The Annals of Statistics
1979, Vol. 7, No. 1, 232–235

DIFFERENTIAL RELATIONS, IN THE ORIGINAL PARAMETERS, WHICH DETERMINE THE FIRST TWO MOMENTS OF THE MULTIPARAMETER EXPONENTIAL FAMILY[1]

BY RICHARD A. JOHNSON, J. LADALLA AND S. T. LIU

University of Wisconsin, Madison

We study general multiparameter exponential families of distribution and obtain differential equations relating the first two moments of the sufficient statistics to the normalization constant. Another result illuminates the structure of both the second order partial derivatives of the likelihood and their expected values.

1. Introduction. It is well known (cf. Lehmann (1959), page 58), that the first two moments of a natural parameter exponential distribution can be expressed as derivatives of the normalization constant. More generally, any moment of an exponential family can be represented by taking further derivatives of the normalization constant (see Bildikar and Patil (1968)). So far, however, these relations seem to be well known only for the natural parameter families and the corresponding general relationship has been overlooked. These relationships, of the first two moments to the derivatives of the normalization constant, are given below in terms of the original parameter space. This latter formulation usually has a more explicit connection to the statistical inference problem of interest. Although in most instances it may be easier to convert to the natural parameters in order to find the covariances of the sufficient statistics, it can still be instructive to learn of the exact relationship between moments and derivatives in terms of the original parameters.

Our representations have partially or implicitly appeared several times in previous literature on exponential families. Neyman (1941), in his study of similar regions, derived differential equations of a form similar to (6) below, except his are for the log likelihood function. However, he did not specify the relation to moments. Lehmann (1947) further indicated that the relation presented by Neyman could be used as a characterization of the exponential family. Efron (1975), in his study of the statistical curvature, also obtained a formula in the log likelihood function similar to (6), for a single parameter exponential family.

The main relationship between second moments is presented as Theorem 2 and Section 3 presents a result related to curvature of the likelihood. The explicit relations (6) and (7) below for second moments and derivatives have proved especially valuable in the study of the asymptotic behavior of posterior distributions by Ladalla (1976) and in a study of sequential conditional probability ratio tests by Liu (1976).

Received August 1977; revised December 1977.

[1]Sponsored by the National Science Foundation under Grant No. MCS77-09574.

AMS 1970 *subject classifications*. Primary 62B99, 62E15; Secondary 62F99.

Key words and phrases. Exponential families, moments and derivatives, likelihood, curvature.

2. Main results. A general parameter exponential family has a pdf of the form

$$f(\mathbf{y}|\boldsymbol{\theta}) = \exp\{\Sigma_{i=1}^{m}\omega_i(\boldsymbol{\theta})y_i - \kappa(\boldsymbol{\theta})\} \tag{1}$$

$$= \exp\{\omega(\boldsymbol{\theta})\mathbf{y}' - \kappa(\boldsymbol{\theta})\} \tag{2}$$

where $\omega(\boldsymbol{\theta}) = (\omega_1(\boldsymbol{\theta}), \cdots, \omega_m(\boldsymbol{\theta}))$ and the dominating measure $\nu(\cdot)$ is some σ-finite measure on R^m. Here $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_m)$ takes values in R^m and the h-dimensional parameter $\boldsymbol{\theta}$ ranges in a subset

$$\Theta \subset \Theta(\nu) = \{\boldsymbol{\theta} : \kappa(\boldsymbol{\theta}) = \log \textstyle\int \exp\{\omega(\boldsymbol{\theta})\mathbf{y}'\}\, d\nu(\mathbf{y}) < \infty\}.$$

Here $\kappa(\boldsymbol{\theta})$ depends on $\boldsymbol{\theta}$ through ω, so $\kappa(\boldsymbol{\theta}) = \psi(\omega(\boldsymbol{\theta}))$ for some (measurable) $\psi(\cdot)$. The natural parameter family results from setting $\omega_i = \omega_i(\boldsymbol{\theta})$, or

$$f(\mathbf{y}|\omega) = \exp\{\omega\mathbf{y}' - \psi(\omega)\} \tag{3}$$

where

$$\omega \in \Omega = \{\omega : \psi(\omega) = \log \textstyle\int \exp(\omega\mathbf{y}')\, d\nu(\mathbf{y}) < \infty\} \subseteq R^m$$

and Ω is the natural parameter space.

To simplify our presentation of relationships between moments and derivatives, we let one dot on the top of a scalar or matrix valued function of $\boldsymbol{\theta}$ represent the operation of taking the derivative $d/d\boldsymbol{\theta}'$ (cf. MacRae (1974) for this and related definitions). Two dots represent taking the second derivative $d^2/d\boldsymbol{\theta}\, d\boldsymbol{\theta}'$.

The result relating first moments and derivatives of $\kappa(\boldsymbol{\theta})$ follows immediately upon an application of the chain rule of differentiation.

THEOREM 1. *In the general exponential family* (2), *if* $\omega(\boldsymbol{\theta})$ *is differentiable with respect to* $\boldsymbol{\theta}$, *then*

$$\dot{\kappa}(\boldsymbol{\theta}) = \frac{d\kappa(\boldsymbol{\theta})}{d\boldsymbol{\theta}'} = \dot{\omega}(\boldsymbol{\theta})E_{\boldsymbol{\theta}}\mathbf{Y}' \tag{4}$$

where

$$\dot{\omega}(\boldsymbol{\theta}) = \frac{d\omega(\boldsymbol{\theta})}{d\boldsymbol{\theta}'} = \left(\frac{\partial\omega_i(\boldsymbol{\theta})}{\partial\theta_j}\right)_{h\times m}.$$

REMARK. In the natural parameter case $m = h$ and $\omega_i(\boldsymbol{\theta}) = \theta_i$, so $\dot{\omega}(\boldsymbol{\theta}) = I_m$ and relation (4) reduces to $\dot{\kappa}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\mathbf{Y}'$ as in Lehmann (1959).

In order to obtain a matrix expression for the second order relationship, we let $*$ be the star product defined by MacRae (1974) as

$$\underset{(p\times q)}{A} * \underset{(ps\times qt)}{B} = \underset{(s\times t)}{\Sigma_{i=1}^{p}\Sigma_{j=1}^{q}a_{ij}B_{ij}} \tag{5}$$

where B is partitioned into blocks as $B = (B_{ij})$, $i = 1, \cdots, p$ and $j = 1, \cdots, q$.

Entry by entry chain rule differentiation of (4) followed by a collection of terms, according to the product (5), yields

THEOREM 2. *Assume that* $\omega(\boldsymbol{\theta})$ *has second order derivatives. Let* $\ddot{\kappa}(\boldsymbol{\theta}) = d^2\kappa(\boldsymbol{\theta})/d\boldsymbol{\theta}\, d\boldsymbol{\theta}'$ *and* $D_{\boldsymbol{\theta}}(\mathbf{Y})$ *be the covariance matrix of* $\mathbf{Y}$ *under* $\boldsymbol{\theta}$. *Then*

$$\ddot{\kappa}(\boldsymbol{\theta}) = \dot{\omega}(\boldsymbol{\theta})D_{\boldsymbol{\theta}}(\mathbf{Y})\dot{\omega}(\boldsymbol{\theta})' + E_{\boldsymbol{\theta}}\mathbf{Y} * \ddot{\omega}(\boldsymbol{\theta}). \tag{6}$$

Here $\ddot{\omega}(\boldsymbol{\theta})$ is partitioned as $(d\omega_k(\boldsymbol{\theta})/d\boldsymbol{\theta}\, d\boldsymbol{\theta}')$, $k = 1, 2, \cdots, m$.

REMARK. For $\omega(\boldsymbol{\theta}) = \boldsymbol{\theta}$, the natural parameter case, $\ddot{\omega}(\boldsymbol{\theta}) = 0$ and $\dot{\omega}(\boldsymbol{\theta}) = I_m$, so we obtain the well-known result $\ddot{\kappa}(\boldsymbol{\theta}) = D_{\boldsymbol{\theta}}(\mathbf{Y})$.

REMARK. For computational purposes, the following component form of (6) may be more convenient:

$$\frac{\partial^2 \kappa(\boldsymbol{\theta})}{\partial \theta_i \, \partial \theta_j} = \frac{\partial \omega(\boldsymbol{\theta})}{\partial \theta_i} D_{\boldsymbol{\theta}}(\mathbf{Y}) \frac{\partial \omega(\boldsymbol{\theta})'}{\partial \theta_j} + E_{\boldsymbol{\theta}} \mathbf{Y} \frac{\partial^2 \omega(\boldsymbol{\theta})'}{\partial \theta_i \, \partial \theta_j}. \tag{7}$$

3. Expected derivatives of the log likelihood function. In Efron (1975), expectations of the first and second derivatives of a single parameter log likelihood function are computed in order to define statistical curvature. Here we generalize part of his computation to the multivariate case. Let $\mathbf{Y}, \mathbf{Y}_1, \cdots, \mathbf{Y}_n$ be $n + 1$ i.i.d. observations from the distribution defined in (2), and let

$$\begin{aligned} L_i(\boldsymbol{\theta}) &= \log f(\mathbf{Y}_i | \boldsymbol{\theta}) \\ &= \omega(\boldsymbol{\theta}) \mathbf{Y}_i' - \kappa(\boldsymbol{\theta}). \end{aligned} \tag{8}$$

Then, the *average* log likelihood $\bar{l}(\boldsymbol{\theta})$ is given by

$$\begin{aligned} \bar{l}(\boldsymbol{\theta}) &= \frac{1}{n} \Sigma_{i=1}^n L_i(\boldsymbol{\theta}) \\ &= \omega(\boldsymbol{\theta}) \overline{\mathbf{Y}}' - \kappa(\boldsymbol{\theta}) \end{aligned} \tag{9}$$

where

$$\overline{\mathbf{Y}} = \left(\overline{Y}_1, \cdots, \overline{Y}_j, \cdots, \overline{Y}_m\right) = \frac{1}{n} \Sigma_{i=1}^n \mathbf{Y}_i.$$

Then from (4) and (9),

$$\dot{\bar{l}}(\boldsymbol{\theta}) = \dot{\omega}(\boldsymbol{\theta}) \left(\overline{\mathbf{Y}} - E_{\boldsymbol{\theta}} \mathbf{Y}\right), \tag{10}$$

and from (6)

$$\ddot{\bar{l}}(\boldsymbol{\theta}) = \left(\overline{\mathbf{Y}} - E_{\boldsymbol{\theta}} \mathbf{Y}\right) * \ddot{\omega}(\boldsymbol{\theta}) - \dot{\omega}(\boldsymbol{\theta}) D_{\boldsymbol{\theta}}(\mathbf{Y}) \dot{\omega}(\boldsymbol{\theta})'. \tag{11}$$

THEOREM 3. *Let $\bar{l}(\boldsymbol{\theta})$ be defined as in* (9) *for a sample of size n, and assume that $\omega(\boldsymbol{\theta})$ has second order derivatives. Then*

$$E\left[\dot{\bar{l}}(\boldsymbol{\theta})\right] = \mathbf{0}, \; E\left[\ddot{\bar{l}}(\boldsymbol{\theta})\right] = -\dot{\omega}(\boldsymbol{\theta}) D_{\boldsymbol{\theta}}(\mathbf{Y}) \dot{\omega}(\boldsymbol{\theta})'$$

$$\operatorname{Var}\left[\dot{\bar{l}}(\boldsymbol{\theta})\right] = \frac{1}{n} \dot{\omega}(\boldsymbol{\theta}) D_{\boldsymbol{\theta}}(\mathbf{Y}) \dot{\omega}(\boldsymbol{\theta})'$$

$$E\left[\ddot{\bar{l}}(\boldsymbol{\theta}) \ddot{\bar{l}}(\boldsymbol{\theta})'\right] = \frac{1}{n} D_{\boldsymbol{\theta}}(\mathbf{Y}) * \left[\ddot{\omega}(\boldsymbol{\theta})' \ddot{\omega}(\boldsymbol{\theta})\right] + \left[\dot{\omega}(\boldsymbol{\theta}) D_{\boldsymbol{\theta}}(\mathbf{Y}) \dot{\omega}(\boldsymbol{\theta})'\right]^2,$$

and

$$E\left[\ddot{\bar{l}}(\boldsymbol{\theta}) \, \dot{\bar{l}}(\boldsymbol{\theta})'\right] = \frac{1}{n} D_{\boldsymbol{\theta}}(\mathbf{Y}) * \left[\ddot{\omega}(\boldsymbol{\theta})' \dot{\omega}(\boldsymbol{\theta})\right],$$

where $\ddot{\omega}(\boldsymbol{\theta})'\ddot{\omega}(\boldsymbol{\theta})$ *is partitioned as*

$$\left(\frac{d^2\omega_i(\boldsymbol{\theta})}{d\boldsymbol{\theta}\, d\boldsymbol{\theta}'} \cdot \frac{d^2\omega_j(\boldsymbol{\theta})}{d\boldsymbol{\theta}\, d\boldsymbol{\theta}'}\right) \qquad \text{for} \quad i, j = 1, \cdots, m$$

and

$$\ddot{\omega}(\boldsymbol{\theta})'\dot{\omega}(\boldsymbol{\theta}) \quad \text{as} \quad \left(\frac{d^2\omega_i(\boldsymbol{\theta})}{d\boldsymbol{\theta}\, d\boldsymbol{\theta}'} \cdot \frac{d\omega_j(\boldsymbol{\theta})}{d\boldsymbol{\theta}'}\right) \qquad \text{for} \quad i, j = 1, 2, \cdots, m$$

in the definition of the $*$ *product* (5).

REFERENCES

[1] BILDIKAR, S. and PATIL, G. P. (1968). Multivariate exponential-type distributions. *Ann. Math. Statist.* **39** 1316–1326.

[2] EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.* **3** 1189–1217.

[3] LADALLA, JOSEPH (1976). The large sample behavior of posterior distributions when sampling from multiparameter exponential family models, and allied results. Unpublished Ph.D. dissertation, Univ. of Wisconsin.

[4] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses.* Wiley, New York.

[5] LEHMANN, E. L. and SCHEFFE, H. (1947). On the problem of similar regions. *Proc. Nat. Acad. Sci.* **33** 382–386.

[6] LIU, S. T. (1976). Sequential conditional probability ratio tests in the multiparameter exponential family. Unpublished Ph.D. thesis, Univ. of Wisconsin.

[7] MACRAE, E. C. (1974). Matrix derivatives with an application to an adaptive linear decision problem. *Ann. Statist.* **2** 337–346.

[8] NEYMAN, J. (1941). On a statistical problem arising in routine analyses and in sampling inspection of mass production. *Ann. Math. Statist.* **12** 46–76.

MATHEMATICS RESEARCH CENTER
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706