PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 1, Fasc. 1 (1980), p. 95-98

A DISTANCE THEOREM FOR EXPONENTIAL FAMILIES

BY

BRADLEY EFRON (STANFORD, CALIFORNIA)

Abstract. Two members of an exponential family can be represented as two points in the natural parameter space of that family or as two points in the expectation parameter space. The theorem describes a simple relation between the interpoint distances in the two spaces, also relating to the symmetrical Kullback-Leibler distance between the two distributions.

1. The Theorem. A k-dimensional exponential family $\mathcal{G}$ has density functions of the form

$$g_\alpha(x) \equiv e^{\alpha' x - \psi(\alpha)}, \qquad \alpha \in A,$$

with respect to some carrier measure $\nu(x)$ on the sample space $\mathcal{X}$, which is contained in $\mathcal{R}^k$. Here α is the *natural parameter* and A is the *natural parameter space*, the convex set in $\mathcal{R}^k$ consisting of all α having the normalizing function

$$e^{\psi(\alpha)} \equiv \int e^{\alpha' x}\, d\nu(x)$$

less than infinity. In order that $\mathcal{G}$ not reduce to a lower-dimensional exponential family, we assume that $\nu(x)$ does not concentrate all its mass on any $(k-1)$-dimensional hyperplane.

The expectation vector and the covariance matrix of x,

$$\beta \equiv \mathrm{E}_\alpha x \quad \text{and} \quad \Sigma_\alpha \equiv \mathrm{Cov}_\alpha x,$$

exist finitely in the interior of A and can be obtained by differentiation of ψ. Let V_α be the gradient operator

$$\left(\frac{\partial}{\partial\alpha(1)}, \frac{\partial}{\partial\alpha(2)}, \ldots, \frac{\partial}{\partial\alpha(k)}\right)'.$$

Then

$$\nabla_\alpha \psi = \beta \quad \text{and} \quad \nabla_\alpha \psi \nabla_\alpha' = \Sigma_\alpha. \tag{1.1}$$

This last result can be written as $\beta \nabla_\alpha' = \Sigma_\alpha$ or, more evocatively, as

$$d\beta = \Sigma_\alpha \, d\alpha. \tag{1.2}$$

The mapping $\alpha \to \beta$, from the natural parameter α to the *expectation parameter* β, is one-to-one inside A, mapping A into the *expectation parameter space* B (not necessarily convex; see [2]). The matrix Σ_α, which is positive definite, will also be denoted by Σ_β when convenient, this being understood to mean $\Sigma_{\beta(\alpha)}$. See [2] and [3], Chapter 2, for general properties of exponential families.

Fig. 1 shows two points α_0 and α_1 inside A and the corresponding points β_0 and β_1 in B. The straight-line segment L_A connecting α_0 to α_1 maps into a curve $\beta(L_A)$ in B, while the straight-line segment L_B connecting β_0 to β_1 maps into a curve $\alpha(L_B)$ in A. It is assumed that L_B is inside B, though a weakened version of the theorem which follows can be proved without this assumption. In all the usual cases, B is convex, so this assumption is automatically fulfilled.

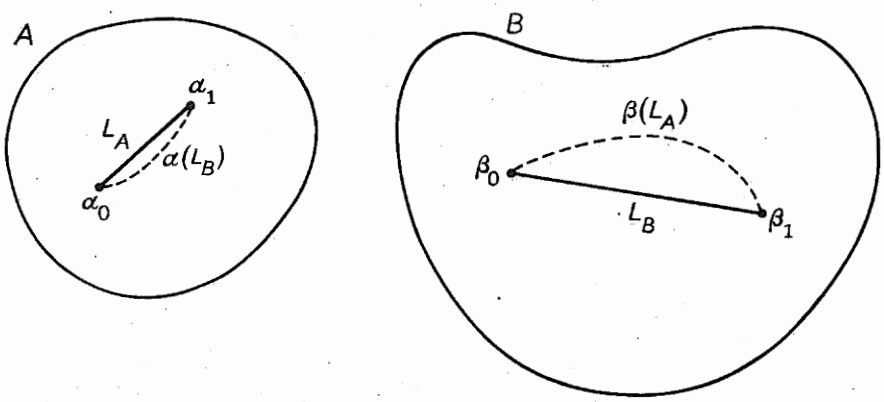

Fig. 1
A – the natural parameter space, *B* – the expectation parameter space

If α_1 is infinitesimally close to α_0, then (1.2) gives

$$(\alpha_1 - \alpha_0)' \Sigma_0 (\alpha_1 - \alpha_0) \approx (\beta_1 - \beta_0)' \Sigma_0^{-1} (\beta_1 - \beta_0), \qquad \Sigma_0 \equiv \Sigma_{\alpha_0} \equiv \Sigma_{\beta_0}.$$

What if α_1 is not close to α_0? Define the averages

$$\bar{\Sigma}_{0,1} \equiv \int_0^1 \Sigma_{\alpha_0 + \theta(\alpha_1 - \alpha_0)} d\theta \quad \text{and} \quad \overline{\Sigma_{0,1}^{-1}} \equiv \int_0^1 \Sigma_{\beta_0 + \theta(\beta_1 - \beta_0)}^{-1} d\theta$$

of Σ_α along L_A and of Σ_β^{-1} along L_B, respectively. Define also the

Kullback-Leibler distance

$$I(i,j) \equiv \mathrm{E}_{\alpha_i} \log \frac{g_{\alpha_i}(x)}{g_{\alpha_j}(x)} = (\alpha_i - \alpha_j)' \beta_i - [\psi(\alpha_i) - \psi(\alpha_j)]. \tag{1.3}$$

THEOREM. *The following four quantities are equal*:
(1) $(\alpha_1 - \alpha_0)' \bar{\Sigma}_{0,1} (\alpha_1 - \alpha_0)$,
(2) $(\beta_1 - \beta_0)' \overline{\Sigma_{0,1}^{-1}} (\beta_1 - \beta_0)$,
(3) $(\alpha_1 - \alpha_0)' (\beta_1 - \beta_0)$,
(4) $I(0,1) + I(1,0)$.

Remark 1. If $\alpha_1 \neq \alpha_0$, then all four quantities are positive. In particular, $(\alpha_1 - \alpha_0)' (\beta_1 - \beta_0) > 0$, implying that the mapping $\alpha \to \beta$ is monotone in a certain obvious sense.

Remark 2. The theorem is true in the wider framework of smooth convex dual mappings $\alpha \to \beta$, as remarked in Section 7 of [2], following the lead of Barndorff-Nielsen [1].

2. Proof of the Theorem. Let ∇_β be the gradient operator

$$\left(\frac{\partial}{\partial \beta(1)}, \frac{\partial}{\partial \beta(2)}, \ldots, \frac{\partial}{\partial \beta(k)} \right)'.$$

Then, because of (1.2), $\nabla_\beta \alpha' = \Sigma_\beta^{-1}$, and for any function h we obtain $\nabla_\beta h = \Sigma_\beta^{-1} \nabla_\alpha h$, h being thought of as defined in both A and B. In particular,

$$\nabla_\beta \psi = \Sigma^{-1} \beta. \tag{2.1}$$

From definition (1.3) we derive

$$\nabla_\beta I(\beta_0, \beta) = \Sigma_\beta^{-1} (\beta - \beta_0).$$

Integrating along the straight-line segment

$$L_B = \{\beta_0 + \theta(\beta_1 - \beta_0) : 0 \leqslant \theta \leqslant 1\},$$

with $d\beta = (\beta_1 - \beta_0) d\theta$, gives

$$I(\beta_0, \beta_1) = \int_0^1 (\beta_1 - \beta_0)' \Sigma_{\beta_0 + \theta(\beta_1 - \beta_0)}^{-1} \theta (\beta_1 - \beta_0) d\theta. \tag{2.2}$$

Integrating (2.1) along L_B gives

$$\psi(\beta_1) - \psi(\beta_0) = \int_0^1 (\beta_1 - \beta_0)' \Sigma_{\beta_0 + \theta(\beta_1 - \beta_0)}^{-1} [\beta_0 + \theta(\beta_1 - \beta_0)] d\theta. \tag{2.3}$$

Subtracting (2.2) from (2.3) yields

$$[\psi(\beta_1) - \psi(\beta_0)] - I(\beta_0, \beta_1) = (\beta_1 - \beta_0)' \overline{\Sigma_{0,1}^{-1}} \beta_0. \tag{2.4}$$

An interchange of β_0 and β_1 in (2.4) gives

$$[-\psi(\beta_1)+\psi(\beta_0)]-I(\beta_1,\beta_0) = (\beta_0-\beta_1)'\overline{\Sigma_{0,1}^{-1}}\beta_1, \tag{2.5}$$

which added to (2.4) results in

$$(\beta_1-\beta_0)'\overline{\Sigma_{0,1}^{-1}}(\beta_1-\beta_0) = I(0,1)+I(1,0). \tag{2.6}$$

Now put

$$\varphi(\alpha) \equiv \alpha'\beta-\psi(\alpha).$$

The differentiation results (1.1) and (1.2) show that

$$\nabla_\alpha\varphi = \Sigma_\alpha\alpha.$$

They also imply, by (1.3), that

$$\nabla_\alpha I(\alpha,\alpha_0) = \Sigma_\alpha(\alpha-\alpha_0).$$

The same argument as in (2.2)-(2.6), now integrating along $L_A = \{\alpha_0 + +\theta(\alpha_1-\alpha_0)\}$, results in

$$(\alpha_1-\alpha_0)'\overline{\Sigma}_{0,1}(\alpha_1-\alpha_0) = I(0,1)+I(1,0). \tag{2.7}$$

Finally, comparing (1.3) with (2.4) gives

$$(\alpha_1-\alpha_0)'\beta_0 = (\beta_1-\beta_0)'\overline{\Sigma_{0,1}^{-1}}\beta_0.$$

Interchanging the arguments 1 and 0 and subtracting from (2.3), we obtain

$$(\alpha_1-\alpha_0)'(\beta_1-\beta_0) = (\beta_1-\beta_0)'\overline{\Sigma_{0,1}^{-1}}(\beta_1-\beta_0). \tag{2.8}$$

The Theorem follows from (2.6), (2.7), and (2.8).

REFERENCES

[1] O. Barndorff-Nielsen, *Exponential families, exact theory*, Aarhus University, Various Publications Series 19 (1970).

[2] B. Efron, *The geometry of exponential families*, Ann. Statist. 6 (1978), p. 362-376.

[3] E. L. Lehmann, *Testing statistical hypotheses*, John Wiley, New York 1959.

Department of Statistics
Stanford University
Stanford, California 94 305, U. S. A.

Received on 8. 5. 1979