



HAL
open science

Defining Exponential Mixture Family with Applications.

Han-Ping Li

► **To cite this version:**

| Han-Ping Li. Defining Exponential Mixture Family with Applications.. 1998. hal-00129626

HAL Id: hal-00129626

<https://hal.science/hal-00129626>

Preprint submitted on 8 Feb 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Defining Exponential Mixture Family with Applications

Han-Ping LI

Université Louis Pasteur, Strasbourg, France

Résumé

Quite different to the classic notion of mixture and quite natural, the so-called exponential mixture is defined, and the exponential mixture family is obtained in this way. Different properties of the exponential mixture family are studied. The associated problem of estimating the mixture coefficient is studied by the likelihood approach and empirical function approach. The exponential mixture family gives us also another classification approach. Finally, a kind of local exponential mixture family approximation is studied.

Key Words: Exponential family, Mixture, Exponential mixture, Kullback information, Classification.

AMT Codes: 62A25, 62B10, 62E10, 62F10, 62H30

1 Introduction

Given $(d+1)$ probability distributions $\mathbf{P}_0, \mathbf{P}_1, \dots, \mathbf{P}_d$ defined on measurable space (Ω, \mathcal{A}) , one can make up a mixture family in a classical way:

$$\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d, \quad \text{with} \quad \alpha_i \geq 0, \quad \sum_{i=1}^d \alpha_i \leq 1;$$

$$\tilde{\mathbf{P}}_\alpha = \sum_{i=1}^d \alpha_i \mathbf{P}_i + (1 - \sum_{i=1}^d \alpha_i) \mathbf{P}_0.$$

This mixture has of cause some advantages:

- i) it is very simple;
- ii) it has a Bayesian interpretation;
- iii) we know how to generate the random variables with this distribution.

It has also some inconvenient, for example:

$\tilde{\mathbf{P}}_\alpha$ is no longer normal distribution even the \mathbf{P}_i are. This is not supprising because the normal distributions possede a log-linear structure. With this point of view in mind, we can define another mixture in the following way:

Let μ be a dominating measure ($\mu = \sum_{i=0}^d \mathbf{P}_i$ for example), $f_i(x) = \frac{d\mathbf{P}_i}{d\mu}(x)$, $i = 0 \cdots d$, be their densities, make a convention as following: $\log 0 = 0$, denote the set:

$$\mathbf{D} = \left\{ \alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d \mid \int \exp \left\{ \sum_{i=1}^d \alpha_i \log(f_i(x)) + (1 - \sum_{i=1}^d \alpha_i) \log(f_0(x)) \right\} \mu(dx) < \infty \right\}.$$

For all element α of \mathbf{D} , we can define

$$\begin{aligned} \Psi(\alpha) &= \log \left(\int \exp \left\{ \sum_{i=1}^d \alpha_i \log(f_i(x)) + (1 - \sum_{i=1}^d \alpha_i) \log(f_0(x)) \right\} \mu(dx) \right) \\ f_\alpha(x) &= \exp \left\{ \sum_{i=1}^d \alpha_i \log(f_i(x)) + (1 - \sum_{i=1}^d \alpha_i) \log(f_0(x)) - \Psi(\alpha) \right\} \\ \mathbf{P}_\alpha(B) &= \int_B f_\alpha(x) \mu(dx) \quad \forall B \in \mathcal{A}. \end{aligned}$$

Definition 1 *The family $\{\mathbf{P}_\alpha, \alpha \in \mathbf{D}\}$ is called the exponential mixture family of $\mathbf{P}_0, \mathbf{P}_1, \dots, \mathbf{P}_d$.*

Why give-we it this name? The reasons are very simple. First of all, it is another kind of mixture, quite different to the classic notion of mixture and rather natural. In the classic definition, the mixture is made of a convex combination. Here, we just take a normalized product:

If we denote for all element α of \mathbf{D} , $C = \int \prod_{i=1}^d f_i(x)^{\alpha_i} f_0(x)^{(1-\sum_{i=1}^d \alpha_i)} \mu(dx) = \exp(\Psi(\alpha)) < \infty$, then

$$f_\alpha(x) = \frac{1}{C} \prod_{i=1}^d f_i(x)^{\alpha_i} f_0(x)^{(1-\sum_{i=1}^d \alpha_i)}.$$

Secondly, there is a tight link between the exponential family and the exponential mixture family defined here. As a matter of fact, we have the following results:

Theorem 1

If $\mathbf{P}_i(\mathrm{d}x) = \exp(\langle \theta_i, T(x) \rangle - K(\theta_i)) \nu(\mathrm{d}x)$, for $i = 0, \dots, d$; then their exponential mixture is given by

$$\mathbf{P}_\alpha(\mathrm{d}x) = \exp(\langle \sum_{i=0}^d \alpha_i \theta_i, T(x) \rangle - K(\sum_{i=0}^d \alpha_i \theta_i)) \nu(\mathrm{d}x).$$

2. Every full exponential family of distributions is an exponential mixture family.
3. If among $\mathbf{P}_0, \mathbf{P}_1, \dots, \mathbf{P}_d$ there is a distribution which dominates all others, then $\{\mathbf{P}_\alpha, \alpha \in \mathbf{D}\}$ is an exponential family of distributions.

Proof:

- 1) Trivial.
- 2) Let $\{\mathbf{P}_\theta, \theta \in \Theta\}$ be an exponential family of distributions :

$$\mathbf{P}_\theta(\mathrm{d}x) = \exp(\langle \theta, T(x) \rangle - K(\theta)) \nu(\mathrm{d}x),$$

where $\Theta = \{\theta : \int \exp(\langle \theta, T(x) \rangle) \nu(\mathrm{d}x) < \infty\}$ is a convex of \mathbb{R}^d , $K(\theta) = \log\left(\int \exp(\langle \theta, T(x) \rangle) \nu(\mathrm{d}x)\right)$. Let $\theta_0, \theta_1, \dots, \theta_d$ be $d+1$ points of Θ such that $\theta_1 - \theta_0, \dots, \theta_d - \theta_0$, are linearly independent. Then for all $\theta \in \Theta$, il existe $\alpha_1, \dots, \alpha_d$ such that $\theta - \theta_0 = \sum_{i=1}^d \alpha_i (\theta_i - \theta_0)$, i.e. $\theta = \sum_{i=0}^d \alpha_i \theta_i$ with $\alpha_0 = 1 - \sum_{i=1}^d \alpha_i$. Let $f_i(x)$ be the density function of \mathbf{P}_i with respect to ν ,

$$f_i(x) = \frac{\mathrm{d}\mathbf{P}_i}{\mathrm{d}\nu}(x) = \exp(\langle \theta_i, T(x) \rangle - K(\theta_i)),$$

then

$$\begin{aligned} \mathbf{P}_\theta(\mathrm{d}x) &= \exp(\langle \sum_{i=1}^d \alpha_i \theta_i, T(x) \rangle - K(\sum_{i=1}^d \alpha_i \theta_i)) \nu(\mathrm{d}x) \\ &= \exp(\sum_{i=0}^d \alpha_i \log(f_i(x)) - \Psi(\alpha)) \nu(\mathrm{d}x), \end{aligned}$$

with

$$\Psi(\alpha) = \log \left(\int \exp \left(\sum_{i=0}^d \alpha_i \log(f_i(x)) \right) \mu(d x) \right) = K(\theta) - \sum_{i=0}^d \alpha_i K(\theta_i).$$

That is, this is also an exponential mixture family.

3) Conversely, suppose that $\mathbf{P}_i \ll \mathbf{P}_0, i = 1 \dots d$, we have then $\{x \in \Omega | f_i(x) > 0\} \subset \{x \in \Omega | f_0(x) > 0\}, i = 1 \dots d$. If we denote $T(x) = \left(\log \left(\frac{f_1(x)}{f_0(x)} \right), \dots, \log \left(\frac{f_d(x)}{f_0(x)} \right) \right)$, with the convention: $\log 0 = 0$, then,

$$\mathbf{P}_\alpha(d x) = \exp(\langle \alpha, T(x) \rangle - \Psi(\alpha)) f_0(x) \mu(d x) = \exp(\langle \alpha, T(x) \rangle - \Psi(\alpha)) \nu(d x).$$

Hence $\{\mathbf{P}_\alpha, \alpha \in \mathbf{D}\}$ is an exponential family of distributions. \square

Theorem 2 *The family $\{\mathbf{P}_\theta, \theta \in \Theta\}$ is a full exponential family of distributions if and only if for any finite number of members of the family $\mathbf{P}_0, \mathbf{P}_1, \dots, \mathbf{P}_d$, their exponential mixture \mathbf{P}_α , for all $\alpha \in \mathbf{D}$ still belongs to the family.*

Remarks:

1. It is clear that $\mathbf{P}_i \in \{\mathbf{P}_\alpha, \alpha \in \mathbf{D}\}, i = 0, \dots, d$ which correspond to $\mathbf{0} = (0, \dots, 0) \in \mathbf{D}$ and $\mathbf{e}_i = (0, \dots, 1, \dots, 0) \in \mathbf{D}, i = 1, \dots, d$.
2. For multidimensional normal distributions $\mathbf{P}_i = N(\theta_i, \Sigma), i = 0, \dots, d$, we have simply $\mathbf{P}_\alpha = N(\theta, \Sigma)$ for $\theta = \sum_{i=0}^d \alpha_i \theta_i$.
For two Poisson distributions $\mathbf{P}_i = \mathcal{P}(\lambda_i), i = 0, 1$, we have $\mathbf{P}_\alpha = \mathcal{P}(\lambda)$ for $\lambda = \lambda_1^\alpha \lambda_0^{(1-\alpha)}$.
3. Let $\Omega = \{a_0, a_1, \dots, a_d\}, \mathcal{A} = \mathcal{P}(\Omega)$. Then the classe of all distributions defined on it is given by:

$$\mathcal{P} = \{(p_0, p_1, \dots, p_d) | p_i \geq 0, \sum_{i=0}^d p_i = 1\}$$

\mathcal{P} is an exponential family:

$$\mathbf{P}(d x) = \exp \left(\sum_{i=0}^d \log(p_i) I_{a_i} \right) \nu(d x) \quad \text{with} \quad \nu(\{d x\}) = 1.$$

\mathcal{P} is also an exponential mixture family of distributions :

$$\frac{d\mathbf{P}}{d\nu}(x) = \exp\left(\sum_{i=0}^d \alpha_i \log(f_i) - \Psi(\alpha)\right)$$

with $\mu(a_i) = 1/e$, $f_i(x) = eI_{a_i}(x)$, $\alpha_i = \log(p_i) + \frac{1 - \sum_{i=0}^d \log(p_i)}{d+1}$, $0 = 1, \dots, d$

and $\Psi(\alpha) = \frac{\sum_{i=0}^d \exp(\alpha_i)}{e}$.

4. The expression is invariant with respect to the dominating measure: if μ and μ' are two dominating measures , with $\mu' \ll \mu$, we have $f'_i(x) = \frac{d\mathbf{P}}{d\mu'} = \frac{d\mathbf{P}}{d\mu}(x) \frac{d\mu}{d\mu'}(x) = f_i(x) \frac{d\mu}{d\mu'}(x)$, then

$$\begin{aligned} f'_\alpha(x) &= \exp\left\{\sum_{i=1}^d \alpha_i \log(f_i(x)) + (1 - \sum_{i=1}^d \alpha_i) \log(f_0(x)) - \Psi(\alpha)\right\} \exp\left(\log\left(\frac{d\mu}{d\mu'}\right)\right) \\ &= \exp\left\{\sum_{i=1}^d \alpha_i \log(f'_i(x)) + (1 - \sum_{i=1}^d \alpha_i) \log(f'_0(x)) - \Psi(\alpha)\right\}. \end{aligned}$$

Here come some examples.

Example: Let $f_0(x)$ be a density with respect to the Lebesgue measure on \mathbb{R} and let $\Phi(t) = \int \exp(tx) f_0(x) dx$ be the Laplace transforme . For all $t \in \mathbf{D} = \{t \in \mathbb{R} | \Phi(t) < \infty\}$, one can define another density on \mathbb{R} by :

$$f_t(x) = \frac{\exp(tx) f_0(x)}{\Phi(t)}.$$

If we take $f_1(x) = \frac{\exp(t_1 x) f_0(x)}{\Phi(t_1)}$ for a $t_1 \in \mathbf{D}$, $t_1 \neq 0$, then

$$f_t(x) = \frac{\exp(tx) f_0(x)}{\Phi(t)} = \exp\left(\frac{t}{t_1} \log(f_1(x)) + \left(1 - \frac{t}{t_1}\right) \log(f_0(x)) - \log\left(\Phi\left(\frac{t}{t_0}\right)\right)\right) :$$

this is an exponential mixture family.

Example: Let $f_0(x)$ be a density with respect to the Lebesgue measure on \mathbb{R}^d and let $\Phi(\mathbf{t}) = \int \exp(\langle \mathbf{t}, x \rangle) f_0(x) dx$ be the Laplace transform. For all $\mathbf{t} \in \mathbf{D} = \{\mathbf{t} \in \mathbb{R}^d | \Phi(\mathbf{t}) < \infty\}$, one can define another density on \mathbb{R}^d by :

$$f_{\mathbf{t}}(x) = \frac{\exp(\langle \mathbf{t}, x \rangle) f_0(x)}{\Phi(\mathbf{t})}.$$

Since \mathbf{D} is convex and contains the origin, we can take $\mathbf{t}_0 = 0 \in \mathbf{D}$, $\mathbf{t}_i \in \mathbf{D}$, $i = 1, \dots, d$ such that $(\mathbf{t}_i, i = 1, \dots, d)$ is linearly independent. Therefore for all $\mathbf{t} \in \mathbf{D}$, we have $\mathbf{t} = \sum_{i=1}^d \alpha_i \mathbf{t}_i$. If we take $f_i(x) = \frac{\exp(\langle \mathbf{t}_i, x \rangle) f_0(x)}{\Phi(\mathbf{t}_i)}$, $i = 0, \dots, d$; then

$$\begin{aligned} f_{\mathbf{t}}(x) &= \frac{\exp(\langle \mathbf{t}, x \rangle) f_0(x)}{\Phi(\mathbf{t})} \\ &= \exp\left(\sum_{i=1}^d \alpha_i \log(f_i(x)) + (1 - \sum_{i=1}^d \alpha_i) \log(f_0(x)) - \left(\log(\Phi(\mathbf{t})) - \sum_{i=1}^d \alpha_i \log(\Phi(\mathbf{t}_i))\right)\right). \end{aligned}$$

this is a exponential mixture family.

Example: For normal distributions, $\mathbf{P}_i = N(\theta_i, \Sigma)$, $i = 0, \dots, d$, we have $\mathbf{P}_\alpha = N(\theta, \Sigma)$ for $\theta = \sum_{i=0}^d \alpha_i \theta_i$.

Example: For two Poisson distributions $\mathbf{P}_i = \mathcal{P}(\lambda_i)$, $i = 0, 1$, we have $\mathbf{P}_\alpha = \mathcal{P}(\lambda)$ for $\lambda = \lambda_1^\alpha \lambda_0^{(1-\alpha)}$.

Example: Let $\mathbf{P}_i(dx) = f_i(x)dx = \frac{1}{\pi(1+(x-a_i)^2)}dx$, $i = 0, 1$; be two Cauchy distributions, then for all $\alpha \in [0, 1]$,

$$\mathbf{P}_\alpha(dx) = \exp\left(\alpha \log(f_1(x)) + (1 - \alpha) \log(f_0(x)) - \psi(\alpha)\right)(dx)$$

defines an exponential mixture family which contains the two Cauchy distributions. We all know the difference between the binomial distribution and the hypergeometric distribution. We can easily obtain an exponential mixture family which contains these two distributions.

Consider two statistical models

$(\Omega^{(1)}, \mathcal{A}^{(1)}, \{\mathbf{P}_\theta | \theta \in \Theta\})$ and $(\Omega^{(2)}, \mathcal{A}^{(2)}, \{\mathbf{Q}_\theta | \theta \in \Theta\})$. Suppose there exists two Markov kernels Π_{12} and Π_{21} such that

$$\mathbf{P}_\theta \Pi_{12} = \mathbf{Q}_\theta \quad \text{and} \quad \mathbf{Q}_\theta \Pi_{21} = \mathbf{P}_\theta,$$

for all $\theta \in \Theta$. Then they are called equivalent.

Chensov gave the following interesting result: (c.f.[2, 1982])

Theorem 3 (*Chensov*)

The operation of the mixture

$$(\{\mathbf{P}_i, i = 0, \dots, k\}) \longrightarrow \mathbf{P}_\alpha$$

is stable by Markov isomorphism: Consider two families of distributions $\{\mathbf{P}_i, i = 0, \dots, k\}$ and $\{\mathbf{Q}_i, i = 0, \dots, k\}$. Suppose there exists two Markov kernels Π_{12} and Π_{21} such that $\mathbf{P}_i \Pi_{12} = \mathbf{Q}_i$ and $\mathbf{Q}_i \Pi_{21} = \mathbf{P}_i, i = 0, \dots, k$. we have then,

$$\mathbf{P}_\alpha \Pi_{12} = \mathbf{Q}_\alpha \quad \text{and} \quad \mathbf{Q}_\alpha \Pi_{21} = \mathbf{P}_\alpha \quad \forall \alpha \in \mathbf{D},$$

that is, the two two statistical models generated by the exponential mixture family are equivalent.

2 Properties linked with the Kullback information and the Fisher information

Let us first of all take some notations:

1) Let $\mathbf{P}(dx) = f(x)\mu(dx)$ and $\mathbf{Q}(dx) = g(x)\mu(dx)$ be two probability distributions with respect to the dominating measure μ , the Kullback information is defined by:

$$K(\mathbf{P}, \mathbf{Q}) = \int \log\left(\frac{f(x)}{g(x)}\right) f(x)\mu(dx).$$

2) Let X be a random variable and denote $\mathbf{E}_\alpha(X)$ whenever the distribution of X is given by \mathbf{P}_α . Let $\{f_i(x), i = 0, \dots, d\}$ be $d+1$ densities with respect to a dominate measure μ , we define for all $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbf{D}$,

$$f_\alpha(x) = \exp\left\{\sum_{i=1}^d \alpha_i \log(f_i(x)) + (1 - \sum_{i=1}^d \alpha_i) \log(f_0(x)) - \Psi(\alpha)\right\},$$

$$\Psi(\alpha) = \log\left(\int \exp\left\{\sum_{i=1}^d \alpha_i \log(f_i(x)) + (1 - \sum_{i=1}^d \alpha_i) \log(f_0(x))\right\} \mu(dx)\right).$$

Recall that $\mathbf{0} = (0, \dots, 0)$ belongs to \mathbf{D} and $\mathbf{e}_i = (0, \dots, 1, \dots, 0)$ is in \mathbf{D} , for all $i = 1, \dots, d$.

Theorem 4

- 1) $\frac{\partial \Psi(\alpha)}{\partial \alpha_i} \Big|_{\alpha=\mathbf{e}_i} = K(\mathbf{P}_i, \mathbf{P}_0),$
- 2) $\frac{\partial \Psi(\alpha)}{\partial \alpha_i} \Big|_{\alpha=\mathbf{0}} = -K(\mathbf{P}_0, \mathbf{P}_i),$
- 3) $\frac{\partial^2 \Psi(\alpha)}{\partial \alpha_i \partial \alpha_j} = \mathbf{E}_\alpha \left(\left(\frac{\partial \log(f_\alpha(X))}{\partial \alpha_i} \right) \left(\frac{\partial \log(f_\alpha(X))}{\partial \alpha_j} \right) \right).$

Proof: In fact,

$$\exp(\Psi(\alpha)) = \int \exp \left\{ \sum_{i=1}^d \alpha_i \log \left(\frac{f_i(x)}{f_0(x)} \right) \right\} f_0(x) \mu(dx),$$

$$\exp(\Psi(\alpha)) \frac{\partial \Psi(\alpha)}{\partial \alpha_i} = \int \log \left(\frac{f_i(x)}{f_0(x)} \right) \exp \left\{ \sum_{i=1}^d \alpha_i \log \left(\frac{f_i(x)}{f_0(x)} \right) \right\} f_0(x) \mu(dx),$$

therefore

$$\frac{\partial \Psi(\alpha)}{\partial \alpha_i} = \mathbf{E}_\alpha \left(\log \left(\frac{f_i(x)}{f_0(x)} \right) \right).$$

$$K(\mathbf{P}_i, \mathbf{P}_0) = \mathbf{E}_{\mathbf{e}_i} \left(\log \left(\frac{f_i(x)}{f_0(x)} \right) \right) = \frac{\partial \Psi(\alpha)}{\partial \alpha_i} \Big|_{\alpha=\mathbf{e}_i}.$$

$$K(\mathbf{P}_0, \mathbf{P}_i) = -\mathbf{E}_0 \left(\log \left(\frac{f_i(x)}{f_0(x)} \right) \right) = -\frac{\partial \Psi(\alpha)}{\partial \alpha_i} \Big|_{\alpha=\mathbf{0}}. \quad \square$$

Theorem 5 Denote $V_\alpha = (\mathbf{E}_\alpha \left(\log \left(\frac{f_1(X)}{f_0(X)} \right) \right), \dots, \mathbf{E}_\alpha \left(\log \left(\frac{f_d(X)}{f_0(X)} \right) \right))'$ we have

- 1) $\mathbf{E}_\alpha \left(\log \left(\frac{f_\beta(X)}{f_\gamma(X)} \right) \right) = \langle \beta - \gamma, V_\alpha \rangle - \Psi(\beta) + \Psi(\gamma).$
- 2) $K(\mathbf{P}_\alpha, \mathbf{P}_\beta) = \langle \alpha - \beta, V_\alpha \rangle - \Psi(\alpha) + \Psi(\beta).$
- 3) $K(\mathbf{P}_\gamma, \mathbf{P}_\alpha) - K(\mathbf{P}_\gamma, \mathbf{P}_\beta) = K(\mathbf{P}_\beta, \mathbf{P}_\alpha) + \langle \beta - \alpha, V_\gamma - V_\beta \rangle.$
- 4) $K(\mathbf{P}_\alpha, \mathbf{P}_\gamma) - K(\mathbf{P}_\beta, \mathbf{P}_\gamma) = K(\mathbf{P}_\alpha, \mathbf{P}_\beta) + \langle \beta - \gamma, V_\alpha - V_\beta \rangle.$

Proof: Trivial computation. \square

Theorem 6

- 1) The mapping $\alpha \longrightarrow K(\mathbf{P}_0, \mathbf{P}_\alpha)$ is convex on \mathbf{D} .
 - 2) For all $\alpha \in \mathbf{D}, \alpha \neq \mathbf{0}$, the mapping $t \longrightarrow K(\mathbf{P}_0, \mathbf{P}_{t\alpha})$ is strictly increasing on $[0, 1]$.
- In particular, When $d = 1$, then the mapping $t \longrightarrow K(\mathbf{P}_0, \mathbf{P}_t)$ is strictly increasing on $[0, 1]$.

Proof: In fact, we have

$$f_\alpha(x) = \exp\left\{\sum_{i=1}^d \alpha_i \log\left(\frac{f_i(x)}{f_0(x)}\right) - \Psi(\alpha)\right\} f_0(x),$$

$$\log\left(\frac{f_0(x)}{f_\alpha(x)}\right) = \sum_{i=1}^d \alpha_i \log\left(\frac{f_0(x)}{f_i(x)}\right) + \Psi(\alpha).$$

Taking expectation with respect to \mathbf{P}_0 , we see that

$$K(\mathbf{P}_0, \mathbf{P}_\alpha) = \sum_{i=1}^d \alpha_i K(\mathbf{P}_0, \mathbf{P}_i) + \Psi(\alpha).$$

Now if we calculate the partial derivative with respect to the i th component of α , that is α_i , we have

$$\frac{\partial K(\mathbf{P}_0, \mathbf{P}_\alpha)}{\partial \alpha_i} = K(\mathbf{P}_0, \mathbf{P}_i) + \frac{\partial \Psi(\alpha)}{\partial \alpha_i}.$$

Hence,

$$\left(\frac{\partial^2 K(\mathbf{P}_0, \mathbf{P}_\alpha)}{\partial \alpha_i \partial \alpha_j}\right) = \left(\frac{\partial^2 \Psi(\alpha)}{\partial \alpha_i \partial \alpha_j}\right)$$

is a definite positive matrix because of the convexity of the function $\Psi(\alpha)$.

Similarly,

$$\log\left(\frac{f_0(x)}{f_{t\alpha}(x)}\right) = t \sum_{i=1}^d \alpha_i \log\left(\frac{f_0(x)}{f_i(x)}\right) + \Psi(t\alpha).$$

Taking expectation with respect to \mathbf{P}_0 , we see that

$$K(\mathbf{P}_0, \mathbf{P}_{t\alpha}) = t \sum_{i=1}^d \alpha_i K(\mathbf{P}_0, \mathbf{P}_i) + \Psi(t\alpha).$$

Here we calculate the derivative with respect to t , we have

$$\begin{aligned} \frac{dK(\mathbf{P}_0, \mathbf{P}_{t\alpha})}{dt} &= \sum_{i=1}^d \alpha_i K(\mathbf{P}_0, \mathbf{P}_i) + \sum_{i=1}^d \alpha_i \frac{\partial \Psi(t\alpha)}{\partial \alpha_i}. \\ \frac{d^2 K(\mathbf{P}_0, \mathbf{P}_{t\alpha})}{dt^2} &= \sum_{i=1}^d \sum_{j=1}^d \alpha_i \alpha_j \frac{\partial^2 \Psi(t\alpha)}{\partial \alpha_i \partial \alpha_j} > 0. \end{aligned}$$

Therefore the mapping $t \rightarrow \frac{dK(\mathbf{P}_0, \mathbf{P}_{t\alpha})}{dt}$ is a strictly increasing function, for all $t \in [0, 1]$. Consequently, for all $t \in [0, 1]$

$$\frac{dK(\mathbf{P}_0, \mathbf{P}_{t\alpha})}{dt} > \left. \frac{dK(\mathbf{P}_0, \mathbf{P}_{t\alpha})}{dt} \right|_{t=0}$$

But

$$\begin{aligned} \left. \frac{dK(\mathbf{P}_0, \mathbf{P}_{t\alpha})}{dt} \right|_{t=0} &= \sum_{i=1}^d \alpha_i K(\mathbf{P}_0, \mathbf{P}_i) + \sum_{i=1}^d \alpha_i \left. \frac{\partial \Psi(t\alpha)}{\partial \alpha_i} \right|_{t=0} \\ &= \sum_{i=1}^d \alpha_i K(\mathbf{P}_0, \mathbf{P}_i) - \sum_{i=1}^d \alpha_i K(\mathbf{P}_0, \mathbf{P}_i) = 0. \end{aligned}$$

Therefore for all $t \in [0, 1]$

$$\frac{dK(\mathbf{P}_0, \mathbf{P}_{t\alpha})}{dt} > 0,$$

that is, the mapping $t \rightarrow K(\mathbf{P}_0, \mathbf{P}_{t\alpha})$ is a strictly increasing function. \square

3 Estimations

Suppose that $\{\log(\frac{f_i}{f_0}((X_i))), i = 1, \dots, k\}$ are linearly independent (c.f.[7, 1982], [8, 1985]) for some sufficient conditions). then α is unique: $\{\mathbf{P}_\alpha, \alpha \in \mathbf{D}\}$ is identifiable. Suppose that (X_1, \dots, X_n) be a sample from the distribution \mathbf{P}_α belonging to $\{\mathbf{P}_\alpha, \alpha \in \mathbf{D}\}$ an exponential mixture family of $\mathbf{P}_0, \mathbf{P}_1, \dots, \mathbf{P}_d$. Consider the problem of estimating the mixture coefficient α .

3.1 The maximum of likelihood method

Since the likelihood is given by

$$L_\alpha(X_1, \dots, X_n) = \exp\left\{\sum_{j=1}^n \sum_{i=1}^d \alpha_i \log\left(\frac{f_i}{f_0}(X_j)\right) - n\Psi(\alpha)\right\} \prod_{j=1}^n f_0(X_j)$$

so the maximum likelihood estimator $\hat{\alpha}$ of α must satisfy the following equation:

$$\frac{\partial \Phi(\alpha)}{\partial \alpha_i} = \frac{1}{n} \sum_{j=1}^n \log\left(\frac{f_i}{f_0}(X_j)\right).$$

Remarks:

- 1) Since the function $\Phi(\alpha)$ is convex, the solution, if it exists, is unique.
- 2) It may exist when n is sufficient large:

$$\frac{1}{n} \sum_{j=1}^n \log\left(\frac{f_i}{f_0}(X_j)\right) \in \mathbf{E}_\alpha\left(\frac{f_i}{f_0}(X)\right) = \frac{\partial \Phi(\alpha)}{\partial \alpha_i}.$$

Example: Let $P_i = N(\theta_i, \Sigma)$, $i = 0, \dots, d$; then we have $P_\alpha = N(\sum_{i=0}^d \alpha_i \theta_i, \Sigma)$. Since $f_i(x) = \exp\left(\langle \Sigma^{-1} \theta, x \rangle - \frac{1}{2} \theta'_i \Sigma^{-1} \theta_i + \frac{1}{2} \theta'_0 \Sigma^{-1} \theta_0 - C\right)$, so

$$\log\left(\frac{f_i}{f_0}(x)\right) = (\theta_i - \theta_0)' \Sigma^{-1} x - \frac{1}{2} \theta'_i \Sigma^{-1} \theta_i + \frac{1}{2} \theta'_0 \Sigma^{-1} \theta_0.$$

We can easily see that $K(\theta) = \frac{1}{2} \theta' \Sigma^{-1} \theta - C$, hence

$$\Psi(\alpha) = K\left(\sum_{i=0}^d \alpha_i \theta_i\right) - \sum_{i=0}^d \alpha_i K(\theta_i),$$

$$\begin{aligned} \frac{\partial \Psi(\alpha)}{\partial \alpha_i} &= (\theta_i - \theta_0)' \frac{\partial K(\sum_{i=0}^d \alpha_i \theta_i)}{\partial \alpha_i} - (K(\theta_i) - K(\theta_i)) \\ &= (\theta_i - \theta_0)' \Sigma^{-1} \left(\sum_{i=0}^d \alpha_i \theta_i\right) - \frac{1}{2} \theta'_i \Sigma^{-1} \theta_i + \frac{1}{2} \theta'_0 \Sigma^{-1} \theta_0, \end{aligned}$$

$$\text{and } \frac{1}{n} \sum_{j=1}^n \log\left(\frac{f_i}{f_0}(X_j)\right) = (\theta_i - \theta_0)' \Sigma^{-1} \bar{X} - \frac{1}{2} \theta'_i \Sigma^{-1} \theta_i + \frac{1}{2} \theta'_0 \Sigma^{-1} \theta_0.$$

Therefore the maximum likelihood estimator $\hat{\alpha}$ of α will satisfy

$$(\theta_i - \theta_0)' \Sigma^{-1} \left(\sum_{i=0}^d \hat{\alpha}_i \theta_i \right) = (\theta_i - \theta_0) \Sigma^{-1} \bar{X}.$$

It is clear that $\{\log(\frac{f_i}{f_0}((X_i))), i = 1, \dots, k\}$ are linearly independent if and only if $(\theta_1 - \theta_0), \dots, (\theta_k - \theta_0)$ are linearly independent. As a consequence, the maximum likelihood estimator $\hat{\alpha}$ of α will satisfy

$$\left(\sum_{i=0}^d \hat{\alpha}_i \theta_i \right) = \bar{X}.$$

In particular, if $d = 2$, $\theta_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\theta_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\theta_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, then we have

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = \begin{pmatrix} \bar{X}_1 \\ \bar{X}_2 \end{pmatrix}.$$

When $d = 1$, $\theta_1 = b$, $\theta_0 = a$, $a \neq b$, then $\hat{\alpha}_1 = \frac{\bar{X} - a}{b - a}$.
They are quite logic!

3.2 The minimum distance method:

Let \mathcal{F} be a class of integrable functions, μ be a signed measure, we define

$$\|\mu\| = \sup_{f \in \mathcal{F}} \left| \int f d\mu \right|.$$

Let (X_1, \dots, X_n) be a sample from the distribution \mathbf{P}_α , $\hat{\mathbf{P}}_n(x) = \frac{1}{n} \sum_{j=1}^n I_{\{X_j \leq x\}}$ be the empirical distribution function. We then define the estimator $\tilde{\alpha}$ of $\alpha \in \mathbf{D}$.

$$\tilde{\alpha} = \arg \inf_{\alpha \in \mathbf{D}} \|\mathbf{P}_\alpha - \hat{\mathbf{P}}_n\|.$$

We know also that if X_1, \dots, X_n are i.i.d. then $\|\hat{\mathbf{P}}_n - \mathbf{P}_\alpha\|$ converge to zero for any V.C. class \mathcal{F} . We know by the Glivenko-Cantelli theorem that the V.C. class is much richer than the class $\{I_{(-\infty, a]}, a \in \mathbb{R}\}$. (c.f.[9, 1984])

4 Classification

Suppose that one makes a number of measurements on an individual and wishes to classify the individual into one of several categories or populations on the basis of the information gotten from these measurements. We assume that there are only a finite number of categories or populations from which the individual may have come and each population is characterized by a probability distribution of measurements. Thus an individual is considered as a random variable from this population. The problem is how to determine the population from which the individual may have come?

Let $\mathbf{P}_0, \dots, \mathbf{P}_d$ be $(d+1)$ probability distributions defined on a measurable space (Ω, \mathcal{A}) . Let (X_1, \dots, X_n) be a random sample having come from one of these populations. Let $\{\mathbf{P}_\alpha, \alpha \in \mathbf{D}\}$ be the exponential mixture family of $\mathbf{P}_0, \dots, \mathbf{P}_d$. And let $\hat{\alpha} = (\hat{\alpha}_0, \dots, \hat{\alpha}_d)'$ be an estimator of $\alpha = (\alpha_0, \dots, \alpha_d)'$ based on (X_1, \dots, X_n) . Consider the following testing hypotheses problems:

$$H_0^{(i)} : \alpha_i \leq \hat{\alpha}_i \quad \text{against the alternative} \quad H_1^{(i)} : \alpha_i > \hat{\alpha}_i,$$

$i = 0, \dots, d$. Since $\{\mathbf{P}_\alpha, \alpha \in \mathbf{D}\}$ is a multiparameter exponential family, we know that there are uniformly most powerful unbiased tests ϕ_i for testing the hypotheses $H_0^{(i)}$ against the alternative $H_1^{(i)}$, $i = 0, \dots, d$. (c.f.[4, 1986]) Let β_i be the power of ϕ_i at the point $\alpha_i = 1$. We then decide to classify X into the i th population if

$$\beta_i = \min_{j=0}^d \beta_j.$$

When $d = 1$, \mathbf{D} is just an interval of \mathbb{R} .

$$f_\alpha(x) = \exp\left\{\alpha \log(f_1(x)) + (1-\alpha) \log(f_0(x)) - \Psi(\alpha)\right\} = \exp\left\{\alpha \log\left(\frac{f_1(x)}{f_0(x)}\right) - \Psi(\alpha)\right\} f_0(x),$$

where $\alpha_1 = \alpha$, $\alpha_0 = 1 - \alpha$. Consider the following testing hypotheses problems:

$$H_0^{(i)} : \alpha_i \leq \hat{\alpha}_i \quad \text{against the alternative} \quad H_1^{(i)} : \alpha_i > \hat{\alpha}_i,$$

$i = 0, \dots, 1$. So the uniformly most powerful tests ϕ_i is given by

$$\phi_1 = \begin{cases} 1 & \text{if } \frac{f_1(x)}{f_0(x)} > C_1 \\ 0 & \text{if } \frac{f_1(x)}{f_0(x)} < C_1 \end{cases}$$

$$\phi_0 = \begin{cases} 1 & \text{if } \frac{f_0(x)}{f_1(x)} > C_0 \\ 0 & \text{if } \frac{f_0(x)}{f_1(x)} < C_0 \end{cases}$$

and $\beta_i \approx \mathbf{P}_i\left(\frac{f_i(x)}{f_j(x)} > C_i\right)$ for $j \neq i$.

In particular, suppose that $f(x, \theta) = \exp(x\theta - K(\theta))$ is a symmetric function when $\theta = 0$ (for example, the one-dimensional normal distribution $\mathcal{N}(\theta, c^2)$) and that $\mathbf{P}_1(dx) = f(x - b)dx$, $\mathbf{P}_0(x) = f(x - b)dx$, with $b > a$; we have $\mathbf{P}_\alpha(dx) = f\left(x - (\alpha b + (1 - \alpha)a)\right)dx$, and $\hat{\alpha} = \frac{\bar{x} - a}{b - a}$.

We then have

$$C_1 - \bar{x} = -(C_0 - \bar{x}),$$

$$\beta_1 = \mathbf{P}_1(\bar{x} - b > C_1 - b), \quad \beta_0 = \mathbf{P}_0(\bar{x} - a < C_0 - a).$$

It follows that

$$\beta_1 < \beta_0 \Leftrightarrow (C_1 - b > a - C_0) \Leftrightarrow (C_1 - b > a - 2\bar{x} + C_1) \Leftrightarrow (b - \bar{x} < \bar{x}).$$

That is \bar{x} is nearer to b than a .

So the classification rule will be

$$d = \begin{cases} f_1 & \text{if } b - \bar{x} \leq \bar{x} - a \\ f_0 & \text{if } b - \bar{x} > \bar{x} - a \end{cases}$$

5 Local Approximation by an exponential mixture family of distributions

Let $\{\mathbf{P}_\theta | \theta \in \Theta\}$, $\Theta \subset \mathbb{R}^k$ be a parametric family of distributions. Let $\theta^* \in \Theta$ be a point of the parameter space. Suppose that we are asked to approximate the family locally in a neighbourhood of θ^* by a nice family. Under some regular conditions, the family $\{\mathbf{P}_\theta | \theta \in \Theta\}$ can be considered as a Riemannian manifold (c.f.[1, 1985]). The only reasonable Riemannian metric is given by Fisher information matrix, and the only reasonable affine connexion are given by Chentsov-Amari α -connexions. (c.f.[5, 1986], [6, 1993]). If we take Chentsov-Amari 1-connexions ($\alpha = 1$), then the exponential family is a flat-space, that is, a space on which the Riemann-Christoffel curvature vanishes identically, just like a hyper-plane in an Euclidean space. Therefore, we are often asked to approximate the family locally in a neighbourhood of θ^* by an exponential family.

We propose the following approach:

Choose $(d + 1)$ points $\theta_0, \dots, \theta_d$ in a neighbourhood of θ^* . Then project the family $\{\mathbf{P}_\theta | \theta \in \Theta\}$ on the exponential mixture family of $\mathbf{P}_{\theta_0}, \mathbf{P}_{\theta_1}, \dots, \mathbf{P}_{\theta_d}$. Let the

$(d + 1)$ points $\theta_0, \dots, \theta_d$ trend to θ^* , and the limit of the projection family obtained in this way is then what we wish.

Intuitively, let μ be a dominating σ -finite measure, denote $f(x, \theta_i) = \frac{d\mathbf{P}_{\theta_i}}{d\mu}(x)$, $\theta_0 = (\theta_{01}, \dots, \theta_{0k})'$, and

$$\nabla l(\theta_0) = \left(\frac{\partial \log(f(x, \theta_0))}{\partial \theta_{01}}, \dots, \frac{\partial \log(f(x, \theta_0))}{\partial \theta_{0k}} \right)'$$

. Then

$$\begin{aligned} f_\alpha(x) &= \frac{d\mathbf{P}_\alpha}{d\mu}(x) \\ &= \exp \left\{ \sum_{i=1}^d \alpha_i \log(f(x, \theta_i)) + \left(1 - \sum_{i=1}^d \alpha_i\right) \log(f(x, \theta_0)) - \Psi(\alpha) \right\} \\ &= \exp \left\{ \sum_{i=1}^d \alpha_i \left(\log(f(x, \theta_i)) - \log(f(x, \theta_0)) \right) - \Psi(\alpha) \right\} f(x, \theta_0) \\ &= \exp \left\{ \sum_{i=1}^d \alpha_i \left\langle \theta_i - \theta_0, \nabla l(\theta_0) \right\rangle - \Psi(\alpha) \right\} f(x, \theta_0) + o_{P_{\theta_0}}(1) \end{aligned}$$

We can also see this by the following consideration:

The family $\{\mathbf{P}_\theta, \theta \in \Theta\}$ is called **localement asymptotiquement quadratique** at point θ^* , $i = 1, \dots, d$. (c.f.[3, 1990]) if :

- 1) $\theta^* \in \text{int}\Theta$,
- 2) There exists a sequence of matrices which trends toward to 0 δ_n , for all bounded sequence \mathbf{t}_n : $\sup_n |\mathbf{t}_n' \mathbf{t}_n| < \infty$, we have

$$\mathbf{P}_{\theta + \delta_n \mathbf{t}_n}^{(n)} \triangleleft \triangleright \mathbf{P}_\theta^{(n)},$$

that is, for all event B , $\mathbf{P}_{\theta + \delta_n \mathbf{t}_n}^{(n)}(B)$ if and only if $\mathbf{P}_\theta^{(n)}(B)$.

- 3) The logarithm of the likelihood ratio can be written as :

$$\Lambda_{\theta + \delta_n \mathbf{t}_n, \theta} = \log \left(\frac{d\mathbf{P}_{\theta + \delta_n \mathbf{t}_n}^{(n)}}{d\mathbf{P}_\theta^{(n)}} \right) = \langle S_n, \mathbf{t}_n \rangle - \frac{1}{2} \mathbf{t}_n' K_n \mathbf{t}_n + o_{P_\theta}(1).$$

where S is a random vector and $P_\theta^{(n)}(K_n \text{ is definite positive}) = 1$.

We know that if $\mathbf{P}_\theta^{(n)}$ is the joint distribution of n independent variables:

$\mathcal{L}(X_1, \dots, X_n) = \mathbf{P}_\theta^{(n)}$, and if the distribution of each variable is absolutely continuous with respect to a dominating measure ν : $\mathcal{L}(X_i)(dx) = f_i(x, \theta)\nu(dx)$

and if the square root of the density is differentiable in quadratic mean, then, $\{\mathbf{P}_\theta^{(n)}, \theta \in \Theta\}$ is locally asymptotically quadratic, and that the matrices K_n trend toward to a constant matrix K . (c.f.[3, 1990]).

Under these conditions, $\{\mathbf{P}_\theta, \theta \in \Theta\}$ is locally approximately an exponential mixture family:

Let $\mathbf{t}_0 = 0$, and let $\mathbf{t}_1, \dots, \mathbf{t}_d$ be a base of \mathbb{R}^d . then for all $\mathbf{t} = \sum_{i=1}^d \alpha_i \mathbf{t}_i$, we have:

$$\begin{aligned} \Lambda_{\theta+\delta_n \mathbf{t}, \theta} - \sum_{i=1}^d \alpha_i \Lambda_{\theta+\delta_n \mathbf{t}_i, \theta} &= -\left(\frac{1}{2} \mathbf{t}' K_n \mathbf{t} - \sum_{i=1}^d \alpha_i \frac{1}{2} \mathbf{t}_i' K_n \mathbf{t}_i\right) + o_{P_\theta}(1) \\ &= -\left(\frac{1}{2} \mathbf{t}' K \mathbf{t} - \sum_{i=1}^d \alpha_i \frac{1}{2} \mathbf{t}_i' K \mathbf{t}_i\right) + o_{P_\theta}(1). \end{aligned}$$

Therefore

$$P_{\theta+\delta_n \mathbf{t}}^n(d x) \simeq \exp\left(\sum_{i=0}^d \alpha_i \log\left(\frac{dP_{\theta+\delta_n \mathbf{t}_i}^n}{d\nu}\right) - \Psi(\alpha)\right) \nu(d x)$$

with $\Psi(\alpha) = \left(\frac{1}{2} \mathbf{t}' K \mathbf{t} - \sum_{i=1}^d \alpha_i \frac{1}{2} \mathbf{t}_i' K \mathbf{t}_i\right)$.

Références

- [1] Amari, Shun-Ichi (1985)
Differential-Geometrical Methods in Statistics
Lecture Notes in Statistics, 28, Springer-Verlag, New York-Berlin-Heidelberg.
- [2] Chensov N.N. (1982)
Statistical Decision Rules and Optimal Inference
Amer. Math.Soc., Providence Rhode Island.
- [3] Le Cam, L. and Yang, G. (1990)
Asymptotics in Statistics, Some Basic Concepts
Springer series in statistics, Springer-Verlag, New York-Berlin-Heidelberg.

- [4] Lehmann, E.L. (1986)
 Testing in Statistical Hypotheses
Wiley series in probability and mathematical statistics, John Wiley & Sons.
- [5] Li, H.P. (1986)
 1) l'Etude de la règle de métrique riemannienne de Fisher-Rao et de la règle de connexion affine de Chentsov-Amari;
 2) l'Approximation de densité par projection poursuite
Thèse de Doctorat en Mathématique à l'Université de Paris-sud, Centre Scientifique d'Orsay
- [6] Li, H.P. (1993)
 Determination of geometrical structure on statistical experiments
Probability and Statistics, Rencontres Franco-Chinoises en Probabilités et Statistiques: (ed. A. Badrikian, P-A Meyer, J-A Yan); p. 132-148, World Scientific
- [7] Pfanzagl, J. and Wefelmeyer, W. (1982)
 Contributions to a General Asymptotic Statistical Theory
Lecture Notes in Statistics, 13, Springer-Verlag, New York-Berlin-Heidelberg.
- [8] Pfanzagl, J. and Wefelmeyer, W. (1985)
 Asymptotic Expansions for General Statistical Models
Lecture Notes in Statistics, 31, Springer-Verlag, New York-Berlin-Heidelberg.
- [9] Pollard, D. (1984)
 Convergence of Stochastic Processes
Springer series in statistics, Springer-Verlag, New York-Berlin-Heidelberg.
- [10] Strasser, H. (1985) Mathematical Theory of Statistics
De Gruyter studies in mathematics, De Gruyter, Berlin

Han-Ping LI, Institut de Recherche Mathématique Avancée,
 Université Louis Pasteur et C.N.R.S.,

7 rue René-Descartes, F-67084 Strasbourg Cedex, France
E-mail: lihp@math.u-strasbg.fr