KYBERNETIKA — VOLUME 22 (1986), NUMBER 2

ON THE UNIQUENESS OF THE M. L. ESTIMATES IN CURVED EXPONENTIAL FAMILIES

ANDREJ PÁZMAN

Curved families (cf. Efron [3]) imbedded in exponential families having full rank differentiable sufficient statistics are considered. It is proved that if the rank is less or equal to the dimension of the sample space, then the maximum likelihood estimate is unique. Examples: the gaussian nonlinear regression, the gaussian family with unknown mean and variance, the β-distribution. Generalized curved exponential families are considered as well.

1. INTRODUCTION AND EXAMPLE

Curved exponential families, i.e. statistical models which can be imbedded in the well known exponential families of probability distributions, have been extensively studied recently in several papers; probably the best known are [3] and [1]. The importance of such families is in the possibility to obtain very general results for a large class of different statistical problems with the aid of geometrical considerations (see the examples given below for some special cases). In the curved exponential families the importance of the maximum likelihood estimates is emphasized once more, however the question of the uniqueness of such estimates seems to be still open.

Let $\mathscr{P} = \{\mathrm{P}_\theta : \boldsymbol{\theta} \in \Theta\}$ be an exponential family of probability distributions given by the densities

$$\frac{\mathrm{d}\mathrm{P}_\theta(\mathbf{x})}{\mathrm{d}\nu} = \exp\{\boldsymbol{\theta}'\, \mathbf{t}(\mathbf{x}) - \varkappa(\boldsymbol{\theta})\} \; ; \quad (\boldsymbol{\theta} \in \Theta) \tag{1}$$

with respect to a carrier measure ν on the sample space $\mathscr{X}$. Denote by k the dimension of the vector $\boldsymbol{\theta}$, and suppose that Θ has a nonvoid interior in $\mathbb{R}^k$, $\operatorname{int} \Theta \neq \emptyset$.

The function $\varkappa(\boldsymbol{\theta}) = \ln \int_{\mathscr{X}} \exp\{\boldsymbol{\theta}'\, \mathbf{t}(\mathbf{x})\}\, \nu(\mathrm{d}\mathbf{x})$ is differentiable on $\operatorname{int} \Theta$ and the mean and the covariance of the statistic $\mathbf{t}(\mathbf{x})$ defined by Eq. (1) are equal to

$$\mathrm{E}_\theta(\mathbf{t}) = \nabla_\theta \varkappa(\boldsymbol{\theta}) := \left(\frac{\partial \varkappa}{\partial \theta_1}, \ldots, \frac{\partial \varkappa}{\partial \theta_k}\right)'$$

$$\mathrm{D}_{\boldsymbol{\theta}}(\mathbf{t}) = \nabla_{\boldsymbol{\theta}} \nabla'_{\boldsymbol{\theta}} \varkappa(\boldsymbol{\theta}) := \left\{ \frac{\partial^2 \varkappa}{\partial \theta_i \, \partial \theta_j} \right\}_{i,j=1}^{k}$$

(cf. [2], Chapter 8). In the case that $\mathrm{D}_{\boldsymbol{\theta}}(\mathbf{t})$ is nonsingular, from $\nabla_{\boldsymbol{\theta}} \mathrm{E}_{\boldsymbol{\theta}}(\mathbf{t}') = \mathrm{D}_{\boldsymbol{\theta}}(\mathbf{t}')$ we obtain by the inverse function theorem that the mapping $\boldsymbol{\theta} \mapsto \mathrm{E}_{\boldsymbol{\theta}}(\mathbf{t})$ is one-to--one on int Θ.

Let Γ be an open subset of R^m (with $m < k$) and let $\boldsymbol{\eta}: \gamma \in \Gamma \mapsto \boldsymbol{\eta}(\gamma) \in \text{int } \Theta$ be a mapping with continuous second order derivatives $\partial^2 \boldsymbol{\eta} / \partial \gamma_i \, \partial \gamma_j$ and with linearly independent vectors of first order derivatives $\partial \boldsymbol{\eta} / \partial \gamma_1, \ldots, \partial \boldsymbol{\eta} / \partial \gamma_m$. The family

(2) $$\{\mathsf{P}_{\boldsymbol{\eta}(\gamma)}: \gamma \in \Gamma\}$$

is called a curved exponential family of dimension m (cf. [3] and [1]).

The log-likelihood function of the family (2) is equal to

$$l_\gamma(\mathbf{t}) := \boldsymbol{\eta}'(\gamma) \, \mathbf{t} - \varkappa[\boldsymbol{\eta}(\gamma)] \, .$$

Given the data point $\mathbf{x} \in \mathscr{X}$, the M. L. estimate (if it exists) is defined by

$$\hat{\gamma} := \hat{\gamma}(\mathbf{x}) := \operatorname*{Arg\,max}_{\gamma \in \Gamma} \, l_\gamma[\mathbf{t}(\mathbf{x})] \, ,$$

and it is a solution of the normal equations:

(3) $$0 = \frac{\partial l_\gamma[\mathbf{t}(\mathbf{x})]}{\partial \gamma_i} = [\mathbf{t}(\mathbf{x}) - \boldsymbol{\beta}(\gamma)]' \, \frac{\partial \boldsymbol{\eta}(\gamma)}{\partial \gamma_i} \, ; \quad (i = 1, \ldots, m)$$

where we denoted $\boldsymbol{\beta}(\gamma) := \mathrm{E}_{\boldsymbol{\eta}(\gamma)}(\mathbf{t})$.

Let

(4) $$\{\{\mathsf{P}_{\boldsymbol{\eta}^i(\gamma)}: \gamma \in \Gamma_i\} \, ; \, (i \in J)\}$$

be a finite or countable set of curved exponential families of (generally different) dimensions $m(i)$; $(i \in J)$ which are imbedded in the same exponential family $\mathscr{P}$. The union of these families will be called the generalized curved exponential family (briefly: GCEF).

That means, it is the family

$$\mathscr{P}_{\mathscr{E}} = \{\mathsf{P}_{\boldsymbol{\theta}}: \mathsf{P}_{\boldsymbol{\theta}} \in \mathscr{P}, \boldsymbol{\theta} \in \mathscr{E}\}$$

where

(5) $$\mathscr{E} = \bigcup_{i \in J} \mathscr{E}_i$$

and where $\mathscr{E}_i = \{\boldsymbol{\eta}^i(\gamma^i): \gamma^i \in \Gamma_i\}$; $(i \in J)$.

The M. L. estimate in this family is

(6) $$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{x}) = \operatorname*{Arg\,max}_{\boldsymbol{\theta} \in \mathscr{E}} \, [\boldsymbol{\theta}' \, \mathbf{t}(\mathbf{x}) - \varkappa(\boldsymbol{\theta})] \, .$$

In contrast to the family (2), in a GCEF we can ensure the existence of the M. L. estimate, e.g. by supposing that $\mathscr{E}$ is compact (i.e. closed and bounded). Examples of compact sets given by Eq. (5) are: finite or bounded countable subsets of int Θ, closed intervals, closed spheres contained in int Θ, etc.

Example 1. Take $\mathscr{X} = \{x_0, x_1, x_2\}$ and let $\mathscr{P}$ be the set of all strictly positive probability measures on $\mathscr{X}$. $\mathscr{P}$ is an exponential family, which can be verified by setting

$$\nu(x_j) = 1 \,,$$
$$\theta_i = \ln \left[\mathsf{P}_\theta(x_i) / \mathsf{P}_\theta(x_0) \right] ,$$
$$t_i(x_j) = \delta_{ij} \; ; \quad (i = 1, 2, \; j = 0, 1, 2) \, ,$$
$$\varkappa(\boldsymbol{\theta}) = \ln \left[1 + e^{\theta_1} + e^{\theta_2} \right] ,$$
$$\boldsymbol{\Theta} = \mathbb{R}^2$$

into Eq. (1). Obviously, $\mathrm{d}\mathsf{P}_\theta(x_j)/\mathrm{d}\nu = \mathsf{P}_\theta(x_j)$ and $\mathrm{E}_\theta(t_i) = \mathsf{P}_\theta(x_i)$.

Consider the curve

$$\boldsymbol{\beta}: \gamma \in (0, \tfrac{1}{2}) \mapsto \begin{pmatrix} \frac{1}{2} \\ \gamma \end{pmatrix} .$$

If the data point is equal to x_1, then every $\gamma \in (0, \frac{1}{2})$ is a M. L. estimate. Indeed, we have

$$l_\gamma[\mathbf{t}(x_1)] = \ln \mathsf{P}_{\eta(\gamma)}(x_1) = \ln \beta_1(\gamma) = \ln (\tfrac{1}{2}) \geqq \ln [\mathbf{t}(x_j)] \, ; \quad (j \neq 1) \, .$$

Hence

$$\mathsf{P}_{\eta(\gamma)}\{x : \hat{\gamma}(x) \text{ is not unique}\} \geqq \mathsf{P}_{\eta(\gamma)}(x_1) = \tfrac{1}{2} \, .$$

Example 2. (The nonlinear regression.) Take in Eq. (1): $\mathscr{X} = \mathbb{R}^N$, $\Theta = \mathbb{R}^N$, $\mathbf{t}(\mathbf{x}) = \mathbf{K}^{-1}\mathbf{x}$, where $\mathbf{K}$ is a positive definite $N \times N$ matrix, $\nu(\mathrm{d}\mathbf{x}) = (2\pi)^{-N/2} \det^{-1/2} (\mathbf{K}) \times \exp \{(-\frac{1}{2}) \, \mathbf{x}'\mathbf{K}^{-1}\mathbf{x}\} \, \lambda(\mathrm{d}\mathbf{x})$, $\varkappa(\boldsymbol{\theta}) = (\frac{1}{2}) \, \boldsymbol{\theta}'\mathbf{K}^{-1}\boldsymbol{\theta}$. Then the curved family given by Eq. (2) has the density

$$\frac{\mathrm{d}\mathsf{P}_{\eta(\gamma)}(\mathbf{x})}{\mathrm{d}\lambda} = \frac{1}{(2\pi)^{N/2} \det^{1/2} \mathbf{K}} \exp \{ -\tfrac{1}{2} (\mathbf{x} - \boldsymbol{\eta}(\gamma))' \, \mathbf{K}^{-1} (\mathbf{x} - \boldsymbol{\eta}(\gamma)) \}$$

with respect to the Lebesgue measure λ. The M. L. estimate coincides here with the least-squares estimate

$$\hat{\gamma} = \operatorname{Arg} \min_{\gamma} \, (\mathbf{x} - \boldsymbol{\eta}(\gamma))' \, \mathbf{K}^{-1} (\mathbf{x} - \boldsymbol{\eta}(\gamma)) \, .$$

The uniqueness of this estimate (with probability one) has been proved by the author in [5].

In this paper we shall prove that the M. L. estimate in a GCEF is unique with probability one, provided that the embedding family $\mathscr{P}$ has the properties:

A) $\mathscr{X}$ is an open subset on $\mathbb{R}^N$, $N \geqq k$.

B) The statistic $\mathbf{t}: \mathscr{X} \mapsto \mathbb{R}^k$ defined by Eq. (1) has continuous first order derivatives and the rank of the matrix $\nabla'_{\mathbf{x}} \mathbf{t}(\mathbf{x})$ (with entries $\partial t_i(\mathbf{x})/\partial x_j$; $(i = 1, \ldots, k, j = 1, \ldots, N)$) is equal to k.

C) The family $\mathscr{P}$ is dominated by the Lebesgue measure λ (i.e. we can suppose that $\nu \ll \lambda$).

We note that B) implies that $t_1(\mathbf{x}) - \mathrm{E}_\theta[t_1(\mathbf{x})], \dots, t_k(\mathbf{x}) - \mathrm{E}_\theta[t_k(\mathbf{x})]$ are linearly independent functions on $\mathscr{X}$, hence $\mathbf{D}_\theta(\mathbf{t})$ is a nonsingular matrix.

The family in Example 2 has the properties A)–C). Other examples are

Example 3. (The gaussian family with unknown mean and variance.) Take $\mathscr{X} = \mathbb{R}^N$, $N \geqq 2$, $\Theta = (-\infty, \infty) \times (-\infty, 0)$,

$$\mathbf{t}(\mathbf{x}) = \left(\sum_{i=1}^N x_i, \sum_{i=1}^N x_i^2\right), \quad \nu(\mathrm{d}\mathbf{x}) = (2\pi)^{-N/2}\, \lambda(\mathrm{d}\mathbf{x}),$$

$\varkappa(\theta_1, \theta_2) = -(N/2) \ln(-2\theta_2) + N(\theta_1^2/2\theta_2)$. Then the density (1), but with respect to λ, is equal to

$$\prod_{i=1}^N \frac{1}{\sqrt{(2\pi)}\,\sigma}\, \mathrm{e}^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

i.e. $\theta_1 = \mu/\sigma^2$, $\theta_2 = -1/(2\sigma^2)$. Evidently

$$[\nabla_{\mathbf{x}} \mathbf{t}'(\mathbf{x})]' = \begin{pmatrix} 1, & \dots, 1 \\ 2x_1, & \dots, 2x_N \end{pmatrix}.$$

Hence rank $[\nabla_{\mathbf{x}} \mathbf{t}'(\mathbf{x})] = k = 2$ if we modify $\mathscr{X}$ taking for $\mathscr{X}$ the set $\mathbb{R}^N - \{\mathbf{x} \in \mathbb{R}^N, \exists_{i \neq j}\, x_i = x_j\}$ which is evidently open, and has probability one for every $\mathsf{P} \in \mathscr{P}$.

Example 4. (The beta-distribution.) Take $N \geqq 2$, $\mathscr{X} = \{\mathbf{x}: \mathbf{x} \in (0,1)^N, x_i \neq x_j; (i \neq j)\}$, $\Theta = (-1, \infty) \times (-1, \infty)$, $\mathbf{t}(\mathbf{x}) = \left(\sum_{i=1}^N \ln x_i, \sum_{i=1}^N \ln(1 - x_i)\right)$, $\nu(\mathrm{d}\mathbf{x}) = \lambda(\mathrm{d}\mathbf{x})$, $\varkappa(\theta_1, \theta_2) = N\{\ln \Gamma(\theta_1 + 1) + \ln \Gamma(\theta_2 + 1) - \ln \Gamma(\theta_1 + \theta_2 + 2)\}$, where $\Gamma(\ )$ is the gamma function. Then the density (1) corresponds to the product of beta-densities

$$\prod_{i=1}^N \frac{\Gamma(\mu_1 + \mu_2)}{\Gamma(\mu_1)\,\Gamma(\mu_2)}\, x_i^{\mu_1 - 1} (1 - x_1)^{\mu_2 - 1}$$

$(\mu_1 = \theta_1 + 1, \mu_2 = \theta_2 + 1)$. Obviously,

$$[\nabla_{\mathbf{x}} \mathbf{t}'(\mathbf{x})]' = \begin{pmatrix} 1/x_1, & \dots, & 1/x_N \\ -1/(1 - x_1), & \dots, & -1/(1 - x_N) \end{pmatrix},$$

and rank $[\nabla_{\mathbf{x}} \mathbf{t}'(\mathbf{x})] = 2$ on $\mathscr{X}$.

2. THE UNIQUENESS

Let us denote

$$\mathscr{T} := \{\mathbf{t}(\mathbf{x}): \mathbf{x} \in \mathscr{X}\}.$$

Proposition 1. Under the assumptions A) and B) the set $\mathscr{T}$ is an open subset of $\mathbb{R}^k$ and

$$F \subset \mathscr{T}, \quad \lambda(F) = 0 \Rightarrow \lambda[\mathbf{t}^{-1}(F)] = 0.$$

The proof follows from Proposition A 3. □

Denote by $\mathsf{P}^{\mathbf{t}}$ the measure induced from the measure P by the mapping $\mathbf{t}$.

Corollary. Under the assumptions A) – C) we have

$$\mathsf{P}^{\mathbf{t}} \ll \lambda$$

for every $\mathsf{P} \in \mathscr{P}$.

Denote the vector of first order derivatives and the matrix of second order derivatives of the log-likelihood function $l_\gamma(\mathbf{t})$ by $\nabla_\gamma l_\gamma(\mathbf{t})$, resp. by $\nabla_\gamma \nabla'_\gamma l_\gamma(\mathbf{t})$. If $\boldsymbol{\theta} \in \mathscr{E}$, i.e. $\boldsymbol{\theta} \in \mathscr{E}_i$ for some i, then by $l_\gamma(\mathbf{t})|_{\boldsymbol{\theta}}$ resp. by $\nabla_\gamma l_\gamma(\mathbf{t})|_{\boldsymbol{\theta}}$ we denote the log-likelihood function at $\boldsymbol{\theta} = \boldsymbol{\eta}^i(\gamma^i)$, $\gamma^i = \gamma$, resp. the vector of the first order derivatives with respect to $\gamma_1^i, \ldots, \gamma_{m(i)}^i$ (i.e. we omit the superscript i).

Proposition 2. Under the assumption A) and B) we have that

$$\lambda\{\mathbf{t} \colon \mathbf{t} \in \mathscr{T}\,,\quad \exists_{\boldsymbol{\theta} \in \mathscr{E}} \nabla_\gamma l_\gamma(\mathbf{t})|_{\boldsymbol{\theta}} = 0\,,\quad \det\left[\nabla_\gamma \nabla'_\gamma l_\gamma(\mathbf{t})\right]_{\boldsymbol{\theta}} = 0\} = 0\,.$$

Proof. From the second equality in Eq. (3) we obtain

$$\det\left[\nabla_\gamma \nabla'_\gamma l_\gamma(\mathbf{t})\right] = \tag{7}$$

$$= \det\left[\left\{(\mathbf{t} - \boldsymbol{\beta}(\gamma))' \frac{\partial^2 \boldsymbol{\eta}(\gamma)}{\partial \gamma_i\, \partial \gamma_j} - \frac{\partial \boldsymbol{\beta}'(\gamma)}{\partial \gamma_i} \frac{\delta \boldsymbol{\eta}(\gamma)}{\partial \gamma_j}\right\}_{i,j=1}^m\right].$$

Denote

$$\mathscr{L}_\gamma := \left\{\{\mathbf{z} \colon \mathbf{z} \in \mathbb{R}^k\,,\quad \mathbf{z}' \frac{\partial \boldsymbol{\eta}(\gamma)}{\partial \gamma_i} = 0\,;\quad (i = 1, \ldots, m)\right\}.$$

Any orthonormal basis $\mathbf{w}_1(\gamma), \ldots, \mathbf{w}_{k-m}(\gamma)$ of $\mathscr{L}_\gamma$ is defined by the equations

$$\mathbf{w}'_l(\gamma) \frac{\partial \boldsymbol{\eta}(\gamma)}{\partial \gamma_j} = 0\,,$$

$$\mathbf{w}'_l(\gamma)\, \mathbf{w}_i(\gamma) = \delta_{il}\,;\quad (i, l = 1, \ldots, k - m\,,\quad j = 1, \ldots, m)\,.$$

Let $U_1, \ldots, U_S$ be the subsets of Γ such that on each U_i a fixed $m \times m$ submatrix of $\nabla_\gamma \boldsymbol{\eta}'(\gamma)$ is nonsingular. Evidently, the vectors $\mathbf{w}_1(\gamma), \ldots, \mathbf{w}_{k-m}(\gamma)$ can be chosen as differentiable functions of γ on each U_i. By differentiation with respect to γ_i we obtain

$$\mathbf{w}'_l(\gamma) \frac{\partial^2 \boldsymbol{\eta}(\gamma)}{\partial \gamma_i\, \partial \gamma_j} = - \frac{\partial \mathbf{w}'_l(\gamma)}{\partial \gamma_i} \frac{\partial \boldsymbol{\eta}(\gamma)}{\partial \gamma_j} \tag{8}$$

Take $\mathbf{t} \in \mathscr{T}$, $\gamma \in \Gamma$ so that

$$[\mathbf{t} - \boldsymbol{\beta}(\gamma)]'\, \partial \boldsymbol{\eta}(\gamma)/\partial \gamma_j = 0\,;\quad (j = 1, \ldots, m)\,.$$

Then

$$\mathbf{t} - \boldsymbol{\beta}(\gamma) = \sum_{l=1}^{k-m} c_l\, \mathbf{w}_l(\gamma) \quad \text{for some } c_i\,.$$

Fix $c_1, \ldots, c_{k-m}$ and define $\mathbf{t}(\tilde{\gamma})$ by

$$\mathbf{t}(\tilde{\gamma}) := \sum_l c_l\, \mathbf{w}_l(\tilde{\gamma}) + \boldsymbol{\beta}(\tilde{\gamma})\,.$$

For $\|\tilde{\gamma} - \gamma\|$ sufficiently small, we have $\mathbf{t}(\tilde{\gamma}) \in \mathscr{T}$. From Eq. (8) we obtain:

$$[\mathbf{t}(\tilde{\gamma}) - \boldsymbol{\beta}(\tilde{\gamma})]' \frac{\partial^2 \boldsymbol{\eta}(\tilde{\gamma})}{\partial \gamma_i \, \partial \gamma_j} = - \sum_l c_l \frac{\partial \mathbf{w}_l(\tilde{\gamma})}{\partial \gamma_i} \frac{\partial \boldsymbol{\eta}(\tilde{\gamma})}{\partial \gamma_j}$$

hence from (7) it follows that

$$\det [\nabla_\gamma \nabla_\gamma' l_\gamma(\mathbf{t})] =$$

$$= \det \left[\left\{ \left(- \sum_{l=1}^{k-m} c_l \frac{\partial \mathbf{w}_l'(\gamma)}{\partial \gamma_i} - \frac{\partial \boldsymbol{\beta}'(\gamma)}{\partial \gamma_i} \right) \frac{\partial \boldsymbol{\eta}(\gamma)}{\partial \gamma_j} \right\}_{i,j=1}^{m} \right].$$

Denote

$$\mathbf{W} := (\mathbf{w}_1(\gamma), \ldots, \mathbf{w}_{k-m}(\gamma)),$$

$$\mathbf{D} := \nabla_\gamma \boldsymbol{\eta}'(\gamma),$$

$$\mathbf{G} := \nabla_\gamma [\sum_l c_l \, \mathbf{w}_l'(\gamma) + \boldsymbol{\beta}'(\gamma)].$$

We have

$$\det [\nabla_\gamma \nabla_\gamma' \, l_\gamma(\mathbf{t})] = \det(-\mathbf{G}\mathbf{D}')$$

$$= \det \begin{pmatrix} -\mathbf{G}\mathbf{D}', & -\mathbf{G}\mathbf{W} \\ \mathbf{W}'\mathbf{D}', & \mathbf{W}'\mathbf{W} \end{pmatrix}$$

$$= \det \begin{pmatrix} -\mathbf{G} \\ \mathbf{W}' \end{pmatrix} (\mathbf{D}', \mathbf{W})$$

since $\mathbf{W}'\mathbf{W} = \mathbf{I}$, $\mathbf{W}'\mathbf{D}' = \mathbf{0}$. Since the matrix $(\mathbf{D}', \mathbf{W})$ has a full rank, we have that, under the assumption $\nabla_\gamma l_\gamma(\mathbf{t}) = 0$,

$$\det [\nabla_\gamma \nabla_\gamma' l_\gamma(\mathbf{t})] = 0 \Leftrightarrow \det(-\mathbf{G}', \mathbf{W}) = 0.$$

Consider the mappings

$$h_i: (\gamma_1, \ldots, \gamma_m, c_1, \ldots, c_{k-m}) \in U_i \times \mathbb{R}^{k-m} \mapsto \sum_{l=1}^{k-m} c_l \, \mathbf{w}_l(\gamma) + \boldsymbol{\beta}(\gamma).$$

We have $\mathscr{T} \subset \bigcup_{i=1}^{S} h_i(U_i \times \mathbb{R}^{k-m})$, since to every $\mathbf{t} \in \mathscr{T}$ there are $\gamma \in \Gamma$, $i \in \{1, \ldots, S\}$ and $\mathbf{c} \in \mathbb{R}^{k-m}$ such that $\mathbf{t} = \sum c_l \, \mathbf{w}_l(\gamma) + \boldsymbol{\beta}(\gamma)$. The set $h_i^{-1}(\mathscr{T})$ is open since $\mathscr{T}$ is open and h_i is continuous. Denote by $\tilde{h}_i$: $h_i^{-1}(\mathscr{T}) \mapsto \mathscr{T}$ the restriction of h_i to the set $h_i^{-1}(\mathscr{T})$. We have

$$\nabla_{(\gamma,\mathbf{c})} \tilde{h}_i(\gamma, \mathbf{c}) = \nabla_{(\gamma,\mathbf{c})} h_i(\gamma, \mathbf{c}) = (\mathbf{G}, \mathbf{W}).$$

Hence $\det [\nabla_{(\gamma,\mathbf{c})} \tilde{h}(\gamma, \mathbf{c})] = 0 \Leftrightarrow \det [\nabla_\gamma \nabla_\gamma' l_\gamma(\mathbf{t})] = 0$. Thus the needed statement follows from Proposition A 4. □

Proposition 3. Under the assumptions A) and B) we obtain

$$\lambda\{\mathbf{x}: \mathbf{x} \in \mathscr{X}, \ \exists_{\tilde{\theta}, \bar{\theta} \in \mathscr{E}} \nabla_\gamma l_\gamma[\mathbf{t}(\mathbf{x})]|_{\tilde{\theta}} = 0, \ \nabla_\gamma l_\gamma[\mathbf{t}(\mathbf{x})]_{\bar{\theta}} = 0,$$
$$l_\gamma[\mathbf{t}(\mathbf{x})]_{\tilde{\theta}} = l_\gamma[\mathbf{t}(\mathbf{x})]_{\bar{\theta}}\} = 0.$$

Proof. Denote

$$\mathscr{G} := \{\mathbf{t}: \mathbf{t} \in \mathscr{T}, \quad \exists_{\theta\in\mathscr{E}} \nabla_\gamma l_\gamma(\mathbf{t})|_\theta = 0, \quad \det[\nabla_\gamma \nabla'_\gamma l_\gamma(\mathbf{t})]_\theta = 0\}.$$

Further denote

$$\Gamma_{rs} := \{(\tilde{\gamma}, \bar{\gamma}): \tilde{\gamma} \in \Gamma_r, \bar{\gamma} \in \Gamma_s, \boldsymbol{\eta}^r(\tilde{\gamma}) \neq \boldsymbol{\eta}^s(\bar{\gamma})\}.$$

According to Propositions 1 and 2, it is sufficient to prove that for every $r, s \in J$

$$\lambda\{\mathbf{t}: \mathbf{t} \in \mathscr{T} - \mathscr{G}, \quad \exists_{(\tilde{\gamma},\bar{\gamma})\in\Gamma_{rs}} \nabla_\gamma l_{\tilde{\gamma}}(\mathbf{t}) = 0,$$

$$\nabla_\gamma l_{\bar{\gamma}}(\mathbf{t}) = 0, \quad l_{\tilde{\gamma}}(\mathbf{t}) = l_{\bar{\gamma}}(\mathbf{t})\} = 0.$$

Denote by m, n the dimensions of Γ_s and Γ_r. Define a mapping

$$\varrho: \Gamma_{rs} \times (\mathscr{T} - \mathscr{G}) \mapsto \mathbb{R}^{m+n+k}$$

by

$$\varrho(\tilde{\gamma}, \bar{\gamma}, \mathbf{t}) := (\nabla'_\gamma l_{\tilde{\gamma}}(\mathbf{t}), \nabla_\gamma l_{\bar{\gamma}}(\mathbf{t}), l_{\tilde{\gamma}}(\mathbf{t}) - l_{\bar{\gamma}}(\mathbf{t}), \mathbf{t}^{(i)'})$$

where $\mathbf{t}^{(i)'} := (t_1, \ldots, t_{i-1}, t_{i+1}, \ldots, t_k)$ and where i is choosen so that $\{\boldsymbol{\eta}^r(\tilde{\gamma})\}_i \neq \{\boldsymbol{\eta}^s(\bar{\gamma})\}_i$ (There is such a subscript i since $\boldsymbol{\eta}^r(\tilde{\gamma}) \neq \boldsymbol{\eta}^s(\bar{\gamma})$).

The Jacobian of ϱ is

$$\nabla_{(\tilde{\gamma},\bar{\gamma},\mathbf{t})}\varrho(\tilde{\gamma}, \bar{\gamma}, \mathbf{t}) = \begin{pmatrix} \nabla_\gamma \nabla'_\gamma l_{\tilde{\gamma}}(\mathbf{t}), & \mathbf{0}, & \nabla_\gamma l_{\tilde{\gamma}}(\mathbf{t}), & \mathbf{0} \\ \mathbf{0}, & \nabla_\gamma \nabla'_\gamma l_{\bar{\gamma}}(\mathbf{t}), & -\nabla_\gamma l_{\bar{\gamma}}(\mathbf{t}), & \mathbf{0} \\ \nabla_t \nabla'_\gamma l_{\tilde{\gamma}}(\mathbf{t}), & \nabla_t \nabla'_\gamma l_{\bar{\gamma}}(\mathbf{t}), & \boldsymbol{\eta}^r(\tilde{\gamma}) - \boldsymbol{\eta}^s(\bar{\gamma}), & \mathbf{I}(i) \end{pmatrix}$$

where $\mathbf{I}(i)$ is the $k \times (k-1)$ matrix obtained from the identity matrix by removing the ith column.

On the set

$$Z := \{(\tilde{\gamma}, \bar{\gamma}, \mathbf{t}): \varrho(\tilde{\gamma}, \bar{\gamma}, \mathbf{t}) = (\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{t}^{(i)}), \mathbf{t} \in \mathscr{T} - \mathscr{G}\}$$

we have

$$\det[\nabla_{\tilde{\gamma},\bar{\gamma},\mathbf{t}}\varrho(\tilde{\gamma}, \bar{\gamma}, \mathbf{t})] =$$

$$= \det[\nabla_\gamma \nabla'_\gamma l_{\tilde{\gamma}}(\mathbf{t})] \det[\nabla_\gamma \nabla'_\gamma l_{\bar{\gamma}}(\mathbf{t})] \det[\boldsymbol{\eta}^r(\tilde{\gamma}) - \boldsymbol{\eta}^s(\bar{\gamma}), \mathbf{I}(i)] \neq 0$$

since $\det[\boldsymbol{\eta}^r(\tilde{\gamma}) - \boldsymbol{\eta}^s(\bar{\gamma}), \mathbf{I}(i)] = \{\boldsymbol{\eta}^r(\tilde{\gamma})\}_i - \{\boldsymbol{\eta}^s(\bar{\gamma})\}_i \neq 0$.

Thus ϱ is a diffeomorphism on Z, and

$$\dim Z = \dim\{\mathbf{y} \in \mathbb{R}^{k+m+n}: \mathbf{y} = (\mathbf{0}, \mathbf{0}, \mathbf{0}, \mathbf{t}^{(i)}); \mathbf{t} \in \mathscr{T} - \mathscr{G}\} \leqq k - 1.$$

It follows that

$$\dim\{\mathbf{t}: \mathbf{t} \in \mathbb{R}^k, \quad \exists_{(\tilde{\gamma},\bar{\gamma})\in\Gamma_{rs}} (\tilde{\gamma}, \bar{\gamma}, \mathbf{t}) \in Z\} \leqq k - 1,$$

hence

$$\lambda\{\mathbf{t}: \quad \exists_{(\tilde{\gamma},\bar{\gamma})\in\Gamma_{rs}} (\tilde{\gamma}, \bar{\gamma}, \mathbf{t}) \in Z\} = 0. \qquad \square$$

A direct consequence of Proposition 3 is the following theorem.

Theorem. Under the assumptions A) – C), for any GCEF imbedded in $\mathscr{P}$ and for any $\mathrm{P} \in \mathscr{P}$ we have

$$\mathrm{P}\{\mathbf{x}: \text{the M. L. estimate } \hat{\theta}(\mathbf{x}) \text{ is not unique}\} = 0.$$

APPENDIX

Let

$$\psi: X \mapsto \mathbb{R}^s$$

be a mapping defined on an open set $X \subset \mathbb{R}^r$ ($r \geqq s$). Suppose that ψ has continuous first order derivatives (the elements of the $r \times s$ matrix $\nabla_{\mathbf{x}}\psi'(\mathbf{x})$).

Proposition A1. (The inverse function theorem.) If $r = s$ and $\det[\nabla_{\mathbf{x}}\psi'(\mathbf{x})] \neq 0$ on X, then ψ is a diffeomorphism (i.e. it is one-to-one and ψ^{-1}: $\psi(X) \mapsto \mathbb{R}^r$ is continuously differentiable).

For the proof cf. [4] or [7].

Proposition A2. If ψ is a diffeomorphism, then

$$F \subset X\,, \quad \lambda(F) = 0 \Rightarrow \lambda[\psi(F)] = 0\,.$$

Proof. F is either bounded, or it is a countable union of bounded sets. For F bounded we have

$$\lambda[\psi(F)] = \int_{\psi(F)} \lambda(\mathrm{d}\mathbf{y}) = \int_F |\det[\nabla_{\mathbf{x}}\psi'(\mathbf{x})]|\, \lambda(\mathrm{d}\mathbf{x}) = 0\,. \qquad \square$$

Proposition A3. If $r \geqq s$ and $\operatorname{rank}[\nabla_{\mathbf{x}}\psi'(\mathbf{x})] = s$ then the set $\psi(X)$ is an open subset of $\mathbb{R}^s$, and

$$F \in \psi(X),\, \lambda(F) = 0 \Rightarrow \lambda[\mathbf{t}^{-1}(F)] = 0\,.$$

Proof. The set X is a finite union of sets $U_1, \ldots, U_S$ such that on each U_i a fixed $s \times s$ submatrix of $\nabla_{\mathbf{x}}\psi'(\mathbf{x})$ is nonsingular (i.e. its determinant is nonzero). Denote by $\mathbf{p}$: $\mathbb{R}^r \mapsto \mathbb{R}^s$ and by $\mathbf{q}$: $\mathbb{R}^r \mapsto \mathbb{R}^{r-s}$ the projections $\mathbf{p}(\mathbf{x}) := (x_1, \ldots, x_s)$, $\mathbf{q}(\mathbf{x}) := (x_{s+1}, \ldots, x_r)$. Suppose that $\nabla_{\mathbf{p}(\mathbf{x})}\psi'(\mathbf{x})$ is nonsingular. Define Ψ: $U_1 \mapsto \mathbb{R}^r$ by

$$\Psi(\mathbf{x}) = (\psi(\mathbf{x}), \mathbf{q}(\mathbf{x}))\,.$$

Ψ is a diffeomorphism on U_1 since the matrix

$$\nabla_{\mathbf{x}}\Psi'(\mathbf{x}) = \begin{pmatrix} \nabla_{\mathbf{p}(\mathbf{x})}\psi'(\mathbf{x}), & \nabla_{\mathbf{q}(\mathbf{x})}\psi'(\mathbf{x}) \\ \mathbf{0} \quad , & \mathbf{I} \end{pmatrix}$$

is nonsingular (Proposition A 1). Denote $V_1 = \psi(U_1)$. Obviously, $\psi(\mathbf{x}) = \mathbf{p} \circ \Psi(\mathbf{x})$, hence $V_1 := \mathbf{p} \circ \Psi(U_1)$ is an open set, and $\psi(X) = \bigcup_i V_i$ is open as well.

Further

$$\lambda[\mathbf{p}^{-1}(F \cap V_1)] = \lambda[(F \cap V_1) \times \mathbb{R}^{r-s}] = 0\,,$$

hence, according to Proposition A 2,

$$\lambda[\psi^{-1}(F \cap V_1)] = \lambda[\Psi^{-1} \circ \mathbf{p}^{-1}(F \cap V_1)] = 0\,.$$

Consequently, $\lambda[\psi^{-1}(F)] = \sum_{i=1}^{S} \lambda[\psi^{-1}(F \cap V_i)] = 0\,.$ $\square$

Proposition A4. If $r = s$ and X is bounded, then

$$\lambda\{\psi(\mathbf{x}): \det[\nabla_{\mathbf{x}}\psi'(\mathbf{x})] = 0\} = 0\,.$$

The proof is given in [7], Lemma 3. 2.

(Received January 24, 1985.)

REFERENCES

[1] S. Amari: Differential geometry of curved exponential families — curvatures and information loss. Ann. Statist. *10* (1982), 357—387.

[2] O. Barndorff-Nielsen: Information and Exponential Families in Statistical Theory. Wiley, Chichester 1979.

[3] B. Efron: The geometry of exponential families. Ann. Statist. *6* (1978), 362—376.

[4] V. Jarník: Diferenciální počet II (Differential Calculus). NČSAV, Praha 1956.

[5] A. Pázman: Nonlinear least squares — uniqueness versus ambiguity. Math. Operationsforsch. Statist. Ser. Statist. *15* (1984), 323—336.

[6] A. Pázman: Probability distribution of the multivariate nonlinear least squares estimates. Kybernetika *20* (1984), 209—230.

[7] S. Sternberg: Lectures on Differential Geometry. Second printing. Prentice-Hall, Englewood Cliffs 1965.

RNDr. Andrej Pázman, DrSc., Matematický ústav SAV (Mathematical Institute — Slovak Academy of Sciences), ul. Obrancov mieru 49, 814 73 Bratislava. Czechoslovakia.