ELSEVIER

Statistical manifold as an affine space: A functional equation approach

Jun Zhang[a],*, Peter Hästö[b,1]

[a]*Department of Psychology, 525 East University Ave., University of Michigan, Ann Arbor, MI 48109-1109, USA*
[b]*Department of Mathematics and Statistics, P. O. Box 68, 00014 University of Helsinki, Finland*

Abstract

A statistical manifold $\mathcal{M}_\mu$ consists of positive functions f such that $f\,d\mu$ defines a probability measure. In order to define an atlas on the manifold, it is viewed as an affine space associated with a subspace of the Orlicz space L^Φ. This leads to a functional equation whose solution, after imposing the linearity constrain in line with the vector space assumption, gives rise to a general form of mappings between the affine probability manifold and the vector (Orlicz) space. These results generalize the exponential statistical manifold and clarify some foundational issues in non-parametric information geometry.
© 2005 Published by Elsevier Inc.

Keywords: Probability density function; Exponential family; Non-parameteric; Information geometry

1. Introduction

Information geometry investigates the differential geometric structure of the manifold of probability density functions with continuous support (or probability distributions with discrete support). Treating a family of parametric probability density functions as a smooth manifold with their parameters as its coordinates originated from Rao (1945), who identified Fisher information to be the Riemannian metric on this manifold. Later Amari (1982), motivated by (Efron, 1975; David, 1975), and also independently Čencov (1972/1982), established a family of affine connections with dualistic properties on the manifold of parametric density functions. A manifold endowed with a Fisher metric and dualistic (conjugate) affine connections is abstracted as a statistical manifold (Lauritzen, 1987). Built upon these foundational work, parametric information geometry has found many applications in theoretical statistics, information

theory, stochastic process, neural computation, machine learning, Bayesian statistics, and other related fields (Amari, 1985; Amari & Nagaoka, 1993/2000). Recently, an interest in the geometry associated with non-parametric probability densities has arisen (Pistone & Sempi, 1995; Giblisco & Pistone, 1998; Pistone & Rogantin, 1999; Grasselli, 2005). Non-parametric statistical models are important in a wider range of areas including psychological measurement and perception (Townsend, Solomon, & Smith, 2001) and model selection and testing (Karabatsos, in press). Unlike in the parametric case where the manifold of probability density functions inherits a Euclidean topology from the space of its natural parameters, a major challenge for the non-parametric case is to define a suitable topology and develop a corresponding notion of convergence. Fortunately, this obstacle was overcome by the introduction of an exponential statistical manifold by Pistone and Sempi (1995). These authors, among other things, gave an explicit formula for a chart of the manifold formed by all density functions absolutely continuous with respect to a given one. The topology on this infinite-dimensional manifold is induced via this chart; it is metric and can be defined based on a notion of exponential convergence of sequences (Pistone & Sempi, 1995; Pistone, 2001).

In this paper we address some foundational issues of non-parametric information geometry, with the goal of

---

*Corresponding author. Fax: +1 313s 763 7480.

E-mail addresses: junz@umich.edu (J. Zhang), peter.hasto@helsinki.fi (P. Hästö).

URLs: http://www-personal.umich.edu/~junz/, http://www.helsinki.fi/~hasto/.

extending the Pistone–Sempi formula that maps the manifold of probability density functions to a subset in an Orlicz space L^Φ (a special Banach space, see Appendix A). The Banach-space valued map is the infinite-dimension extension of the coordinate functions which take values in $\mathbb{R}^n$. Recall that $\mathbb{R}^n$ has both a topological structure, in terms of its canonical topology, an algebraic structure as a vector space, and a geometric or affine structure as a set of points. Our approach is to exploit this affine structure and view the statistical manifold as an affine space associated with the vector space L^Φ.

An affine space is a homogeneous set of points such that no point stands out in particular. Affine spaces differ from vector spaces in that no origin has been selected. So affine space is fundamentally a geometric structure—an example being the plane. The structure of an affine space is given by an operation $\oplus : A \times U \to A$ which associates to a point a in the affine space A and a vector u in the vector space U another point $a \oplus u$ in A. We think of this as a translation of a point a in its space A by a vector u. Notice that it makes no sense to add two points of A in this setting.

One advantage of modelling a statistical manifold as a generalized affine space is to address the issue of representation of probability measures, i.e., through the use of an extended vector space U as opposed to the probability simplex A. If the operation $\oplus$ exists and is continuous and differentiable, a global mapping is established between U and A. This allows the setting up of a global atlas at any given point of A (recall that the usual differentiable manifold only assumes the existence of an atlas locally).

The structure of the rest of this paper is as follows: In the next section we review the Pistone–Sempi framework, and consider an obvious generalization of their original formulation. In Section 3 we introduce the affine structure on the probability manifold and derive a corresponding functional equation. We then solve this equation and derive a general expression for the $\oplus$ operation. This generalizes the exponential model in a natural way. Discussion of possible applications of these results is given in Section 4, followed by conclusions in Section 5. The reader who needs some refreshing of their Orlicz space theory can start by consulting Appendix A.

2. The Pistone–Sempi framework revisited

Pistone and Sempi (1995) introduced a non-parametric exponential statistical manifold consisting of all densities that are absolutely continuous with respect to a given one. In this section we review the parts of that article most relevant to our considerations.

Let X be a set and μ a σ-finite probability measure on X, in other words $\mu(X) = 1$. We will consider the set

$$\mathcal{M}_\mu := \{p \in L^1(X, \mu) : p > 0 \text{ a.e., } \|p\|_{L^1(X, \mu)} = 1\}.$$

This set will be endowed with a structure of a differentiable manifold. Since X and μ will be fixed, we abbreviate $L^1 = L^1(X, \mu)$ and $L^1(p) = L^1(X, p\mu)$, and similarly for other function spaces. The expectation operator $E_p(f)$ over a function f on X is defined as

$$E_p(f) = \int_X fp \, d\mu = \|f\|_{L^1(p)}.$$

Recall that a differentiable manifold $\mathcal{M}$ of a set of functions is defined as follows: there exists a system of charts $\{\Omega_i, \phi_i\}_{i \in \mathbb{N}}$, collectively called an atlas, such that

(i) each Ω_i is an open set on $\mathcal{M}$ and the union $\cup_{i \in \mathbb{N}} \Omega_i$ covers $\mathcal{M}$;

(ii) the associated mappings $\phi_i : \Omega_i \to B$ are all homeomorphisms (here B is some Banach space) and have the properties that $\phi_i(\Omega_i)$ is open in B and that whenever Ω_i and Ω_j have non-empty intersection then the mapping $\phi_j \circ \phi_i^{-1} : \phi_i(\Omega_i \cap \Omega_j) \to \phi_j(\Omega_i \cap \Omega_j)$ is differentiable up to certain order.

The functions ϕ_i themselves are called coordinate functions. In the finite-dimensional case, the coordinate functions are valued in (subsets of) $\mathbb{R}^n$. However, in the construction of Pistone and Sempi, the coordinate functions ϕ_i are valued on subsets of an "exponential Orlicz space" (which is a Banach space) defined by the Young function $\Phi(t) = \exp(|t|) - 1$; we denote this Orlicz space by $\exp L(p) = \exp L(X, p\mu)$.

A major achievement of Pistone and Sempi (1995) is to provide an atlas ϕ_p centered at a point $p \in \mathcal{M}_\mu$ and mapping its neighborhood to $\exp L_0(p)$:

$$\phi_p(q) = \log \frac{q}{p} - E_p\left(\log \frac{q}{p}\right), \tag{1}$$

where $\exp L_0(p)$ is a sub-space of $\exp L(p)$ consisting of the functions with zero mean:

$$\exp L_0(p) = \{u \in \exp L(p) : E_p(u) = 0\}.$$

In other words, the space $\exp L_0(p)$ contains all random variables with zero expectation value. Pistone and Sempi (1995) showed that in an open neighborhood surrounding p, the $E_p\{\cdot\}$ term in (1) is always finite and hence ϕ_p is a well-defined chart for any reference point $p \in \mathcal{M}_\mu$, with $\phi_p(p) = 0$. (Note that the chart is not global because, in the infinite-dimension setting, the second term in (1) may be unbounded for certain $q \in \mathcal{M}_\mu$.) The inverse of the coordinate mapping ϕ_p^{-1}, which maps (a subset of) $\exp L_0(p) \to \mathcal{M}_\mu$, gives the exponential family of density functions

$$\phi_p^{-1}(u) = \frac{pe^u}{E_p(e^u)}. \tag{2}$$

Again this formula is well-defined only when $E_p(e^u)$ is finite. For this reason, Pistone and Sempi require that u lies in the closed unit ball $B^{\exp}(p)$ of $\exp L(p)$. Under this condition, the coordinate transformation $\phi_{p_2} \circ \phi_{p_1}^{-1}$

between two atlases centered at different points p_1, p_2 of $\mathscr{M}_\mu$ is simply

$$u \mapsto u + \log \frac{p_1}{p_2} - \mathrm{E}_{p_2}\left(u + \log \frac{p_1}{p_2}\right).$$

The construction of the mapping (1) from $\mathscr{M}_\mu$ to the space of zero-mean random variables can be understood in the following way. Suppose we start from the entire Orlicz space $\exp L(p)$, as opposed to $\exp L_0(p)$, and restrict the coordinate function $u \in \exp L(p)$ to lie in the closed unit ball $B^{\exp}$, i.e., $\|u\|_{\exp L(p)} \leqslant 1$. By the definition of the norm and the monotonicity of the Young function, this implies that

$$\int_X \exp(|u|)\, p\, d\mu \leqslant \int_X \exp(|u|/\|u\|_{\exp L(p)})\, p\, d\mu = 1,$$

i.e., $\exp |u|$ is in $L^1(p)$. Thus $\exp u$ is also in $L^1(p)$ since

$$\int_X \exp(u) p\, d\mu \leqslant \int_X \exp(|u|) p\, d\mu.$$

So there is an obvious way to get a function of unit integral (and hence an element of $\mathscr{M}_\mu$) using $\exp u$, namely

$$u \mapsto \frac{p \exp u}{\mathrm{E}_p(\exp u)} \tag{3}$$

which is just (2). It is clear that in this case $\phi_p^{-1} : B^{\exp} \to M_\mu$ as defined above is many-to-one. Specifically, $\phi_p^{-1}(u) = \phi_p^{-1}(v)$ if and only if $\exp u = c \exp v$, or expressed differently, $u = v + \log c$, for some positive constant c. Because of this extra degree of freedom, we can define a foliation of $\exp L(p) = \bigcup_c \exp L_c(p)$ where

$$\exp L_c(p) = \{u \in \exp L(p) : \mathrm{E}_p(u) = \log c\}.$$

We can then require that $\phi_p^{-1}(q)$ be defined on a particular leaf of this foliation of the Banach space. When $c = 1$ this leads to the Pistone–Sempi's formula of ϕ_p, i.e., (1).

We now investigate whether this construction can be extended to an arbitrary Orlicz space L^Φ, using an arbitrary class of Young function Φ, instead of from $\exp L$, using the particular Young function $\exp(|\cdot|) - 1$. Introduce a strictly increasing function $\Phi : \mathbb{R} \to (0, \infty)$ that is convex on $[0, \infty)$ and satisfies $\Phi(-t) = (\Phi(t))^{-1}$ (and $\Phi(0) = 1$). Now $\Psi(t) = \Phi(|t|) - 1$ is a Young function. With an abuse of notation, we still denote the corresponding Orlicz space as L^Φ rather than L^Ψ. The arguments of the above paragraph, which was made to $\Phi(t) = e^t$, can be now made to a general Φ. Everything works fine up until (3). When we require, in general, that $\phi_p^{-1}(u) = \phi_p^{-1}(v)$ if and only if $\Phi(u) = c\Phi(v)$, we run into a problem. Let us define $u_c = \Phi^{-1}(c\Phi(u))$, a function on X. The problem is that we do not know if for a fixed $u \in B^\Phi$ (the unit ball in L^Φ) there exists a positive constant c such that

$$\mathrm{E}_p(u_c) = 0.$$

Hence we must look for another foliation procedure. Since our goal is to turn the multiplicative constant into an additive one, it is natural to look at the logarithm. We easily see that

$$\mathrm{E}_p(\log(\Phi(u_c))) = \log(c) + \mathrm{E}_p(\log \Phi(u)). \tag{4}$$

In order for this formula to make sense, we need to prove that the integral involved in the right-hand side is absolutely convergent. It follows from our assumption $\log \Phi(-t) = -\log \Phi(t)$ that

$$\mathrm{E}_p(|\log \Phi(u)|) = \mathrm{E}_p(\log \Phi(|u|)).$$

An application of Jensen's inequality yields

$$\mathrm{E}_p(\log \Phi(|u|)) \leqslant \log \mathrm{E}_p(\Phi(|u|)) < \infty,$$

where we used also $u \in L^\Phi$. That $\mathrm{E}_p(|\log \Phi(u)|)$ (and hence $\mathrm{E}_p(\log \Phi(u))$) is finite allows us to impose, with reference to (4), the foliation

$$\mathrm{E}_p(\log(\Phi(u_c))) = 0.$$

Therefore the Pistone–Sempi approach will work if we use any positive function $\Phi(t)$ that is strictly increasing and convex on $[0, \infty)$, and satisfies $\Phi(t)\Phi(-t) = 1$. The corresponding map is

$$u \mapsto \frac{p\, \Phi(u)}{\mathrm{E}_p(\Phi(u))}. \tag{5}$$

We call this the pseudo-exponential map. This generalization of Pistone–Sempi is somewhat straightforward because it still uses the same kind of normalization for mapping B^Φ to $\mathscr{M}_\mu$ and follows their same insights into the representation of probability density functions using Orlicz space functions. We turn to another approach in the next section.

3. The affine space model of statistical manifolds

An affine space is a set of points in which each point can be "translated" to any other point through an associated vector space. More precisely, the set A is an affine space associated to the vector space U if there exists an operation $\oplus : A \times U \to A$, called "right translation" or "translation" for short, through which U acts on A transitively. Writing out the axioms, this means that

(i) $(p \oplus u) \oplus u' = p \oplus (u + u')$ for all $p \in A$ and $u, u' \in U$.
(ii) $(p \oplus 0) = p$ for all $p \in A$.
(iii) The restricted mapping $f_u(p) = p \oplus u$ is surjective for every fixed $u \in U$.

It can be easily verified that the exponential model satisfies the above axioms with

$$p \oplus u = \frac{pe^u}{\mathrm{E}_p(e^u)}.$$

The Pistone–Sempi formula (2) can be viewed as defining the right translation for the exponential family. Of course, the affine structure of the exponential map, with log-likelihoods as score functions, is well known, e.g., (Amari, 1985; Murray & Rice, 1993). So one way to generalize the

exponential model is to see what other representations ⊕ could have.

Let us denote $p \oplus u$ by $F(p, u)$. Then the first axiom above can be written as $F(F(p, u), u') = F(p, u + u')$. This is a functional equation known as the "translation equation" (Aczél, 1966, 8.2.2). Below we derive a general form for the solution of this equation in the infinite dimensional case, following the finite-dimensional solution in (Aczél, 1966, 8.2.2, Theorem 1). In the next theorem we assume that the Banach space splits as a direct sum of the type $B = V \oplus \mathbb{R} u_0$ (here the direct sum ⊕ is not to be confused with the right-translation operator above). We use the notation $u = [v, c]$ to denote the splitting of individual elements of the Banach space B accordingly.

Theorem 1. *Let P and B be Banach spaces and suppose that $F: P \times B \to P$ is a function satisfying*

$$F(F(p, u), u') = F(p, u + u') \tag{6}$$

for all $p \in P$ and $u, u' \in B$. Assume that B splits as $V \oplus \{cu_0 : c \in \mathbb{R}\}$ for a fixed $u_0 \in B$, and there exists $\pi \in P$ so that $G_c(\cdot) = F(\pi, [\cdot, c]): V \to P$ is a bijection for every fixed $c \in \mathbb{R}$. Then all continuous solutions of (6) are of the form

$$F(p, u) = G_0(G_0^{-1}(p) + K(u)), \tag{7}$$

where $K: B \to V$ is linear.

Proof. First, it is clear from (6) along with the bijectivity assumption that $F(p, 0) \equiv p$. Fix $\pi \in P$ so that $G_c(\cdot) = F(\pi, [\cdot, c]): V \to P$ is a bijection for every fixed $c \in \mathbb{R}$. For $v \in V$, $G_0(v) = F(\pi, [v, 0])$ so that $G_0(0) = \pi$ with inverse $G_0^{-1}(\pi) = 0$.

We next set $p = \pi$, $u = [v, c]$ and $u' = [v', -c]$ in (6):

$$F(F(\pi, [v, c]), [v', -c]) = F(\pi, [v + v', 0])$$
$$= G_0(v + v').$$

We denote $q = F(\pi, [v, c]) = G_c(v)$, and note that then $G_c^{-1}(q) = v$. We have thus derived that

$$F(q, [v', -c]) = G_0(G_c^{-1}(q) + v') \tag{8}$$

for every $q \in P$ (here we use that G_c is a surjection) and $[v', -c] \in B$.

We substitute this expression for F in the original functional equation (with $u = [v, c]$ and $u' = [v', c']$) and get

$$G_0(G_{-c'}^{-1}(G_0(G_{-c}^{-1}(\pi) + v)) + v')$$
$$= G_0(G_{-(c+c')}^{-1}(\pi) + v + v').$$

Cancelling the outermost G_0 and defining $l: \mathbb{R} \to V$ by $l(c) = G_{-c}^{-1}(\pi)$ yields

$$G_{-c'}^{-1}(G_0(l(c) + v)) = l(c + c') + v. \tag{9}$$

Setting $v = -l(c)$ in (9), rearranging, and recalling $G_0(0) = \pi$ gives

$$l(c') + l(c) = l(c + c').$$

Due to the assumed continuity of G_c^{-1}, l is continuous with $l(0) = G_0^{-1}(\pi) = 0$, so the above Cauchy equation has only

linear solutions

$$l(c) = \kappa c,$$

where κ is an arbitrary element in V. Finally, setting $c = 0$ in (9) gives

$$G_{-c'}^{-1}(G_0(v)) = l(c') + v,$$

or, after denoting $q = G_0(v)$

$$G_{-c'}^{-1}(q) = l(c') + G_0^{-1}(q).$$

Substituting into (8) gives

$$F(p, u) = G_0(G_0^{-1}(p) + K(u)),$$

where $K([v, c]) = v + l(c) = v + c\kappa$ is a linear transformation mapping B to V. It is easy to see that every F of this form is a solution. □

The previous theorem is not directly applicable to our setting. The problem is that, in $F(p, u)$, the range in the second variable u depends on the value of the first variable $p \in \mathcal{M}_\mu$, because u is in the unit ball of $L^\Phi(p)$. The way to get around this is to notice that $L^\infty \subset L^\Phi(p)$ for every $p \in \mathcal{M}_\mu$, and, moreover, it is dense. So we consider only functions $F: \mathcal{M}_\mu \times L^\infty \to \mathcal{M}_\mu$. Since we know all continuous solutions in a dense subset of our space, we obviously know all continuous solutions in the whole space, as well.

With the understanding that we restrict ourselves to dense subsets when necessary, we can see immediately how Theorem 1 relates to the charting of a statistical manifold. By defining $q \mapsto \phi_p(q) \in V$ as

$$\phi_p(q) = G_0^{-1}(q) - G_0^{-1}(p),$$

we get a chart mapping a region (centered on the reference point p) of the manifold $\mathcal{M}_\mu$ to $V \subset B$. Note that K is an identity map when restricted to V: $K(\phi_p(q)) = \phi_p(q)$. So for all p, q

$$F(p, \phi_p(q)) = q,$$

effecting a right-translation from $p \in \mathcal{M}_\mu$ to $q \in \mathcal{M}_\mu$. In the above, the coordinate function $\phi_p(\cdot)$ is valued in V. In general, if it is valued in the entire B, then there is an additional term $c(u_0 - \kappa)$ in the expression of $\phi_p(q)$, where the constant c is arbitrary.

Let us next see how the exponential model fits into this general scheme. Recall that Theorem 1 involves a co-dimension 1 splitting of the Banach space B in the form $B = V \oplus \{cu_0 : c \in \mathbb{R}\}$. Set $u_0 = 1$, the constant function, so that the splitting becomes $\exp L = \exp L_0 \oplus \mathbb{R}$, where $\exp L_0$ is (as defined in Section 1) the space of exponentially integrable functions of zero mean, and $\mathbb{R}$ denotes the space of constant functions. Then choose $\kappa = 0$, which amounts to requiring $F(\pi, u) = \pi, \forall u \in \mathbb{R}$, i.e., for all constant functions u. Consequently, $K(u) \equiv 0$ if and only if u is a constant function. Choose $G_0^{-1}(p) = K(\log p)$. Then

$$p \oplus u = \exp(K^*(K(\log p) + K(u)))$$
$$= \exp(K^*(K(\log(p) + u))),$$

where K^* denotes the pseudo-inverse of K, i.e., a mapping such that $K \circ K^*$ is the identity, which will be specified next. We know that $K^*(K(u)) = u + c$ for some constant c, since constants are the only elements in the kernel of K. Let us define a functional (interpreting the constant function as a real number) by $c[u] = K^*(K(u)) - u$. Then our previous equation gives

$$p \oplus u = \exp(u + \log(p) + c[u + \log p])$$
$$= pe^u e^{c[\log(p \exp u)]}.$$

Thus we get the Pistone–Sempi model if we choose K^* so that

$$c[u] = -\log \int_X \exp u \, d\mu. \tag{10}$$

Let us look at some obvious generalizations: we fix a non-zero $u_0 \in L^\Phi$ and define L to be a linear transformation so that $K(u) = 0$ if and only if $u = cu_0$. We define $G_0^{-1}(p) = K(\Phi^{-1}(p))$. We require that G_0^{-1} be an injection, so $\Phi^{-1}(p) = \Phi^{-1}(q) + cu_0$ if and only if $p = q$ and $c = 0$. We find that

$$p \oplus u = \Phi(u + \Phi^{-1}(p) + c[u + \Phi^{-1}(p)]),$$

where the functional $c[u] = K^*(K(u)) - u$ is as before. Now we should choose

$$c[v] = \min\left\{ t \in \mathbb{R} : \int_X \Phi(v + t) \, d\mu = 1 \right\},$$

in order for $p \oplus u$ to be in the manifold. For $\Phi(t) = e^t$, the above reduces to (10).

To conclude this section, we examine whether or not our previous generalization (5) investigated in Section 2 satisfies the right-translation property. Imposing (6) on (5) yields

$$\Phi(u_1)\Phi(u_2) = c\,\Phi(u_1 + u_2),$$

where c is some constant. So $\log(c^{-1}\Phi)$ satisfies the Cauchy equation

$$\log(c^{-1}\Phi(u_1)) + \log(c^{-1}\Phi(u_2)) = \log(c^{-1}\Phi(u_1 + u_2)),$$

with solution $\log(c^{-1}\Phi(u)) = \beta u$ (where β is another constant), or $\Phi(u) = ce^{\beta u}$. Therefore the exponential model studied by Pistone and Sempi (1995) is the only common element in the two generalized forms of charts for statistical manifolds, as investigated here.

4. Potential applications and discussions

An atlas on a manifold provides a systematic way of representing points on the manifold by coordinate functions. In the case of a statistical manifold $\mathcal{M}_\mu$ where points are themselves probability density functions (positive functions in L^1), the coordinates are Banach-space valued functions as well (e.g., valued in L^Φ), with normalization and positivity constraints removed. Each individual chart centered at $p \in \mathcal{M}_\mu$, denoted $\phi_p^{-1}(u)$ in Section 2 and $F(p, u)$ in Section 3, contains a bijective mapping from an open subset of L^Φ to an open subset of $\mathcal{M}_\mu$ that includes the point p.

Let us give an example for the use of the differential geometric notion of chart/atlas for computations. Our generalized expression (5) based on the normalization approach has a natural interpretation as the Bayes formula

$$q = \frac{pl}{E_p(l)},$$

where p is the prior distribution and $l = c\Phi(u)$ is the likelihood function. In other words, the Orlicz-space functions $u = \Phi^{-1}(l/c)$ are just the Φ^{-1}-transformed likelihoods on the support X (here c is a certain constant). The Pistone–Sempi model (2) is then the Bayes formula using the log-likelihood function

$$u = \log l - E_p(\log l).$$

Viewed in this way, the meaning of the bijective transformation between $\exp L(p)$, or $L^\Phi(p)$ in general, and $\mathcal{M}_\mu$ as given by any p-referenced chart is that the posterior distribution q, with respect to the prior distribution p, is in one-to-one correspondence to (a subspace of) the Orlicz space where likelihood functions (after proper transformation) are defined. So trajectories in $\mathcal{M}_\mu$ can be mapped to trajectories in $\exp L(p)$ or in $L^\Phi(p)$; aggregation of posterior distributions in $\mathcal{M}_\mu$ can be carried out by concatenating likelihood functions on a vector space. The upshot is that, given the prior distribution p, the space of posterior distributions and the space of likelihood functions are in one-to-one correspondence as established by the chart ϕ_p, so that computations involve posterior distributions can be effectively performed using p-centered likelihood functions.

Note that our approaches in extending the exponential model is not an all inclusive one. Burdet, Combe, and Nencka (2001) gave the following chart for mapping the open unit ball of $L^2(p)$ to $\mathcal{M}_\mu$:

$$p \oplus u = p\left(u + \sqrt{1 - E_p(u^2)}\right)^2$$

with the inverse given (for fixed p) by

$$\phi_p(q) = \sqrt{\frac{q}{p}} - E_p\left(\sqrt{\frac{q}{p}}\right).$$

Clearly for such $\oplus$

$$(p \oplus u) \oplus u' \neq p \oplus (u + u');$$

hence this model does not fit into the scheme of Section 3. As a generalization of the translation equation (6), one should consider the so-called "transformation equation" (Aczél, 1966, 8.2.1)

$$(p \oplus u) \oplus u' = p \oplus (u \circ u')$$

where $\circ$ is some associative operator

$$(u \circ u') \circ u'' = u \circ (u' \circ u'').$$

The corresponding functional equation is

$$F(F(p, u), v) = F(p, G(u, v)),$$

where $G: B \times B \to B$ is associative. This is the subject for future research.

5. Conclusions

We have proposed two extensions to Pistone–Sempi's exponential model (2) as an atlas on the manifold of non-parametric probability density functions. The first approach constructs coordinate functions in a general Orlicz space using a Young function whose logarithm is assumed to be odd, and yields (5) as a generalized form for normalization-based models. The second approach views the statistical manifold as an affine space that obeys a "right-translation" property, and derives (7) as the general representation of an atlas. Exponential map is shown to be the only common element from these two approaches, highlighting its importance in charting statistical manifolds.

Appendix A. Orlicz spaces

In this appendix we give a short recap of the theory of Orlicz spaces. For more details the reader is referred to one of the many books on the subject, e.g., (Krasnosel'skiĭ & Rutickiĭ, 1961; Rao & Ren, 1991).

Recall that a Young function is a convex increasing function $\Phi: [0, \infty) \to [0, \infty)$ with $\Phi(0) = 0$. For a Young function Φ we define a *modular* on the set of measurable functions by

$$\varrho_\Phi(u) := \int_X \Phi(|u(x)|)\, d\mu(x).$$

Using the modular we can define the *Luxemburg norm* on the same set by

$$\|u\|_\Phi := \inf\{t > 0: \varrho_\Phi(u/t) \leqslant 1\}.$$

Using these concept we define the Orlicz space by

$$L^\Phi(X) := \{u: \|u\|_\Phi < \infty\}.$$

The best-known example of an Orlicz space is given by the Young functions $\Phi(t) = t^p$ for some real number p greater than or equal to 1. In this case we have

$$\varrho_\Phi(u) := \int_X |u(x)|^p\, d\mu(x).$$

and the norm is just the Lebesgue norm:

$$\begin{aligned}
\|u\|_\Phi &:= \inf\{t > 0: \varrho_\Phi(u/t) \leqslant 1\} \\
&= \left(\int_X |u(x)|^p\, d\mu(x)\right)^{1/p}.
\end{aligned}$$

Another important case is $\Phi(t) = \exp(|t|) - 1$, and the corresponding Orlicz space $\exp L$ is the space of exponentially integrable functions.

References

Aczél, J. (1966). *Lectures on functional equations and their applications.* New York, London: Academic Press.

Amari, S. (1982). Differential geometry of curved exponential families—curvatures and information loss. *Annals of Statistics*, *10*, 357–385.

Amari, S. (1985). *Differential-geometrical methods in statistics.* Lecture Notes in Statistics (Vol. 28). New York: Springer.

Amari, S., & Nagaoka, H. (1993/2000). *Methods of information geometry.* Translations of Mathematical Monographs, (Vol. 191). Oxford: Oxford University Press.

Burdet, G., Combe, Ph., & Nencka, H. (2001). On real Hilbertian info-manifolds. *Disordered and complex systems* London, 2000 (pp. 153–158), AIP Conference Proceedings Vol. 553, Melville, NY: American Institute of Physics.

Čencov, N.N. (1972/1982). *Statistical decision rules and optimal inference.* Providence, RI, USA: American Mathematical Society, 1982 (Originally published in Russian, Nauka, Moscow, 1972).

David, A. P. (1975). Discussion to Efron's paper. *Annals of Statistics*, *3*, 1231–1234.

Efron, B. (1975). Defining the curvature of a statistical problem (with application to second order efficiency) (with discussion). *Annals of Statistics*, *3*, 1189–1242.

Giblisco, P., & Pistone, G. (1998). Connections on non-parametric statistical manifolds by Orlicz space geometry. *Infinite Dimensional Analysis Quantum Probability, and Related Topics*, *1*(2), 325–347.

Grasselli, M. R. (2005). Dual connections in nonparametric classical information geometry. *Annals of the Institute for Statistical Mathematics* (in press).

Karabatsos, G. Bayesian nonparametric model selection and model testing. *Journal of Mathematical Psychology*, in press.

Krasnosel'skiĭ, M. A., & Rutickiĭ, Ya. B. (1961). *Convex functions and orlicz spaces.* Translated by Boron, L.F. Groningen, The Netherlands: P. Noordhoff, Ltd.

Lauritzen, S. (1987). Statistical manifolds. In: S. Amari, O. Barndorff-Nielsen, R. Kass, S. Lauritzen, C. R. Rao (Eds.), *Differential geometry in statistical inference*, *IMS Lecture Notes* (Vol 10) (pp. 163–216). Hayward, CA.

Murray, M. K., & Rice, J. W. (1993). *Differential geometry and statistics.* London: Chapman & Hall.

Pistone, G. (2001). New ideas in non-parametric estimation. *Disordered and complex systems* (pp. 159–164) (London, 2000), AIP Conference Proceedings, Vol. 553, Melville, NY: American Institute of Physics.

Pistone, G., & Rogantin, M. P. (1999). The exponential statistical manifold: mean parameters, orthogonality and space transformations. *Bernoulli*, *5*(4), 721–760.

Pistone, G., & Sempi, C. (1995). An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Annals of Statistics*, *23*(5), 1543–1561.

Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of the Calcutta Mathematical Society*, *37*, 81–91.

Rao, M., & Ren, Z. (1991). *Theory of Orlicz spaces.* Monographs and Textbooks in Pure and Applied Mathematics: Vol. 146. New York: Marcel Dekker, Inc.

Townsend, J. T., Solomon, B., & Smith, J. S. (2001). The perfect Gestalt: Infinite dimensional Riemannian face spaces and other aspects of face perception. In M. J. Wenger, & J. T. Townsend (Eds.), *Computational, geometric and process perspectives of facial cognition: Contexts and challenges* (pp. 39–82). Mahwah, NJ: Lawrence Erlbaum Associates.