The Annals of Statistics
1995, Vol. 23, No. 5, 1543–1561

AN INFINITE-DIMENSIONAL GEOMETRIC STRUCTURE ON THE SPACE OF ALL THE PROBABILITY MEASURES EQUIVALENT TO A GIVEN ONE

BY GIOVANNI PISTONE AND CARLO SEMPI

Politecnico di Torino and Università di Lecce

Let $\mathscr{M}_\mu$ be the set of all probability densities equivalent to a given reference probability measure μ. This set is thought of as the maximal regular (i.e., with strictly positive densities) μ-dominated statistical model. For each $f \in \mathscr{M}_\mu$ we define (1) a Banach space L_f with unit ball $\mathscr{V}_f$ and (2) a mapping s_f from a subset $\mathscr{U}_f$ of $\mathscr{M}_\mu$ onto $\mathscr{V}_f$, in such a way that the system $(s_f, \mathscr{U}_f, f \in \mathscr{M}_\mu)$ is an affine atlas on $\mathscr{M}_\mu$. Moreover each parametric exponential model dominated by μ is a finite-dimensional affine submanifold and each parametric statistical model dominated by μ with a suitable regularity is a submanifold. The global geometric framework given by the manifold structure adds some insight to the so-called geometric theory of statistical models. In particular, the present paper gives some of the developments connected with the Fisher information metrics (Rao) and the Hilbert bundle introduced by Amari.

0. Introduction. The aim of the present note is to give a construction of an infinite-dimensional geometric structure on each class of equivalent probability measures. Many authors have dealt with just this problem in the finite-dimensional (i.e., parametric) case; here we limit ourselves to mentioning Rao (1949), Efron (1975), various contributions published in the IMS book by Amari, Barndorff-Nielsen, Kass, Lauritzen and Rao (1987), and the paper by Barndorff-Nielsen and Jupp (1989). We believe that the framework we propose is interesting on its own merits, is fairly simple, modulo an even cursory knowledge of Orlicz spaces, and yet may turn out to be sufficiently general to accommodate many interesting statistical ideas and models. The knowledge of Orlicz spaces that is required is entirely given below when it is needed, if one is willing to accept the standard facts (for which, however, references are given). The only Orlicz space we explicitly use is built on an exponential function that strongly suggests the well-known notions connected with moment generating functions and statistical exponential families. What we present below claims to be nothing more than a first step: a brief mention of some applications to the general theory of statistical models is given in the closing section. No new applied methodology is given but "nothing is so practical as a good theory": in fact, we think that if a manifold structure is at hand, then the full force of the methods of global analysis can be put to work for interesting statistical problem; as an example, see the application of the

Received August 1992; revised March 1995.
AMS 1991 *subject classification*. 62A25.
Key words and phrases. Nonparametric statistical manifolds, Orlicz spaces.

"implicit function theorem" in Groeneboom and Wellner (1992). This circle of ideas may even be useful in the Bayesian approach, where the role of the parameters is played by a priori distributions; it is then useful to have a geometric structure on distributions. We intend to pursue this line of investigation in order to deepen it and to compare it with other approaches, notably those mentioned above.

The starting point of the research here reported was a suggestion contained in a comment by Dawid (1975) on one of the papers by Efron (1975). Precisely, Dawid suggested the introduction of a nonparametric model and identified also the relevant tangent structure. This point was further developed in the definition of Hilbert bundle by Amari [see Amari, Barndorff-Nielsen, Kass, Lauritzen and Rao (1987) and also the comments in Barndorff-Nielsen and Jupp (1989)].

1. The statistical model and its topology.

1.1. *The statistical model and contents of the paper.* Let $(\Omega, \mathscr{B}, \mu)$ be a probability space, and let us denote by $\mathscr{M}_\mu$ the set of the densities of all the probability measures equivalent to μ,

$$\mathscr{M}_\mu = \{f \in L^1(\mu) : f > 0 \ \mu\text{-a.s.}, E(f) = 1\}. \tag{1}$$

The probability measure whose density is f will be denoted by $f \cdot \mu$ and its expectation by $E_f(\cdot)$, where $E(\cdot)$ denotes the expectation with respect to the reference measure μ.

The set $\mathscr{M}_\mu$ can be thought of as the *maximal regular μ-dominated statistical model*: here "regular" means that all probability measures in the model are equivalent to the reference probability measure μ. The more obvious "geometry" of $\mathscr{M}_\mu$ comes from the fact that it is a convex subset of $L^1(\mu)$. One of the key ideas in the geometric theory of statistical models is that the embedding in $L^1(\mu)$ is not "natural" for statistical purposes [see, e.g., Čentsov (1971)]. In the finite sample space case it is shown in that reference that a natural geometry is given by the embedding $\mathscr{M}_\mu \ni f \mapsto f^{1/2} \in L^2(\mu)$ of the set of densities in the Hilbert space $L^2(\mu)$. [See the review article by Kass (1989) for a simple exposition.]

Two topologies are usually considered for $\mathscr{M}_\mu$: that of convergence in μ-measure and that of the Banach space $L^1(\mu)$. These coincide, as is easy to check by looking at the argument used in the proof of Scheffé's theorem [Scheffé (1947), but see also Sempi (1989)]; moreover, if Ω is a metric space, then this topology implies weak convergence (*convergence étroite*). They also coincide with the topology induced by the mapping $f \mapsto \sqrt{f} \in L^2(\mu)$, which is used in connection with the Hellinger distance.

The construction of the mentioned geometric structure will be achieved by the definition of an atlas on the metric space $\mathscr{M}_\mu$ so as to give it the structure of a manifold modeled on Banach spaces. The relevant Banach spaces are the subspaces, denoted in the present paper by L_f, of centered random variables in the Orlicz spaces $L^{\phi}(f \cdot \mu)$ based on the exponential Young function $\phi(x) = \cosh x - 1$ to be defined in Section 2. The existence of the atlas and the characterization of its tangent bundle is then established in Section 3. Section 4 discusses the connection of the geometry with the statistical applications.

The topology induced on the space $\mathscr{M}_\mu$ by the atlas will be stronger than the L^1-topology and will reduce in the case of sequences to the convergence defined in the following subsection.

1.2. *Exponential convergence* (*e-convergence*). Let $(g_n)_{n \in \mathbb{N}}$ be a sequence and let g be a density both in $\mathscr{M}_\mu$.

DEFINITION 1.1. The sequence $(g_n)_{n \in \mathbb{N}}$ in $\mathscr{M}_\mu$ is e-convergent (exponentially convergent) to g if $(g_n)_{n \in \mathbb{N}}$ tends to g in μ-probability as $n \to \infty$, and, moreover, the sequences $(g_n/g)_{n \in \mathbb{N}}$ and $(g/g_n)_{n \in \mathbb{N}}$ are eventually bounded in each $L^p(g)$, $p > 1$, that is,

$$(2) \quad \forall\, p > 1, \qquad \limsup_{n \to \infty} E_g\left[\left(\frac{g_n}{g}\right)^p\right] < +\infty, \qquad \limsup_{n \to \infty} E_g\left[\left(\frac{g}{g_n}\right)^p\right] < +\infty.$$

The previous definition can be extended in an obvious way from the sequences to other types of limit. In particular, a function g mapping a real interval I into $\mathscr{M}_\mu$ is said to be e-continuous at a point $t_0 \in I$ if $\lim_{t \to t_0} g(t) = g(t_0)$ in μ-probability and

$$\forall\, p > 1 \qquad \limsup_{t \to t_0} E_g\left[\left(\frac{g(t)}{g(t_0)}\right)^p\right] < +\infty, \qquad \limsup_{t \to t_0} E_g\left[\left(\frac{g(t_0)}{g(t)}\right)^p\right] < +\infty.$$

Let us remark that e-convergence is equivalent to the convergence of g_n/g and g/g_n to 1 with respect to all p-seminorms $h \mapsto E_g[|h|^p]$, $p > 1$.

As the boundedness condition in Definition 1.1. is given with respect to seminorms which depend on the limit g, the following proposition will be useful in the sequel.

PROPOSITION 1.2. *Let $a > 0$ be given, and consider $f_1, f_2 \in \mathscr{M}_\mu$ with*

$$(3) \qquad \frac{f_1}{f_2} \in L^{1+a}(f_2 \cdot \mu).$$

If the sequence $(h_n)_{n \in \mathbb{N}}$ in $\mathscr{M}_\mu$ is eventually bounded in all $L^p(f_2 \cdot \mu)$, $p > 1$, then it is also eventually bounded in all $L^p(f_1 \cdot \mu)$, $p > 1$.

PROOF. It suffices to apply Hölder's inequality:

$$\begin{aligned} \int |h_n|^p f_1 \, d\mu &= \int |h_n|^p \left(\frac{f_1}{f_2}\right) f_2 \, d\mu \\ &\le \left(\int \left(\frac{f_1}{f_2}\right)^{1+a} f_2 \, d\mu\right)^{1/(1+a)} \left(\int h_n^{p(1+a)/a} f_2 \, d\mu\right)^{a/(1+a)}. \qquad \square \end{aligned}$$

REMARK 1.3. It follows from Proposition 1.2 that condition (2) in the definition of e-convergence is a consequence of the condition

$$\limsup_{n\to\infty} E_f\left[\left(\frac{g_n}{g}\right)^p\right] < +\infty, \qquad \limsup_{n\to\infty} E_f\left[\left(\frac{g}{g_n}\right)^p\right] < +\infty, \tag{4}$$

where $f \in \mathscr{M}_\mu$ is such that $g/f \in L^{1+a}(f\cdot\mu)$ for some $a > 0$.

The notion of e-convergence is quite strong, but nevertheless the exponential arcs in $\mathscr{M}_\mu$ are continuous except at the endpoints for that convergence, as the following proposition shows.

PROPOSITION 1.4. *Let two densities f_0 and f_1 be given in $\mathscr{M}_\mu$.*

(a) *The function*

$$C: \mathbb{R} \ni t \mapsto C(t) = E\left(f_1^t f_0^{1-t}\right) \in \mathbb{R}_+ \cup \{+\infty\}$$

is convex, and its proper domain contains the closed interval $[0,1]$; *the exponential arc from f_0 to f_1 in $\mathscr{M}_\mu$, defined by*

$$f: [0,1] \ni t \mapsto f_t = \frac{f_1^t f_0^{1-t}}{C(t)}, \tag{5}$$

is continuous in μ-probability on the closed interval $[0,1]$ *and e-continuous on the open interval* $]0,1[$.

(b) *If, moreover, C is finite on an open set containing* $[0,1]$, *then the exponential arc* (5) *is e-continuous.*

PROOF. The normalizing constant $C(t)$ can be written

$$C(t) = E_{f_0}\left[\exp\left(t \ln \frac{f_1}{f_0}\right)\right].$$

This shows convexity. Moreover, it is finite and continuous on $[0,1]$ because of the inequality $f_1^t f_0^{1-t} \le (1-t)f_0 + tf_1$: this is easily verified by taking the logarithm of both sides and using the concavity of the logarithmic function. This in turn implies that the exponential arc is continuous in μ-probability and in $L^1(\mu)$.

Let t be any point in $[0,1]$, let s be in $[0,1]$ and let $s \to t$. To verify the e-convergence of f_s to f_t, we have to show the boundedness, as $s \to t$, of the μ-expectation of both

$$\begin{aligned} \left(\frac{f_s}{f_t}\right)^p f_t &= \left(\frac{C(t)}{C(s)}\right)^p \frac{1}{C(t)} f_1^{t+p(s-t)} f_0^{1-[t+p(s-t)]}, \\ \left(\frac{f_t}{f_s}\right)^p f_t &= \left(\frac{C(s)}{C(t)}\right)^p \frac{1}{C(t)} f_1^{t+p(t-s)} f_0^{1-[t+p(t-s)]}. \end{aligned} \tag{6}$$

If s is such that both $(t + p(s - t))$ and $(t + p(t - s))$ are in the proper domain of C, then we can write

$$E_{f_t}\left[\left(\frac{f_s}{f_t}\right)^p\right] = \left(\frac{C(t)}{C(s)}\right)^p \frac{C[t + p(s - t)]}{C(t)},$$

$$E_{f_t}\left[\left(\frac{f_t}{f_s}\right)^p\right] = \left(\frac{C(s)}{C(t)}\right)^p \frac{C[t + p(t - s)]}{C(t)}.$$

If $t \in]0, 1[$, then s satisfies $0 \le t + p(s - t) \le 1$, and $0 \le t + p(t - s) \le 1$ in a neighborhood of t in $[0, 1]$ and the boundedness condition of the e-continuity follows.

In the case $t = 0$ or $t = 1$, the value $t + p(t - s)$ is not in $[0, 1]$ and the conclusion follows from the fact that the proper domain of C is a neighborhood of $[0, 1]$ in $\mathbb{R}$. □

REMARK 1.5. The previous proposition shows in particular that $\mathscr{M}_\mu$ is not connected by e-continuous exponential arcs unless the reference probability space $(\Omega, \mathscr{B}, \mu)$ is finite. In fact in such a case it is not difficult to produce an example where the e-continuity fails on one of the boundary points of the interval $[0, 1]$: take $f_0 = 1$ and $f_1 \in L^1(\mu)$ but $f_1 \notin L^p$, $p > 1$. The set of all densities connected by an e-continuous exponential arc to a given $f \in \mathscr{M}_\mu$ consists of the set of all densities g such that both functions

$$t \mapsto E_f\left[\exp\left(t \ln \frac{g}{f}\right)\right], \quad \text{and} \quad s \mapsto E_g\left[\exp\left(s \ln \frac{f}{g}\right)\right]$$

are finite in a neighborhood of 0. It is easily verified that this is equivalent to the requirement that f and g belong to the same one-dimensional exponential model. The construction developed in the present paper will show that this property can be thought of as the construction of the maximal exponential model containing f (see Section 4).

2. Moment generating functions and Orlicz spaces. In the present section we are going to present the construction of the Banach spaces that will play the role of generalized parameters or coordinates for maximal nonparametric exponential models. Notice that all the construction is local in the sense that this constructions will be done with reference to a neighborhood of a particular density f in the model $\mathscr{M}_\mu$. Note also that, while the notion of norm is derived from functional analysis, the vector space itself is an object well known in the theory of moment generating functions. Moreover, the resulting distance will be related with the more common L^2-norm in Section 4.

For each real random variable u on $(\Omega, \mathscr{B}, f \cdot \mu)$, the moment generating function of u, that is, the Laplace transform of the distribution of the random variable u with respect to the probability measure $f \cdot \mu$, is the function $\hat{u}_f$

with values in $\overline{\mathbb{R}}_+ = [0, +\infty]$ defined by

$$\hat{u}_f(t) = \int \exp(tu) f \, d\mu = E_f(\exp(tu)), \qquad t \in \mathbb{R}. \tag{7}$$

We recall a few useful results on moment generating functions.

PROPOSITION 2.1. *The moment generating function* $\hat{u}_f$ (a) *is convex and lower-semicontinuous*; (b) *is analytic in the interior of its proper domain* $D(\hat{u}_f)^0$; *and* (c) *its derivatives are obtained by differentiating under the integral sign.*

PROOF. See Widder (1941).

DEFINITION 2.2. For each density $f \in \mathscr{M}_\mu$, we consider the set of all random variables u such that the following hold:

(a) The moment generating function $\hat{u}_f$ is defined in a neighborhood of the origin 0.

(b) The expectation of u is zero [condition (a) implies the existence of a finite expectation].

This set is easily shown to be a vector space and will be denoted by L_f, that is,

$$L_f = \left\{ u \in L^1(f \cdot \mu) : 0 \in D(\hat{u}_f)^0, E_f(u) = 0 \right\}. \tag{8}$$

It follows from Proposition 2.1(b, c) that all the moments of every $u \in L_f$ exist and that they are the values at 0 of the derivatives of $\hat{u}_f$.

The next step is now to show that L_f is actually a Banach space because it is a closed subspace of a particular Banach space, the Orlicz space based on an exponentially growing function shortly to be introduced. For a general reference on Orlicz spaces, see, for example, Krasnosel'skii and Rutickii (1958) or Rao and Ren (1991).

We say that a real function ϕ is a *Young function* if $\phi(0) = 0$, ϕ is even, strictly increasing, convex and $\lim_{t \to \infty} t^{-1}\phi(t) = +\infty$. Young functions generalize the family of functions $u \mapsto (1/p)u^p$, $p > 1$; the Orlicz space of the Young function ϕ, denoted by $L^\phi(f \cdot \mu)$ is the corresponding generalization of the Lebesgue space of p-integrable functions $L^p(f \cdot \mu)$. Such a generalization is not straightforward because it is not enough to assume ϕ-integrability, that is, a condition such as $E_f[\phi(u)] < +\infty$. Such a condition is not sufficient to define an element of $L^\phi(f \cdot \mu)$ unless a further assumption on the growth of ϕ is introduced, which is always satisfied in the case of the usual Lebesgue spaces.

We define the set $L^\phi(f \cdot \mu)$ as the collection of all the (μ-equivalence classes of) random variables u for which there exists a positive number a such that $E_f[\phi(u/a)]$ is finite. This class is a Banach space for the norm defined by

$$\|u\|_{\phi,f} = \inf\{a > 0 \colon E_f[\phi(u/a)] \le 1\}. \tag{9}$$

Such a norm is the norm induced by the convex balanced set $\{u\colon E_f[\phi(u)] \le 1\}$: this means that it is the unique norm for which such a set is the closed unit ball.

Let us remark that the definition of norm we have given implies that for a given $\varepsilon > 0$ the distance between u and v is less than ε if and only if there exists a constant $p > \varepsilon^{-1}$ such that

$$E_f[\phi(p|u - v|)] \le 1. \tag{10}$$

By the same argument, one shows that $u_n \to u$ in $L^{\phi}(f\cdot\mu)$, that is, $\lim_{n\to+\infty}\|u_n - u\|_{\phi,f} = 0$ if and only if, for all $p > 1$,

$$\limsup_{n\to\infty} E_f[\phi(p|u_n - u|)] \le 1. \tag{11}$$

In particular [see Krasnosel'skii and Rutickii (1958), Section 9.14], the convergence $\lim_{n\to+\infty}\|u_n - u\|_{\phi,f} = 0$ implies that $\int\phi(u_n - u)f\,d\mu \to 0$.

From now on we shall use a specific Orlicz space by considering the Young function $\phi(u) = \cosh u - 1$, and we shall let $L^{(\cosh-1)}(f\cdot\mu)$ denote the Orlicz space, and $\|\cdot\|_f$ its norm.

PROPOSITION 2.3.

(a) *The set* L_f *defined in Definition* 2.2 *coincides with the closed subspace of the space* $L^{(\cosh-1)}(f\cdot\mu)$ *of zero expectation random variables.*

(b) *The set* L_f *is a vector subspace of the space* $L_0^p(f\cdot\mu)$ *of centered p-integrable random variables for every* $p \ge 1$ *with continuous and dense embedding; it contains the space* $L_0^{\infty}(f\cdot\mu)$ *of centered essentially bounded random variables, with continuous embedding (where the symbol "*$\hookrightarrow$*" marks the continuity of the embeddings, which is more than simply being a subspace)*

$$L_0^{\infty}(f\cdot\mu) \hookrightarrow L_f \hookrightarrow \bigcap_{p>1} L_0^p(f\cdot\mu). \tag{12}$$

PROOF. (a) Assume $u \in L^{\phi}(f\cdot\mu)$ and $E_f(u) = 0$. Then there exists $a > 0$ such that $E_f(\exp(u/a) + \exp(-u/a))$ is finite, and this implies, from the convexity of the exponential function, the finiteness of the moment generating function in the interval $]-a^{-1}, a^{-1}[$; therefore u is in L_f. Conversely, let us assume $u \in L_f$. Then there exists t such that both t and $-t$ are in the domain of the moment generating function $\hat{u}_f$, and this means that $E_f(\exp(tu) + \exp(-tu))$ is finite, so that $u \in L^{\phi}(f\cdot\mu)$.

(b) The first inclusion in (12) depends on the fact that, for every $t \in \mathbb{R}$, $\exp(tu)$ is bounded whenever u is. The second inclusion depends on the inequality

$$\phi(|u|) = \cosh|u| - 1 = \sum_{n=1}^{\infty} \frac{|u|^{2n}}{(2n)!} \geq \frac{|u|^{2p}}{(2p)!}, \qquad p \in \mathbb{N}.$$

Denseness follows from the denseness of L^{∞} in L^p, $p \geq 1$. □

Let B_f denote the open unit ball of $L^{(\cosh - 1)}(f \cdot \mu)$, and by $\mathscr{V}_f$ its trace on L_f, namely,

$$B_f = \{u \in L^{(\cosh - 1)}(f \cdot \mu) : \|u\|_f < 1\}, \qquad \mathscr{V}_f = B_f \cap L_f. \tag{13}$$

The following proposition is the "nonparametric" counterpart of Proposition 2.1.

PROPOSITION 2.4. *The moment generating functional* $\Phi_f: L^{(\cosh - 1)}(f \cdot \mu) \to \overline{\mathbb{R}}_+ = [0, +\infty]$ *defined by*

$$\Phi_f(u) = E_f(\exp u) \tag{14}$$

satisfies the following:

(a) *takes the value* 1 *when evaluated at the function that is identically* 0, *elsewhere it is greater than* 1, *is convex and lower semicontinuous and its proper domain* $D(\Phi_f) = \{u \in L^{(\cosh - 1)}(f \cdot \mu): \Phi_f(u) < \infty\}$ *contains the open unit ball* B_f;

(b) *is infinitely Gâteaux-differentiable in the interior of its proper domain*; *the nth derivative in the direction v is given by*

$$v \mapsto E_f(v^n \exp u); \tag{15}$$

(c) *is bounded and infinitely Fréchet-differentiable on the open unit ball* B_f.

PROOF. (a) As for convexity, there is nothing to prove if either $E_f(\exp u) = +\infty$ or $E_f(\exp v) = +\infty$, or both; otherwise, it is an immediate consequence of the convexity of the exponential function.

In order to prove that Φ_f is lower semicontinuous, one has to show that the level set $C_a = \{u \in L^{(\cosh - 1)}(f \cdot \mu): \Phi_f(u) = E_f(\exp u) \leq a\}$ is closed for every $a > 0$. Thus let $\{u_n\}$ be a sequence in C_a such that $\|u_n - u\|_{(\cosh - 1), f} \to 0$. This, in turn, implies that $u_n \to u$ in $f \cdot \mu$-measure [Sempi (1986)]. The sequence $\{\exp(u_n)\}$ converges in $(f \cdot \mu)$-measure to $\exp u$. Therefore there exists a subsequence $(n(k): k \in \mathbb{N})$ such that $\exp u_{n(k)} \to \exp u$, $k \to \infty$, μ-a.s. Fatou's lemma then gives $E_f(\exp u) = E_f(\liminf_k \exp[u_n(k)]) \leq \liminf_k E_f(\exp[u_n(k)]) \leq a$, so that $u \in C_a$. If $u \in B_f$, then $\int(\cosh u) f \, d\mu \leq 2$ so that $E_f(\exp u) < +\infty$, namely, u is in the proper domain of Φ. Therefore $B_f \subset D(\Phi_f)$.

(b) Since Φ_f is convex and lower semicontinuous, it is continuous in the interior of its proper domain $D(\Phi_f)^0$ [Ekeland and Temam (1974), I.2.5]. Thus, if $u \in D(\Phi_f)^0$, then, for every $v \in L^{(\cosh - 1)}(f \cdot \mu)$, $u + tv$ is again in $D(\Phi_f)^0$ for t small enough. By Proposition 2.1(b) the mapping $t \mapsto \Phi_f(u + tv)$

is analytic and its derivatives are

$$\left.\frac{d^n \Phi_f(u + tv)}{dt^n}\right|_{t=0} = E_f(v^n \exp u).$$

(c) As for Fréchet-differentiability, one has to prove the continuity of the multilinear mapping $v \mapsto E_f(v^n \exp u)$. Let r be the radius of a ball centered at the origin and entirely contained in the unit ball B_f. For every unit vector v, rv is again in B_f and

$$E_f\left[\frac{(rv)^n}{n!}\exp u\right] \le 2E_f[\cosh(rv + u)] \le 4.$$

Hence $E_f(v^n \exp u) \le 4(n!/r^n)$. □

The following proposition is the nonparametric version of well-known properties of the exponential models. Note that it is important to restrict consideration to the class of centered random variables in order to ensure the positivity properties of the function Ψ_f.

PROPOSITION 2.5. *For each $u \in L_f$ let the cumulant generating functional Ψ_f be defined by*

(16) $$\Psi_f(u) = \ln \Phi_f(u).$$

Then Ψ_f has proper domain $D(\Phi_f) \cap L_f \supset \mathscr{V}_f$ and the following properties:

(a) *Ψ_f is null at 0, elsewhere it is strictly positive; it is convex, lower semicontinuous, infinitely Gâteaux-differentiable in the interior of its proper domain, infinitely Fréchet-differentiable on $\mathscr{V}_f$;*

(b) *$\forall\, u \in \mathscr{V}_f$, $g = \exp[u - \Psi_f(u)] \cdot f$ is a probability density in $\mathscr{M}_\mu$ and the value of the nth derivative of Ψ_f at u in the direction v is the nth cumulant of v with respect to g, that is,*

(17) $$d_u^n \Psi_f(v^{0n}) = \left.\frac{d^n \ln E_g(\exp tv)}{dt^n}\right|_{t=0};$$

(c) *in particular,*

(18) $$d_u^2 \Psi_f(v_1, v_2) = E_g(v_1 v_2) - E_g(v_1)E_g(v_2).$$

PROOF. Positivity follows from Jensen's inequality: $E_f(\exp u) \ge \exp[E_f(u)] = \exp 0 = 1$; differentiability follows from Proposition 2.4 and the chain rule; the rest are simple computations. □

DEFINITION 2.6. Let us define the *symmetrized cumulant generating functional* $\tilde{\Psi}_f$ by

$$\begin{aligned}\tilde{\Psi}_f(u) &= \tfrac{1}{2}\left[\Psi_f(u) + \Psi_f(-u)\right]\\ &= \ln\left[\Phi_f(u)\Phi_f(-u)\right]^{1/2}, \qquad u \in L_f.\end{aligned}$$

It is a positive, convex, balanced function on L_f with proper domain $D(\Psi_f) \cap -D(\Psi_f) \supset \mathscr{V}_f$; we denote the induced norm by $\|\cdot\|_{\tilde{\Psi}_f}$, that is,

$$\|u\|_{\tilde{\Psi}_f} = \inf\left\{a > 0 : \tilde{\Psi}_f(u/a) \le 1\right\}, \qquad u \in L_f. \tag{19}$$

The relations between this new norm and the previously defined norm $\|\cdot\|_f$, based on the functional $u \mapsto E_f[\cosh u - 1]$, will be derived from the following inequalities:

(a) It follows from $\ln x \le x - 1$ that

$$\forall\, u \in L_f, \qquad \tilde{\Psi}_f(u) \le E_f[\cosh u - 1]. \tag{20}$$

(b) From Jensen's inequality $[E_f(\exp u)]^c \le E_f(\exp cu)$, $u \in L_f$, $c > 1$, it follows that

$$\forall\, u \in L_f, c > 1, \qquad c\tilde{\Psi}_f(u) \le \tilde{\Psi}_f(cu).$$

(c) If $a, b \ge 1$, then $a(b-1) > b - 1$; if $\ln(ab)^{1/2} \le c$, $c > 1$, then $(a+b)/2 \le 1 + \exp 2c)/2$; using $\Phi_f(u)$, $\Phi_f(-u) \ge 1$ and the definition of $\tilde{\Psi}_f(u)$, it follows that

$$\tilde{\Psi}_f(u) \le c \quad \Rightarrow \quad E_f(\cosh u - 1) \le \tfrac{1}{2}(\exp 2c - 1).$$

PROPOSITION 2.7. *The following inequalities between $E_f(\cosh u - 1)$ and $\tilde{\Psi}_f(u)$ hold for all $u \in L_f$:*

$$\tilde{\Psi}_f(u) \le E_f(\cosh u - 1) \le \exp\left[2\tilde{\Psi}_f(u)\right], \qquad u \in L_f.$$

In particular, the norm $\|\cdot\|_f$ dominates the norm $\|\cdot\|_{\tilde{\Psi}_f}$.

PROOF. The domination property follows from the first inequality, which in turn is (a). The second inequality follows from (b) and (c). □

EXAMPLE 2.8. It is interesting to show how the previously defined norms reduce to the usual L^2-norm in a Gaussian case. Let us consider centered Gaussian random variables of the form $u = \int_0^1 a(s)\, d\beta(s)$, where $\beta(t)$, $t \in [0,1]$, is the standard Wiener process for the probability measure $f \cdot \mu$ and $a \in L^2([0,1])$. We easily compute [from $p \int_0^1 a(s)\, d\beta(s) \sim N(0, p^2 \int_0^1 a^2(s)\, ds)$]:

$$\Phi_f\left(p\int_1^0 a(s)\, d\beta(s)\right) = \exp\left(\frac{p^2}{2}\int_0^1 a^2(s)\, ds\right);$$

$$E_f\left[\cosh\left(p\int_0^1 a(s)\, d\beta(s)\right) - 1\right] = \exp\left(\frac{p^2}{2}\int_0^1 a^2(s)\, ds\right) - 1;$$

$$\overline{\Psi}_f\left(p\int_0^1 a(s)\, d\beta(s)\right) = \frac{p^2}{2}\int_0^1 a^2(s)\, ds.$$

Then it follows that

$$(21)\quad \begin{aligned} \left\| \int_0^1 a(s)\, d\beta(s) \right\|_f &= \sqrt{\frac{1}{2\ln 2}\int_0^1 a^2(s)\, ds} = \frac{\|a\|_2}{\sqrt{2\ln 2}}, \\ \left\| \int_0^1 a(s)\, d\beta(s) \right\|_{\tilde{\Psi}_f} &= \frac{1}{2}\sqrt{\frac{1}{2}\int_0^1 a_2(s)\, ds} = \frac{1}{2}\|a\|_2. \end{aligned}$$

It follows that both the norms we have defined on L_f are proportional to the $L^2(f \cdot \mu)$-norm on the vector space of Wiener integrals.

Now we have completed the study of the space of coordinates L_f, which is a topological B-space for each one of the norm we have defined in the present section. The next topic will be the construction of an atlas with such a coordinate space.

3. The manifold of equivalent probability measures. Our treatment of infinite-dimensional differential geometry is elementary in the sense that we shall limit our discussion to the existence of an atlas (Theorem 3.6) and to the characterization of the tangent bundle (Section 3.3); we refer to the coordinate-free presentation of infinite-dimensional manifold of Lang (1972). The metric theory we present does not fit into the standard theory of Riemannian manifolds because our manifold is based on Banach spaces, which are not in general Hilbert spaces. Nevertheless, we define a continuous bilinear form on the manifold (Section 3.4) together with continuous multilinear forms.

3.1. *Atlases and manifolds.* Let us consider the following map:

$$(22)\qquad \mathscr{V}_f \ni u \mapsto \exp\left[u - \Psi_f(u)\right] \cdot f \in \mathscr{M}_\mu,$$

where $\Psi_f(u) = \ln E_f(\exp u) = \ln \Phi_f(u)$ is the cumulant generating functional of Proposition 2.5 and Φ_f is the moment generating functional of Proposition 2.4. This mapping is one-to-one, as equality of its values at u and v implies that $(u - v)$ is constant and 0 is the only constant contained in $\mathscr{V}_f$. We shall denote the range of the mapping defined in (22) by $\mathscr{U}_f$, and its inverse on $\mathscr{U}_f$ will be denoted by s_f. Such an inverse, $s_f\colon \mathscr{U}_f \to \mathscr{V}_f$, is easily computed, for $g \in \mathscr{U}_f$, as

$$(23)\qquad s_f(g) = \ln\frac{g}{f} - E_f\left[\ln\frac{g}{f}\right].$$

The functions s_f, $f \in \mathscr{M}_\mu$, will be the coordinate mappings of our manifold in the sense that, locally around each $f \in \mathscr{M}_\mu$, each $g \in \mathscr{U}_f$ will be "parameterized" by its centered log-likelihood (23).

REMARKS. The value $s_f(g)$ of the mapping $s_f(\cdot)$ defined in (23) is the log-likelihood of g with respect to f plus the Kullback–Leibler information $E_f(\ln(f/g))$.

The mapping s_f is connected to the maximum likelihood estimator as follows: the *maximum expected* (*at* g) *log-likelihood*, that is, the maximum of the function

$$\mathscr{V}_f \ni u \mapsto E_g\left[\ln \frac{\exp[u - \Psi_f(u)] \cdot f}{f}\right] = E_g(u) - \Psi_f(u),$$

is obtained, by differentiating and using Proposition 2.5(b), at that point $\hat{u}$ for which $E_g(\cdot) = E_{\exp[u - \Psi_f(u)] \cdot f}(\cdot)$, that is, at $\hat{u} = s_f(g)$.

Let us compute now the change-of-coordinates formula: if f_1 and f_2 are two points in $\mathscr{M}_\mu$ such that $\mathscr{U}_{f_1} \cap \mathscr{U}_{f_2} \neq \varnothing$, then the composite (transition) mapping

$$s_{f_2} \circ s_{f_1}^{-1} : s_{f_1}(\mathscr{U}_{f_1} \cap \mathscr{U}_{f_2}) \to s_{f_2}(\mathscr{U}_{f_1} \cap \mathscr{U}_{f_2}) \tag{24}$$

simplifies to

$$s_{f_2} \circ s_{f_1}^{-1}(u) = u + \ln \frac{f_1}{f_2} - E_{f_2}\left[u + \ln \frac{f_1}{f_2}\right], \tag{25}$$

where the algebraic computations are done in the space of μ-classes of measurable functions and the expectation is well defined as long as $\mathscr{U}_{f_1} \cap \mathscr{U}_{f_2} \neq \varnothing$ because this implies $u + \ln(f_1/f_2) \in \mathscr{V}_{f_2}$.

We shall show that the sets $\mathscr{U}_f$, with $f \in \mathscr{M}_\mu$, are sequentially e-open. In order to prove this point, the following two propositions will be useful. In fact it is shown that e-convergence is local with respect to each $\mathscr{U}_f$, in the sense that the L^p-boundedness condition does not depend on the limit, but it depends on the set $\mathscr{U}_f$.

PROPOSITION 3.1. *Let $g \in \mathscr{M}_\mu$; if $g \in \mathscr{U}_f$, which implies $\|s_f(g)\|_f < 1/(1 + \delta)$ for some $\delta > 0$. Then*

$$\frac{g}{f} \in L^{1+\delta}(f \cdot \mu), \qquad \frac{f}{g} \in L^{2+\delta}(g \cdot \mu). \tag{26}$$

As a consequence, a sequence $(h_n)_{n \in \mathbb{N}}$ is eventually bounded in each $L^p(f \cdot \mu)$, $p > 1$, if and only if it is eventually bounded in each $L^p(g \cdot \mu)$, $p > 1$.

PROOF. Let $s_f(g) = u \in \mathscr{V}_f$, that is, $g = \exp[u - \Psi_f(u)] \cdot f$. Therefore,

$$\int \left(\frac{g}{f}\right)^{1+\delta} f\,d\mu + \int \left(\frac{f}{g}\right)^{2+\delta} g\,d\mu \int \left[\left(\frac{g}{f}\right)^{1+\delta} + \left(\frac{f}{g}\right)^{1+\delta}\right] f\,d\mu$$

$$= \int \big[\exp[u(1+\delta)]\exp\{-\Psi_f(u)(1+\delta)\}$$

$$+ \exp[-u(1+\delta)]\exp\{\Psi_f(u)(1+\delta)\}\big] f\,d\mu$$

$$\leq 2\exp\{|\Psi_f(u)|(1+\delta)\} \int \cosh\{u(1+\delta)\} f\,d\mu$$

(because of the assumption)

$$\leq 4\exp\{|\Psi_f(u)|(1+\delta)\}.$$

The last part now follows from Proposition 1.2. □

PROPOSITION 3.2. *Let $g \in \mathscr{U}_f$, and let $(g_n)_{n \in \mathbb{N}}$ be a sequence in $\mathscr{M}_\mu$ that converges to g in probability [and in $L^1(\mu)$]. Then the sequence is e-convergent if and only if both sequences g_n/g and g/g_n are eventually bounded for each $L^p(f \cdot \mu)$, $p > 1$.*

PROOF. The proposition follows from Proposition 3.1. □

DEFINITION 3.3. Let us now recall basic definitions from Lang (1972). The set $\mathscr{M}_\mu$ is covered by $(\mathscr{U}_f, s_f)_{f \in \mathscr{M}_\mu}$ and the maps $s_f: \mathscr{U}_f \to \mathscr{V}_f \subset L_f$ are one-to-one on the unit ball $\mathscr{V}_f$ of a Banach space L_f. Such an object is called an *atlas of class* C^∞ if, moreover, the following two conditions are satisfied:

(a) $\forall\, f, g \in \mathscr{M}_\mu$, $s_f(\mathscr{U}_f \cap \mathscr{U}_g)$ is open in L_f;
(b) $\forall\, f, g \in \mathscr{M}_\mu$ the *transition mapping*

$$s_g \circ s_f^{-1}: s_f(\mathscr{U}_f \cap \mathscr{U}_g) \to s_g(\mathscr{U}_f \cap \mathscr{U}_g)$$

is a C^∞-isomorphism.

In fact, in the following subsection we shall show much more, that is, the covering is a sequentially open covering for e-convergence, the mappings s_f are homeomorphisms for e-convergence, and the atlas is actually *affine*, that is, the transition mappings are affine functions.

3.2. *Main result.* The following proposition shows that each $\mathscr{U}_f$ is e-open and that the coordinate mappings s_f are sequentially continuous from e-convergence to L_f-convergence.

PROPOSITION 3.4. *Let us assume that the sequence $(g_n)_{n \in \mathbb{N}}$ is e-convergent to g as $n \to \infty$, and that $g \in \mathscr{U}_f$. Then the sequence $(g_n)_{n \in \mathbb{N}}$ is eventually in $\mathscr{U}_f$, and the corresponding sequence of coordinates $u_n = s_f(g_n)$ converges to $u = s_f(g)$ in $\mathscr{V}_f$.*

PROOF. Let us define $\tilde{u}_n$ by $g_n = \exp[\tilde{u}_n - \Psi_f(u)] \cdot f$, $n \in \mathbb{N}$. From the assumptions we have $u \in \mathscr{V}_f$ but on $\tilde{u}_n = \ln(g_n/f) - \Psi_f(u)$ we do not have any integrability assumption. From the e-convergence of the sequence $(g_n)_{n \in \mathbb{N}}$ to g there follows the convergence of the sequence $(u_n)_{n \in \mathbb{N}}$ to u in μ-measure. Moreover, by the definition of e-convergence and Proposition 3.2 we know that, $\forall\, p > 1$,

$$\frac{1}{2}\left\{E_f\left[\left(\frac{g_n}{g}\right)^p + \left(\frac{g}{g_n}\right)^p\right]\right\} - 1 = E_f[\cosh p(\tilde{u}_n - u) - 1] \tag{27}$$

tends to 0, and this in turn implies the eventual convergence of $(u_n)_{n \in \mathbb{N}}$ to u in the Orlicz space $L^{(\cosh - 1)}(f \cdot \mu)$ [see (10)]. The sequence of functions $u_n = \tilde{u}_n - E_f(\tilde{u}_n)$ is eventually well-defined in L_f and eventually belongs to $\mathscr{V}_f$. In such a case $g_n = \exp[u_n - \Psi_f(u_n)] \cdot f$. □

Now we show the sequential continuity of the inverse coordinate mappings s_f^{-1} from L_f-convergence to e-convergence.

PROPOSITION 3.5. *Let u_n, $u \in \mathscr{V}_f$, $n \in \mathbb{N}$, and assume $u_n \to u$ in L_f. If $u_n = s_f(g_n)$, $u = s_f(g)$, $n \in \mathbb{N}$, then g_n e-converges to g.*

PROOF. The sequence (g_n) is μ-convergent to g and

$$E_f[\phi(p(u_n - u))] \tag{28}$$
$$= \frac{1}{2}\left\{E_f\left[\left(\frac{g_n}{g}\right)^p\right]\left(\frac{\Phi_f(u_n)}{\Phi_f(u)}\right)^p + E_f\left[\left(\frac{g}{g_n}\right)^p\right]\left(\frac{\Phi_f(u)}{\Phi_f(u_n)}\right)^p\right\} - 1,$$

where Φ_f was defined in Proposition 2.4. This shows that the condition of Proposition 3.2 is satisfied. □

We are now ready to state the main result of the present paper.

THEOREM 3.6. *The collection of pairs $\{(\mathscr{U}_f, s_f): f \in \mathscr{M}_\mu\}$ is an affine C^∞-atlas on $\mathscr{M}_\mu$. The induced topology on sequences is equivalent to e-convergence.*

PROOF. We have to verify conditions (a) and (b) of Definition 3.3.

(a) The set $s_f(\mathscr{U}_f \cap \mathscr{U}_g)$ is open because it coincides with $s_f(\mathscr{U}_g)$, $\mathscr{U}_g$ is sequentially e-open and s_f^{-1} is sequentially continuous and defined on an open set $\mathscr{V}_f$ of a Banach space L_f.

(b) The transition mapping $s_g \circ s_f^{-1}$ is continuous because it is the composition of two continuous mappings; moreover, it is affine, as formula (25) shows. Then it is affine and C^∞. □

3.3. *The tangent bundle.* In the previous section we constructed a manifold structure where the value of the coordinate (parameter value) is the log-likelihood. Now we consider the tangent manifold structure and we show that the geometric notion of tangent vector corresponds to an equally well known statistical notion, that is, the score.

Let f be a density in $\mathscr{M}_\mu$, and let a curve through f be given, that is, a one-dimensional parametric statistical model $I \ni t \mapsto g(t)$, I open real interval, $t_0 \in I$, $g(t_0) = f$. If $(\mathscr{U}_{f_1}, s_{f_1})$ is a chart for f, then $g(t) = \exp\{u_1(t) - \Psi_{f_1}[u_1(t)]\} \cdot f_1$, $u_1(t) = s_{f_1}[g(t)]$, $t \in I$. With respect to this chart the tangent vector to the curve at f is $u_1'(t_0)$. In a different chart $(\mathscr{U}_{f_2}, s_{f_2})$ we will get $u_2'(t_0)$, and the two vectors will be connected by the equivalence relation $(s_{f_2} \circ s_{f_1}^{-1})'[s_{f_1}(f)]u_1'(t_0) = u_2'(t_0)$. The collection of tangent vectors at f, modulo this equivalence relation, is called the tangent space at f of the manifold $\mathscr{M}_\mu$. It is a vector space and has the same topology as any one of the Banach spaces L_g for all $g \in \mathscr{M}_\mu$ such that $f \in \mathscr{U}_g$.

This abstract construction in our case has a concrete realization. In fact, the manifold $\mathscr{M}_\mu$ is affine [see (25)], and we get $(s_{f_2} \circ s_{f_1}^{-1})' \circ s(u) = u - E_{f_2}[u]$, so that $u_1 \in L_{f_1}$ is equivalent to $u_2 \in L_{f_2}$ (i.e., they both represent the same tangent vector) if and only if $u_2 - u_1$ is a constant. The tangent space at f is denoted by T_f, and usually we shall identify it with L_f (i.e., the coordinate space at f). With such an identification the tangent vector of a one-dimensional statistical model $g(t) = \exp\{u_f(t) - \Psi_f[u(t)]\} \cdot f$ is $u'(t) - E_{g(t)}(u'(t))$. However, from $\Psi_f(u(t)) = \ln E_f(\exp[u(t)])$ it follows that $[d\Psi_f(u(t))/dt] = E_{g(t)}[u'(t)]$. Then the tangent vector will be represented by $d \ln[g(t)/f]/dt$, that is, the score function at t. Notice again the importance of considering centered u's.

The model $\mathscr{M}_\mu$ endowed with the collection of all the tangent spaces is called the tangent bundle and denoted by $T(\mathscr{M}_\mu)$. An atlas of the tangent bundle is given by the mappings $T(\mathscr{U}_f) \ni (g, u) \mapsto (s_f(g), u - E_f(u))$. This new manifold has transition functions

$$\begin{aligned} &\mathscr{U}_{f_1} \times L_{f_1} \ni (u_1, v_1) \\ (29) \qquad &\mapsto \left(u_1 + \ln \frac{f_1}{f_2} - E_{f_2}\left[u_1 + \ln \frac{f_1}{f_2}\right], v_1 - E_{f_2}[v_1]\right) \in \mathscr{U}_{f_2} \times L_{f_2}. \end{aligned}$$

In our opinion the preceding definitions fill the "technical details" of the original suggestion by Dawid (1975). It is our hope that the explicit definition of tangent bundle shall be relevant in the computation of approximation of parametric statistical models [cf. Barndorff-Nielsen and Jupp (1989)]. Another connected important issue we will not discuss here is the Hilbert bundle as defined by Amari [see the references in Amari, Barndorff-Nielson, Kass, Lauritzen and Rao (1987); see the next section also].

3.4. *The cumulant forms, the Fisher information and the definition of a Riemannian metric on* $\mathscr{M}_\mu$. In this section we show how the most important item of the geometric theory of statistical models, that is, the Fisher information metrics and the corresponding Riemannian metrics, fits in our theory. The construction of a nonparametric extension of the Fisher information metrics, as a quadratic form on the space of square-integrable random variables, is due to Amari in the case of parametric statistical models. Here we give a generalization that works for the full nonparametric model; each parametric case is then obtained by restriction of the structure to a submanifold in a standard way.

Let a point of the tangent bundle be given, $(f, v) \in T(\mathscr{M}_\mu)$, and let us consider two different charts for it, centered respectively at f_1 and f_2. Let (u_1, v_1) and (u_2, v_2) be the coordinates.

Let us denote by $K_f(v_i^{0n}) = d_{v_i}^n \Psi_f(u_i)$, $u_i = s_{f_i}(f)$, the cumulant n-form at f of the random variables u_i, $i = 1, 2$ (see Proposition 2.5). Because v_i, $i = 1, 2$, represent the same tangent vector, their difference is a constant and the cumulant n-forms differ only for $n = 1$. In other words the cumulant n-forms for $n \geq 2$ are an intrinsic (that is, not depending on the coordinate system) object of the tangent bundle. In particular, the first one has an explicit representation as the scalar product (covariance) in $L^2(f) \cap L_f$. The tangent bundle endowed with the system of cumulant n-forms, $n \geq 2$, could be called *cumulant bundle* as an extension of the Hilbert bundle by Amari.

4. Application to the geometry of parametric models. We conclude by discussing briefly the connection between the structure of affine manifold of Theorem 3.6 and the geometry of exponential models.

For every f in $\mathscr{M}_\mu$ the *maximal exponential model at* f is defined to be the family of densities

$$\mathscr{E}(f) = \left\{\exp\left[u - \Psi_f(u)\right] \cdot f \colon u \in D(\Phi_f)^0, E_f(u) = 0\right\} \tag{30}$$

(cf. Proposition 2.4). The mapping

$$L_f \supset D(\Psi_f)^0 \ni u \mapsto \exp\left[u - \Psi_f(u)\right] \cdot f \in \mathscr{M}_\mu$$

is the likelihood mapping when the maximal exponential model is parameterized by L_f. Such a mapping is continuous (cf. Theorem 3.6).

If $g_1, g_2 \in \mathscr{E}(f)$ are two densities in the maximal exponential model, we have

$$g_i = \exp\left[u_i - \Psi_f(u_i)\right] \cdot f, \qquad u_i \in D(\Psi_f)^0, i = 1,2,$$

$$E\left[g_1^{1-\theta} g_2^{\theta}\right] = \frac{\Phi_f((1-\theta)u_1 + \theta u_2)}{\Phi_f^{1-\theta}(u_1)\Phi_f^{\theta}(u_2)}, \qquad \theta \in [0,1],$$

then they are connected by an e-continuous exponential arc by Proposition 2.4(a) on the convexity of $(D\Phi_f)^0$ and by Proposition 1.4(b). In other words, the maximal exponential model $\mathscr{E}(f)$ is a connected subset of the manifold $\mathscr{M}_\mu$. Notice that in the present discussion *connected* means *connected by e-continuous arcs.*

The following theorem gives a more precise result on the way the maximal exponential model (30) is related to the manifold structure.

THEOREM 4.1. *The maximal exponential model* (30) *is the connected component containing f of the manifold $\mathscr{M}_\mu$.*

PROOF. We know from the preceding remark that $\mathscr{E}(f)$ is indeed connected; we are left to prove maximality. In other words, we have to show that, given a finite sequence $g_0 = f, g_1, \dots, g_n$ of densities in $\mathscr{M}_\mu$, such that g_{i-1} is connected to g_i by an e-continuous exponential arc, $i = 1, \dots, n$, then g_n belongs to $\mathscr{E}(f)$. It is enough to prove the following: if $g_1 \in \mathscr{E}(f)$ and g_1 is connected to g_2 by an e-continuous exponential arc, then $g_2 \in \mathscr{E}(f)$, because then the previous proposition follows induction on n.

Let us consider again the exponential arc connecting g_1 and g_2 in $\mathscr{E}(f)$, namely,

$$g(\theta) = \frac{g_1^{1-\theta} g_2^{\theta}}{E\left[g_1^{1-\theta} g_2^{\theta}\right]}, \qquad \theta \in [0,1].$$

As it is a continuous mapping from the closed interval $[0,1]$ to $\mathscr{M}_\mu$, then its image $\{g(\theta)\colon \theta \in [0,1]\}$ is covered by a finite number of open sets of the form

$$\left\{\exp\left[v - \Psi_{g(\theta_i)}(v)\right] \cdot g(\theta_i)\colon \|v\|_{g(\theta_i)} < r_i\right\}, \qquad i = 1,\dots,n,\ \theta_i \in [0,1].$$

These sets are overlapping and all the relevant spaces $L_{g(\theta_i)}$ are isomorphic to L_f under the mapping $u \mapsto u - E_f(u)$ as topological vector spaces, so that we can assume that the covering has the form $\{\exp[v - \Psi_{g(\theta_i)}(v)] \cdot g(\theta_i)\colon \|v - E_f(v)\|_f < r_i'\}$.

Assume again by induction that the proposition is true, that is, $g_2 \in \mathscr{E}(f)$ for a covering with $n-1$ open sets. We are then reduced to proving the following: if the set $\{\exp[v - \Psi_{g_2}(v)] \cdot g_2\colon v \in L_{g_2}, \|v - E_f(v)\|_f < r\}$ contains a density g_1 in $\mathscr{E}(f)$, then $g_2 \in \mathscr{E}(f)$, where r is such that the r-ball centered at g_1 is contained in $D(\Psi_f)^0$.

Finally, let us consider that

$$g_1 = \exp\left[u - \Psi_f(u)\right] \cdot f = \exp\left[v - \Psi_{g_2}(v)\right] \cdot g_2, \qquad \|v - E_f(v)\|_f < r,$$

simplifies to

$$g_2 = \exp\left\{u - v + E_f(u-v) - \left[\Psi_f(u) - \Psi_{g_2}(v) + E_f(u-v)\right]\right\} \cdot f.$$

Because $[\Psi_f(u) - \Psi_{g_2}(v) + E_f(u-v)]$ is finite and

$$\left\{u - v + E_f(u-v)\colon \|v - E_f(v)\|_f < r\right\}$$

is a neighborhood of $u - (v - E_f(v))$ contained in $D(\Psi_f)^0$, then g_2 belongs to $\mathscr{E}(f)$. □

Let us consider now a parametric exponential model of the form

$$\left\{g_\theta = \exp\left[\langle \theta \mid T\rangle - \Psi_f(\theta)\right] \cdot f\colon \theta \in \Theta\right\}, \tag{31}$$

where Θ is an open set in $\mathbb{R}^d$, $d \in \mathbb{N}$. We can assume also, without loss of generality, that $E_f(T) = 0$. The mapping $\Theta \ni \theta \mapsto g_\theta \in \mathscr{M}_\mu$ has, with respect to the chart s_h, where h is in the parametric model (31), the local representation $\theta \mapsto \langle \theta - \theta_0 \mid T\rangle \in L_{g\theta_0}$. The subspace generated by T splits L_f. Then the parametric exponential model is a submanifold of the manifold $\mathscr{M}_\mu$, but such a manifold has also an affine structure that fits with the affine structure of the atlas we have constructed.

In our framework a parametric model is just a finite-dimensional submanifold. Between the parametric models some models have a more stringent finiteness condition, that is, they are submodels (i.e., submanifolds) of an exponential model: they are the *curved exponential models* as defined originally by Efron (1975). A curved exponential model is embedded in a finite-dimensional geometry, where all the technicalities connected with the infinite-dimensional Banach spaces of coordinates disappear as all the coordinates are real vectors; on the other hand a general parametric model is finite-dimensional and is described by real parameters, but all the analytic computations are much more difficult because it cannot be treated as a surface in a finite-dimensional space.

Our remarks are quite general, but they are strictly connected with many practical computational problems. We limit oursleves to mentioning two items that are the object of current research, namely, the problem of computing approximate models and the problem of finite-dimensional filters in filtering theory.

Acknowledgments. Various preliminary versions of this article have been presented in talk form on many occasions at seminars (U. Lecce 1989, U. Milan 1992, U. l'Aquila 1992 and U. Pisa 1993), and the authors have tried to do their best to incorporate comments and criticism in the present version. In particular, the first author thanks Professor U. Bruzzo (U. Genova) and Professor J. Pejsachovicz (U. Torino) for many conversations on the general theory of infinite-dimensional manifolds. The present version contains some further developments and improvements suggested by the anonymous referees, whom we wish to thank. S. Zabell has kindly communicated to us that similar ideas were presented in an unpublished set of lectures (1985–1986) in collaboration with J. Reeds.

REFERENCES

AMARI, S.-I., BARNDORFF-NIELSEN, O. E., KASS, R. E., LAURITZEN, S. L. and RAO, C. R. (1987). *Differential Geometry and Statistical Inference.* IMS, Hayward, CA.

BARNDORFF-NIELSEN, O. E. and JUPP, P. E. (1989). Approximating exponential models. *Ann. Inst. Statist. Math.* **41** 247–267.

ČENTSOV, N. N. (1971). *Statistical Decision Rules and Optimal Inference.* Nauka, Moscow. [In Russian; English translation (1982), Amer. Math. Soc., Providence, RI.]

DAWID, A. P. (1975). Discussion of "Defining a curvature of a statistical problem (with applications to second-order efficiency)" by B. Efron. *Ann. Statist.* **3** 1231–1234.

DAWID, A. P. (1977). Further comments on some comments on a paper by Bradley Efron. *Ann. Statist.* **5** 1249.

EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second-order efficiency) (with discussion). *Ann. Statist.* **3** 1189–1242.

EKELAND, I. and TEMAM, R. (1974). *Analyse Convexe et Problèmes Variationnels.* Dunod, Gauthier-Villars, Paris.

GROENEBOOM, P. and WELLNER, J. A. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhäuser, Boston.

HALMOS, P. R. (1950). *Measure Theory.* Van Nostrand Reinhold, New York. [New printing issued by Springer, Berlin (1974).]

KASS, R. E. (1989). The geometry of asymptotic inference (with discussion). *Statist. Sci.* **4** 188–234.

KRASNOSEL'SKII, M. A. and RUTICKII, YA. B. (1958). *Convex Functions and Orlicz Spaces.* Fizmatgiz, Moskva. [In Russian; English translation (1961), Noordhoff, Groningen.]

LANG, S. (1972). *Differential Manifolds.* Addison-Wesley, Reading, MA.

NEVEU, J. (1972). *Martingales à Temps Discrets*. Masson, Paris.
RAO, C. R. (1949). On the distance between two populations. *Sankhyā* **9** 246–248.
RAO, M. M. and REN, Z. D. (1991). *Theory of Orlicz Spaces*. Dekker, New York.
SCHEFFÉ, H. (1947). A useful convergence theorem for probability distributions. *Ann. Math. Statist.* **18** 434–438.
SEMPI, C. (1986). Orlicz metrics for weak convergence of distribution functions. *Riv. Mat. Univ. Parma (4)* **12** 289–292.
SEMPI, C. (1989). Variations on a theme by Scheffé. *Rend. Istit. Mat. Univ. Trieste* **21** 219–223.
VAN DER VAART, A. W. (1991). On differentiable functionals. *Ann. Statist.* **19** 178–205.
WIDDER, D. V. (1941). *The Laplace Transform*. Princeton Univ. Press.

DIPARTIMENTO DI MATEMATICA
POLITECNICO DI TORINO
CORSO DUCA DEGLI ABRUZZI 24
10129 TORINO
ITALY

DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DI LECCE
VIA ARNESANO
73100 LECCE
ITALY