# ON THE STATISTICAL INFERENCE FOR INFINITE PARAMETER EXPONENTIAL FAMILIES

B. L. S. Prakasa Rao and Prem S. Puri*

Indian Statistical Institute, New Delhi
and
Purdue University

Technical Report #84-28

Department of Statistics
Purdue University

August 1984

*This research work was carried out at the Indian Statistical Institute, New Delhi, where this author spent his sabbatical leave from Purdue University for the year 1983-84. He is grateful to the Institute and its staff for providing a stimulating atmosphere for carrying out this research. This research was also supported in parts by U. S. National Science Foundation Grants numbers MCS-8102733 and INT-8305474.

# ON THE STATISTICAL INFERENCE FOR INFINITE PARAMETER EXPONENTIAL FAMILIES

B. L. S. Prakasa Rao and Prem S. Puri*

Indian Statistical Institute, New Delhi

and

Purdue University

## 0. INTRODUCTION

The exponential families of distributions involving finite number of parameters are commonly used for modeling purposes in statistical literature. Consequently extensive research efforts have been devoted to the study of their properties and inference problems related to them (see Lehmann (1959), Berk (1972), Barndorf-Nielsen (1978)). There are however great number of practical situations, where one needs to consider instead exponential families involving infinite number of parameters. For instance, all the discrete distributions with countably infinite support, hereafter referred to as infinitomial distributions, fall in this category (see example 2 of next section). One of the early papers which deals with infinitomial distributions is due to Rao (1958). Also, there are distributions which do not form an exponential family in the classical framework and yet they could be expressed as members of infinite parameter exponential (IPE) families (see example 1 of next section for Cauchy family). In his study of nonparametric estimation of densities, Crain (1974) has considered finite parameter exponential families

*This research work was carried out at the Indian Statistical Institute, New Delhi, where this author spent his sabbatical leave from Purdue University for the year 1983-84. He is grateful to the Institute and its staff for providing a stimulating atmosphere for carrying out this research. This research was also supported in parts by U. S. National Science Foundation Grants numbers MCS-8102733 and INT-8305474.

as an approximation to IPE families (see also Section 6). Among others, these are the key considerations that motivated our present work. Some earlier works dealing with some probabilistic aspects of IPE families are due to Soler (1977) and Johansen (1977).

In this paper, we introduce IPE families in Section 1 along with some examples. Section 2 deals with some properties of such families. Asymptotic properties of the estimators of the parameters are discussed in Sections 3-5. An application of these results to nonparametric density estimation is given in Section 6. Finally, it may be remarked that our present approach deals with the simultaneous estimation of all the infinite parameters involved. In order to accomplish this aim we have adopted the Hilbert-Space approach and have used the properties of probability measures on Hilbert-Spaces (see Grenander (1963)). Our approach to the statistical aspect of the problem is similar to that of Berk (1972) for finite parameter exponential families.

## 1. PRELIMINARIES.

Consider a measurable space $(\mathcal{X}, \mathcal{B})$ with a $\sigma$-finite measure $\mu$ defined on it. Let $L_2(\mu)$ denote the space of measurable square integrable functions. If $\mu(\mathcal{X}) = \infty$, let $\{H_i(x), i \geq 1\}$ be a complete orthonormal basis (CONB) for $L_2(\mu)$. If, on the other hand, $\mu(\mathcal{X}) < \infty$, let $\{H_i(x), i \geq 0\}$ with $H_0(x) \equiv 1$ form a CONB for $L_2(\mu)$. Suppose $\nu$ is another $\sigma$-finite measure on $(\mathcal{X}, \mathcal{B})$ equivalent to $\mu$. Let $\ell_2$ be the space of all real square summable sequences. Define

$$(1) \qquad \Omega = \{\underset{\sim}{\xi} \in \ell_2 : \int_{\mathcal{X}} e^{\sum_{i=1}^{\infty} \xi_i H_i(x)} d\nu(x) < \infty\}$$

and

$$(2) \qquad C(\underset{\sim}{\xi}) = \ell n \int_{\mathcal{X}} e^{\sum_{i=1}^{\infty} \xi_i H_i(x)} d\nu(x) \leq \infty, \quad \text{for } \underset{\sim}{\xi} \in \ell_2.$$

Observe that

$$(3) \qquad \Omega = \{\underset{\sim}{\xi} \in \ell_2: \ C(\xi) < \infty\}.$$

It can be easily shown, using Holder inequality, that $\Omega$ is a convex set in $\ell_2$. We shall only consider the case when $\Omega$ has a non empty interior $\Omega^0$.

<u>DEFINITION</u>. A random variable X defined on $(\mathcal{X},\mathcal{B})$ is said to have distribution belonging to an <u>infinite</u> <u>parameter</u> <u>exponential</u> (IPE) <u>family</u> with respect to $(\mu,\nu)$ if it has a density with respect to measure $\nu$ given by

$$(4) \qquad f(x,\underset{\sim}{\xi}) = \exp\{\sum_{i=1}^{\infty} \xi_i H_i(x) - C(\xi)\},$$

where $\underset{\sim}{\xi} \in \Omega$.

This family will be denoted by $\mathcal{E}(\mu,\nu)$. Evidently in the case where $\mu(\mathcal{X}) < \infty$, $\log f(x,\underset{\sim}{\xi}) \in L_2(\mu)$ for every $\underset{\sim}{\xi} \in \Omega$. Furthermore the family of densities $\{f(x,\underset{\sim}{\xi}), \underset{\sim}{\xi} \in \Omega\}$ is identifiable as we shall observe later (see Proposition 1).

Let us now consider the special case when $\mathcal{X}$ is countable, $\mathcal{B}$ is the $\sigma$-algebra of all subsets of $\mathcal{X}$ and $\mu$ is the counting measure on $(\mathcal{X},\mathcal{B})$. Let $\{J_i(x), i \geq 0\}$ be a CONB for $L_2(\mu)$. Then

$$(5) \qquad \sum_{i=1}^{\infty} \xi_i J_i(x) \in L_2(\mu) \Leftrightarrow \underset{\sim}{\xi} = (\xi_1, \xi_2, \ldots) \in \ell_2.$$

Consider

$$(6) \qquad \Omega = \{\underset{\sim}{\xi} \in \ell_2 : \sum_{\mathcal{X}} \exp\{\sum_{i=1}^{\infty} \xi_i J_i(x)\} < \infty\}.$$

Then, for $\underset{\sim}{\xi} \in \Omega$,

$$(7) \qquad f(x, \underset{\sim}{\xi}) = \exp[\sum_{i=1}^{\infty} \xi_i J_i(x) - C(\underset{\sim}{\xi})],\ x \in \mathcal{X},$$

is a probability function where

$$(8) \qquad C(\underset{\sim}{\xi}) = \ln[\sum_{\mathcal{X}} \exp\{\sum_{i=1}^{\infty} \xi_i J_i(x)\}].$$

The class of distributions as defined by (7) for $\underset{\sim}{\xi} \in \Omega$ is a subset of the class of all infinitomial distributions on $\mathcal{X}$ (see example 2 below).

We illustrate the IPE families with the following few examples.

Example 1. (Cauchy) Let

$$(9) \qquad f(x, \lambda) = \frac{1}{\pi(1+(x-\lambda)^2)} \sqrt{2\pi}\ e^{\frac{1}{2} x^2},\ -\infty < x < \infty,$$

which is the Cauchy density with respect to $\nu = N(0,1)$, the standard normal probability measure. Let $\mu = N(0,1)$. It is easy to check that $\ln f(x;\lambda) \in L_2(\mu)$ so that

(10) $$\ln f(x,\lambda) = \tfrac{1}{2}\ln\left(\tfrac{2}{\pi}\right) + \tfrac{1}{2}x^2 - \ln(1+(x-\lambda)^2)$$

$$= \sum_{i=0}^{\infty} a_i(\lambda)\, H_i(x)$$

for all x, where $H_0(x) \equiv 1$, $\{H_i(x),\ i \geq 0\}$ are the Hermite polynomials and

(11) $$\xi_i = a_i(\lambda) = \int_{-\infty}^{\infty} H_i(x)\ \ln f(x,\lambda)\ d\mu(x),\ i \geq 1.$$

Thus

(12) $$f(x,\lambda) = \exp\{\sum_{i=1}^{\infty} a_i(\lambda)\ H_i(x) + a_0(\lambda)\},\ -\infty < \lambda < \infty,$$

form an IPE family of densities.

Example 2. (Infinitomial) Let $\mathcal{X} = \{x_0, x_1, x_2, \ldots\}$ and $\mu$ and $\nu$ be both counting measures. Suppose X has a discrete distribution with

$$P(X = x_i) = p_i, \quad i \geq 0,$$

where $p_i \geq 0$ for all $i \geq 0$ and $\sum_{i=0}^{\infty} p_i = 1$.

Define, for $i \geq 0$,

(13) $$J_i(x) = \begin{cases} 1 & \text{if } x = x_i \\ 0 & \text{if } x \neq x_i \end{cases} .$$

The $\{J_i(x),\ i \geq 0\}$ form a CONB for $\ell_2$. The probability function of X can be written in the form

$$(14)\qquad \prod_{i=0}^{\infty} p_i^{J_i(x)} = \exp\{\sum_{i=0}^{\infty} \xi_i J_i(x)\},\ x \in \mathcal{X}\ ,$$

where $\xi_i = \ln p_i$, $i \geq 0$. In (14), we adopt the convention that $0 \times (-\infty) = 0$. Here

$$(15)\qquad \Omega = \{\underset{\sim}{\xi} = (\xi_0, \xi_1, \ldots):\ \sum_{i=0}^{\infty} e^{\xi_i} = 1,\ \sum_{i=0}^{\infty} \xi_i^2 < \infty\}.$$

Example 3. Let X be a random vector defined on a measurable space $(\mathcal{X}, \mathcal{B})$ with density f with respect to a $\sigma$-finite measure $\nu$. Suppose $\ln f \in L_2(\mu)$ where $\mu$ is a $\sigma$-finite measure equivalent to $\nu$. Let $\{H_i(x),\ i \geq 1\}$ be a CONB for $L_2(\mu)$. Then

$$(16)\qquad \ln\ f(x) = \sum_{i=1}^{\infty} a_i(f)\ H_i(x),\quad x \in \mathcal{X},$$

where

$$(17)\qquad a_i(f) = \int_{\mathcal{X}} H_i(x)\ \ln\ f(x)\ d\mu(x),\ i \geq 1.$$

Hence

$$(18)\qquad f(x) = \exp[\sum_{i=1}^{\infty} a_i(f)\ H_i(x)],$$

which shows that f is a member of IPE family. This representation of f can be used for a nonparametric estimation of f. Similar approaches for nonparametric density estimation have been adopted in literature in the past where the authors have instead assumed that $f \in L_2(\nu)$ (see Prakasa Rao [1983)).

Example 4. (Gaussian processes) Let $\{X(t),\ 0 \leq t \leq T\}$ be a Gaussian process with mean function $m(t)$ and continuous covariance function $R(s,t)$. It is known that the process $X(t)$ can be represented in the form

$$X(t) = m(t) + \sum_{j=1}^{\infty} \lambda_j^{\frac{1}{2}} \phi_j(t)\, Y_j\,, \tag{19}$$

where $\lambda_j$ and $\phi_j(\cdot)$ satisfy the integral equation

$$\lambda_j \phi_j(s) = \int_0^T R(s,t)\, \phi_j(t)\, dt,\ 0 \leq s \leq T \tag{20}$$

and $Y_j$ are i.i.d. $N(0,1)$, $j = 1,2,\ldots$ (see Basawa and Prakasa Rao (1980), p. 167 and Grenander (1981), p. 64). Let, for $j = 1,2,\ldots,$

$$a_j = \int_0^T m(t)\, \phi_j(t)\, dt \tag{21}$$

and

$$Z_j = a_j + \lambda_j^{\frac{1}{2}} Y_j. \tag{22}$$

Then $Z_j$, $j = 1,2,\ldots$ are independent $N(a_j,\lambda_j)$. The process $\{X(t),\ 0 \leq t \leq T\}$ is completely determined by the sequence $\{Z_j,\ j \geq 1\}$ and vice versa. Let $\mu_m$ denote the probability measure induced by the process $X(t)$ on $L_2([0,T])$ and $\mu_0$ denote the probability measure induced by the same process when $m(t) \equiv 0$. It is known that $\mu_m << \mu_0$ if $\sum_{j=1}^{\infty} (a_j^2/\lambda_j) < \infty$ (see Chatterjee and Mandrekar (1978)). In fact, the Randon-Nikodym derivative can be written in the form

$$(23) \qquad \frac{d\mu_m}{d\mu_0} = \exp\left\{ \sum_{j=1}^{\infty} Z_j \frac{a_j}{\lambda_j} - \frac{1}{2} \sum_{j=1}^{\infty} \frac{a_j^2}{\lambda_j} \right\}.$$

Here the space $\mathcal{X}$ is the space of sequences $(z_1, z_2, \ldots)$, $\mathcal{B}$ is the $\sigma$-algebra generated by finite dimensional cylinder sets and $\nu = \mu_0$. Let us also choose $\mu = \mu_0$. Then it is clear from the earlier construction that a basis for $L_2(\mathcal{X}, \mathcal{B}, \mu_0)$ is the completion under $L_2$-norm of all functions

$$H_{i_1}\left(\frac{z_{i_1}}{\lambda_{i_1}^{\frac{1}{2}}}\right), \ldots, H_{i_k}\left(\frac{z_{i_k}}{\lambda_{i_k}^{\frac{1}{2}}}\right),$$

where $\{H_i(x),\ i \geq 1\}$ with $H_0(x) \equiv 1$ is the sequence of Hermite polynomials. In this framework, the density is given by

$$(24) \qquad \frac{d\mu_m}{d\mu_0} = \exp\left\{ \sum_{j=1}^{\infty} \frac{a_j}{\lambda_j^{\frac{1}{2}}} H_1\left(\frac{z_j}{\lambda_j^{\frac{1}{2}}}\right) - \frac{1}{2} \sum_{j=1}^{\infty} \frac{a_j^2}{\lambda_j} \right\}$$

and the family

$$\left\{ \frac{d\mu_m}{d\mu_0}:\ m(\cdot) \text{ such that } \sum_{j=1}^{\infty} \frac{a_j^2}{\lambda_j} < \infty \right\}.$$

is an IPE family for a fixed $\{\lambda_j\}$ or equivalently for a fixed covariance function $R(s,t)$.

Example 5. Let $\{X_i,\ i \geq 1\}$ be i.i.d. $N(0,1)$ and $\{Y_i,\ i \geq 1\}$ be independent random variables with $Y_i$ distributed as $N(\mu_i, \sigma_i^2)$ such that

(i) $0 < c_1 \leq \sigma_i \leq c_2 < \infty$,

(ii) $\sum_{i=1}^{\infty} |1-\sigma_i^2| < \infty$, and

$$\text{(iii)} \quad \sum_{i=1}^{\infty} \mu_i^2 < \infty.$$

Let $P_{\underset{\sim}{\mu},\underset{\sim}{\sigma}}$ and $P_{\underset{\sim}{0},\underset{\sim}{1}}$ be the probability measures induced by the processes $\{Y_i, i \geq 1\}$ and $\{X_i, i \geq 1\}$ respectively on $(R^{\infty}, \mathcal{B}_{\infty})$. It is known that $P_{\underset{\sim}{\mu},\underset{\sim}{\sigma}}$ and $P_{\underset{\sim}{0},\underset{\sim}{1}}$ are equivalent for every $(\underset{\sim}{\mu},\underset{\sim}{\sigma})$ satisfying (i) - (iii) (see Chatterjee and Mandrekar (1978)), so that the Radon-Nikodym derivative is given by

$$\text{(25)} \qquad \frac{dP_{\underset{\sim}{\mu},\underset{\sim}{\sigma}}}{dP_{\underset{\sim}{0},\underset{\sim}{1}}} = \prod_{i=1}^{\infty} \left[\exp\left\{-\tfrac{1}{2}\frac{(z_i-\mu_i)^2}{\sigma_i^2} + \tfrac{1}{2} z_i^2\right\}\sigma_i^{-1}\right]$$

$$= \exp\left\{-\tfrac{1}{2}\sum_{i=1}^{\infty} z_i^2 \left(\frac{1}{\sigma_i^2} - 1\right) + \sum_{i=1}^{\infty} \mu_i z_i - c(\underset{\sim}{\mu},\underset{\sim}{\sigma})\right\},$$

where $c(\underset{\sim}{\mu},\underset{\sim}{\sigma})$ is the appropriate normalizing constant. This in turn can be rewritten as an IPE family of distributions in terms of $\{H_i(x)\}$ as was done in the last example.

## $\underset{\sim}{2}$. PROPERTIES OF IPE FAMILIES.

In this section, we state and prove some important properties of IPE families defined through (1) - (4).

Proposition 1. (i) The natural parameter space $\Omega$ is convex.

(ii) The family $\mathcal{E}(\mu,\nu)$ is identifiable.

(iii) Let $\phi(\cdot)$ be any bounded measurable function on $(\mathcal{X}, \mathcal{B})$. Then the integral

$$\text{(26)} \qquad \int_{\mathcal{X}} \phi(x) \exp\left[\sum_{i=1}^{\infty} \xi_i H_i(x)\right] d\nu(x)$$

considered as a function of the complex variables $\xi_j = \theta_j + i\,\delta_j$ $(j = 1,2,\dots,k)$ with $k < \infty$ while $\xi_j$'s for $j > k$ considered as real and fixed with $\sum_{j=k+1}^{\infty} \xi_j^2 < \infty$, is an analytic function in each of these variables in the region R of parameter points for which $(\theta_1,\dots,\theta_k)$ is an interior point of $\Omega \cap (\underset{\sim}{\xi}: \xi_j$ for $j > k$ fixed with $\sum_{j=k\_1}^{\infty} \xi_j^2 < \infty)$ considered as a set in $R^k$. In particular, this finite dimensional analyticity property holds at every point in the interior of $\Omega$.

(iv) In view of (iii), the partial derivatives of all orders with respect to $\xi$'s of the integral (26) can be computed under the integral sign.

PROOF. (i) The proof is standard and follows from Holder inequality.

(ii) Let $\underset{\sim}{\xi}$ and $\underset{\sim}{\xi}'$ both in $\Omega$ be such that

$$f(x,\underset{\sim}{\xi}) = f(x,\underset{\sim}{\xi}'),\ a{\cdot}e[\nu],$$

or equivalently from (4).

$$\sum_{i=1}^{\infty} (\xi_i - \xi_i')\, H_i(x) = C(\underset{\sim}{\xi}) - C(\underset{\sim}{\xi}'),\ a{\cdot}e[\mu]. \tag{27}$$

Since left side and hence the right side of (27) is square integrable with respect to $\mu$, it follows that $\mu(\mathcal{X}) < \infty$. Consequently as per our definition of IPE family, $\{H_i(x),\ i \geq 0\}$ with $H_0(x) \equiv 1$ is a CONB for $L_2(\mu)$. From this fact and the relation (27), it follows that $\xi_i = \xi_i'$, $i \geq 1$.

The proofs for (iii) and (iv) follow the standard arguments for the finite dimensional exponential families (see Lehmann (1959)). □

Let $T_i = H_i(X)$, $i \geq 1$ and $\underset{\sim}{T} = (T_1, T_2, \ldots)$, where the distribution of X belongs to IPE family defined by (1) - (4). Then the following result is given without proof as it follows from standard arguments with minor changes (see Lehmann (1959), p. 52).

PROPOSITION 2. (i) The distribution of $\underset{\sim}{T}$ for any $\underset{\sim}{\xi} \in \Omega$ is of an IPE form given by

$$\text{(28)} \qquad dP_{\underset{\sim}{\xi}}^{\underset{\sim}{T}}(\underset{\sim}{t}) = \exp\left[\sum_{i=1}^{\infty} \xi_i t_i - C(\xi)\right] \cdot d\lambda(\underset{\sim}{t}),$$

for some measure $\lambda$ defined on the Hilbert space $\ell_2$.

(ii) For any proper subset A of $Z^+ = \{1,2,\ldots\}$, the conditional distribution of $\{T_i,\ i \in A\}$ given $\{T_i,\ i \in Z^+ - A\}$ exists and is again a member of an IPE family.

PROPOSITION 3. The family of distributions of T given by (28) for $\underset{\sim}{\xi} \in \Omega$ is complete provided the interior of $\Omega$ is nonempty.

PROOF. By making an appropriate translation of the parameter space, we may assume without loss of generality that $\underset{\sim}{0} \in \Omega^0$ and hence, for every $k \geq 1$, there exists a constant $a_k > 0$ such that

$$I_k = \{(\xi_1, \xi_2, \ldots, \xi_k, 0, 0, \ldots) : -a_k < \xi_j < a_k,\ j = 1, 2, \ldots, k; \xi_j = 0 \text{ for } j \geq k\} \subset \Omega^0.$$

Let $f(\underset{\sim}{t})$ be a measurable function of $\underset{\sim}{t} \in \ell_2$ such that

$$\text{(29)} \qquad E_{\underset{\sim}{\xi}}\, f(\underset{\sim}{T}) = 0, \quad \text{for all } \underset{\sim}{\xi} \in \Omega.$$

Then, for every $k \geq 1$, for all $\underset{\sim}{\xi} \in I_k$,

(30) $$\int \exp\left(\sum_{j=1}^{k} \xi_j t_j\right) f^+(\underset{\sim}{t})\, d\lambda(\underset{\sim}{t}) = \int \exp\left(\sum_{j=1}^{k} \xi_j t_j\right) f^-(\underset{\sim}{t})\, d\lambda(\underset{\sim}{t}),$$

where $f(\underset{\sim}{t}) = f^+(\underset{\sim}{t}) - f^-(\underset{\sim}{t})$. Since $\underset{\sim}{0} \in I_k$, the relation (30) implies that

(31) $$\int f^+(\underset{\sim}{t})\, d\lambda(\underset{\sim}{t}) = \int f^-(\underset{\sim}{t})\, d\lambda(\underset{\sim}{t}).$$

If the common value of (31) is zero, then $f(\underset{\sim}{t}) = 0$, a$\cdot$e$[\lambda]$ and the result is proved. If the common value is positive, then dividing (30) by this common value, we may write

(32) $$\int \exp\left(\sum_{j=1}^{k} \xi_j t_j\right) dP^+(\underset{\sim}{t}) = \int \exp\left(\sum_{j=1}^{k} \xi_j t_j\right) dP^-(\underset{\sim}{t}),$$

where

(33) $$dP^+(\underset{\sim}{t}) = f^+(\underset{\sim}{t})\, d\lambda(\underset{\sim}{t}),\ dP^-(\underset{\sim}{t}) = f^-(\underset{\sim}{t})\, d\lambda(\underset{\sim}{t})$$

are probability measures on $\ell_2$. Following the standard argument (see Lehmann (1959), p. 132-33), it can be shown that

(34) $$e^{i\sum_{j=1}^{k} \theta_j t_j}\, dP^+(\underset{\sim}{t}) = e^{i\sum_{j=1}^{k} \theta_j t_j}\, dP^-(\underset{\sim}{t}),$$

for all real $\theta_j$, $j = 1,\ldots,k$, and for every $k \geq 1$. From (34), by letting $k \to \infty$, and using bounded convergence theorem, it follows that for every

$\underset{\sim}{\theta} = (\theta_1, \theta_2, \ldots)$ subject to $||\theta|| = (\sum_{i=1}^{\infty} \theta_i^2)^{\frac{1}{2}} < \infty$. We have

$$(35) \qquad \int e^{i\sum_{j=1}^{\infty} \theta_j t_j} \, dP^+(\underset{\sim}{t}) = \int e^{i\sum_{j=1}^{\infty} \theta_j t_j} \, dP^-(\underset{\sim}{t}),$$

and the characteristic functionals of the probability measures $P^+$ and $P^-$ on Hilbert space $\ell_2$ coincide. Thus, using Theorem 6.2.1 of Grenander ((1963), p. 130), it follows that $P^+$ and $P^-$ are identical. Consequently, from (33), we have $f^+(\underset{\sim}{t}) = f^-(\underset{\sim}{t})$ or equivalently $f(\underset{\sim}{t}) = 0$, a·e[$\lambda$]. □

<u>PROPOSITION 4</u>. The statistic $\underset{\sim}{T}(X) = (H_1(X), H_2(X), \ldots)$ is a sufficient statistic for the infinite parameter exponential family $\mathcal{E}(\mu,\nu)$ given by (4).

The above proposition is an immediate consequence of Halmos-Savage Factorization theorem. We omit the proof.

Let

$$(36) \qquad \Omega_1 = \{\underset{\sim}{\xi} \in \Omega: E_{\underset{\sim}{\xi}} \, ||T|| < \infty\}.$$

Note that $\Omega_1$ is still convex. Assume that $\Omega_1^0$ is nonempty. Then, by Proposition 1, the partial derivative vector

$$\dot{\underset{\sim}{C}}(\underset{\sim}{\xi}) = \left( \frac{C(\underset{\sim}{\xi})}{\partial \xi_1}, \quad \frac{C(\underset{\sim}{\xi})}{\partial \xi_2}, \ldots \right)$$

exists for $\underset{\sim}{\xi} \in \Omega_1^0$ and

$$(37) \qquad \dot{\underset{\sim}{C}}(\underset{\sim}{\xi}) = (\eta_1(\underset{\sim}{\xi}), \eta_2(\underset{\sim}{\xi}), \ldots), \; \underset{\sim}{\xi} \in \Omega_1^0 \quad ,$$

where

$$\eta_j(\underset{\sim}{\xi}) = E_{\underset{\sim}{\xi}}[H_j(X)], \quad j = 1,2,\dots . \tag{38}$$

We extend the definition of $\dot{\underset{\sim}{C}}(\underset{\sim}{\xi})$, to $\Omega_1$, through the relation (37) since $\eta_j$'s exist for all $\underset{\sim}{\xi} \in \Omega_1$.

<u>LEMMA 1</u> (i) For all $\underset{\sim}{\xi}_1 \neq \underset{\sim}{\xi}_2$ in $\Omega_1$,

$$\underset{\sim}{\xi}_1' \; \underset{\sim}{\eta}(\underset{\sim}{\xi}_2) - C(\underset{\sim}{\xi}_1) < \underset{\sim}{\xi}_2' \; \underset{\sim}{\eta}(\underset{\sim}{\xi}_2) - C(\underset{\sim}{\xi}_2). \tag{39}$$

(ii) The mapping $\underset{\sim}{\eta}(\underset{\sim}{\xi}) = \dot{\underset{\sim}{C}}(\underset{\sim}{\xi})$ is 1-1 on $\Omega_1$.

(iii) $C(\underset{\sim}{\xi})$ is strictly convex, lower semi-continuous ($\ell$.s.c) on $\Omega_1$ and hence a continuous function for $\underset{\sim}{\xi} \in \Omega_1^0$.

<u>PROOF</u>. By Jensen inequality, we have for $\underset{\sim}{\xi}_1, \underset{\sim}{\xi}_2 \in \Omega_1$,

$$E_{\underset{\sim}{\xi}_2}[\ell n \; \{f(X,\underset{\sim}{\xi}_1)/f(X,\underset{\sim}{\xi}_2)\}] \leq 0 \tag{40}$$

equality holding if and only if $\underset{\sim}{\xi}_1 = \underset{\sim}{\xi}_2$. From this and the fact that $E_{\underset{\sim}{\xi}} \; ||T|| < \infty$, for $\underset{\sim}{\xi} \in \Omega_1$, the relation (39) follows. (ii) is now an easy consequence of (i). Again in (iii), the convexity of $C(\underset{\sim}{\xi})$ follows from Holder inequality. To prove the lower semi-continuity of $C(\underset{\sim}{\xi})$, it is sufficient to prove that

$$\psi(\underset{\sim}{\xi}) = \int \exp(\underset{\sim}{\xi}'H) \; d\nu \; ,$$

where $\underset{\sim}{H}' = (H_1, H_2, \ldots)$, is $\ell$.s.c. For this, let $\underset{\sim}{\xi}_0 \in \Omega_1^0$. Then, for $\underset{\sim}{\xi} \in \Omega_1^0$,

$$(41) \qquad \psi(\underset{\sim}{\xi}) - \psi(\underset{\sim}{\xi}_0) = \int \exp(\underset{\sim}{\xi}_0'\underset{\sim}{H})(\exp((\underset{\sim}{\xi}-\underset{\sim}{\xi}_0)'H)-1)d\nu$$

$$\geq \int \exp(\underset{\sim}{\xi}_0'\underset{\sim}{H})((\underset{\sim}{\xi}-\underset{\sim}{\xi}_0)'H)\, d\nu.$$

However

$$|\int \exp(\underset{\sim}{\xi}_0'\underset{\sim}{H})((\underset{\sim}{\xi}-\underset{\sim}{\xi}_0)'\underset{\sim}{H})\, d\nu|$$

$$\leq ||\underset{\sim}{\xi}-\underset{\sim}{\xi}_0|| \int ||\underset{\sim}{H}|| \exp(\underset{\sim}{\xi}_0'\underset{\sim}{H})\, d\nu \leq \varepsilon$$

for

$$||\underset{\sim}{\xi}-\underset{\sim}{\xi}_0|| \leq \varepsilon[\int ||\underset{\sim}{H}|| \exp(\underset{\sim}{\xi}_0'\underset{\sim}{H})\, d\nu]^{-1}.$$

Thus it follows from (41) that for such $\underset{\sim}{\xi}$

$$\psi(\underset{\sim}{\xi}) \geq \psi(\underset{\sim}{\xi}_0) - \varepsilon$$

and hence $C(\underset{\sim}{\xi})$ is $\ell$.s.c. in $\Omega_1$. Finally the continuity of $C(\underset{\sim}{\xi})$ in $\Omega_1^0$ follows from a result in Roberts and Varberg ((1973), p. 112). □

Let, for $\underset{\sim}{y} \in \ell_2$ and $\underset{\sim}{\xi} \in \Omega_1$,

$$(42) \qquad q(\underset{\sim}{y}|\underset{\sim}{\xi}) = \underset{\sim}{\xi}'\underset{\sim}{y} - C(\underset{\sim}{\xi}),$$

(43) $$q(\underset{\sim}{y}|V) = \sup\{q(\underset{\sim}{y}|\underset{\sim}{\xi}): \underset{\sim}{\xi} \in V\},$$

and

(44) $$B(V) = \{\underset{\sim}{y}: q(\underset{\sim}{y}|V) < \infty\},$$

for $V \subset \Omega_1$. The following lemma will be useful in the next section where consistency properties of certain estimators of $\underset{\sim}{\xi}$ are studied.

LEMMA 2. (i) For every $V \subset \Omega_1$, the function $q(\cdot|V)$ is convex, $\ell$.s.c. and hence continuous.

(ii) The function $q(\underset{\sim}{y}|\cdot)$ is strictly concave and u.s.c. over $\Omega_1$.

PROOF: (i) The properties that $q(\cdot|V)$ is convex and $\ell$.s.c. follow from its definition and that (42) is linear in $\underset{\sim}{y}$. These two properties in turn imply its continuity (see Roberts and Varberg (1973), p. 112). Part (ii) is a consequence of Lemma 1 (iii). □

## 3. STRONG CONSISTENCY.

In this section, we shall study the uniqueness and strong consistency of the maximum likelihood estimator (MLE) of $\underset{\sim}{\xi}$. Let $X_1,\ldots,X_n$ be i.i.d. with common density belonging to $\mathcal{E}(\mu,\nu)$. Let

(45) $$\overline{\underset{\sim}{H}}_n = (n^{-1}\sum_{j=1}^{n} H_1(X_j),\ n^{-1}\sum_{j=1}^{n} H_2(X_j),\ldots).$$

Note that maximizing the likelihood function (based on the sample $X_1,\ldots,X_n$) with respect to $\underset{\sim}{\xi}$ is equivalent to maximizing the function $q(\overline{\underset{\sim}{H}}_n|\underset{\sim}{\xi})$ over $\underset{\sim}{\xi}$. The following lemma will be used to prove the main result.

LEMMA 3. Let $\Omega_1$ be locally compact. Then, for every $\xi \in \Omega_1^0$ and sufficiently small $\varepsilon > 0$,

$$\sup_{s \in V} q(\eta(\xi)|s) < q(\eta(\xi)|\xi) \tag{46}$$

where $V = \Omega_1 - S$,

$$S = \{s \in \Omega_1 : ||s-\xi|| \leq \varepsilon\} \tag{47}$$

and $q(\eta|s)$ is as defined in (42).

PROOF. Fix $\xi \in \Omega_1^0$. For $s \neq \xi$, by Lemma 1, we have

$$s'\eta(\xi) - C(s) < \xi'\eta(\xi) - C(\xi), \quad s \in \Omega_1. \tag{48}$$

If $C(s) = \infty$, then (48) holds trivially. Hence for all $s \neq \xi$ and $s \in \overline{\Omega}_1$, we have

$$s'\eta(\xi) - C(s) < q(\eta(\xi)|\xi) \tag{49}$$

Note that $q(\eta(\xi)|\cdot)$ is upper semi-continuous (u.s.c.) in view of Lemma 1 (iii) and (42). Furthermore $\overline{V}$ is compact. Consequently $q(\eta(\xi)|s)$ attains its supremum for $s \in \overline{V}$ (see Ash (1972), p. 389). The inequality (46) now follows from (49) and the fact that $\xi \notin \overline{V}$. □

THEOREM 1. Let $\Omega_1$ be locally compact and $\eta(\xi) = E_{\xi}H(X) \in B^0(\Omega_1)$, where $B(\cdot)$ is as defined in (44) and $\xi \in \Omega_1^0$ be the "true" parameter. Then there exists a measurable sequence $\{\hat{\xi}_n\}$ such that $\hat{\xi}_n \in \Omega_1$ and

$$(50) \qquad ||\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}|| \to 0, \text{ a.e as } n \to \infty.$$

Moreover this sequence is a unique MLE for $\underset{\sim}{\xi}$ for large enough n.

PROOF. $\Omega_1$ being locally compact there exists an $\varepsilon > 0$ such that the set

$$S = \{\underset{\sim}{s} \in \Omega_1 : ||\underset{\sim}{s} - \underset{\sim}{\xi}|| \leq \xi\}$$

is compact. Define $V = \Omega_1 - S$. As a first step, we shall show that for every sample point w, there exists n(w) such that

$$(51) \qquad \sup_{\underset{\sim}{s} \in S} q(\underset{\sim}{\bar{H}}_n | \underset{\sim}{s}) > \sup_{\underset{\sim}{s} \in V} q(\underset{\sim}{\bar{H}}_n | \underset{\sim}{s}) = q(\underset{\sim}{\bar{H}}_n | V),$$

for $n \geq n(w)$. By strong law of large numbers (with strong convergence) for Hilbert-space valued random elements (see Grenander (1963), Theorem 6.4.2, p. 144) it follows that $||\underset{\sim}{\bar{H}}_n - \underset{\sim}{\eta}(\underset{\sim}{\xi})|| \to 0$, a.e, as $n \to \infty$. Consequently, since $\underset{\sim}{\eta}(\underset{\sim}{\xi}) \in B^0(\Omega_1)$ by hypothesis, for every w, there exists $n_1(w)$ such that

$$(52) \qquad \underset{\sim}{\bar{H}}_n \in B^0(\Omega_1), \text{ for } n \geq n_1(w).$$

Again since $q(\cdot|V)$ is continuous by Lemma 2,

$$(53) \qquad q(\underset{\sim}{\bar{H}}_n | V) \to q(\underset{\sim}{\eta}(\xi) | V), \text{ a.e.}$$

Analogously we also have

$$(54) \qquad q(\underset{\sim}{\bar{H}}_n | \underset{\sim}{\xi}) \to q(\underset{\sim}{\eta}(\underset{\sim}{\xi}) | \underset{\sim}{\xi}), \text{ a.e.}$$

From (52), (53), (54) and (46), it follows that for almost every w, there exists n(w) such that for $n \geq n(w)$

$$(55) \qquad q(\underset{\sim}{\overline{H}}_n | \underset{\sim}{\xi}) > q(\underset{\sim}{\overline{H}}_n | V).$$

Since $\underset{\sim}{\xi} \in S$, (51) follows from (55). Thus, eventually, the global supremum of $q(\underset{\sim}{\overline{H}}_n|\cdot)$ is same as that over S. However, since S is compact and $q(\underset{\sim}{\overline{H}}_n|\cdot)$ is u.s.c., it follows from a known selection theorem (see Himmelberg et.al. (1976), p. 391), that there exists for every n, a measurable function $\hat{\xi}_n$ with values in S such that

$$(56) \qquad \sup_{\underset{\sim}{s} \in S} q(\underset{\sim}{\overline{H}}_n | \underset{\sim}{s}) = q(\underset{\sim}{\overline{H}}_n | \hat{\xi}_n).$$

Since $\varepsilon > 0$ is arbitrary and $\hat{\underset{\sim}{\xi}}_n \in S$, the existence of $\{\hat{\underset{\sim}{\xi}}_n\}$ which is a MLE for large enough n and the relation (50) now follow from (51) and (56). Finally the uniqueness of the estimator follows from the strict concavity of $q(\underset{\sim}{\overline{H}}_n|\cdot)$ (see Lemma 2) and the fact that $\Omega_1$ is convex. □

The following theorem gives a method of obtaining the MLE via the likelihood equation.

<u>THEOREM 2</u>. Let C(s) be Fréchet-differentiable for all $s \in \Omega_1^0$. Then, for almost every sample point w, the likelihood equation

$$(57) \qquad \underset{\sim}{\overline{H}}_n = \dot{\underset{\sim}{C}}(\underset{\sim}{s})$$

has a unique solution $\underset{\sim}{\xi}_n^*$ for n large enough, which maximizes the likelihood function over $\Omega_1^0$ provided that

$$\mathcal{D}(\Omega_1^0) = \{\underset{\sim}{y}:\ \underset{\sim}{y} = \underset{\sim}{\eta}(\underset{\sim}{s}) \text{ for some } \underset{\sim}{s} \in \Omega_1^0\}$$

is open.

PROOF. Since $||\overline{\underset{\sim}{H}}_n - \underset{\sim}{\eta}(\underset{\sim}{\xi})|| \to 0$, a$\cdot$e, as $n \to \infty$, where $\underset{\sim}{\xi} \in \Omega_1^0$ is the "true" parameter, for almost every sample point w, there exists n(w) such that for $n \geq n(w)$, $\overline{\underset{\sim}{H}}_n \in \mathcal{D}(\Omega_1^0)$. Consequently, since $\dot{\underset{\sim}{C}}(\underset{\sim}{s}) = \underset{\sim}{\eta}(\underset{\sim}{s})$ is 1-1 from $\Omega_1^0$ to $\mathcal{D}(\Omega_1^0)$, equation (57) has a unique solution for $n \geq n(w)$. Denote this solution by $\underset{\sim}{\xi}_n^*$. Since $C(\underset{\sim}{s})$ is convex on the open convex set $\Omega_1^0$ and the Fréchet-derivative $C(\underset{\sim}{\xi}_n^*)$ exists, it follows from Theorem A in Roberts and Varberg ((1973), p. 98) that

$$C(\underset{\sim}{s}) - C(\underset{\sim}{\xi}_n^*) \geq (\underset{\sim}{s} - \underset{\sim}{\xi}_n^*)'\ \dot{\underset{\sim}{C}}(\underset{\sim}{\xi}_n^*), \text{ for } \underset{\sim}{s} \in \Omega_1^0 \tag{58}$$

or equivalently

$$\underset{\sim}{s}'\ \dot{\underset{\sim}{C}}(\underset{\sim}{\xi}_n^*) - C(\underset{\sim}{s}) \leq \underset{\sim}{\xi}_n^{*\prime}\ \dot{\underset{\sim}{C}}(\underset{\sim}{\xi}_n^*) - C(\underset{\sim}{\xi}_n^*), \text{ for } \underset{\sim}{s} \in \Omega_1^0. \tag{59}$$

Using (57) for $\underset{\sim}{s} = \underset{\sim}{\xi}_n^*$ in (59), yields

$$\underset{\sim}{s}'\ \overline{\underset{\sim}{H}}_n - C(\underset{\sim}{s}) \leq \underset{\sim}{\xi}_n^{*\prime}\overline{\underset{\sim}{H}}_n - C(\underset{\sim}{\xi}_n^*), \text{ for } \underset{\sim}{s} \in \Omega_1^0 \tag{60}$$

which proves that $\underset{\sim}{\xi}_n^*$ maximizes the likelihood function over $\Omega_1^0$. □

## 4. WEAK CONSISTENCY AND ASYMPTOTIC NORMALITY.

In this section, we shall show that the MLE $\hat{\xi}_n$ is weakly consistent and asymptotically normally distributed under some conditions. For this we shall need the following lemma which is a simple consequence of Proposition 1 (iv).

LEMMA 4. Let $\xi \in \Omega^0$,

$$C_{ij}(\xi) = \frac{\partial^2 C(\xi)}{\partial \xi_i \partial \xi_j}, \tag{61}$$

and

$$\sigma_{ij}(\xi) = E_\xi\{H_i(X)H_j(X)\} - \eta_i(\xi)\eta_j(\xi), \tag{62}$$

for i, j = 1,2,..., with

$$\Sigma(\xi) \equiv (\sigma_{ij}(\xi)); \; \ddot{C}(\xi) \equiv (C_{ij}(\xi)). \tag{63}$$

Then

$$\Sigma(\xi) = \ddot{C}(\xi). \tag{64}$$

THEOREM 3. Let $\mathcal{D}(\Omega_1^0)$ be open, where

$$\mathcal{D}(\Omega_1^0) = \{y: y = \eta(s), \text{ for some } s \in \Omega_1^0\}. \tag{65}$$

Also assume that for every $\xi \in \Omega_1^0$, $C(\xi)$ is twice Fréchet differentiable and that there exists a neighbourhood $N(\xi)$ such that

(66) $$0 < \inf_{\underset{\sim}{\theta} \in N(\underset{\sim}{\xi})} \rho(\ddot{\underset{\sim}{C}}(\underset{\sim}{\theta})) = K < \infty$$

where $\rho(\ddot{\underset{\sim}{C}}(\underset{\sim}{\theta}))$ is the infimum of the eigenvalues of the Hermitian operator $\ddot{\underset{\sim}{C}}(\underset{\sim}{\theta})$ in $\ell_2$. Then, given $\varepsilon > 0$ and $\underset{\sim}{\xi} \in \Omega_1^0$, there exists a set $A(n,\varepsilon)$ such that

(i) $P_{\underset{\sim}{\xi}}(A(n,\varepsilon)) > 1 - \varepsilon$, for $n \geq n(\varepsilon,\underset{\sim}{\xi})$, and

(ii) for every $w \in A(n,\varepsilon)$ and $n \geq n(\varepsilon,\underset{\sim}{\xi})$, the equation

(67) $$\overline{\underset{\sim}{H}}_n = \dot{\underset{\sim}{C}}(\underset{\sim}{s})$$

has a unique solution $\underset{\sim}{\xi}_n^*$.

(iii) For $n \geq n(\varepsilon,\underset{\sim}{\xi})$, define $\hat{\underset{\sim}{\xi}}_n = \underset{\sim}{\xi}_n^*$ for $w \in A(n,\varepsilon)$ and $\hat{\underset{\sim}{\xi}}_n = 0$ for $w \notin A(n,\varepsilon)$. Then $\hat{\underset{\sim}{\xi}}_n$ is measurable and

(68) $$P_{\underset{\sim}{\xi}}(||\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}|| > \varepsilon K^{-1}) \leq \varepsilon, \text{ for } n \geq n(\varepsilon,\underset{\sim}{\xi}).$$

(iv) $\hat{\underset{\sim}{\xi}}_n$ maximizes the likelihood function over $\Omega_1^0$ for $w \in A(n,\varepsilon)$.

PROOF. Fix $\underset{\sim}{\xi} \in \Omega_1^0$. By weak law of large numbers (with strong convergence) for Hilbert-Space valued random elements (see Grenander (1963)), for every $\varepsilon > 0$, there exists $n(\varepsilon,\underset{\sim}{\xi})$ such that

(69) $$P_{\underset{\sim}{\xi}}(||\overline{\underset{\sim}{H}}_n - \dot{\underset{\sim}{C}}(\underset{\sim}{\xi})|| > \varepsilon) \leq \varepsilon.$$

Define $A(n,\varepsilon) = [||\overline{\underset{\sim}{H}}_n - \dot{\underset{\sim}{C}}(\underset{\sim}{\xi})|| \leq \varepsilon]$. Choose $\varepsilon$ sufficiently small so that

(70) $\{\underset{\sim}{y}: ||\underset{\sim}{y} - \dot{\underset{\sim}{C}}(\underset{\sim}{\xi})|| \leq \varepsilon\} \subset \mathscr{D}(\Omega_1^0)$.

This is possible since $\mathscr{D}(\Omega_1^0)$ is open. For $n \geq n(\varepsilon,\underset{\sim}{\xi})$ and $w \in A(n,\varepsilon)$, since $\overline{\underset{\sim}{H}}_n \in \mathscr{D}(\Omega_1^0)$ and since $\dot{\underset{\sim}{C}}(\underset{\sim}{\xi})$ is one to one from $\Omega_1^0$ to $\mathscr{D}(\Omega_1^0)$ (see Lemma 1 (ii)), the equation $\overline{\underset{\sim}{H}}_n = \dot{\underset{\sim}{C}}(\underset{\sim}{s})$ has a unique solution for $\underset{\sim}{s}$, say $\underset{\sim}{\xi}_n^*$. This proves (i) and (ii). For (iii) we proceed as follows

For $w \in A(n,\varepsilon)$ and $n \geq n(\varepsilon,\underset{\sim}{\xi})$

(71)
$$\begin{aligned}\varepsilon &\geq ||\overline{\underset{\sim}{H}}_n - \dot{\underset{\sim}{C}}(\underset{\sim}{\xi})|| \\ &= ||\dot{\underset{\sim}{C}}(\hat{\underset{\sim}{\xi}}_n) - \dot{\underset{\sim}{C}}(\underset{\sim}{\xi})|| \\ &= ||\ddot{\underset{\sim}{C}}(\hat{\underset{\sim}{s}}_n)(\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi})||\end{aligned}$$

where $\hat{\underset{\sim}{s}}_n = \underset{\sim}{\xi} + t\,(\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi})$ for some $0 < t < 1$. The last equality follows from the mean value theorem for functions defined on normed linear spaces (see Roberts and Varberg (1973), p. 71). On the other hand using (66) we have

(72) $$||\ddot{\underset{\sim}{C}}(\hat{\underset{\sim}{s}}_n)(\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi})|| \geq K\,||\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}||.$$

It therefore follows from (71), (72) and (69) that

(72a)
$$\begin{aligned}P_{\underset{\sim}{\xi}}(||\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}|| > \varepsilon K^{-1}) &\leq P_{\underset{\sim}{\xi}}(||\overline{\underset{\sim}{H}}_n - \dot{\underset{\sim}{C}}(\underset{\sim}{\xi})|| > \varepsilon) \\ &\leq \varepsilon.\end{aligned}$$

This proves the weak consistency of the sequence $\{\hat{\underset{\sim}{\xi}}_n\}$. Finally (iv), namely that $\hat{\underset{\sim}{\xi}}_n$ maximizes the likelihood function over $\Omega_1^0$ for $w \in A(n,\varepsilon)$ follows from the lines of argument as used for (58) - (60) in the proof of Theorem 2. □

It may be remarked here that the operator $\ddot{\underset{\sim}{C}}(\underset{\sim}{\xi})$ has an inverse whenever $0 < \rho(\ddot{\underset{\sim}{C}}(\underset{\sim}{\xi})) < \infty$ is satisfied. Again, in the following theorem, the symbol $\mathcal{N}(\underset{\sim}{m},\underset{\sim}{S})$ will stand for the Gaussian probability measure on $\ell_2$ with mean $\underset{\sim}{m}$ and covariance operator $\underset{\sim}{S}$ (see Grenander (1963), p. 140 for definition).

THEOREM 4. Let

(73) $$\Omega_2 = \{\underset{\sim}{\xi} \in \Omega: \ E_{\underset{\sim}{\xi}}||\underset{\sim}{H}(X)||^2 < \infty\},$$

and the conditions of Theorem 3 hold with $\Omega_1^0$ replaced by $\Omega_2^0$. Furthermore let

(74) $$||\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\xi})\ \ddot{\underset{\sim}{C}}(\underset{\sim}{\theta}_n) - \underset{\sim}{I}|| \xrightarrow{p} 0$$

whenever

(75) $$||\underset{\sim}{\theta}_n - \underset{\sim}{\xi}|| \xrightarrow{p} 0, \text{ as } n \to \infty.$$

Then the MLE $\hat{\xi}_n$ of Theorem 3 is weakly consistent (with strong convergence) and

(76) $$n^{\frac{1}{2}}(\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}) \xrightarrow{\mathcal{L}} \mathcal{N}(\underset{\sim}{0},\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\xi})) \text{ as } n \to \infty.$$

PROOF. The results of Theorem 3 continue to hold even when $\Omega_1$ is replaced here by $\Omega_2$. By the central limit theorem for Hilbert-space valued random elements (see Grenander (1963), Theorem 6.5.1, p. 145), we have

$$n^{\frac{1}{2}}(\bar{H}_n(X) - \eta(\xi)) \xrightarrow{\mathscr{L}} \mathcal{N}(0, \Sigma(\xi)), \tag{77}$$

as $n \to \infty$ whenever $\xi \in \Omega_2$. Again using the mean value theorem for functions defined on normed linear spaces, we have, following the notations in Theorem 3, for $n \geq n(\varepsilon, \xi)$ and $w \in A(n,\varepsilon)$,

$$\dot{C}(\hat{\xi}_n) = \dot{C}(\xi) + \ddot{C}(s_n^*)(\hat{\xi}_n - \xi), \tag{78}$$

with

$$||s_n^* - \xi|| \leq ||\hat{\xi}_n - \xi||. \tag{79}$$

However for $w \in A(n,\varepsilon)$, we also have

$$\bar{H}_n = \dot{C}(\hat{\xi}_n). \tag{80}$$

Thus, for $n \geq n(\varepsilon, \xi)$, we have

$$\begin{aligned} (81)\qquad & n^{\frac{1}{2}} \Sigma^{-1}(\xi)\ (\bar{H}_n - \dot{C}(\xi)) \\ &= n^{\frac{1}{2}} \Sigma^{-1}(\xi)\ \ddot{C}(s_n^*)(\hat{\xi}_n - \xi)\ I(A(n,\varepsilon)) \\ &\quad + n^{\frac{1}{2}} \Sigma^{-1}(\xi)(\bar{H}_n - \dot{C}(\xi))\ I(A^c(n,\varepsilon)) \end{aligned}$$

where I(A) is the indicator function of the event A. Since $I(A(n,\varepsilon)) \xrightarrow{p} 1$, as $n \to \infty$, the second term on the right side of (81) tends to zero in probability. Thus the limit distribution of the first term on the right side of (81) is same as that of the left side, which in view of (77) is $\mathcal{N}(\underset{\sim}{0},\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\xi}))$. However, in view of (79) and the condition (74) the first term on the right side of (81) has the same limit distribution as that of $n^{\frac{1}{2}}(\hat{\underset{\sim}{\xi}}_n-\underset{\sim}{\xi})$, which proves the theorem. Here at the end we have used Slutsky type theorem for Hilbert-space valued random elements. □

## 5. WEAK CONVERGENCE

In this section as an alternative to the classical approach adopted in section 4, we shall consider the weak convergence of the log-likelihood ratio process and thereby study the asymptotic distribution of the MLE. Let

(82) $$\psi_n(\underset{\sim}{s}) = n(\underset{\sim}{s}'\,\overline{\underset{\sim}{H}}_n - C(\underset{\sim}{s}))$$

be the log-likelihood function based on a sample $X_1,\ldots,X_n$ from $\mathcal{E}(\mu,\nu)$, defined in section 1 but with parameter $\underset{\sim}{s}\in\Omega_2^0$. Let $X_n(\underset{\sim}{t})$, $\underset{\sim}{t}\in\ell_2$ be the log-likelihood ratio process defined by

(83) $$\begin{aligned} X_n(\underset{\sim}{t}) &= \psi_n(\underset{\sim}{\xi}+\underset{\sim}{t}\,n^{-\frac{1}{2}}) - \psi_n(\underset{\sim}{\xi}) \\ &= n\{n^{-\frac{1}{2}}\underset{\sim}{t}'\,\overline{\underset{\sim}{H}}_n - C(\underset{\sim}{\xi}+\underset{\sim}{t}\,n^{-\frac{1}{2}}) + C(\underset{\sim}{\xi})\} \end{aligned}$$

which evidently takes values in the space of continuous functions of $\underset{\sim}{t}\in\ell_2$.

Let

(84) $$Y_n(\underset{\sim}{t}) = X_n(\underset{\sim}{t}) - E_{\underset{\sim}{\xi}}[X_n(\underset{\sim}{t})]$$

$$= n^{\frac{1}{2}}\, \underset{\sim}{t}'(\overline{\underset{\sim}{H}}_n - \dot{\underset{\sim}{C}}(\underset{\sim}{\xi})).$$

Since $\underset{\sim}{\xi} \in \ell_2$, for every $\underset{\sim}{t} \in \ell_2$, by the central limit theorem for Hilbert-space valued random elements (see Grenander (1963), Theorem 6.5.1, p. 145),

(85) $$Y_n(\underset{\sim}{t}) \xrightarrow{\mathscr{L}} \underset{\sim}{t}'\underset{\sim}{Z},$$

as $n \to \infty$, where $\underset{\sim}{Z}$ is distributed as $\mathcal{N}(\underset{\sim}{0}, \underset{\sim}{\Sigma}(\underset{\sim}{\xi}))$.

<u>THEOREM 5</u>. The process $Y_n(\underset{\sim}{t})$ for $\underset{\sim}{t} \in \ell_2$, $||\underset{\sim}{t}|| \leq \delta$, converges weakly, for every $\delta > 0$, as $n \to \infty$, to the process

(86) $$Y(\underset{\sim}{t}) = \underset{\sim}{t}'\underset{\sim}{Z},\ ||\underset{\sim}{t}|| \leq \delta.$$

<u>PROOF</u>. Since $\sum_{i=1}^{k} \lambda_i Y_n(\underset{\sim}{t}_i)$ converges weakly to $\sum_{i=1}^{k} \lambda_i Y(\underset{\sim}{t}_i)$ for every real $(\lambda_1, \ldots, \lambda_k)$ and arbitrary $\underset{\sim}{t}_i$, $i = 1, \ldots, k$ with $||\underset{\sim}{t}_i|| \leq \delta$ and $k \geq 1$, it follows from Cramér-Wold theorem that the finite dimensional distributions of the process $Y_n(\underset{\sim}{t})$ converge to those of the process $Y(\underset{\sim}{t})$. The theorem will follow once we prove the tightness condition

(87) $$\lim_{d \to 0} \underline{\lim}_{n \to \infty} P_{\underset{\sim}{\xi}}\Big( \sup_{||\underset{\sim}{t}_1 - \underset{\sim}{t}_2|| \leq d} |Y_n(\underset{\sim}{t}_1) - Y_n(\underset{\sim}{t}_2)| \leq \varepsilon \Big) = 1,$$

for every $\varepsilon > 0$. This can be shown by using (84) and (85) as follows:

(88) $$\lim_{d\to 0} \underline{\lim}_{n\to\infty} P_{\underset{\sim}{\xi}}\Big(\sup_{||\underset{\sim}{t}_1-\underset{\sim}{t}_2||\le d} |Y_n(\underset{\sim}{t}_1)-Y_n(\underset{\sim}{t}_2)| \le \varepsilon\Big)$$

$$\ge \lim_{d\to 0} \underline{\lim}_{n\to\infty} P_{\underset{\sim}{\xi}}(d||n^{\frac{1}{2}}(\overline{\underset{\sim}{H}}_n-\dot{\underset{\sim}{C}}(\underset{\sim}{\xi}))|| \le \varepsilon)$$

$$\ge \lim_{d\to 0} P_{\underset{\sim}{\xi}}(d||\underset{\sim}{Z}|| \le \varepsilon) = 1. \qquad \square$$

<u>THEOREM 6</u>. Let $C(\underset{\sim}{\xi})$ be twice continuously Frechét-differentiable at every $\underset{\sim}{\xi} \in \Omega_2^0$. Then the process $X_n(\underset{\sim}{t})$, $||\underset{\sim}{t}|| \le \delta$, converges weakly for every $\delta > 0$ to the process

(89) $$X(\underset{\sim}{t}) = \underset{\sim}{t}'\underset{\sim}{Z} - \tfrac{1}{2}(\underset{\sim}{\Sigma}(\underset{\sim}{\xi})\underset{\sim}{t}, \underset{\sim}{t}), \quad ||\underset{\sim}{t}|| \le \delta,$$

where $(\cdot,\cdot)$ denotes the inner product in $\ell_2$.

<u>PROOF</u>. The result follows from (64), (84), Theorem 5 and the fact that

(90) $$E_{\underset{\sim}{\xi}} X_n(\underset{\sim}{t}) = -\tfrac{1}{2}(\ddot{\underset{\sim}{C}}(\underset{\sim}{\xi}+\alpha\underset{\sim}{t}\, n^{-\frac{1}{2}})\underset{\sim}{t}, \underset{\sim}{t})$$

for some $0 < \alpha < 1$, which in turn follows from Taylor expansion of $C(\underset{\sim}{\xi} + \underset{\sim}{t}\, n^{\frac{1}{2}})$(see Roberts and Varberg (1973), p. 70). $\square$

<u>THEOREM 7</u>. Let, for every $\underset{\sim}{\xi} \in \Omega_2^0$ and for the $\psi_n(\underset{\sim}{s})$ as defined in (82), the following conditions hold:

(91) (i) $$\lim_{\delta\to\infty} \overline{\lim}_{n\to\infty} P_{\underset{\sim}{\xi}}\Big(\sup_{||\underset{\sim}{s}-\underset{\sim}{\xi}|| \ge \delta n^{-\frac{1}{2}}} \psi_n(\underset{\sim}{s}) > \psi_n(\underset{\sim}{\xi})\Big) = 0.$$

(ii) $C(\underset{\sim}{\xi})$ is twice continuously Fréchet-differentiable.

(iii) $\ddot{\underset{\sim}{C}}(\underset{\sim}{\xi})$ is a positive definite operator.

Then a measurable MLE $\hat{\underset{\sim}{\xi}}_n$ exists and

(92) $$n^{\frac{1}{2}}(\hat{\underset{\sim}{\xi}}_n-\underset{\sim}{\xi}) \xrightarrow{\mathcal{L}} \eta(\underset{\sim}{0},\underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\xi})), \text{ as } n \to \infty.$$

<u>PROOF</u>. Condition (91) guarantees that the likelihood function is maximized globally over $\Omega_2^0$ at a point within the region

(93) $$B_\delta = \{\underset{\sim}{s}: ||\underset{\sim}{s}-\underset{\sim}{\xi}|| \leq \delta\, n^{-\frac{1}{2}}\}$$

for large enough $\delta$ and with probability converging to one as $n \to \infty$, where $\underset{\sim}{\xi}$ is the "true" parameter. On the other hand, the continuity of the function $\psi_n(\underset{\sim}{s})$ and the compactness of the set $B_\delta$ implies the existence of a measurable MLE $\hat{\xi}_n$ over this set. Thus we restrict our attention to the parameter set $B_\delta$ in order to study the limit distribution of $\hat{\underset{\sim}{\xi}}_n$. The log-likelihood ratio process $X_n(\underset{\sim}{t})$ with $||\underset{\sim}{t}|| \leq \delta$, is the same as the process $\psi_n(\underset{\sim}{s}) - \psi_n(\underset{\sim}{\xi})$, for $\underset{\sim}{s} \in B_\delta$, under the transformation

(94) $$\underset{\sim}{s} = \underset{\sim}{\xi} + \underset{\sim}{t}\, n^{-\frac{1}{2}}.$$

By Theorem 6, the process $X_n(\underset{\sim}{t})$ with $||\underset{\sim}{t}|| \leq \delta$, converges weakly to the process $X(\underset{\sim}{t})$ as defined in (89) and hence for any continuous functional $h(\cdot)$ which maps the space of continuous functions on the space $||\underset{\sim}{t}|| \leq \delta$, $\underset{\sim}{t} \in \ell_2$ to the space $\ell_2$, the following holds (Billingsley (1969)):

$$(95) \qquad h(X_n) \xrightarrow{\mathscr{L}} h(X), \text{ as } n \to \infty.$$

In particular, it follows that $n^{\frac{1}{2}}(\underset{\sim}{\xi}_n - \underset{\sim}{\xi})$ has a limiting distribution which is the same as the distribution of the location of the supremum of the limiting process $X(\underset{\sim}{t})$. Again using the condition (iii), it can be verified directly that

$$(96) \qquad \underset{\sim}{t}'\underset{\sim}{Z} - \tfrac{1}{2}(\underset{\sim}{\Sigma}(\underset{\sim}{\xi})\, \underset{\sim}{t}, \underset{\sim}{t}) \leq \hat{\underset{\sim}{t}}\, \underset{\sim}{Z} - \tfrac{1}{2}(\underset{\sim}{\Sigma}(\underset{\sim}{\xi})\, \hat{\underset{\sim}{t}}, \hat{\underset{\sim}{t}})$$

for all $\underset{\sim}{t} \in \ell_2$, where

$$(97) \qquad \hat{\underset{\sim}{t}} = \underset{\sim}{\Sigma}(\underset{\sim}{\xi})^{-1}\, \underset{\sim}{Z},$$

thereby establishing the result (92). □

## 6. DENSITY ESTIMATION

In continuation to example 3 of section 1, we can estimate the density $f(x,\underset{\sim}{\xi})$ belonging to $\mathcal{E}(\mu,\nu)$ family, by $f(x,\hat{\underset{\sim}{\xi}}_n)$, where $\hat{\xi}_n$ is a MLE as discussed in the previous sections. More specifically, let

$$(98) \qquad f(x,\underset{\sim}{\xi}) = \exp\{\sum_{i=1}^{\infty} \xi_i H_i(x) - C(\underset{\sim}{\xi})\},\ \underset{\sim}{\xi} \in \Omega_2^0,$$

where $\mu$ is a probability measure and $\{H_i(x),\ i \geq 0\}$ with $H_0(x) \equiv 1$ be a CONB for $L_2(\mu)$. Let $\hat{\underset{\sim}{\xi}}_n$ be a MLE of $\underset{\sim}{\xi}$ based on a sample of size n from (98). Then the estimator $f(x,\hat{\underset{\sim}{\xi}}_n)$ converges to $f(x,\underset{\sim}{\xi})$ almost surely or in probability according as $\hat{\underset{\sim}{\xi}}_n$ is strongly or weakly consistent. One measure of the closeness of the estimator $f(x,\hat{\underset{\sim}{\xi}}_n)$ to $f(x,\underset{\sim}{\xi})$ may be taken to be

(99) $$n \int_{\mathcal{X}} [\ln f(x,\hat{\underset{\sim}{\xi}}_n) - \ln f(x,\underset{\sim}{\xi})]^2 \, d\mu(x),$$

which is equal to

(100) $$n \,||\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}||^2 + n(C(\hat{\underset{\sim}{\xi}}_n) - C(\underset{\sim}{\xi}))^2.$$

Under the assumption that $C(\underset{\sim}{\xi})$ is twice continuously Fréchet differentiable and $n^{\frac{1}{2}} (\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}) = 0_p(1)$, it can be shown that

(101) $$n\{[C(\hat{\underset{\sim}{\xi}}_n) - C(\underset{\sim}{\xi})]^2 - (\underset{\sim}{J}(\underset{\sim}{\xi})(\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}), (\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}))\} \xrightarrow{p} 0$$

as $n \to \infty$ where

(102) $$\underset{\sim}{J}(\underset{\sim}{\xi}) = \dot{\underset{\sim}{C}}(\underset{\sim}{\xi}) \, \dot{\underset{\sim}{C}}(\underset{\sim}{\xi})'.$$

Consequently, the limit distribution of (100), whenever it exists, is same as that of the quadratic form

(103) $$n((\underset{\sim}{I} + \underset{\sim}{J}(\underset{\sim}{\xi}))(\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi}), (\hat{\underset{\sim}{\xi}}_n - \underset{\sim}{\xi})),$$

where $\underset{\sim}{I}$ is the identity operator. Assuming that (92) holds the limit distribution of (103) is identical with the distribution of

(104) $$((\underset{\sim}{I} + \underset{\sim}{J}(\underset{\sim}{\xi})) \underset{\sim}{Z}, \underset{\sim}{Z})$$

where $\underset{\sim}{Z}$ is distributed as $\mathcal{N}(0, \underset{\sim}{\Sigma}^{-1}(\underset{\sim}{\xi}))$. Again the distribution of (104) is known (see, for instance, Gikhman and Skorokhod (1974), p. 352).

## 7. CONCLUDING REMARKS

(a) The weak convergence of the log-likelihood ratio process as discussed in section 5 has been found useful in literature on inference problems for the finite parameter case. More specifically, it has been used for the study of asymptotic theory of Bayes estimators (see Ibragimov and Hasminskii (1981) and Basawa and Prakasa Rao (1980)). The results of section 5 can be put to use for similar investigations in the infinite parameter case.

(b) It will be of interest to obtain sufficient but easily verifiable conditions for the condition (91) to hold. The condition (91) is also related to the problem of finding the rate of convergence of the MLE in our present case (see Prakasa Rao (1968) and Ibragimov and Hasminskii (1984)).

(c) The condition $\sum_{i=1}^{\infty} \xi_i^2 < \infty$ on the parameter space arises naturally in our present approach. However there are infinite parameter families where such a condition may not hold, for instance, see example 2 for the infinitomial case. In order to include such cases one needs to study appropriate extensions of the present IPE families such as the extension to general topological vector spaces. The probabilistic frame work for one such extension is given by Soler (1977) (see also Johansen (1977)).

(d) Another useful extension of the present IPE families is when the $\xi_i$'s are themselves known functions of infinite number of unknown parameters $\theta_j$. Moreover the function $H_i(\cdot)$ need not form a CONB of a Hilbert-space. Such extensions are useful for statistical inference for stochastic processes such as the problem of estimation of drift parameter in linear stochastic differential equation, etc.

(e) The present work was restricted to the estimation problem. The problem of testing hypothesis concerning $\underset{\sim}{\xi}$ has also been studied and shall be dealt with elsewhere.

## REFERENCES

Ash, R. B. (1972) Real Analysis and Probability, Academic Press, New York.

Barndorf-Nielsen, O. (1978) Information and Exponential Families in Statistical Theory, Wiley, New York.

Basawa, I. V. and Prakasa Rao, B. L. S. (1980) Statistical Inference for Stochastic Processes, Academic Press, London.

Berk, R. H. (1972) Consistency and asymptotic normality of MLE's for exponential models. Ann. Math. Statist. 43, 193-204.

Billingsley, P. (1969) Convergence of Probability Measures. Wiley, New York.

Chatterji, S. D. and Mandrekar, V. (1978) Equivalence and singularity of Gaussian measures and applications. In Probabilistic Analysis and Related Topics Vol. 1 (Ed. A.T. Bharucha-Reid) Academic Press, New York, 169-195.

Crain, B. R. (1974) Estimation of distributions using orthogonal expansions. Ann. Statistics 2, 454-463.

Gikhman, I. and Skorokhod, A. (1974) The Theory of Stochastic Processes Vol. I, Springer-Verlag, Berlin.

Grenander, U. (1963) Probabilities on Algebraic Structures, Wiley, New York.

Grenander, U. (1981) Abstract Inference, Wiley, New York.

Himmelberg, C. J., Parthasarathy, T. and Van Vleck, F. S. (1976). Optimal plans for dynamic programming problems. Mathematics of Operations Research 1, 390-394.

Ibragimov, I. A. and Hasminskii, R. Z. (1981) Statistical Estimation: Asymptotic Theory, Springer-Verlag, New York.

Johansen, S. (1977) Homomorphisms and general exponential families. In Recent Developments in Statistics. (Ed. J. R. Barra et. al.) North-Holland, Amsterdam, 489-499.

Lehmann, E. (1959) Testing Statistical Hypotheses, Wiley, New York.

Prakasa Rao, B. L. S. (1968) Estimation of the location of the cusp of a continuous density. Ann. Math. Statist. 39, 76-87.

Prakasa Rao, B. L. S. (1983) Nonparametric Functional Estimation, Academic Press, New York.

Rao, C. R. (1958) Maximum likelihood estimation for the multinomial distribution with an infinite number of cells. Sankhyā 20, 211-218.

Roberts. A. W. and Varberg, D. E. (1973) Convex Functions, Academic Press, New York.

Soler, J.-L. (1977) Infinite dimensional exponential type statistical spaces (Generalized exponential families). In Recent Developments in Statistics (Ed. J. R. Barre et. al.) North-Holland, Amsterdam, 269-284.