Multivariate quadratic natural exponential families

Muriel Casalis and Gérard Letac*

These lectures are devoted to the richest of the natural exponential families in $\mathbb{R}^n$, namely those with a quadratic variance function. Lecture one gives mainly definitions and a few easy results. Lectures 2 and 3 are for the simple quadratic NEF, their properties and characterizations, lectures 4 and 5 study Wishart distributions from the point of view of NEF.

1 Natural exponential families

The frame of these lectures is the study of some statistical exponential families concentrated on a finite dimensional real linear space E. Although in practical situations, $E = \mathbb{R}^n$, it is better to be not linked to a particular basis if we do not need it. This gives us the freedom to choose in due time a basis adapted to the geometry of the considered problem, and to avoid tedious and useless changes of basis. However, sometimes the price to pay for this free coordinate approach is too high, specially in higher derivations questions, and we shall use $\mathbb{R}^n$ in these circumstances. Similarly, we do not assume in general that a particular Euclidean structure on E is given: there are many advantages to see clearly what belongs to E and what belongs to E^*, i.e. the dual of E or the space of the linear forms on E. If θ is such a linear form, we denote its action on the vector x of E by $<\theta, x>$, rather than by $\theta(x)$, and the map $(\theta, x) \mapsto <\theta, x>$ from $E^* \times E$ to $\mathbb{R}$ is called the canonical bilinear map. The linear space of linear operators $a : E^* \to E$ such that for all α and β in E^*, one has $<\alpha, a(\beta)> = <\beta, a(\alpha)>$ is the space of symmetric operators from E^* to E and is denoted by $L_s(E^*, E)$. This space is isomorphic to the space of the quadratic forms on E^* but is more easy to deal with.

Now, given a positive measure μ on E, its Laplace transform is the function defined on E^* and valued in $(0, +\infty]$ defined by

$$L_\mu(\theta) = \int_E e^{<\theta, x>} \mu(dx).$$

Let us insist on the fact that the measure μ is not necessarily bounded. We have now the simple statement:

Proposition 1.1 : Consider a positive measure μ on E. Then the set

$$D(\mu) = \{\theta \in E^*; L_\mu(\theta) < \infty\}$$

*Laboratoire de Statistique et Probabilités, Université Paul Sabatier, 31062 Toulouse, France

is convex, the cumulant function $k_\mu = \log L_\mu$ is convex on $D(\mu)$, and is strictly convex if and only if μ is not concentrated on some affine subspace of E.

Proof : Let θ and θ' be two distinct elements of $D(\mu)$. For λ in $(0,1)$, define $p = 1/\lambda$ and $q = 1/(1-\lambda)$, and apply the Hölder's inequality to the measured space (E, μ), to the pair (p, q) and to the functions f and g defined by

$$f(x) = e^{\lambda<\theta,x>}, \quad g(x) = e^{(1-\lambda)<\theta',x>}.$$

Clearly, these functions belong to L_p and L_q respectively. Thus their product is in L_1, and this gives the convexity of $D(\mu)$. The Hölder's inequality itself provides the convexity of the function k_μ. Finally, Hölder's inequality becomes an equality if and only if $|f|^p$ and $|g|^q$ are proportional μ almost everywhere. In our particular case, the convexity inequality :

$$k_\mu(\lambda\theta + (1-\lambda)\theta') \leq \lambda k_\mu(\theta) + (1-\lambda)k_\mu(\theta')$$

becomes an equality if and only if there exists a positive number C such that

$$e^{<\theta,x>} = Ce^{<\theta',x>}$$

μ almost everywhere. In other terms, equality occurs if and only if μ is concentrated on the affine hyperplane which is the set of x in E such that

$$< \theta - \theta', x >= \log C.$$

Definition 1.1 : $\mathcal{M}(E)$ is the set of positive measures μ on E such that the interior $\Theta(\mu)$ of $D(\mu)$ is not empty and such that μ is not concentrated on some affine hyperplane of E.

For instance, $\mathcal{M}(\mathbb{R})$ is the set of measures on $\mathbb{R}$ which are not concentrated on one point and for which there exist θ_0 in $\mathbb{R}$ and $a > 0$ such that for $|\theta - \theta_0| \leq a$ one has

$$\int_{-\infty}^{+\infty} e^{\theta x}\mu(dx) < \infty.$$

The measures $\delta_0 + \delta_1$ and $1_{(0,\infty)}(x)dx$ are in $\overline{M}(\mathbb{R})$, the measures dx, $dx/(1+x^2)$ or $e^{-\sqrt{|x|}}dx$ are not.

Definition 1.2 : Let μ be in $\mathcal{M}(E)$. Then the set of probabilities on E defined for θ in $\Theta(\mu)$ by

$$P(\theta, \mu)(dx) = \exp(< \theta, x > -k_\mu(\theta))\mu(dx) \tag{1}$$

is called the natural exponential family (NEF) generated by μ. We denote it by $F(\mu)$.

This definition calls for numerous comments:

1) $F(\mu) = F(\mu')$ occurs if and only if there exists (α, b) in $E^* \times \mathbb{R}$ such that $\mu'(dx) = e^{<\alpha,x>+b}\mu(dx)$: this is an easy exercise to prove this. Note that in particular a NEF is generated by any of its elements. It has other generators, which are often easier to use, simply because there are ...less partial. For instance, in $\mathbb{R}$, $1_{(0,\infty}(x)dx$ is a good choice to generate the NEF of exponential distributions

$$F = \{e^{-x/\sigma}1_{(0,\infty}(x)\frac{dx}{\sigma} \ ; \ \sigma > 0\},$$

while choosing $\mu(dx) = 2e^{-2x}1_{(0,\infty}(x)\,dx$ to generate F seems artificial.

2) We separate from common usage by declaring that elements of $D(\mu)$ which are not in $\Theta(\mu)$ do not provide elements of the NEF $F(\mu)$. For instance, consider for $p > 0$

$$\mu(dx) = \frac{p}{\sqrt(2\pi)} x^{-3/2} e^{-p^2/2x} 1_{(0,\infty}(x) dx \tag{2}$$

Here, this measure is a so called stable distribution of type 1/2. Then clearly $D(\mu) = (-\infty, 0]$. By differentiation under the integral sign, one can prove that for $\theta \leq 0$, one has

$$L_\mu(\theta) = e^{-p\sqrt(-2\theta)}$$

and the NEF $F(\mu)$ is a family of inverse Gaussian distributions. However, the stable distribution μ itself, which corresponds to the limiting case $\theta = 0$, according to our definition, does not belong to the family. We do this to have clearer statements of the theorems: statistical literature is sometimes sloppy with this question.

3) One can get the feeling that our definition of exponential families, while restricting to the NEF, is very narrow. Our claim is that is the one which is useful, and that all other kinds of exponential families are actually relying on it. To be clear, let us define what we call a general exponential family (GEF). They are families of probabilities (or "models"), living not on a linear space E, but on a measurable space Ω and given by the data of two objects : first a positive measure ν on Ω, and second a measurable map t from Ω to a finite dimensional real linear space E such that the image μ in E of the measure ν on Ω by the map t is an element of $\mathcal{M}(E)$, as in Def.1. With these notations and hypothesis, the GEF generated by the pair (ν, t) is the set of probabilities given for θ in $\Theta(\mu)$ by

$$P(\theta, \nu, t)(d\omega) = \exp(<\theta, t(\omega)> - k_\mu(\theta))\nu(d\omega) \tag{3}$$

Let also call $F(\mu)$ the NEF associated to this GEF. As we shall see, all information which is obtained on the GEF about the estimation of the unknown parameter θ will be obtained through the associated NEF build with the so called "sufficient statistic" t.

The most famous example of a GEF is the set of Gaussian distributions on $\mathbb{R}$:

$$N_{m,v}(dz) = \frac{1}{\sqrt(2\pi v)} e^{-(z-m)^2/2}\, dz. \tag{4}$$

where both the mean m and the variance v are unknown. Here the abstract space Ω is $\mathbb{R}$, the measure ν can be taken as $dz/\sqrt(2\pi)$ and t maps $\mathbb{R}$ to $E = \mathbb{R}^2$ by $z \mapsto x = (z, z^2/2)$. Hence the measure μ on $\mathbb{R}^2$ has not a density with respect to the Lebesgue measure of $\mathbb{R}^2$, but is singular, being concentrated on the parabola $y = z^2/2$.

4) There are also in the literature exponential families which appear to be still more general than the GEF: they simply replace θ in (1.3) by a fonction f of an other parameter, say α, valued in some other manifold F. If F has the dimension of E and if f is a diffeomorphism, we have an artificial generality, which does not bring anything useful and new (this reparametrization of the GEF may however come from practical or historical reasons). If the dimension of F is smaller than the dimension of E, we are in the different

world of "curved exponential families" where the methods of the differential geometry have to be used, the problems are harder, and cannot be solved with the sole algebraic tools that we are going to consider here.

5) One frequently meets in the statistical literature the statement : "This distribution belongs to THE exponential family". We think that this is misleading: a distribution should mean a definite probability measure on a space, not a whole model on this space, as intended in the sentence. Confusing a measure and a family of measures implies a confusion at the upper level between a model and a family of models. Finally, having a probability measure P –say, on the line– which belongs to some NEF is not a rare event: it suffices that P belongs to $\mathcal{M}(\mathbb{R})$.

Before studying NEF, let us recall some properties of the Laplace transforms. For convenience, we take $E = \mathbb{R}^n$, for using partial derivatives instead of differentials. Denote by $\mathbb{N}$ the set of non negative integers. If $a = (a_1, ..., a_n)$ belongs to $\mathbb{N}^n$ and if $x = (x_1, ..., x_n)$ is in $\mathbb{R}^n$, , we denote

$$a! = a_1!...a_n!, \quad x^a = x_1^{a_1}...x_n^{a_n},$$

and

$$(\frac{\partial}{\partial x})^a = (\frac{\partial}{\partial x_1})^{a_1}...(\frac{\partial}{\partial x_n})^{a_n}.$$

Proposition 1.2 : Let μ be in $\mathcal{M}(\mathbb{R}^n)$ and θ be in $\Theta(\mu)$. Then L_μ is analytic on $\Theta(\mu) + i\mathbb{R}^n$ and, for a in $\mathbb{N}^n$, one has

$$(\frac{\partial}{\partial \theta})^a L_\mu(\theta) = \int x^a e^{<\theta,x>} \mu(dx). \tag{5}$$

Furthermore, for i and j in $\{1, ..., n\}$

$$\frac{\partial}{\partial \theta_i} k_\mu(\theta) = \int x_i P(\theta, \mu)(dx) = m_i, \tag{6}$$

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} k_\mu(\theta) = \int (x_i - m_i)(x_j - m_j) P(\theta, \mu)(dx). \tag{7}$$

Proof : Without loss of generality we may assume that $0 \in \Theta(\mu)$ and prove the analyticity in 0. It is shown by the dominated convergence theorem, by using the inequality : for $|\theta_i| \leq a$

$$e^{<\theta,x>} \leq 2^n \cosh |ax_1| ... \cosh |ax_n|$$

This also gives (1.5). Equalities (1.6) and (1.7) are easy to obtain.

A consequence is that the Laplace transform characterizes the measure :

Corollary 1.3 : Let μ and μ' be in $\mathcal{M}(\mathbb{R}^n)$ such that $I = \Theta(\mu) \cap \Theta(\mu')$ is not empty and such that $L_\mu = L_{\mu'}$ on I. Then $\mu = \mu'$.

Proof : Let us fix θ_0 in I. Since L_μ and $L_{\mu'}$ are analytic on $I + i\mathbb{R}^n$ and coincide on the open set I, they are equal on the whole strip $I + i\mathbb{R}^n$, in particular on $\theta_0 + i\mathbb{R}^n$. Thus $P(\theta_0, \mu)$ and $P(\theta_0, \mu')$ have the same Fourier transform. From the theorem of uniqueness of Fourier transforms, the result is proved.

Note also that Prop.1.2 implies that $P(\theta,\mu)$ has moments af all orders if θ is in $\Theta(\mu)$. This would not be necessarily true for a θ in $D(\mu)\setminus\Theta(\mu)$. We now reformulate the equalities (1.6) and (1.7) in a coordinate free context. For μ in $\overline{M}(E)$ and θ in $\Theta(\mu)$, and for α and β in E^* :

$$k'_\mu(\theta)=\int_E xP(\theta,\mu)(dx)=m, \tag{8}$$

$$k''_\mu(\theta)(\alpha,\beta)=\int_E <\alpha,x-m><\beta,x-m>P(\theta,\mu)(dx). \tag{9}$$

In (1.8) and (1.9), k'_μ and k''_μ denote the first and second differentials of the cumulant transform k_μ. Remark that k'_μ must be a linear form on E^* thus an element of E, and that k''_μ belongs to $L_s(E^*,E)$. They are respectively the mean and the covariance of the probability $P(\theta,\mu)$.

2 The parametrization by the mean, and the variance function

We now exploit the equality (1.8). For this, we introduce the following set:

Definition 2.1 : Let μ be in $\mathcal{M}(E)$. Then the subset $M_F=k'_\mu(\Theta(\mu))$ of E is called the domain of the means of the NEF $F=F(\mu)$.

It is easy to see that M_F depends only on F itself, and not on a particular μ used to generate it. Its name comes obviously from (1.8). We are going to use M_F as a new parametrization set of the NEF F in what follows :

Proposition 2.1 : Let μ be in $\mathcal{M}(E)$, θ and θ' be in $\Theta(\mu)$ such that $k'_\mu(\theta)=k'_\mu(\theta')$. Then $\theta=\theta'$, and k'_μ is an analytic diffeomorphism from $\Theta(\mu)$ onto $M_{F(\mu)}$.

Proof : If θ and θ' are distinct, the function on $[0,1]$ defined by $t\mapsto k_\mu(\theta+t(\theta'-\theta))$ is strictly convex, by Prop.1.1 , and its derivative

$$t\mapsto<\theta'-\theta,k'_\mu(\theta+t(\theta'-\theta))>$$

is strictly increasing. Watching it on 0 and 1 provides a contradiction. The fact that k'_μ is an analytic diffeomorphism comes from the implicit function theorem, suitably stated to include analyticity. It is applicable there, since the differential $k''_\mu(\theta)$ is never singular.

We denote by ψ_μ the inverse of k'_μ from M_F onto $\Theta(\mu)$. Thus the map

$$m\mapsto P(\psi_\mu(m),\mu)=P(m,F) \tag{1}$$

is a new parametrization of the NEF F by its domain of the means. It is easily seen that $P(m,F)$ does not depend on μ. In some sense, it is more natural than the θ parametrization, which depends on a somewhat arbitrary choice of a generating measure μ. However, since it depends on the implicit function theorem for the calculation of ψ_μ, usually its explicit form is more involved, and we use it rather as an important theoretical tool.

We make now some remarks about M_F. Since it is the image of the convex open set $\Theta(\mu)$ by a diffeomorphism, it is an open set which is homeomorphic to E. Denote by $C(F)$ the interior of the convex hull of the support of the elements of F (it is the same for each of them). Then (1.8) shows that $M_F \subset C(F)$. Actually, in most of the practical cases, these two sets are equal, but not always. There are examples in $\mathbb{R}^2$ where M_F is not even a convex set (In $\mathbb{R}$, M_F is always an open interval, since it is homeomorphic to $\mathbb{R}$).

Definition 2.2 : A NEF F on E is said to be steep if $M_F = C(F)$.

The following delicate theorem is due to Ole Barndorff-Nielsen, we are not going to prove it and it still lacks of a simple proof. The second part explains why we use the word "steep".

Theorem 2.2 : Let μ be in $\mathcal{M}(E)$ and $F = F(\mu)$.Then

1) F is steep if $D(\mu)$ is open.

2) F is steep if and only if for all θ in $D(\mu) \setminus \Theta(\mu)$ and θ' in $\Theta(\mu)$, then

$$\lim_{t \to 0} < \theta - \theta', k'_\mu(\theta + t(\theta' - \theta)) >= \infty$$

In the statistical literature, one frequently calls "regular" a NEF such that condition 1) of the theorem is fulfilled. Of course, there are steep families which are not regular: the NEF of inverse Gaussian distributions generated by (1.2) provides an example.

We now turn to the definition of the variance function.

Definition 2.3 : Let F be a NEF on E, and for m in M_F, denote by $V_F(m)$ the covariance operator of $P(m, F)$. Then the map $m \mapsto V_F(m)$ from M_F to $L_s(E^*, E)$ is called the variance function of F.

The following proposition provides different ways to compute it and shows that the variance function characterizes the NEF.

Proposition 2.3 : Let $F = F(\mu)$ be a NEF on E. Then

$$V_F(m) = (\psi'_\mu(m))^{-1}, \tag{2}$$

$$k''_\mu(\theta) = V_F(k'_\mu(\theta)). \tag{3}$$

Furthermore, if F and F' are two NEF such that V_F and V'_F coincide on a non void open subset J of $M_F \cap M_{F'}$, then $F = F'$.

Proof : From the definition of $\psi_\mu(m)$, we write $k'_\mu(\psi_\mu(m)) = m$. Taking the differential of both members gives

$$k''_\mu(\psi_\mu(m)) \circ \psi'_\mu(m) = id_E$$

which gives (2.2). Identity (2.3) is obtained by the definition of V_F and (1.9).

Finally, take m_0 in J and μ and μ' generating F and F' respectively. Let $\theta_0 = \psi_\mu(m_0)$ and $\theta'_0 = \psi'_\mu(m_0)$. We choose arbitrarily α in E^* and consider the function

$$y(t) = k'_\mu(\theta_0 + t\alpha),$$

as defined in a neighborhood of 0 in $\mathbb{R}$ and valued in E. Then, from (2.3), it satisfies the ordinary differential autonomous equation

$$y'(t) = V_F(y(t))(\alpha)$$

with initial condition $y(0) = m_0$. From uniqueness in the Cauchy Lipschitz theorem, we deduce that

$$k'_\mu(\theta_0 + t\alpha) = k'_{\mu'}(\theta'_0 + t\alpha)$$

in a neighborhood of 0 for t. Thus there exists a constant b such that $e^b L_\mu(\theta_0 + t\alpha) = L_{\mu'}(\theta'_0 + t\alpha)$. A standard reasoning based on the analyticity of Laplace transforms then shows that

$$\mu'(dx) = e^{<\theta'_0-\theta_0,x>+b}\mu(dx),$$

and this shows that $F = F'$. The proof of the Prop.2.3 is now complete.

We conclude this section by recalling how to estimate by the maximum likelihood the parameter of a NEF :

Proposition 2.3 : Let $F = F(\mu)$ be a NEF on E. Let $x_1, ..., x_n$ be a sample of observations of the distribution $P(\theta, \mu) = P(m, F)$. Then if $\overline{x}_n = (x_1 + ... + x_n)/n$ belongs to M_F, the maximum likelihood estimators of m and θ are $\overline{x}_n$ and $\psi_\mu(\overline{x}_n)$.

Proof : From Prop.1.1, the function on $\Theta(\mu)$:

$$\theta \mapsto <\theta, x_1 + ... + x_n> -nk_\mu(\theta)$$

is a strictly concave function whose differential cancels for $\theta = \psi_\mu(\overline{x}_n)$. Thus the maximum likelihood exists and is reached there.

Comments: 1) Since M_F is an open set and since by the law of large numbers $(x_1 + ... + x_n)/n$ converges almost surely towards an element m of M_F, there exists almost surely an n such that $(x_1 + ... + x_n)/n$ belongs to M_F, thus the necessary condition of the proposition is not that demanding.

2) Suppose that we have to solve the same problem with a GEF, as described in (1.3). Then we apply Prop.2.3 to the associated exponential family, getting the estimator of θ equals to $\psi_\mu(\overline{x}_n)$, with $x_j = t(\omega_j)$. For instance, if we apply this to the Gaussian family (1.4), where both the variance and the mean are unknown, the domain of the means of the associated NEF is the interior of the parabola $y = z^2/2$, the family is steep, $\overline{x}_n$ is in the interior almost surely for $n > 1$, and the estimators of the mean and the variance are the classical ones : $\overline{Z}_n = (Z_1 + ... + Z_n)/n$ and

$$S_n^2 = \frac{1}{n}((Z_1 - \overline{Z}_n)^2 + ... + (Z_n - \overline{Z}_n)^2).$$

3 The Jorgensen set of a NEF and the conditioning

Definition 3.1 : Let μ be in $\mathcal{M}(E)$. The set of $p > 0$ such that there exists μ_p in $\mathcal{M}(E)$ with $\Theta(\mu_p) = \Theta(\mu)$ and $L_{\mu_p} = (L_\mu)^p$ is denoted by $\Lambda(\mu)$ and is called the Jorgensen set of μ.

It is clear that $\Lambda(\mu) \cup \{0\}$ is an closed additive semigroup. It is also clear that $\Lambda(\mu) = \Lambda(\mu')$ if $F(\mu) = F(\mu')$. For this reason, it makes sense to talk about the Jorgensen set $\Lambda(F)$ of the NEF F. One has $\Lambda(F) = (0, \infty)$ if and only if the distributions in F are infinitely divisible. Considering p in $\Lambda(\mu)$, the NEF $F_p = F(\mu_p)$ does not depend on the particular μ chosen to generate F. B.Jorgensen calls the model

$$\bigcup_{p\in\Lambda(F)} F_p$$

an exponential dispersion model. The domain of the means of F_p is pM_F, and the variance function is $pV_F(m/p)$.

In one dimension, finding the Jorgensen set is usually easy. However, in higher dimensions, it can be a harder task, as we shall see later with the Wishart distributions.

Finally, the last result of this introduction to NEF is the following:

Proposition 3.1 : Let μ be in $\mathcal{M}(E)$, $F = F(\mu)$, $p_0, ..., p_n$ be in $\Lambda(F)$ and θ be in $\Theta(\mu)$. Let $X_0, ..., X_n$ be independent random variables in E with respective distributions

$$P(\theta, \mu_0), ..., P(\theta, \mu_n),$$

$S = X_0 + ... + X_n$ and $p = p_0 + ... + p_n$. Then the distribution of S is $P(\theta, \mu_p)$ and the conditional distribution K_s of $(X_1, ..., X_n)$, knowing that $S = s$, does not depend on θ.

Comments : This conditional distribution is sometimes called the Dirichlet distribution of parameters $(p_0, ..., p_n; s)$ associated to the NEF F. In the statement of the proposition, we have written for simplicity μ_j for μ with index p_j.

Proof : Let $G : E \to \mathbb{R}$ and $F : E^n \to \mathbb{R}$ be any functions with compact support. Computing $\mathbb{E}(G(S)F(X_1, ..., X_n))$ in two ways, we get that

$$K_s(dx_1, ..., dx_n)\mu_p(ds) = (\delta_{x_1+...+x_n} * \mu_0)(ds)\mu_1(dx_1)...\mu_n(dx_n),$$

and this gives the proof.

4 Morris class of Natural Exponential Families on $\mathbb{R}$

In the previous lecture, G. Letac has introduced the variance function of a NEF as an important tool to study its statistical properties since it characterizes entirely the NEF. The first one, Morris [16] has isolated the class of NEF with a quadratic variance function (QVF), ie, where $V(m)$ is, at most, a quadratic polynomial function of the mean. Up to affinities and powers, there are only 6 such families. Let us briefly mention what we mean by affinities and powers: Let $F = F(\mu)$ be a NEF on E, $\varphi : x \mapsto ax + b$ be an affinity on E with $a \in GL(E), b \in E$, then $\varphi(F) = \{\varphi_* P(m, F); m \in M_F\}$ is still a NEF generated by $\varphi_*(\mu)$ and such that:

$$M_{\varphi(F)} = \varphi(M_F) \qquad V_{\varphi(F)}(m) = aV_F(\varphi^{-1}m)\ {}^t a$$
$$V_{\varphi(F)}(m) = a^2 V_F(\tfrac{m-b}{a})\ for\ E = \mathbb{R} \qquad (1)$$

Let $\Lambda_F = \left\{\lambda > 0; \exists\ \mu_\lambda \in \mathcal{M}(E); L_{\mu_\lambda}(\theta) = (L_\mu(\theta))^\lambda\right\}$ be the Jorgensen set of F. Recall that if $\lambda \in \Lambda_F, F_\lambda = F(\mu_\lambda)$ is characterized by

$$M_{F_\lambda} = \lambda M_F \qquad V_{F_\lambda}(m) = \lambda V_F(\frac{m}{\lambda}) \qquad (2)$$

One will say that two NEF F and $\tilde{F}$ are of the same type if there exist an affinity φ and a power λ in Λ_F such that $\tilde{F} = \varphi(F_\lambda)$. We have listed the 6 types of QVF-NEF in the following table. Let us briefly indicate how we get these types: let be $V_F(m) = am^2 + bm + c$. Then:

Either $a = b = 0, V_F(m) = c > 0$ and from (2), F is a Normal NEF with variance $\sigma^2 = c$.
Either $a = 0, b \neq 0$, then F is a affinity of the Poisson family (with $\varphi : x \mapsto bx - \frac{c}{b^2}$)
Or $a \neq 0$ and V_F has two, one or zero roots in $I\!R$; up to an affinity or a power (see (1) and (2)), V_F can be written:

- two roots: $V_F = m \pm m^2$ ($+$ for $a > 0$ and - for $a < 0$); hence F is of type Binomial or Negative-Binomial
- one double root $V_F = m^2$ (since $V_F > 0$) and F is of type Gamma
- no root $V_F = m^2 + 1$ (since $V_F > 0$) and F is of type Hyperbolic cosine.

The six types of quadratic NEF

Name	M_F	V_F	$\mu(dx)$	$k_\mu(\theta)$
Normal	$I\!R$	1	$\exp\left(-\frac{x^2}{2}\right)\frac{dx}{\sqrt{2\pi}}$	$\frac{\theta^2}{2}$
Poisson	$(0;+\infty)$	m	$\sum_{n=0}^{+\infty}\frac{\delta_n}{n!}$	$\exp\theta$
Binomial	$(0;1)$	$m-m^2$	$\delta_0+\delta_1$	$log(1+e^\theta)$
N-Binomial	$(0;+\infty)$	$m+m^2$	$\sum_{n=0}^{+\infty}\delta_n$	$-log(1-e^\theta)$
Gamma	$(0;+\infty)$	m^2	$1_{(0;+\infty)}dx$	$-log(-\theta)$
Hyperbolic cosinus	$I\!R$	$1+m^2$	$\frac{1}{2ch\left(\frac{\pi x}{2}\right)}dx$	$-log(cos\theta)$

$P(m,F)$	Λ_F	M_{F_λ}	V_{F_λ}	$P(m,F_\lambda)$
$\exp - \dfrac{(x-m)^2}{2}\dfrac{dx}{\sqrt{2\pi}}$	$(0;+\infty)$	$I\!R$	$\lambda=\sigma^2$	$\dfrac{\exp - \frac{(x-m)^2}{2\sigma^2}}{\sqrt{2\pi\sigma^2}}dx$
$\sum_{n=0}^{+\infty}\dfrac{m^n e^{-m}}{n!}\delta_n(dx)$	$(0;+\infty)$	$(0;+\infty)$	m	$P(m,F)$
$m\delta_0+(1-m)\delta_1$	$I\!N^*$ $(\lambda=N)$	$(0,N)$	$m-\frac{1}{N}m^2$	$\sum_{k=0}^{N} C_N^k\left(\dfrac{m}{N}\right)^k\left(1-\dfrac{m}{N}\right)^{N-k}\delta_k(dx)$
$p\sum_{n=0}^{+\infty} q^n\delta_n(dx)$ $q=\dfrac{m}{1+m}$	$(0;+\infty)$	$(0;+\infty)$	$\dfrac{1}{\lambda}m^2+m$	$\sum_{n=0}^{+\infty}\dfrac{\Gamma(n+\lambda)}{\Gamma(n)n!}\left(\frac{m}{m+\lambda}\right)^n\left(\frac{\lambda}{m+\lambda}\right)^\lambda\delta_n$
$\dfrac{1}{m}e^{-\frac{x}{m}}1_{(0;+\infty)}(x)dx$	$(0;+\infty)$	$(0;+\infty)$	$\dfrac{1}{\lambda}m^2$	$\dfrac{e^{-\frac{x}{m}}}{\Gamma(\lambda)m^\lambda}x^{\lambda-1}1_{(0;+\infty)}(x)dx$
$cos\theta e^{\theta x}\mu(dx)$ $m=tan\theta$	$(0;+\infty)$	$I\!R$	$\lambda+\frac{1}{\lambda}m^2$	(cf Morris [16])

In his paper Morris lists some properties satisfied by the QVF-NEF and presents them in a unified way. They concern infinite divisibility, cumulants, orthogonal polynomials, large deviations, limits in distribution ([16]) as well as some Bayesian estimation and regression structure ([17]).
Actually these 6 types had already been isolated in the statistical literature by different ways. Let us cite:
Meixner[15]: for exponential generating function
Lancaster[12]: idem
Laha & Lukacs[11]: for quadratic regression
Shanbhag[19]: diagonality of Bhattacharya matrices
Feinsilver[7]: orthogonal polynomials

We focus this talk on the characterizations using orthogonal polynomials, that is Meixner's and Feinsilver's ones. We use the terminology of NEF to present them.

5 Meixner and Feinsilver characterizations of the Morris class

5.1 The characterizations

We begin with Meixner's result as it was exposed by A. Koudou in a unpublished work.

Theorem 5.1 *Let F be a NEF on $\mathbb{R}$, μ be a probability of F with mean 0. If (Q_n) is a sequence of $\mu-$orthogonal polynomials such that Q_n is monic of degree n (ie the coefficient of x^n is 1), then the following statements are equivalent:*

i) there exist an open set O of $\mathbb{R}$ and $a, b : O \to \mathbb{R}$ two analytic functions such that for any z in O

$$\sum \frac{z^n}{n!} Q_n = \exp\{a(z)x + b(z)\}$$

ii) F is quadratic.
In this case $a(z) = \psi_\mu(\alpha z)$ and $b(z) = -k_\mu(a(z))$ for some real number α.

The sequence (Q_n) is said to have an exponential generating function.

Feinsilver's characterization precises one particular way to get such a sequence:

Theorem 5.2 *Let F be a NEF on $\mathbb{R}$, μ be a probability of F with mean 0 and let f_μ be the density of $P(m, F)$ with respect to μ, ie for $x \in \mathbb{R}$ and $m \in M_F$:*

$$f_\mu(x, m) = \exp\{\psi_\mu(m)x - k_\mu(\psi_\mu(m))\}$$

Then, if $P_n(x) = \frac{\partial^n}{\partial m^n} f_\mu(x, m)_{|m=0}$, P_n is a polynomial in x of degree n.
Furthermore (P_n) are μ-orthogonal if and only if F is quadratic.

In fact the hypothesis "μ is a probability of mean 0" is not necessary. We have the same result with any probability with mean m_0 in M_F. Indeed if we consider $\varphi : x \mapsto x - m_0$, $\varphi(\mu)$ has mean 0 and

$$\begin{aligned} f_{\varphi(\mu)}(x,m) &= \exp\{\psi_{\varphi(\mu)}(m)x - k_{\varphi(\mu)}\psi_{\varphi(\mu)}(m)\} \\ &= \exp\{\psi_\mu(m_0+m)(x-m_0) - k_\mu\psi_\mu(m_0+m)\} \\ &= f_\mu(x-m_0, m_0+m) \end{aligned}$$

Hence

$$\begin{aligned} f^{(n)}_{\varphi(\mu)}(x,0) &= f^{(n)}_\mu(x-m_0,m_0) \\ P_{\varphi(\mu),n}(x) &= P_{\mu,n}(x-m_0) \end{aligned}$$

Consequently $(P_{\varphi(\mu),n})$ are $\varphi(\mu)$-orthogonal if and only if $(P_{\mu,n})$ are μ-orthogonal and the conclusion of the previous Theorem is the same. We suppose $m_0 = 0$ yet in the proofs.

Before giving their main properties , let us recall what polynomials are involved for the six QVF-NEF:

Classical orthogonal polynomials on $I\!R$:

Name	**Denoted**	**First terms**	**Induction relations**
Hermite	H_n	$H_0 = 1$ $H_1(x) = 2x$	$2xH_n(x) = H_{n+1}(x) + 2nH_{n-1}(x)$
Charlier	C_n^α $(\alpha > 0)$	$C_0^\alpha = 1$ $C_1^\alpha(x) = \dfrac{\alpha - x}{\alpha}$	$xC_n^\alpha(x) = -\alpha C_{n+1}^\alpha(x) + (n+\alpha)C_n^\alpha(x)$ $-nC_{n-1}^\alpha(x)$
Laguerre	L_n^α $(\alpha > -1)$	$L_0^\alpha = 1$ $L_1^\alpha(x) = -x+\alpha+1$	$-xL_n^\alpha(x) = (n+1)L_{n+1}^\alpha(x)$ $-(2n+\alpha+1)L_n^\alpha(x)$ $+(n+\alpha)L_{n-1}^\alpha(x)$
Krawchouk	$K_n^{p,N}$ $(0 < p < 1)$ $(N \in I\!N)$	$K_0^{p,N} = 1$ $K_1^{p,N}(x) = x - pN$	$xK_n^{p,N}(x) = (n+1)K_{n+1}^{p,N}(x)$ $+(n + p(N-2n))K_n^{p,N}(x)$ $+p(1-p)(N-n+1)K_{n-1}^{p,N}(x)$
Meixner (first type)	$M_n^{c,\beta}$ $(c \neq 1)$ $(\beta \in I\!R)$	$M_0^{c,\beta} = 1$ $M_1^{c,\beta}(x) = x - \frac{\beta c}{1-c}$	$xM_n^{c,\beta}(x) = M_{n+1}^{c,\beta}(x)$ $+\dfrac{(1+c)n+\beta c}{1-c}M_n^{\beta,c}(x)$ $+\dfrac{cn(n+\beta-1)^2}{1-c}M_{n-1}^{\beta,c}(x)$
Pollaczek	P_n^λ $(\lambda \in I\!R^*)$	$P_0^\lambda = 1,$ $P_1^\lambda(x) = \frac{x}{\lambda}$	$xP_n^\lambda(x) = (n+\lambda)P_{n+1}^\lambda(x) + nP_{n-1}^\lambda(x)$

Orthogonal polynomials for the Morris class:

Type	m_0	$P(m_0,F)(dx)$	$P_n(x)$
Gaussian $V_F(m)=1$	0	$\exp(-\frac{x^2}{2})\frac{dx}{\sqrt{2\pi}}$	$(\frac{1}{\sqrt{2}})^n H_n(\sqrt{2}x)$
Poisson $V_F(m)=(m-m_0)+m_0$	1	$\sum_{n\geq 0}\dfrac{1}{n!}\exp(-1)\delta_n$	$(-1)^n C_n^1(x)$
Binomial $V_F(m)=-(m-m_0)^2+(1-2m_0)(m-m_0)+m_0-m_0^2$	$\frac{1}{2}$	$\frac{1}{2}(\delta_0+\delta_1)$	$n!(\frac{1}{4})^n K_n^{\frac{1}{2},1}(x)$
Negative-binomial $V_F(m)=(m-m_0)^2+(2m_0+1)(m-m_0)+m_0^2+m_0$	1	$\sum_{n\geq 0}(\dfrac{1}{2})^{x+1}\delta_n$	$2^{-\frac{3n}{2}}M_n^{\frac{1}{2},1}(x-1)$
Gamma $V_F(m)=(m-m_0)^2+2m_0(m-m_0)+m_0^2$	1	$\exp(-x)1\!\!1_{(0,+\infty)}(dx)$	$(-1)^n L_n^0(x+1)$
Hyperbolic	0	$\dfrac{dx}{2ch(\dfrac{\pi x}{2})}$	$n!P_n^1(x)$

5.2 Proofs

Let us begin with a preliminary result:

Lemma 5.1 *Let μ be a probability on $I\!R$ such that $0 \in \Theta(\mu)$ and $\int x\mu(dx) = 0$. Denote by (Q_n) the sequence of orthogonal polynomials with respect to μ such that Q_n is monic of degree n. Let be $r = sup\{\alpha; (-\alpha, \alpha) \subset \Theta(\mu)\}$. Then the entire serie $\sum \frac{z^n}{n!} Q_n(x)$ valued in $L^2(\mu)$ has a radius of convergence $\geq r/2$.*

Proof: We have to prove that for $|z| < r/2$:

$$lim_{N \to +\infty} \int \left(\sum_{n=0}^{N} \frac{z^n}{n!} Q_n(x) \right)^2 \mu(dx) = lim_{N \to +\infty} \sum_{n=0}^{N} \frac{z^{2n}}{(n!)^2} \int Q_n^2(x)\mu(dx) < +\infty$$

Denote $\int Q_n^2(x)\mu(dx) = b_n$ and $\int x^n \mu(dx) = \gamma_n$. Then since $x^n - Q_n(x)$ is of degree $\leq n$, Q_n and $x^n - Q_n$ are orthogonal, so
$\int x^{2n}\mu(dx) = \int (x^n - Q_n(x))^2 \mu(dx) + \int Q_n^2(x)\mu(dx)$ ie: $\gamma_{2n} \geq b_n$.
But $\int e^{zx}\mu(dx) = \sum \frac{z^n}{n!}\gamma_n$ if $|z| \leq \alpha < r$ from the analyticity of L_μ on $\Theta(\mu)$. Hence $\sum \frac{\alpha^{2n}}{(2n)!}\gamma_{2n} < +\infty$ if $\alpha < r$.
Now since $\frac{1}{(2n)!} \sim_{+\infty} \frac{\sqrt{\pi n}}{(n!)^2} \frac{1}{2^{2n}}$, we get $\sum \frac{(\alpha/2)^{2n}}{(n!)^2}\gamma_{2n} < +\infty$ if $\alpha < r$. Hence $\sum \frac{z^{2n}}{(n!)^2} b_n \leq \sum \frac{z^{2n}}{(n!)^2}\gamma_{2n}$ and converges if $|z| \leq \alpha/2 < r/2$.

Let us denote by

(1) $\sum \frac{z^n}{n!} Q_n(x) = \exp\{a(z)x + b(z)\}$

(1') $a(z) = \psi_\mu(\alpha z), b(z) = -k_\mu(a(z))$

(2) F is quadratic with $V_F(m) = am^2 + bm + c$

(3) $P_n(x) = \frac{\partial^{(n)} f_\mu}{\partial m^n}(x, 0)$ are orthogonal polynomials

(3') $xP_n = cP_{n+1} + nbP_n + \{an(n-1) + 1\}P_{n-1}$

To prove Meixner's and Feinsilver's equivalences, we can proceed like this:
First prove (3) $\Longrightarrow$ (2) $\Longrightarrow$ (3') $\Longrightarrow$ (3) (this is Feinsilver's result).
Then since $f_\mu(x, m) = \exp\{\psi_\mu(m)x - k_\mu(\psi_\mu(m))\} = \sum \frac{m^n}{n!} P_n(x)$ the implications (3) $\Longrightarrow$ (1) and (1) with (1') $\Longrightarrow$ (3) are obvious. It will only remain to prove (1) $\Longrightarrow$ (1') to conclude to Meixner's result.

Proof of Feinsilver's result

We begin by proving (3) $\Longrightarrow$ (2).

For small m and $x \in I\!R$, $\sum \frac{m^n}{n!} P_n(x) = \exp\{\psi_\mu(m)x - k_\mu(\psi_\mu(m))\}$. Since $\psi_\mu(0) = 0$, the series $\sum \frac{m^n}{n!} P_n(x)$ is valued in $L^2(\mu)$ and from the orthogonality of the polynomials (P_n), we get for small m and $\tilde{m}$

$$\begin{aligned}
\underbrace{\sum \frac{(m.\tilde{m})^n}{(n!)^2} \int P_n(x)^2 \mu(dx)}_{\exp g(m.\tilde{m})} &= \int \left(\sum \frac{m^n (\tilde{m})^{\tilde{n}}}{n!(\tilde{n})!} P_n(x) P_{\tilde{n}}(x) \right) \mu(dx) \\
&= \int \exp\left\{ (\psi_\mu(m) + \psi_\mu(\tilde{m}))x - k_\mu \psi_\mu(m) - k_\mu \psi_\mu(\tilde{m}) \right\} \mu(dx) \\
&= \exp\left\{ k_\mu\left(\psi_\mu(m) + \psi_\mu(\tilde{m})\right) - k_\mu\left(\psi_\mu(m)\right) - k_\mu\left(\psi_\mu(\tilde{m})\right) \right\}
\end{aligned}$$

We can derive the *log* of this expression with respect to $\tilde{m}$; we get :

$g'(m.\tilde{m})m = k'_\mu\left(\psi_\mu(m) + \psi_\mu(\tilde{m})\right)\psi'_\mu(\tilde{m}) - k'_\mu\left(\psi_\mu(\tilde{m})\right)\psi'_\mu(\tilde{m})$.

At $\tilde{m} = 0$, since $k'_\mu\left(\psi_\mu(\tilde{m})\right) = \tilde{m}$ and $k'_\mu\left(\psi_\mu(\tilde{m})\right)\psi'_\mu(\tilde{m}) = \tilde{m} V_F(\tilde{m})^{-1}$, we have :
$g'(0)m = \psi'_\mu(0) k'_\mu\left(\psi_\mu(m)\right) = V_F(0)^{-1} m$, hence $g'(0) = V_F(0)^{-1}$.

Deriving a second time : $g''(m.\tilde{m})m^2 =$
$k''_\mu\left(\psi_\mu(m) + \psi_\mu(\tilde{m})\right)\psi'_\mu(\tilde{m})^2 + k'_\mu\left(\psi_\mu(m) + \psi_\mu(\tilde{m})\right)\psi''_\mu(\tilde{m}) - V_F(\tilde{m}) - \tilde{m} V'_F(\tilde{m}) V_F(\tilde{m})^{-2}$

At $\tilde{m} = 0 : g''(0)m^2 = (\psi'_\mu(0))^2 V_F(m) + m\psi''_\mu(0) - V_F(0)^{-1}$
Hence $V_F(m)$ is quadratic on a open set around 0 and by extension on $I\!R$.

(2) $\Longrightarrow$ (3'):

We have $V_F(m) = am^2 + bm + c$. Since $\exp\{\psi_\mu(m)x - k_\mu \psi_\mu(m)\} = \sum \frac{m^n}{n!} P_n(x)$ for small m, we have $\exp(\theta x) = \sum \frac{(k'_\mu(\theta))^n}{n!} P_n(x) e^{k_\mu(\theta)}$.

Derive this equality with respect to θ:

$$\begin{aligned}
xe^{\theta x} &= \left(\sum n \frac{(k'_\mu(\theta))^{n-1} k''_\mu(\theta)}{n!} P_n(x) + \sum \frac{(k'_\mu(\theta))^{n+1}}{n!} P_n(x) \right) e^{k_\mu(\theta)} \\
&= \left(\sum \frac{(k'_\mu(\theta))^{n-1}}{(n-1)!} (a(k'_\mu(\theta))^2 + b(k'_\mu(\theta)) + c) P_n(x) + \sum \frac{(k'_\mu(\theta))^{n+1}}{(n)!} P_n(x) \right) e^{k_\mu(\theta)}
\end{aligned}$$

Hence looking at the coefficient of $(k'_\mu(\theta))^n = m^n$, we get

$xP_n(x) = cP_{n+1} + nbP_n + \{an(n-1) + 1\}P_{n-1}$

This is (3').

(3') $\Longrightarrow$ (3)

We here suppose that $P_1(x) = \frac{1}{c}x$ and $xP_n = cP_{n+1} + nbP_n + \{an(n-1) + 1\}P_{n-1}$
Then by induction on n, we can successively prove:

i) $\forall n \neq 0 \int P_n(x)\mu(dx) = 0$

ii) $x^q P_n = \beta_{q,n} P_{n-q} + \sum_{n-q+1 \leq k \leq n+q} \beta^k_{q,n} P_k(x)$ with $\beta_{q,n} = 0$ if $n < q$.

iii) $P_n(x) = \alpha_n x^n + \sum\limits_{q<n} \alpha_q x^q$

We then prove that $\int x^q P_n(x)\mu(dx) = 0$ if $q \neq n$ and consequently $\int P_n P_q \mu = 0$ if $n \neq q$.

Proof of Meixner's result

To conclude to Meixner's result from Feinsilver's one, we have seen that it is enough to prove (1) $\Longrightarrow$ (1') now.

So we suppose the polynomials Q_n monic, μ-orthogonal and with a generating exponential function. Then:

$$\begin{aligned}\int \sum \frac{z^n}{n!} Q_n(x)\mu(dx) &= \int \left(\sum \frac{z^n}{n!} Q_n(x)\right) Q_0(x)\mu(dx) &&= \int Q_0^2(x)\mu(dx) = 1\\ &= \int \exp\{a(z)x + b(z)\}\mu(dx) &&= \exp\{k_\mu(a(z)) + b(z)\}\end{aligned}$$

Hence $b(z) = -k_\mu(a(z))$.

Then $\int \left(\sum \frac{z^n}{n!} Q_n(x)\right) Q_1(x)\mu(dx) = z \int Q_1^2(x)\mu(dx)$ where $Q_1(x) = \alpha_1 x + \beta$. Since $\int Q_1(x)\mu(dx) = \int Q_1(x) Q_0(x)\mu(dx) = 0$, we get $\beta = 0$, hence $\int Q_1^2(x)\mu(dx) = \alpha_1^2 \int x^2 \mu(dx) = \alpha_1^2 V_F(0)$, so that :

$$\int \left(\sum \frac{z^n}{n!} Q_n(x)\right) Q_1(x)\mu(dx) = \alpha_1^2 z V_F(0).$$

But we also have:

$$\int \left(\sum \frac{z^n}{n!} Q_n(x)\right) Q_1(x)\mu(dx) = \int \alpha_1 x \exp\{a(z)x - k_\mu(a(z))\}\mu(dx) = \alpha_1 k'_\mu(a(z))$$

Hence $k'_\mu(a(z)) = \alpha_1 z V_F(0)$, or equivalently if $\alpha = \alpha_1 V_F(0)$, $a(z) = \psi_\mu(\alpha z)$. Finally:

$$\sum \frac{z^n}{n!} Q_n(x) = f_\mu(x, \alpha z) \text{ and } Q_n(x) = \alpha^n f_\mu^{(n)}(x, 0) = \alpha^n P_n(x).$$

6 Quadratic Natural Exponential Families on $I\!R^d$

The sequel is devoted to the generalization of the previous characterizations to $I\!R^d$. For this we first have to define what we mean by QVF-NEF.

Let F be a NEF on $I\!R^d$ with variance-function $V_F : M_F \to L_s(I\!R^d)$ (the space of symmetric endomorphisms of $I\!R^d$). F is said to be quadratic (or a QVF-NEF) if there exist

A: $\quad I\!R^d \times I\!R^d \to L_s(I\!R^d)$ bilinear

B: $\quad I\!R^d \to L_s(I\!R^d)$ linear

C $\quad \in L_s(I\!R^d)$

such that $V_F(m) = A(m,m) + B(m) + C$, i.e if a basis $e^* = (e_i)$ of $I\!R^d$ is specified,

$$V_F(m)_{|e^*} = \sum A_{ij} m_i m_j + \sum B_i m_i + C$$

with A_{ij}, B_i, C symmetric matrices and $m = \sum m_i e_i$.

The class of QVF-NEF on $I\!R^d$ is not entirely determined. Only different elements have been isolated ; in particular it can be proved that :

for A=0, F is an affinity of a Poisson-Gaussian family whose distributions are products of real Poisson or Gaussian distributions(see Letac [13]), see below. In that case F is reducible.

for B=C=0, i.e $V_F(m) = A(m,m)$, F is a Wishart family on a symmetric cone (see Casalis [2])

for $A(m,m) = am \otimes m : x \mapsto am < m, x >$, i.e $V_F(m) = am \otimes m + B(m) + C$, F is said to be a simple quadratic NEF.

In these two last cases, F is said irreducible in the sense that we cannot write it as a product of NEF on subspaces of $I\!R^d$.

The sub-class of simple QVF-NEF plays a particular role among the quadratic class. It firstly constitues an important tool in the determination of some cubic NEF on $I\!R^d$ which generalize the real cubic NEF (see Hassairi [9]). It also satisfies different statistical properties (as a simple UMVU estimator for V_F (see Casalis[3]), a property on Bayesian conjugate prior distributions families (see Consonni and Veronese [6])). In the next section we will see which role it plays in Meixner's and Feinsilver's characterizations.

This sub-class is completely described (see Casalis [5]):

Theorem 6.1 *There are exactly $2d+4$ types of such families on $I\!R^d$.*

They are described as follows:

a)The (d+1) Poisson-Gaussian types $(PG)_k,\ k = 0, .., d$

They are defined from the (d+1) types of NEF characterized by Letac [13] as the only types with an affine variance-function. Up to affinities and powers, there exists k in $\{0, ..d\}$ such that F is the family of the products of k Poisson distributions and $d-k$ normal distributions. Hence, F is characterized in a suitable basis e^* of $I\!R^d$ by:

$M_F = (0; +\infty)^k \otimes I\!R^{d-k}$

$V_F(m)_{|e^*} = diag(m_1, .., m_k, 1, .., 1)$

b) The (d+1) Negative multinomial-Gamma types $(NM-Ga)_k, k = 0, .., d$

We shall first introduce the well-known Negative multinomial distributions on $I\!R^d$ as distributions of a natural exponential family. We adopt the presentation of Letac[14]

Let $\{e_1, .., e_d\}$ denote the canonical basis of $I\!R^d$ and e_0 be the null vector. Then consider the measure $\nu_0(dx) = \sum_{i=1}^{d} \delta_{e_i}$ and for n in $I\!N$, $\nu_0^{\otimes n}$ the n^{th} convolution of ν_0 (with the convention $\nu_0^{\otimes 0} = \delta_{e_0}$). Now let be

$$\nu^{(d)} = \sum_{n=0}^{+\infty} \nu_0^{\otimes n} = (\delta_{e_0} - \nu_0)^{\otimes -1} \tag{3}$$

Clearly the Laplace transform of $\nu^{(d)}$ is given by

$$L_{\nu^{(d)}}(\theta) = (1 - \sum_{i=1}^{d} e^{\theta_i})^{-1} \text{ on } \ \Theta(\nu^{(d)}) = \{\theta \in I\!R^d; \sum_{i=1}^{d} e^{\theta_i} < 1\}$$

The family $F(\nu^{(d)})$ is the analogous of the real family of the geometric distributions. It is composed by the probabilities $P(m,F)$ defined on $\mathbb{N}^d$ by:
if $S = m_1 + .. + m_d$, then

$$P(m,F)(n_1e_1 + .. + n_de_d) = \frac{1}{1+S}\left(\begin{array}{c} n_1+..+n_d \\ n_1,...,n_d \end{array}\right)\left(\frac{m_1}{1+S}\right)^{n_1}..\left(\frac{m_d}{1+S}\right)^{n_d} \qquad (4)$$

The variance-function of $F(\nu^{(d)})$ is given on $M_{F(\nu^{(d)})} = \{m \in \mathbb{R}^d; \forall j \ m_j > 0\}$ by:

$$V_{F(\nu^{(d)})}(m) = m \otimes m + diag(m_1,..,m_d)$$

For $p > 0$, the p^{th} power $\nu_p^{(d)}$ of $\nu^{(d)}$ generates the family of the Negative multinomial distributions with parameter p on $\mathbb{R}^d$. The NEF $F(\nu^{(d)})$ is the $(NM-Ga)_d$ family.

To define the $(NM-Ga)_{d-1}$ family, we consider the following mixture of a $(d-1)$ dimensional Negative multinomial family and a Gamma family on $\mathbb{R}$. Let $\nu^{(d-1)}$ denote the measure given in (3) on $\mathbb{R}^{d-1}$ and for $p > 0$ let γ_p be the following measure on $\mathbb{R}$:

$$\gamma_p(dx) = \frac{1}{\Gamma(p)}x^{p-1}1_{(0,\infty)}(x)dx \qquad (5)$$

We then introduce $\mu^{(d-1)}(dx_1,..,dx_d) = \nu^{(d-1)}(dx_1,..,dx_{d-1})\gamma_{\sum_{i=1}^{d-1}x_i+1}(dx_d)$ with Laplace-transform on $\Theta(\mu^{(d-1)}) = \{\theta \in \mathbb{R}^d;\ \sum_{i=1}^{d-1}e^{\theta_i} + \theta_d < 0\}$

$$L_{\mu^{(d-1)}}(\theta) = \left(-\theta_d - \sum_{i=1}^{d-1}e^{\theta_i}\right)^{-1}$$

The variance function of $F = F(\mu^{(d-1)})$ is defined on $M_F = (0,\infty)^d$ by:

$$V_F(m) = m \otimes m + diag(m_1,..,m_{d-1},0)$$

The powers F_p of F are generated by the measures

$$\mu_p^{(d-1)}(dx_1,..,dx_d) = \nu_p^{(d-1)}(dx_1,..,dx_{d-1})\gamma_{\sum_{i=1}^{d-1}x_i+p}(dx_d)$$

for all $p > 0$. They are composed by the distributions of random variables $(X_1,..X_d)$ where $(X_1,..X_{d-1})$ has a Negative multinomial distribution with parameter p and X_d conditionally to $(X_1,..X_{d-1})$ has a Gamma distribution with parameter $\sum_{i=1}^{d-1}X_i + p$.

The $d-1$ other $(NM-Ga)_k$ families admit a Gaussian part in addition. Let k be in $\{0,..,d-2\}$. We still denote by $\nu^{(k)}$ and by γ_p the measures on $\mathbb{R}^k$ and $\mathbb{R}$ respectively given by (3) and (5). Let $\lambda_p^{(k)}$ be the normal distribution on $\mathbb{R}^k$ with mean 0 and covariance pI_k. Then we put if $k \geq 1$:

$$\mu^{(k)}(dx_1,..,dx_d) = \nu^{(k)}(dx_1,..,dx_k)\ \gamma_{\sum_{i=1}^{k}x_i+1}(dx_{k+1})\ \lambda_{x_{k+1}}^{(d-k-1)}(dx_{k+2},..,dx_d)$$

and $\mu^{(0)}(dx_1,..,dx_d)=\gamma_1(dx_1)\ \lambda^{(d-1)}_{x_1}(dx_2,..,dx_d)$.
We have the following:

$$\begin{array}{lcl}
\Theta(\mu^{(k)}) & = & \left\{\theta\in I\!\!R^d;\ \theta_{k+1}+\frac{1}{2}\sum_{i=k+2}^d\theta_i^2+\sum_{i=1}^k e^{\theta_i}<0\right\} \\
L_{\mu^{(k)}}(\theta) & = & \left(-\theta_{k+1}-\frac{1}{2}\sum_{i=k+2}^d\theta_i^2-\sum_{i=1}^k e^{\theta_i}\right)^{-1} \\
M_{F(\mu^{(k)})} & = & (0,\infty)^{k+1}\times I\!\!R^{d-k-1} \\
V_{F(\mu^{(k)})}(m) & = & m\otimes m+diag(m_1,..,m_k,0,m_{k+1},..,m_{k+1})
\end{array}$$

Here again the powers F_p of $F(\mu^{(k)})$ are generated for all $p>0$ by

$$\mu_p^{(k)}(dx_1,..,dx_d)=\nu_p^{(k)}(dx_1,..,dx_k)\ \gamma_{\sum_{i=1}^k x_i+p}(dx_{k+1})\ \lambda^{(d-k-1)}_{x_{k+1}}(dx_{k+2},..,dx_d)$$

They are composed by the distributions of $(X_1,..,X_d)$ where $(X_1,..,X_k)$ has a Negative multinomial distribution with parameter p, X_{k+1} given $(X_1,..,X_k)$ is Gamma distributed with parameter $\sum_{i=1}^k X_i+p$ and $(X_{k+2},..,X_d)$ given $(X_1,..,X_{k+1})$ are $d-k-1$ real independent Gaussian variables with mean 0 and variance X_{k+1}.

Note that the cases where the three Negative multinomial, Gamma and Gaussian families are mixed do not appear in $I\!\!R^2$. The family $F(\mu^{(1)})$ on $I\!\!R^2$ appears the first time in the paper of Bar-lev, Bshouty, Enis, Letac, Li Lu and Richards [1] as one of the NEF whose margins are in two different Morris families.

c)<u>The multinomial type M</u>
We take again the presentation of the multinomial distributions from Letac [14].
Let $\{e_1,..,e_d\}$ denote the canonical basis of $I\!\!R^d$ and e_o be the null vector. Then the measure on $I\!\!N^d$, $\mu(dx)=\sum_{i=0}^{d}\delta_{e_i}$, generates a NEF F with variance function on $M_F=\{m\in I\!\!R^d;\forall j\ m_j>0,\sum_{j=1}^d m_j<1\}$ equal to:

$$V_F(m)=m\otimes m+diag(m_1,..,m_d)$$

For any p in $I\!\!N-\{0\}$ the p^{th} power F_p of F is the set of probabilities $P(m,F_p)$ defined by

$$P(m,F_p)(n_1e_1+..+n_de_d)=\begin{pmatrix} p \\ n_0,n_1,..,n_d\end{pmatrix}\left(1-\frac{\sum m_j}{p}\right)^{n_0}\sum_{j=1}^d\left(\frac{m_j}{p}\right)^{n_j} \qquad (6)$$

where $n_0,n_1,..,n_d$ are positive integers with sum p.

d)<u>The hyperbolic type H</u>
Similarly to the $(NM-Ga)_{d-1}$ type, this last case is built from the following mixture of Negative multinomial family on $I\!\!R^{d-1}$ and the Morris family of the Hyperbolic cosine distributions on $I\!\!R$.
Let $\nu^{(d-1)}$ be the measure on $I\!\!R^{d-1}$ given in (3), and for $p>0$, α_p be defined by its Laplace-transform on $(-\pi/2,\pi/2)$ equal to $L_{\alpha_p}(\theta)=(cos\theta)^{-p}$. Now we introduce:

$$\mu(dx_1,..,dx_d)=\nu^{(d-1)}(dx_1,..,dx_{d-1})\ \alpha_{\sum_{i=1}^{d-1}x_i+1}(dx_d)$$

Then

$$\begin{aligned}
\Theta(\mu) &= \{\theta \in I\!R^d;\ \sum_{i=1}^{d-1} e^{\theta_i} < cos\theta_d\} \\
L_\mu(\theta) &= \left(cos\theta_d - \sum_{i=1}^{d-1} e^{\theta_i}\right)^{-1} \\
M_F &= (0,\infty)^{d-1} \times I\!R \\
V_F(m) &= m \otimes m + diag(m_1, .., m_{d-1}, \sum_{i=1}^{d-1} m_i + 1)
\end{aligned}$$

The powers F_p of F are generated for all $p > 0$ by the measures

$$\mu_p(dx_1, .., dx_d) = \nu_p^{(d-1)}(dx_1, .., dx_{d-1})\ \alpha_{\sum_{i=1}^{d-1} x_i + p}(dx_d)$$

Therefore F_p is composed with the distributions of $(X_1, .., X_d)$ where $(X_1, .., X_{d-1})$ has a Negative multinomial distribution and X_d conditionally to $(X_1, .., X_{d-1})$ has the Hyperbolic cosine distribution with parameter $\sum_{i=1}^{d-1} X_i + p$.

Note that all the simple quadratic distributions presented here have the remarkable following property: if $(X_1, .., X_d)$ has such a distribution, then the law of X_1 belongs to a Morris family and for any $k = 1, .., d-1$, the law of X_{k+1} conditionally to $(X_1, .., X_k)$ is also a Morris distribution with Jorgensen parameter depending of an affinity of $X_1, .., X_k$. (Indeed, it is easy to check that if $(X_1, .., X_d)$ has the Negative multinomial law (4), then for any $k = 1, ..d-1$, $(X_1, .., X_k)$ has still a Negative multinomial distribution (given by (4) replacing d by k), while X_{k+1} conditionally to $(X_1, .., X_k)$ has a Negative binomial distribution with Jorgensen parameter $1 + \sum_{i=1}^{k} X_i$. The same remark holds true for Multinomial distributions given by (6): X_{k+1} conditionally to $(X_1, .., X_k)$ has a binomial distribution with Jorgensen parameter $p - \sum_{i=1}^{k} X_i$).

The proof of Theorem 6.1 is long and technical. It relies only on algebraic arguments from the three necessary conditions:

i) V_F is symmetric: $< \alpha, V_F(m)\beta > = < \beta, V_F(m)\alpha >$

ii) $V_F(m)$ is positive definite on M_F

iii) $\psi''_\mu(m)(\alpha, \beta) = ((V_F(m))^{-1})'(\alpha, \beta) = -(V_F(m))^{-1}(V'_F(m)\alpha)(V_F(m))^{-1}\beta$ is symmetric in α and β as Hessian. This is equivalent to

$$V'_F(m)(V_F(m)u)v = V'_F(m)(V_F(m)v)u \tag{7}$$

We first prove that there exists m_0 such that $V_F(m_0) = 0$. The translation $x \mapsto x - m_0$ yields to the case where $V_F(m) = A(m,m) + B(m)$. In that case, relation (7) becomes $B(B(m)u)v = B(B(m)v)u$. We then introduce the endomorphisms $Q(u)$ linear in u defined by ${}^tQ(u)m = B(m)u$. They hence commute and consequently have all a triangular Jordan matrix in the same basis of $\mathbb{C}^d$. Distinguishing the cases where the eigenvalues are simple or multiple, real or complex and using linear transformations we isolate the $2d + 4$ previous simple QVF.

7 Meixner's and Feinsilver's characterizations on $\mathbb{R}^d$

Let us first introduce some notations: let $(e_1, .., e_d)$ be the canonical basis of $\mathbb{R}^d$. Classically if $x = \sum x_i e_i$ and $y = \sum y_i e_i$, $< x, y >= \sum x_i y_i$. If $n = {}^t(n_1, \ldots, n_d) = \sum_{i=1}^{d} n_i e_i$ is in $\mathbb{N}^d$, we write $|n| = n_1 + \ldots + n_d$ the *order of* n and we also write $x^n = x_1^{n_1} \ldots x_d^{nd}$ and $n\,! = n_1\,! \ldots n_d\,!$. A polynomial Q in x of order k can be written as: $Q(x) = \sum_{q \in \mathbb{N}^d;\ |q| \leq k} c_q x^q$, where at least one of the real numbers c_q is non zero when $|q| = k$.

The space of all polynomials on $\mathbb{R}^d$ is denoted by Π^d and, for any k in $\mathbb{N}$, Π_k^d is the subspace of the polynomials of Π^d with order less or equal to k.
Let F be a NEF on $\mathbb{R}^d$, μ be one generating measure with mean m_0. Then the probabilities $P(m, F)$ of F have a density with respect to μ equal to

$$f_\mu(x, m) = \exp\{< \psi_\mu(m), x > -k_\mu \psi_\mu(m)\}$$

As on $\mathbb{R}$, we can differentiate $f_\mu(x, m)$ with respect to m in the directions given by the vectors of the basis and so define the functions (P_n) by

$$P_n(x) = \frac{\partial^{|n|}}{\partial m^n} f_\mu(x, m)_{|m=m_0} = f_\mu^{|n|}(x, m_0)(\underbrace{e_1}_{n_1}, ..., \underbrace{e_d}_{n_d})$$

as Feinsilver [8] does, or we can differentiate it in other directions and so introduce for any A in $GL(\mathbb{R}^d)$

$$\begin{aligned} P_{A,n}(x) &= f_\mu^{|n|}(x, m_0)(\underbrace{Ae_1}_{n_1}, ..., \underbrace{Ae_d}_{n_d}) \\ &= f_\mu^{(n)}(x, m_0)(Ae_1, ..., Ae_d) \end{aligned} \tag{8}$$

Actually these $P_{A,n}$ are polynomials :

Proposition 7.1 *Let A be in $GL(\mathbb{R}^d)$ and define $P_{A,n}(x) = f_\mu^{(n)}(x, m_0)(Ae_1, \ldots, Ae_d)$. Then:*

i) there exists a $r > 0$ such that, for all m in the open ball $B(m_0, r)$ and for all x in $\mathbb{R}^d$,

$$f_\mu(x, Am) = \sum_{n \in \mathbb{N}^d} \frac{(m - A^{-1}m_0)^n}{n!} P_{A,n}(x).$$

ii) $P_{A,n}(x)$ is a polynomial in x of degree $|n|$ and the $(P_{A,n})_{n \in \mathbb{N}^d}$ form a basis of Π^d.

Feinsilver [8] proves that under the hypothesis "$m_0 = 0, V_F(m_0)$ is diagonal", the orthogonality of the (P_n) yields to a sub-class of QVF-NEF without specifying it. Labeye-Voisin in an unpublished work shows that this is a sub-class of the simple QVF-NEF.

Pommeret then completes this last result [10],[18]. The introduction of the linear automorphisms A only gives more flexibility on the hypothesis, what is essential in the practical determination of the polynomials since we have not to transform F by an affinity first. Indeed if A is in $GL(\mathbb{R}^d)$, then we have:

$$f_\mu^{(n)}(x, m_0)(Ae_1, \ldots, Ae_d) = f_{A^{-1}(\mu)}^{(n)}(A^{-1}x, A^{-1}m_0)(e_1, \ldots, e_d).$$

as in the real case, so that if $Q_n(x) = f_{A^{-1}(\mu)}^{(n)}(x, A^{-1}m_0)(e_1, \ldots, e_d)$, then the (Q_n) are $A^{-1}(\mu)$ orthogonal if and only if the $(P_{A,n})$ are μ-orthogonal.

The Theorems

Pommeret finally gets the following characterization:

Theorem 7.1 *Let F be an irreducible NEF on $\mathbb{R}^d$ and μ be a probability in F with mean m_0. Let $A = (a_{ij})$ be in $GL(\mathbb{R}^d)$ and let the polynomials $(P_{A,n})_{n \in \mathbb{N}^d}$ be defined as in* (8). *Then the three following statements are equivalent:*

i) the $(P_{A,n})_{n \in \mathbb{N}^d}$ are μ-orthogonal,

ii) F is simple quadratic and $A^{-1}V_F(m_0)\ {}^tA^{-1}$ is diagonal,

iii) there exist real numbers a, $(v_s)_{s=1,\ldots,d}$ and $(b_{ts}^l)_{(t,s,l)\in\{1,\ldots,d\}^3}$ such that:

$$\begin{aligned} x_i P_{A,n}(x) &= \sum_{s=1}^d a_{is}\{v_s P_{A,n+e_s}(x) + \sum_{t,l=1}^d n_l b_{ts}^l P_{A,n+e_t-e_l}(x) \\ &\quad + n_s(1 + a(|n| - 1))P_{A,n-e_s}(x) + (A^{-1}m_0)_s P_{A,n}(x)\}. \end{aligned}$$

In this case we have:

$$V_F(m) = a(m - m_0)\ {}^t(m - m_0) + B(m - m_0) + V_F(m_0),$$

where $(A^{-1}B(Am)\ {}^tA^{-1})_{ij} = \sum_{k=1}^d b_{ij}^k m_k$ *and* $A^{-1}V_F(m_0)\ {}^tA^{-1} = \begin{pmatrix} v_1 & & 0 \\ & \ddots & \\ 0 & & v_d \end{pmatrix}$.

Consequently we deduce the following characterization:

Corollary 7.1 *Let F be an irreducible NEF on $\mathbb{R}^d$. With the notations of Theorem* 7.1, *F is simple quadratic if and only if there exists (A, m_0) in $GL(\mathbb{R}^d) \times M_F$ such that the sequence of polynomials $(P_{A,n})_{n \in \mathbb{N}^d}$ is μ-orthogonal.*

Examples

We give here some examples of polynomials $P_{A,n}$, for the five types of pure quadratic NEF on $\mathbb{R}^2$ described by Casalis [4]. For each NEF and for each m_0 in M_F we take arbitrarily A symmetric and such that $A^2 = V_F(m_0)$.
Since each simple quadratic NEF is a combination of quadratic NEF on $\mathbb{R}$ of the Gaussian, Gamma, Negative binomial, Binomial and Hyperbolic cosine types, we obtain combinations of the associated polynomials on $\mathbb{R}$.

Orthogonal polynomials for the simple quadratic class on $\mathbb{R}^2$:

$V_F(m)$	m_0	A	$P_{A,n}(x)$
Two variables Binomial type $-m\,{}^tm+\begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}$	$\begin{pmatrix} 1/4 \\ 3/8 \end{pmatrix}$	$\frac{\sqrt{3}}{8}\begin{pmatrix} 0 & 2 \\ 2 & -1 \end{pmatrix}$	$(\frac{4}{\sqrt{3}})^{\lvert n\rvert} M_{n_1}^{c_1,\beta_1}(x_2)\, M_{n_2}^{c_2,\beta_2}(x_1+\frac{3}{4})$ $\begin{cases} c_1=-1, \beta_1=x_1-1 \\ c_2=-1/3, \beta_2=n_1-1 \end{cases}$
Two variables Negative Binomial type $m\,{}^tm+\begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}$	$\begin{pmatrix} 1 \\ \sqrt{5}-1 \end{pmatrix}$	$\frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 2 \\ 2 & \sqrt{5}-1 \end{pmatrix}$	$(\frac{1}{\sqrt{2}})^{\lvert n\rvert} M_{n_1}^{c_1,\beta_1}(x_2)\, M_{n_2}^{c_2,\beta_2}(x_1)$ $\begin{cases} c_1=\frac{3-\sqrt{5}}{2}, \beta_1=x_1+1 \\ c_2=1/2, \beta_2=n_1+1 \end{cases}$
Gamma Negative Binomial type $m\,{}^tm+\begin{pmatrix} m_1 & 0 \\ 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 0 & \sqrt{2} \\ \sqrt{2} & \sqrt{2} \end{pmatrix}$	$n_1!(-\frac{1}{\sqrt{2}})^{n_1} L_{n_1}^{x_1}(x_2)\,(\frac{1}{\sqrt{2}})^{n_2} M_{n_2}^{c_2,\beta_2}(x_1)$ $c_2=1/2, \beta_2=n_1+1$
Gamma Gaussian type $m\,{}^tm+\begin{pmatrix} 0 & 0 \\ 0 & m_1 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	Id	$(\sqrt{\frac{x_1}{2}})^{n_2} H_{n_2}(\frac{x_2}{\sqrt{2x_1}})(-1)^{n_1} n_1! L_{n_1}^{n_2}(x_1)$
Negative Binomial Hyperbolic type $m\,{}^tm-\begin{pmatrix} m_1 & m_2 \\ m_2 & -m_1 \end{pmatrix}$	$\begin{pmatrix} 2 \\ 0 \end{pmatrix}$	$\sqrt{2}$ Id	$(\frac{1}{\sqrt{2}})^{n_2}(x_1-1)\dots(x_1-1+n_2) P_{n_2}^{x_1}(x_2)$ $(\frac{1}{\sqrt{2}})^{n_1} M_{n_1}^{c_1,\beta_1}(x_1-1)$ $c_1=1/2, \beta_1=n_2+1$

Pommeret also extends this result to the class of QVF-NEF by defining a "pseudo-orthogonality" for polynomials :

We say that the polynomials (Q_n) *are* pseudo-orthogonal *if* $\int Q_n(x)Q_q(x)\mu(dx) = 0$ *if* $|n| \neq |q|$.

Pommeret gets a decomposition of Π^d in orthogonal subspaces. Recall that Π^d denotes the space of polynomials of $I\!R^d$ and Π^d_k the subspaces of polynomials of order $\leq k$. Introduce E_k as the subspace of Π^d generated by the $(P_{A,n})_{|n|=k}$; it is one supplementary of Π^d_{k-1} in Π^d_k. It is easy to see from the n^{th} linearity of $f_\mu^{(n)}(x, m_0)$ that the (E_k) are independent of A. We can obviously write $\Pi^d = \oplus E_k$ and for simple quadratic NEF $\Pi^d = \overset{\perp}{\oplus} E_k$. However this property is not characteristic of the simple QVF-NEF. We have the following (see Pommeret [18])

Theorem 7.2 *Let F be a NEF on $I\!R^d$ and $\mu = P(m_0, F)$. Then, the following statements are equivalent:*

i) $\Pi^d = \overset{\perp}{\oplus}_{k\in I\!N} E_k$,

ii) the polynomials $P_n(x) = f_\mu^{(n)}(x, m_0)(e_1, \ldots, e_d)$ *are μ-pseudo-orthogonal,*

iii) F is quadratic,

iv) if $A = (a_{ij})$ *is in* $GL(I\!R^d)$*, then there exist real numbers* (v^{ij}_{st}), (b^s_{ji}) *such that:*

$$\begin{aligned} x_i P_{A,n}(x) &= \sum_{s=1}^d a_{is}\left(\sum_{t,u,v=1}^d v^{ti}_{uv}(n_u - \delta_{uv})n_v P_{A,n+e_t-e_u-e_v}(x)\right. \\ &\quad + n_s P_{A,n-e_s}(x) + \sum_{t,u=1}^d n_u b^u_{it} P_{A,n+e_t-e_u}(x) \\ &\quad \left. + (A^{-1}m_0)_s P_{A,n}(x) + \sum_{t=1}^d (V_{A^{-1}F}(m_0))_{ti} P_{A,n+e_t}(x)\right). \end{aligned}$$

We conclude now in characterizing the polynomials $(P_{A,n})$ among all the sequences of (pseudo-)orthogonal polynomials by a condition equivalent to Meixner's one. Let us remark that if, on $I\!R$, up to factors of normalization, there is only one sequence of orthogonal polynomials with respect to a measure μ, the subspaces E_k being one-dimensional, on $I\!R^d$ this is no more true. The dimension of the subspace E_k is $r_k = \begin{pmatrix} k+d-1 \\ k \end{pmatrix}$ and when F is quadratic (resp. simple quadratic) we know some pseudo-orthogonal (resp. orthogonal) basis. These are the $\{P_{A,n};\ |n| = k\}$ for suitable A. But, of course, there are not the only pseudo-orthogonal ones. We have the following (see Pommeret[18]):

Theorem 7.3 *Let F be a NEF on $I\!R^d$ and μ the probability of F with mean m_o. Let $(Q_n)_{n\in I\!N^d}$ be a sequence of μ-pseudo-orthogonal polynomials such that Q_n is of order $|n|$. Then, the two following statements are equivalent:*

i) the generating function of the (Q_n) is exponential, i.e. there exist $r > 0$ and two analytic functions: $a:\ B(0,r) \to I\!R^d$, $\quad b:\ B(0,r) \to I\!R$, such that, for all z in $B(0,r)$,

$$\sum_{n\in I\!N^d} \frac{z^n}{n!} Q_n(x) \ = \ \exp\{\langle a(z), x\rangle + b(z)\},$$

ii) there exists A in $Gl(I\!R^d)$ such that, for all n in $I\!N^d$,

$$Q_n(x) \ = \ Q_0(x) P_{A,n}(x) \ = \ Q_0(x) f_\mu^{(n)}(x, m_0)(Ae_1, \ldots, Ae_d).$$

In this case, $a(z) = \psi_\mu(Az + m_0)$ and $b(z) = -k_\mu(\psi_\mu(Az + m_0))$.

From this theorem we can deduce:

Corollary 7.2 *Let F be a NEF on $I\!R^d$ and μ a probability of F with mean m_0. Then,*

i) there exists a sequence of μ-pseudo-orthogonal polynomials with an exponential generating function if and only if F is quadratic,

ii) if F is irreducible, there exists a sequence of μ-orthogonal polynomials with an exponential generating function if and only if F is simple quadratic.

This Corollary generalizes Meixner's characterization. The proofs of these Theorems are similar to the real case.

References

[1] Bar-Lev, S.K., Bshouty,D., Enis, P., Letac, G., Li Lu, I., Richards, D.(1994). The Diagonal Natural Exponential Families in $I\!R^n$ and their classification. *Journal of Theoretical Probability.* **7**, 883-929.

[2] Casalis, M. (1991). Les familles exponentielles à variance quadratique homogène sont des lois de Wishart sur un cône symétrique. *C.R. Acad. Sc.* **312**, SÚrie I, 537-540.

[3] Casalis, M. (1992). Un estimateur de la variance pour une famille exponentielle naturelle à fonction variance quadratique. *C.R. Acad. Sc.* **314**, SÚrie I, 143-146.

[4] Casalis, M. (1992). Les familles exponentielles sur $I\!R^2$ de fonction-variance $am\otimes m + B(m) + C$. *C.R. Acad. Sc.* **314**, SÚrie I, 635-638.

[5] Casalis, M. (1994). Les 2d+4 types de familles exponentielles naturelles quadratiques simples sur $I\!R^d$. *To appear in Ann. Statist.*

[6] Consonni, G. and Veronese, P. (1992). Conjugate priors for exponential families having quadratic variance functions. *J.Am.Statist.Assoc.* **87**, 1123-1127.

[7] Feinsilver, P. (1986). Some classes of orthogonal polynomials associated with martingales. *Proc. A.M.S.* **98**, 298-302.

[8] Feinsilver, P. (1991). Orthogonal polynomials and coherent states. Symmetries in science: Plenum press, 159-192.

[9] Hassairi, A.(1992). *Classification des familles exponentielles naturelles dans* $I\!R^d$ *de variance cubique de type Mora.* Doctorat d'Etat, Université Paul Sabatier,Toulouse.

[10] Labeye-Voisin, E. and Pommeret, D. (1995). Polyn¶mes orthogonaux associÚs aux familles exponentielles naturelles sur $I\!R^d$. *C.R. Acad. Sc.* **320**, SÚrie I, 79-84.

[11] Laha, R.G. and Lukas, E. (1960). On a problem connected with quadratic regression. *Biometrika.* **47**, 335-343.

[12] Lancaster, H.O. (1975). Joint probability distributions in the Meixner classes. *J. Roy. Statist. Soc.* **37**, 434-443.

[13] Letac, G. (1992). *Lectures on natural exponential families and their variance functions.* Instituto de matemática pura e aplicada: Monografias de matemática. **50**, RIO.

[14] Letac, G. (1988). Le problème de la classification des familles exponentielles naturelles de $I\!R^d$ ayant une fonction variance quadratique. *Lecture Notes in Mathematics, Probability Measures on Groups*, vol IX, Springer Verlag.

[15] Meixner, C.N (1934). Orthogonal Polynomsysteme mit einer besonderen Gestalt der erzengenden Function. *J. London Math. Soc.* **9**, 6-13.

[16] Morris, C.N. (1982). Natural exponential families with quadratic variance functions. *Ann. Statist.* **10**, 65-82.

[17] Morris, C.N. (1983). Natural exponential families with quadratic variance functions: statistical theory. *Ann. Statist.* **11**, 515-529.

[18] Pommeret, D.(1995). Orthogonal polynomials and natural exponential families. *submitted*

[19] Shanbhag, D.N. (1979). Diagonality of the Batthacharyya matrix as a characterization. *Theory of Prob. and Appl.* **24**, 430-433.