podem utilitzar?

Si tenim una connexió afí en la varietat S, podem definir el *valor mitjà* d'$\mathcal{U}$:

$$m \in S, \qquad \int_\chi \exp_m^{-1} \mathcal{U}(x)\, p(x;\theta)\mu(dx) = 0_m.$$

on

$$\begin{aligned} \exp_m : S_m &\to S \\ v &\mapsto \exp_m(v) = \gamma_v(1) \end{aligned}$$

sent γ_v la geodèsica amb origen m i vector tangent en m, v.

Aquest valor mitjà és el que Emery i Mokobodzki [14] denominen *baricentre exponencial*.

A més la *distància de Rao* ens permet donar una mesura de dispersió en S:

$$\int_\chi \rho^2(\mathcal{U}(x), p)\, p(x;\theta)\mu(dx)$$

que podem denominar distància de Rao quadràtic mitjà, on $\rho^2(\mathcal{U}(x), p)$ és la distància de Rao entre l'estimació $\mathcal{U}$ i el veritable valor p (punt corresponent a $p(x;\theta)$).

Si utilitzem la connexió riemanniana associada a la distància de Rao, tenim:

$$E\left(\rho^2(\mathcal{U}, p)\right) \geq \frac{\{\div B - E(\div A)\}^2}{nq} + \|B\|^2,$$

on $A = \exp_p^{-1}(\mathcal{U}(x))$, $B = E(A)$ (biaix intrínsec) i $\div\cdot$ és l'operador divergència, això és, la traça de la derivada covariant.

Podem donar cotes de $\div(A)$, utilitzant el teorema de comparació de Bishop i a partir d'aquí cotes per a $E(\rho^2(\mathcal{U}, p))$, que depenen de les curvatures seccionals de S. En el cas euclidià:

$$E\left(\rho^2(\mathcal{U}, p)\right) \geq \frac{(\div B + q)^2}{nq} + \|B\|^2.$$

Igualment, s'han obtingut cotes pel *risc global* d'un estimador $\mathcal{U}$ utilitzant la distribució *a priori* de Jeffreys. En el cas euclidià:

$$\frac{1}{\mathrm{vol}(S_R)} \int_{S_R} E\left(\rho^2(\mathcal{U}_k, p)\right)\, V(dp) \geq \frac{q}{n}\left(1 - \frac{{}_0F_1\left(q/2+1; nqR^2/4\right)}{{}_0F_1\left(q/2; nqR^2/4\right)}\right)$$

on S_R és l'esfera q–dimensional de radi R i

$$ {}_0F_1(a;z) = \sum_{j=0}^{\infty} \frac{z^j}{a(a+1)\cdots(a+j-1)\cdot j!}. $$

Clàssicament, la *blackwellització* d'un estimador millora la seva precisió. En el nostre context, l'anàleg de l'esperança condicionada serà el valor mitjà condicionat. Definirem el valor mitjà condicionat de $\mathcal{U}$ amb respecte a $\mathcal{D}$ com una aplicació $\mathcal{D}$-mesurable, Z, tal que

$$ E(\exp_Z^{-1}(\mathcal{U}(\cdot))|\mathcal{D}) = 0_Z. $$

Escriurem $\mathfrak{M}\,(\mathcal{U}|\mathcal{D}) = Z$.

Si D és una σ-àlgebra suficient en $\mathcal{X}$ i, si fixat p, $\rho^2(\cdot,p)$ és una funció convexa en S:

$$ E_p(\rho^2(\mathcal{U}_k,p)) \geq E_p\left(\rho^2(\mathfrak{M}\,(\mathcal{U}|\mathcal{D}),p)\right). $$

L'estudi del comportament asimptòtic d'un estimador, que ha requerit l'ús de desenvolupaments tensorials de Taylor, completa el treball realitzat:

- Sota certes condicions de regularitat l'estimador del màxim de versemblança (MLE) és eficient asimptòticament i intrínsecament:

$$ \lim_{n\to\infty} nE(\rho^2(\mathcal{U},p)) = q $$

- Per a qualsevol estimador $\mathcal{U}$ eficient (asimptòticament i intrínsecament), el vector de biaix $B_p = E_p(\exp_p^{-1}(\mathcal{U}))$ té l'expressió asimptòtica

$$ [B\,(\mathcal{U})]^i = S^i(\mathcal{U}) - \frac{1}{4n} g^{lm} g^{ij} T_{jlm} + o(n^{-1}) $$

  on $S^i(\mathcal{U}_k) = o(n^{-1/2})$ i s'anul·la per l'estimador del màxim de versemblança. Per la seva part, T_{jlm} és el tensor de asimetria,

$$ T_{jlm} = E\left(\frac{\partial l}{\partial\theta^j}\frac{\partial l}{\partial\theta^r}\frac{\partial l}{\partial\theta^m}\right); \quad l = \log p(x;\theta) $$

Referències

[1] S. Amari: *Differential-Geometrical Methods in Statistics*, Lecture notes in statistics 28, Springer Verlag, Nova York, 1985.

[2] C. Atkinson i A. F. S. Mitchell: *Rao's distance measure*, Sankhyá, 43 A:345–365, 1981.

[3] O. E. Barndorff-Nielsen: *Likelihood and observed geometries*, Ann. Statist. 14:856–873, 1986.

[4] O. E. Barndorff-Nielsen: *Strings, tensorial combinants, and bartlett adjustments*, Proc. Roy. Soc. Londres, 406 A:127–137, 1986.

[5] O. E. Barndorff-Nielsen i D. R. Cox: *Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood*, J. Royal. Statist. Soc. 46:483–495, 1984.

[6] M. S. Bartlett: *Properties of sufficiency and statistical tests*, Proc. R. Soc. London, 160:268–282, 1937.

[7] J. M. Bernardo: *Non–informative prior distributions: a subjectivist approach*, Bull. Internat. Statist. Inst., 46:94–97, 1975.

[8] J. M. Bernardo: *Reference posterior distributions for bayesian inference*, J. R. Statist. Soc., 41:113–147, 1979.

[9] J. Burbea i C. R. Rao: *Entropy differential metric, distance and divergence measures in probability spaces: a unified approach*, J. Multivariate Anal., 12:575–596, 1982.

[10] M. Calvo i J. M. Oller: *An explicit solution of information geodesic equations for the multivariate normal model*, Statistics and Decisions, 9:119–138, 1990.

[11] N. N. Čencov: *Statistical Decision Rules and Optimal Infererence*, Nauka, Moscou (en rus, versió anglesa, 1982, Math. Monographs, 53, Amer. Math. Soc., Providence), 1972.

[12] J. M. Corcuera: Tesis Doctoral, *Análisis intrínseco de la estimación puntual*, Universitat de Barcelona, 1994.

[13] Bradley Efron: *Defining the curvature of a statistical problem (with applications and second order efficiency)*, Ann. Statist., 3:1189–1242, 1975.

[14] M. Emery i G. Mokobodzki: *Séminaire de Probabilités XXV*, chapter Sur le barycentre d'une probabilité dans une variété, pàg. 220–233. Vol. 1485 of Lecture Notes in Maths, Springer, 1991.

[15] P. S. Eriksen: *Geodesics connected with the Fisher metric on the multivariate normal manifold*. Preprint R86-13, Institute of Electronic Systems, Aalborg University Center, Dinamarca, 1986.

[16] K. R. Gabriel: *The biplot graphic display of matrices with application to principal component analysis*, Biometrika, 58:453–467, 1971.

[17] J. Hartigan: *Invariant prior distributions*, Ann. Math. Statist., 35:836–845, 1964.

[18] H. Hotelling: *Analysis of a complex of statistical variables into principal components*, J. Educ. Psych., 24:417–441, 498–520, 1933.

[19] H. Jeffreys: *An invariant form for the prior probability in estimation problems*, Proc. Royal Soc., 186 A:453–461, 1946.

[20] D. N. Lawley: *A general method for approximating to the distribution of the likelihood ratio criteria*, Biometrika, 43:295–303, 1956.

[21] D. V. Lindley: *The use of prior probability distributions in statistical inference and decisions*, Proc. Berkeley Simp., 1:436–468, 1961.

[22] L. T. Madsen: *The geometry of statistical models. A generalitation of curvature*, Research report, 79-1, Danish Medical Research Council, Statistical Research Unit, 1979.

[23] P. C. Mahalanobis: *On the generalized distance in statistics*, Proc. Nat. Inst. Sci. India, 2(1):49–55, 1936.

[24] K. Matusita: *Decision rule, based on the distance, for the classification problem*, Ann. Inst. Statist. Math., 8:67–77, 1956.

[25] P. McCullagh i D. R. Cox: *Invariants and likelihood ratio statistics*, Ann. Statist., 14:1419–1430, 1986.

[26] C. R. Rao: *Asymptotic efficiency and limiting information*, (J. Neyman, ed.), Proc. Fourth Berkeley Symp. Math. Statist. Prob., 1:531–545, 1961.

[27] C. R. Rao: *Criteria of estimation in large samples*, Sankhyá, 25:189–206, 1963.

[28] C. R. Rao: *Efficient estimates and optimum inference procedures in large samples, with discussion*. J. Royal. Statist. Soc., 24:46–72, 1961.

[29] C. R. Rao: *Information and accuracy attainable in the estimation of statistical parameters*, Bull. Calcutta Math. Soc., 37:81–91, 1945.

[30] C. Villegas: *On haar priors*, Foundations of Statistical Inference, 409–414, 1971.

Departament d'Estadística
Universitat de Barcelona
Gran Via de les Corts Catalanes, 585
08007 Barcelona
corcuera@cerber.mat.ub.es