Exponential families of mixed Poisson distributions

A. Ferrari[a], G. Letac[b,*], J.-Y. Tourneret[c]

[a]*LUAN, Université de Nice Sophia-Antipolis, 06108 Nice, France*
[b]*LSP, Université Paul Sabatier, 31062 Toulouse, France*
[c]*IRIT/ENSEEIHT/TESA, 31071 Toulouse, France*

Abstract

If $I=(I_1,\ldots,I_d)$ is a random variable on $[0,\infty)^d$ with distribution $\mu(d\lambda_1,\ldots,d\lambda_d)$, the mixed Poisson distribution $MP(\mu)$ on $\mathbb{N}^d$ is the distribution of $(N_1(I_1),\ldots,N_d(I_d))$ where $N_1,\ldots,N_d$ are ordinary independent Poisson processes which are also independent of I. The paper proves that if F is a natural exponential family on $[0,\infty)^d$ then $MP(F)$ is also a natural exponential family if and only if a generating probability of F is the distribution of $v_0+v_1Y_1+\cdots+v_qY_q$ for some $q\leqslant d$, for some vectors $v_0,\ldots,v_q$ of $[0,\infty)^d$ with disjoint supports and for independent standard real gamma random variables $Y_1,\ldots,Y_q$.

Keywords: Natural exponential families; Multivariate gamma; Overdispersion

1. Introduction

Consider the Poisson distribution with parameter $\lambda>0$ defined by

$$P(\lambda)(dx)=\sum_{n=0}^{\infty}e^{-\lambda}\frac{\lambda^n}{n!}\delta_n(dx).$$

If we randomize the parameter λ by some probability $\mu(d\lambda)$ on $(0,\infty)$ we get a new probability $MP(\mu)$ on the set $\mathbb{N}$ of nonnegative integers defined by $MP(\mu)=\int_0^\infty P(\lambda)\mu(d\lambda)$. We have the

* Corresponding author. Fax: +335 61 55 60 89.
E-mail address: gerard.letac@tele2.fr (G. Letac).

identifiability property: $P(\mu) = P(\mu')$ if and only if $\mu = \mu'$: just compute the generating function

$$f_{MP(\mu)}(z) = \sum_{n=0}^{\infty} z^n MP(\mu)(\{n\})$$

of $MP(\mu)$ and link it to the Laplace transform $L_\mu(\theta)$ of μ by $f_{MP(\mu)}(z) = L_\mu(z-1)$. One sees $MP(\mu)$ as the distribution of $N(I)$ where $t \mapsto N(t)$ is a standard Poisson process on $\mathbb{N}$ which is independent of the random variable I with distribution μ. One reason of the interest on these mixed Poisson distributions lies on the fact that they are overdispersed, in the sense that their variance is bigger than their mean. However, it should be pointed out that one easily constructs an overdispersed distribution ν concentrated on $\mathbb{N}$ such that no μ with $\nu = MP(\mu)$ can possibly exist. An example is $\sum_{n=0}^{\infty} \nu_n z^n = (1+z+z^2)/(6-3z)$. There is an abundant literature on the topic for which Grandell [5] offers a good synthesis and references.

Suppose now that μ belongs to a natural exponential family (NEF) F concentrated on $[0,\infty)$. Denote by $V_F(m)$ its variance function defined on the domain of the means $M_F \subset (0,\infty)$ in the sense initiated by Morris [10]. We consider the model

$$MP(F) = \{MP(\mu);\ \mu \in F\}.$$

Let us start with the following simple observation: if $\mu \in F$ has mean m then the variance of $MP(\mu)$ is $m + V_F(m)$. It is tantalizing to think that we have created a new natural exponential family G with variance function V_G such that $M_G \supset M_F$ and such that $V_G(m) = m + V_F(m)$ on M_F. This is false: if $p > 1$ is not an integer consider the NEF F such that $M_F = (1,\infty)$ and $V_F(m) = (m-1)^p$. Such a NEF does exist from Bar-Lev and Enis [1] or Jorgensen [7]. Then one sees that $V_G(m) = m + (m-1)^p$ is not a variance function: from Theorems 3.1 and 3.2 of Letac and Mora [8] one should have $M_G = (1,\infty)$ and thus for $m_0 > 1$ we would have the contradiction

$$\int_1^{m_0} \frac{dm}{m + (m-1)^p} = \infty.$$

Furthermore, for some NEF F's the function $m + V_F(m)$ can be actually the variance function of some NEF G with no relation either with $MP(F)$ or with F. A provocative example is $V_F(m) = m^p$ with $p > 1$ is not an integer and $M_F = (0,\infty)$. In this case $V_G(m) = m + m^p$ is the variance function of a NEF such that $M_G = (0,\infty)$ but G is concentrated on the additive semigroup $\mathbb{N} + p\mathbb{N}$. For checking this it is enough is to compute the corresponding cumulant transform and to observe that the elements of G must be infinitely divisible with a discrete Lévy measure concentrated on $\mathbb{N} + p\mathbb{N}$. Finally Bent Jorgensen [7] offers a different construction of mixed Poisson distributions from a NEF (see the remark in Section 3 below for a description of the Jorgensen's manner.)

Thus, a natural question is: if the NEF G exists do we have $G \supset MP(F)$? In Section 3 Theorem 1 says : yes, if and only if F is a gamma family, i.e. when there exists a number $p > 0$ such that $V_F(m) = m^2/p$. In this case G is a negative binomial *NEF*. Section 4 extends the question to $\mathbb{N}^d$. We randomize $(\lambda_1, \dots, \lambda_d)$ in the product $P(\lambda_1)(dx_1)\cdots P(\lambda_d)(dx_d)$ by the probability $\mu(d\lambda)$ on $[0,\infty)^d$ and consider the probability on $\mathbb{N}^d$ defined by

$$MP(\mu) = \int_0^\infty \dots \int_0^\infty P(\lambda_1)\cdots P(\lambda_d)\mu(d\lambda_1,\dots,d\lambda_d).$$

We get a similar characterization (Theorem 2) which is described in the abstract above. The multivariate distributions which occur in Theorem 2 have been recently isolated and characterized by Konstancja Bobecka and Jacek Wesołowski [2] (details are given in Section 4). The proof of Theorem 2 needs some care and the particular case $d = 1$ of Theorem 1 is a preparation to $d \geqslant 2$. Section 2 recalls some facts about NEF.

This study was motivated by statistical optics. Mixed Poisson distributions are commonly used to model data recorded from low flux objects or with short exposure times using photocounting cameras. This physical model arises from the semiclassical theory of statistical optics described in Goodman [4]. In this theory, the classical theory of propagation is used up to the camera, leading to a high flux image. Conditionally to this image, the number of photons counted on the pixels is distributed according to a Poisson distribution whose mean is the high flux intensity.

A common problem for example in astrophysics is to estimate parameters of the wavefront (the mixing distribution) from photocounts recorded on a set of pixels. A description is found in Ferrari et al. [3]. A general assumption is that the vector of complex amplitudes associated to adjacent pixels of the image is a zero mean Gaussian vector, which implies that the vector of the corresponding intensities is distributed according to a multivariate gamma distribution. An important question is to derive conditions ensuring that the associated mixed Poisson distribution belongs to a NEF. This result is important since the computational complexity of most estimation or detection methods is usually reduced when applied to distributions belonging to an NEF.

2. NEF on $\mathbb{R}$ and $\mathbb{R}^d$

This section describes the notations and classical facts about natural exponential families, mainly taken from Morris [10] and Letac and Mora [8]. Denote by

$$L_\nu(\theta) = \int_{\mathbb{R}^d} e^{\langle\theta,\lambda\rangle}\nu(d\lambda) \leqslant \infty$$

the Laplace transform of a positive measure ν defined for $\theta \in \mathbb{R}^d$ not concentrated on any affine hyperplane. The Hölder inequality proves that the set $D(\nu)$ of $\theta \in \mathbb{R}^d$ such that $L_\nu(\theta) < \infty$ is a convex set and that the cumulant function $k_\nu = \log L_\nu$ is a strictly convex function on this set. Denote by $\Theta(\nu)$ the interior of $D(\nu)$ and assume that $\Theta(\nu)$ is not empty. Then k_ν is real analytic on $\Theta(\nu)$. The set $F(\nu)$ of probabilities

$$\nu_\theta(d\lambda) = e^{\langle\theta,\lambda\rangle - k_\nu(\theta)}\nu(d\lambda),$$

where θ runs $\Theta(\nu)$ is called the NEF with generating measure ν. Note that $F(\nu) = F(\nu')$ does not imply $\nu = \nu'$ but implies only the existence of some $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$ such that $\nu(d\lambda) = e^{\langle a,\lambda\rangle + b}\nu'(d\lambda)$. Thus, a member μ of the NEF $F(\nu)$ can always be taken as a generating measure. However, some generating measures are not necessarily probabilities and can even be unbounded. We mention also that $\theta \mapsto \nu_\theta(d\lambda)$ is called a canonical parametrization of the NEF. Other parametrizations of the type

$$t \mapsto \nu_{\alpha(t)}(d\lambda) = e^{\langle\alpha(t),\lambda\rangle + \beta(t)}\nu(d\lambda)$$

with $\beta(t) = -k_\nu(\alpha(t))$, could be considered. Since $\theta \mapsto k_\nu(\theta)$ is a strictly convex function on the open set $\Theta(\nu)$, the map

$$\theta \mapsto m = k'_\nu(\theta) = \int_{\mathbb{R}^d} \lambda\nu_\theta(d\lambda)$$

from $\Theta(\nu)$ to $\mathbb{R}^d$ is one to one. The open set $M_F = k'_\nu(\Theta(\nu))$ is called the domain of the means of F. Denote by $m \mapsto \theta = \psi_\nu(m)$ the inverse function of k'_ν from M_F onto $\Theta(\nu)$. The Hessian matrix $k''_\nu(\theta)$ is the covariance matrix of the probability $\nu_\theta(d\lambda)$. Denoting $V_F(m) = k''_\nu(\psi_\nu(m))$, the map from M_F to the positive definite symmetric matrices of order d defined by $m \mapsto V_F(m)$ is called the variance function of F. It characterizes F since its knowledge leads by integration of a differential equation to the knowledge of the cumulant function of a generating measure of F.

3. The case of real exponential families

For $d = 1$, for p and $a > 0$ the gamma distribution with shape parameter p and scale parameter a is

$$\gamma_{p,a}(d\lambda) = e^{-\lambda/a}\frac{\lambda^{p-1}}{a^p}\mathbf{1}_{(0,\infty)}(\lambda)\frac{d\lambda}{\Gamma(p)}. \tag{1}$$

This is a member of the NEF F generated by $\nu(d\lambda) = \frac{\lambda^{p-1}}{\Gamma(p)}\mathbf{1}(\lambda)d\lambda$. The domain of the means of F is $M_F = (0, \infty)$ and its variance function is $V_F(m) = m^2/p$. The Laplace transform of $\gamma_{p,a}$ is $L(z) = \frac{1}{(1-az)^p}$ with a suitable definition of this analytic function in $\{z \in \mathbb{C};\ \Re z < 1/a\}$: we have simply to impose that it is real on the real axis. We see that the generating function of $MP(\gamma_{p,a})$ is $\frac{1}{(1+a-az)^p} = (\frac{1-q}{1-qz})^p$ with the notation $q = a/(1+a) \in (0, 1)$. Thus, if $N \sim MP(\gamma_{p,a})$ then with the Pochhammer notation $(p)_0 = 1$ and $(p)_{k+1} = (p+k)(p)_k$ we have

$$\Pr(N = k) = \frac{1}{k!}(p)_k(1-q)^p q^k = \frac{1}{k!}(p)_k\frac{a^k}{(1+a)^{p+k}}, \tag{2}$$

a negative binomial distribution. This is a member of the NEF G of negative binomial distributions generated by

$$\sum_{k=0}^{\infty}\frac{1}{k!}(p)_k\delta_k.$$

The domain of the means of G is $M_G = (0, \infty)$ and its variance function is $V_F(m) = m + m^2/p$. Thus, both F and $G = MP(F)$ are NEF in this particular example. We show that this is the only case:

Theorem 1. *If the image of the NEF $F(\nu)$ on $[0, \infty)$ by $\mu \mapsto MP(\mu)$ is still an NEF, then there exists $p > 0$ such that $F(\nu)$ is the family of gamma distributions with fixed shape parameter p.*

Proof. Denote $L_\nu(\theta) = \int_0^\infty e^{\lambda\theta}\nu(d\lambda)$ for $\theta \in \Theta$, where Θ is the interior of the convergence domain of $L_\nu(\theta)$. Note that Θ is either $\mathbb{R}$ or some half line $(-\infty, a)$. Suppose that the image of $F(\nu)$ by $\mu \mapsto MP(\mu)$ is an NEF on $\mathbb{N}$ generated by some measure $\sum_{n=0}^\infty p_n\delta_n$. Consequently, there exists two functions α and β defined on $\Theta + 1$ such that for all n

$$\int_0^\infty e^{\lambda(\theta-1)}\frac{\lambda^n}{n!}\nu(d\lambda) = p_n e^{n\alpha(\theta)+\beta(\theta)}, \tag{3}$$

which can be rewritten

$$L_\nu^{(n)}(\theta - 1) = n!p_n e^{n\alpha(\theta)+\beta(\theta)}. \tag{4}$$

Recall that ν is concentrated on $(0, \infty)$ and thus that L_ν is not a constant. Being a Laplace transform the function L_ν cannot be a polynomial and $L_\nu^{(n)}$ cannot be identically 0. This implies $p_n \neq 0$ for all n. Eq. (**??**) shows that $\alpha(\theta)$ and $\beta(\theta)$ are real-analytic functions on the interval $\Theta + 1$. Indeed $\theta \mapsto L_\nu(\theta - 1)$ is analytic in the half complex plane $\Theta + 1 + i\mathbb{R}$ as well as its nth derivative $L_\nu^{(n)}(\theta - 1)$. Furthermore, since $L_\nu^{(n)}(\theta - 1)$ is positive on $\Theta + 1$ (because $p_n > 0$), its logarithm is real-analytic. Consequently, $n\alpha(\theta) + \beta(\theta)$ and $(n + 1)\alpha(\theta) + \beta(\theta)$ are real-analytic on $\Theta + 1$, which implies by linear combination that $\alpha(\theta)$ and $\beta(\theta)$ are real-analytic on $\Theta + 1$. This proves the existence of $\alpha'(\theta)$ and $\beta'(\theta)$. By taking the logarithms of both sides of (**??**) and differentiating with respect to θ we get

$$\frac{L_\nu^{(n+1)}(\theta - 1)}{L_\nu^{(n)}(\theta - 1)} = n\alpha'(\theta) + \beta'(\theta). \tag{5}$$

We now fix θ. Assume first that $a = \alpha'(\theta) \neq 0$ and denote $p = \beta'(\theta)/\alpha'(\theta)$. Eq. (**??**) can be written $L_\nu^{(n+1)}(\theta - 1) = a(p+n)L^{(n)}(\theta - 1)$ hence $L_\nu^{(n)}(\theta - 1) = L_\nu(\theta - 1)(p)_n a^n$. Since ν is concentrated on $(0, \infty)$ we have $L'_\nu(\theta - 1) = a > 0$. Since $L''_\nu(\theta - 1) = pa^2/2 > 0$ we have $p > 0$. The Taylor formula applied to the analytic function L_ν for small values of h can be written as follows:

$$\begin{aligned} L_\nu(\theta - 1 + h) &= L_\nu(\theta - 1) \sum_{n=0}^{\infty} (p)_n \frac{(ah)^n}{n!} \\ &= L_\nu(\theta - 1)(1 - ah)^{-p}. \end{aligned}$$

The result $\frac{L_\nu(\theta-1+h)}{L_\nu(\theta-1)} = (1 - ah)^{-p}$ is valid for any $h \in (-\infty, 1/a)$, since the Laplace transform is an analytic function. The right-hand side of this expression is the Laplace transform of the gamma distribution $\gamma_{p,a}$. Moreover, the Laplace transform of $\nu_{\theta-1}$ is

$$\int_0^\infty e^{\lambda h}\nu_{\theta-1}(d\lambda) = \int_0^\infty \frac{e^{\lambda(h+\theta-1)}}{e^{k_\nu(\theta-1)}}\nu(d\lambda) = \frac{L_\nu(\theta - 1 + h)}{L_\nu(\theta - 1)},$$

which shows that $\nu_{\theta-1} = \gamma_{p,a}$. In other words, the exponential family for $\{MP(\mu_\theta);\ \theta \in \Theta\}$ is the family of gamma distributions with fixed shape parameter p.

If $\alpha'(\theta) = 0$, (**??**) yields $\frac{L_\nu^{(n+1)}(\theta-1)}{L_\nu^{(n)}(\theta-1)} = \beta'(\theta)$ which leads to

$$\frac{L_\nu(\theta - 1 + h)}{L_\nu(\theta - 1)} = e^{\beta'(\theta)h}.$$

This is the noninteresting case where ν is a Dirac measure concentrated on the point $\beta'(\theta)$. Our definition of NEF excludes this and the proof of Theorem 1 is complete. $\square$

Remark. For clarification it should be pointed out that Jorgensen [7, pp. 166–167] mentions a different object. Given a NEF $F = F(\nu)$ on $(0, \infty)$ and taking the number ψ in a suitable interval, Bent Jorgensen considers a NEF H_ψ on $\mathbb{N}$ with a cumulant function of the form $\theta \mapsto k_\nu(\psi + e^\theta)$. In this case, introducing the reciprocal function $h_\psi(m)$ of the map $z \mapsto zk'_\nu(\psi + z)$ he proves that

$$V_{H_\psi}(m) = m + V_F\left(\frac{m}{h_\psi(m)}\right) h_\psi(m)^2.$$

The Jorgensen's construction seems motivated by the particular case $V_F(m) = m^p$ with $p \geqslant 1$ as considered by Hougaard et al. [6]. This family H_ψ is obtained from F by a Poisson mixing process, but in a slightly complicated way. For describing it adopt the following notation: if μ is a measure on R and $c > 0$ denote $d_c\mu$ the image of the measure μ by the dilation $x \mapsto cx$. Then H_ψ is the set of all $MP(d_c \nu_{\psi+c})$ such that c is in $\Theta(\nu)$. For instance if F is a gamma family generated by $\nu(d\lambda) = \frac{\lambda^{p-1}}{\Gamma(p)}\mathbf{1}_{(0,\infty)}(\lambda)\, d\lambda$ a simple calculation gives that H_ψ is a negative binomial NEF with variance function

$$V_{H_\psi}(m) = m + \frac{\psi}{p}m^2.$$

4. The case of multivariate exponential families

A *line multivariate gamma* distribution governed by a nonzero vector $(a_1, \dots, a_d)$ in $[0, \infty)^d$ and the parameter p is the distribution of the random variable $X = (a_1 Y, \dots, a_d Y)$ where Y is a real random variable with distribution $\gamma_{p,1}$. Its Laplace transform is

$$L_\mu(z) = \mathbb{E}(e^{\langle \theta, X\rangle}) = (1 - a_1\theta_1 - \cdots - a_d\theta_d)^{-p}.$$

Its image by $\mu \mapsto MP(\mu)$ is the negative multinomial distribution on $\mathbb{N}^d$ with generating function

$$\mathbb{E}(z_1^{N_1} \dots z_d^{N_d}) = c^p(1 - c(a_1 z_1 + \cdots + a_d z_d))^{-p}$$

where $c = (1 + a_1 + \cdots + a_d)^{-1}$. If some of the $a_i's$ are zero, (say $a_i > 0$ if and only if $i \leqslant m$) then the half line image of $[0, \infty)$ by $y \mapsto (a_1 y, \dots, a_m y)$ is concentrated on $[0, \infty)^m$ and the corresponding distribution is concentrated on $\mathbb{N}^m$.

For an integer $0 \leqslant q$ consider $q + 1$ subsets of $\{1, \dots, d\}$ denoted by $\{T_0, \dots, T_q\}$ and such that $\{1, \dots, d\} = \cup_{m=0}^q T_m$, $T_m \neq \emptyset, \forall m \geqslant 1$ and $T_i \cap T_j = \emptyset, \forall i \neq j$. Consider a product of line multivariate gamma distributions concentrated on $[0, \infty)^{T_m}$ for $m = 1, \dots, q$ and a Dirac mass on $[0, \infty)^{T_0}$. More specifically, consider a distribution ν on $[0, \infty)^d$ such that there exist nonnegative numbers $a_1, \dots, a_d$ and positive numbers $p_1, \dots, p_q$ such that the Laplace transform of ν is

$$L_\nu(\theta) = e^{\sum_{k \in T_0} a_k\theta_k} \prod_{m=1}^q \left(1 - \sum_{k \in T_m} a_k\theta_k\right)^{-p_m}. \tag{6}$$

Another presentation of these distributions can be helpful. If $v = (v^{(1)}, \dots, v^{(d)})$ is in $[0, \infty)^d$ let us call support of ν the set of $i \in \{1, \dots, d\}$ such that $v^{(i)} > 0$. If we now define $v_m^{(i)} = a_i \mathbf{1}_{T_m}(i)$ then the $q + 1$ vectors $v_0, \dots, v_q$ of $\mathbb{R}^d$ have disjoint supports. Introduce the random variables Y_m with distribution $\gamma_{p_m,1}$ such that $(Y_1, \dots, Y_q)$ are independent. Then ν defined by (**??**) is the distribution of $v_0 + v_1 Y_1 + \cdots + v_q Y_q$. Conversely if $v_0, \dots, v_q$ have disjoint supports the distribution of $v_0 + v_1 Y_1 + \cdots + v_q Y_q$ has type (**??**). These distributions (at least for $v_0 = 0$) have been characterized in Bobecka and Wesołowski [2] as follows: suppose that X and X' are independent random variables of $(0, \infty)^d$. Write

$$\frac{X}{X + X'} = \left(\frac{X_1}{X_1 + X_1'}, \dots, \frac{X_d}{X_d + X_d'}\right)$$

Then they obtain an elegant multivariate version of a theorem due to Lukacs [9]: the random variables $X + X'$ and $\frac{X}{X+X'}$ are independent if and only if there exists a nonzero sequence of vectors $(v_1, \dots, v_q)$ in $[0, \infty)^d$ with disjoint supports and independent standard gamma variables $(Y_1, \dots, Y_q, Y_1', \dots, Y_q')$ such that $X \sim v_1 Y_1 + \dots + v_q Y_q$ and $X' \sim v_1 Y_1' + \dots + v_q Y_q'$.

The real domain $D(\nu)$ of existence of this Laplace transform (**??**) is open and is the set Θ of θ_k's such that $1 - \sum_{k \in T_m} a_k \theta_k > 0$ for all $m = 1, \dots, q$. For $\theta \in \Theta$, the element ν_θ of the natural exponential family F generated by ν has the following Laplace transform:

$$h \mapsto \frac{L_\nu(\theta + h)}{L_\nu(\theta)} = e^{\sum_{k \in T_0} a_k h_k} \prod_{m=1}^{q} \left[1 - r_m \sum_{k \in T_m} a_k h_k \right]^{-p_m},$$

where $r_m = r_m(\theta) = (1 - \sum_{k \in T_m} a_k \theta_k)^{-1}$. Note that ν_θ is also a product of line multivariate gamma distributions. The reader can verify that the family $MP(F) = \{MP(\nu_\theta);\ \theta \in \Theta\}$ is indeed a natural exponential family generated by $MP(\nu)$ (warning: the parametrization $\theta \mapsto MP(\nu_\theta)$ of $MP(F)$ is not the canonical one). The next theorem shows that we have obtained in this way all natural exponential families F such that $MP(F)$ is also an exponential family. It is an extension of the above Theorem 1.

Theorem 2. *If the image of the NEF F on $[0, \infty)^d$ by $\mu \mapsto MP(\mu)$ is still a natural exponential family, then there exists $q + 1$ disjoints subsets of $\{1, \dots, d\}$ denoted by $\{T_0, T_1, \dots, T_q\}$ and there exist nonnegative numbers $a_1, \dots, a_d$ and positive numbers $p_1, \dots, p_q$ such that F has a generating measure ν with Laplace transform* (**??**).

Proof. Similar to the proof of Theorem 1, the case where F is concentrated on a subspace of $\mathbb{R}^d$ such as $(0, \dots, 0) \times \mathbb{R}^q$ with $q < d$ is discarded. This case leads to mixed Poisson distributions concentrated on $(0, \dots, 0) \times \mathbb{N}^q$. Denote by ν an arbitrary generating measure of F and $L(\theta)$ its Laplace transform defined as

$$L(\theta) = \int_{[0,\infty)^d} e^{\langle \lambda, \theta \rangle} \nu(d\lambda),$$

for $\theta \in \Theta$, where Θ is the interior of the domain of convergence of $L(\theta)$. Note that the fact that ν is concentrated on $[0, \infty)^d$ implies that $\Theta + a \subset \Theta$ for any $a = (a_1, \dots, a_d)$ such that $a_i \leqslant 0$. Suppose that the image of $F(\nu)$ by $\mu \mapsto MP(\mu)$ is an NEF on $\mathbb{N}^d$ generated by some measure $\sum_{n \in \mathbb{N}^d} p_n \delta_n$. We write $\mathbf{1} \in \mathbb{R}^d$ for the vector with components equal to 1. We use the standard notations $\lambda^n = \lambda_1^{n_1} \cdots \lambda_d^{n_d}$ and $n! = n_1! \cdots n_d!$, for $n = (n_1, \dots, n_d)$ in $\mathbb{N}^d$ and λ in $[0, \infty)^d$. Similarly $L^{(n)}(\theta)$ means

$$\frac{\partial^{n_1}}{\partial \theta_1^{n_1}} \cdots \frac{\partial^{n_d}}{\partial \theta_d^{n_d}} L(\theta).$$

Thus there exist two functions $\alpha : \Theta + \mathbf{1} \to \mathbb{R}^d$ and $\beta : \Theta + \mathbf{1} \to \mathbb{R}$ such that

$$\int_{[0,\infty)^d} e^{\langle \lambda, (\theta - \mathbf{1}) \rangle} \frac{\lambda^n}{n!} \nu(d\lambda) = p_n e^{\langle n, \alpha(\theta) \rangle + \beta(\theta)},$$

which can be rewritten

$$L^{(n)}(\theta - \mathbf{1}) = n! p_n e^{\langle n, \alpha(\theta) \rangle + \beta(\theta)} \quad \forall n \in \mathbb{N}^d. \tag{7}$$

A discussion similar to that of Theorem 1 shows that $p_n > 0$ for all $n \in \mathbb{N}^d$, since ν is not concentrated on some subspace of type $(0, \dots, 0) \times \mathbb{R}^q$. As a consequence, the real-analyticity of α and β on $\Theta + \mathbf{1}$ can be deduced of the analyticity of L, by imitating again the proof of Theorem 1. Denote $e_i = (0, \ldots, 1, \ldots, 0) \in \mathbb{N}^d$, where the unique 1 is in position i. We also write $\alpha(\theta) = (\alpha_1, \ldots, \alpha_d)$, $\alpha_{ij} = \frac{\partial \alpha_j}{\partial \theta_i}$ and $\beta_i = \frac{\partial \beta}{\partial \theta_i}$. By taking the logarithms of both sides of (**??**) and applying $\frac{\partial}{\partial \theta_i}$ we get

$$\frac{L^{(n+e_i)}(\theta - \mathbf{1})}{L^{(n)}(\theta - \mathbf{1})} = \left\langle n, \frac{\partial \alpha(\theta)}{\partial \theta_i} \right\rangle + \frac{\partial \beta(\theta)}{\partial \theta_i} = \sum_{k=1}^{d} n_k \alpha_{ik} + \beta_i. \tag{8}$$

This last equality implies

$$\alpha_{ij}\alpha_{jk} = \alpha_{ji}\alpha_{ik}, \tag{9}$$
$$\alpha_{ij}\beta_j = \alpha_{ji}\beta_i, \tag{10}$$

for all i, j, k in $\{1, \ldots, d\}$. Indeed, by using (**??**) and $\frac{\partial^2}{\partial \theta_i \partial \theta_j} = \frac{\partial^2}{\partial \theta_j \partial \theta_i}$, the quantity $L^{(n+e_i+e_j)}(\theta - \mathbf{1})$ can be written in two ways:

$$\begin{aligned} L^{(n+e_i+e_j)}(\theta - \mathbf{1}) &= L^{(n)}(\theta - \mathbf{1}) \left(\sum_{k=1}^{d} n_k \alpha_{ik} + \beta_i + \alpha_{ij} \right) \left(\sum_{k=1}^{d} n_k \alpha_{jk} + \beta_j \right), \\ &= L^{(n)}(\theta - \mathbf{1}) \left(\sum_{k=1}^{d} n_k \alpha_{ik} + \beta_i \right) \left(\sum_{k=1}^{d} n_k \alpha_{jk} + \beta_j + \alpha_{ji} \right). \end{aligned}$$

Since this equality holds for all n, (**??**) and (**??**) are easily obtained.

Assume first that $\alpha_{ij} \neq 0$ for all i and j (separating this case is not absolutely necessary but makes the reading easier) and fix $\theta - \mathbf{1}$ (as we did in the proof of Theorem 1). In this case, (**??**) and (**??**) imply that there exist numbers $a_1, \ldots, a_d$ and p such that $a_i = \alpha_{ij}$ for all j and $p = \beta_i / \alpha_{ij}$ for all i and j. Equality (**??**) can then be written

$$\frac{L^{(n+e_i)}(\theta - \mathbf{1})}{L^{(n)}(\theta - \mathbf{1})} = a_i \left(p + \sum_{k=1}^{d} n_k \right).$$

As a consequence, the following result can obtained:

$$\frac{L(\theta - \mathbf{1} + h)}{L(\theta - \mathbf{1})} = \left(1 - \sum_{k=1}^{d} a_k h_k \right)^{-p}, \tag{11}$$

after noting

$$\frac{1}{(1 - \sum_{k=1}^{d} a_k h_k)^p} = \sum_{s=0}^{\infty} (p)_s \frac{1}{s!} \left(\sum_{k=1}^{d} a_k h_k \right)^s = \sum_{s=0}^{\infty} (p)_s \sum_{n;\ n_1 + \cdots + n_d = s} \frac{1}{n!} \prod_{k=1}^{d} (a_k h_k)^{n_k}.$$

Consider now the implications of (**??**) and (**??**) in the general case where some α_{ij} can be 0. For this, consider a directed graph $G = (V, E)$ whose set of vertices $V = \{1, \ldots, d\}$ is such that (i, j) is an edge if and only if $\alpha_{ij} \neq 0$. We also write $i \to j$ instead of $\alpha_{ij} \neq 0$ or $(i, j) \in E$ and $i \leftrightarrow i$ when the loop (i, i) exists (this loop may or may not exist). Suppose that there exists k such that $\alpha_{ki} = 0$ for all i. Eq. (**??**) can then be written

$$\frac{L^{(n+e_k)}(\theta - \mathbf{1})}{L^{(n)}(\theta - \mathbf{1})} = \beta_k.$$

Thus, for any integer n_k, we have $L^{(n_k e_k)}(\theta - \mathbf{1}) = L(\theta - \mathbf{1})(\beta_k)^{n_k}$. After multiplication by $h_k^{n_k}/n_k!$ and summation (with respect to n_k) we obtain

$$L(\theta - \mathbf{1} + h) = L(\theta - \mathbf{1})e^{\beta_k h_k},$$

for $h = h_k e_k$. More generally, denote by T_0 the set of k such that there is no i such that $k \to i$. The above reasoning shows that

$$L(\theta - \mathbf{1} + h) = L(\theta - \mathbf{1})e^{\sum_{k \in T_0} \beta_k \theta_k}, \quad h = \sum_{k \in T_0} h_k e_k.$$

Some definitions about graphs need to be recalled. Consider a directed graph $G = (V, E)$, where V is a finite set and $E \subset V \times V$. The graph $G_1 = (V_1, E_1)$ is called a subgraph of G if $V_1 \subset V$ and $E_1 \subset E \cap (V_1 \times V_1)$. Furthermore, G_1 is the *induced* graph on V_1 if $E_1 = E \cap (V_1 \times V_1)$. The following result can be easily obtained:

Lemma. *Consider the graph G defined as above on $V=\{1, \ldots, d\}$ by the matrix (α_{ij}) satisfying* (**??**)*. Then*

1. *Let i and j be distinct in V. If the induced subgraph G_1 on $V_1 = \{i, j\}$ contains either the subgraph $i \to j \leftrightarrow j$ or the subgraph $i \leftrightarrow j$, then the induced graph is $i \leftrightarrow i \leftrightarrow j \leftrightarrow j$.*
2. *If the induced subgraph G_1 on $V_1 = \{i, j, k\}$ contains the subgraph $i \to j \to k$ then G_1 contains the subgraph $k \leftarrow i \leftrightarrow i \leftrightarrow j \leftrightarrow j \to k$.*
3. *If the induced subgraph G_1 on $V_1 = \{i, j\}$ is either the subgraph $i \leftrightarrow i \to j$ or the subgraph $i \to j$, then $\beta_j = 0$.*

These results are illustrated in Fig. **??**. The proof of the lemma involves the three following cases:

1. If $i \to j \leftrightarrow j$, by setting $k = j$ and $k = i$ in (**??**), we obtain $\alpha_{ji} \neq 0$ and $\alpha_{ii} \neq 0$. If $i \leftrightarrow j$, by setting $k = j$ and $k = i$ in (**??**), we obtain $\alpha_{jj} \neq 0$ and $\alpha_{ii} \neq 0$.
2. (**??**) imply $\alpha_{ji} = \alpha_{jk} = \alpha_{jj} = a_j$ and $\alpha_{ij} = \alpha_{ii} = \alpha_{ik} = a_i$. If $\alpha_{ij} \neq 0$, we obtain $\alpha_{ii} \neq 0$ and $\alpha_{ik} \neq 0$. Similarly, if $\alpha_{jk} \neq 0$, we obtain $\alpha_{jj} \neq 0$ and $\alpha_{ji} \neq 0$.
3. Apply (**??**).

We come back to the proof of Theorem 2. Define the relation $i \sim j$ on $V = \{1, \ldots, d\}$ by either $i = j$ or the induced graph on $\{i, j\}$ is $i \leftrightarrow i \leftrightarrow j \leftrightarrow j$. It is easy to deduce from the Lemma that $\sim$ is an equivalence relation. We remark that this implies that each element of T_0 is alone in its equivalence class. Recall also that the definition of T_0 implies that there are no arrows between two elements of T_0. Denote the other equivalence classes by $T_1, \ldots, T_q$.

Suppose now that there exists $i \in \bigcup_{m=1}^q T_m$ and $k \in T_0$ such that $i \to k$. Then part 3 of the Lemma implies that $\beta_k = 0$. Eq. (**??**) can be used to prove that $L^{(n+e_k)} = 0$ for all $n \in \mathbb{N}^d$. Thus, $h \mapsto L(\theta - \mathbf{1} + h)$ does not depend on h_k. Since $\beta_k = 0$, this implies that μ is concentrated on $\{\lambda \in \mathbb{R}^d;\ \lambda_k = 0\}$, a case which has been excluded from the beginning. There are finally no arrows between $\bigcup_{m=1}^q T_m$ and T_0.

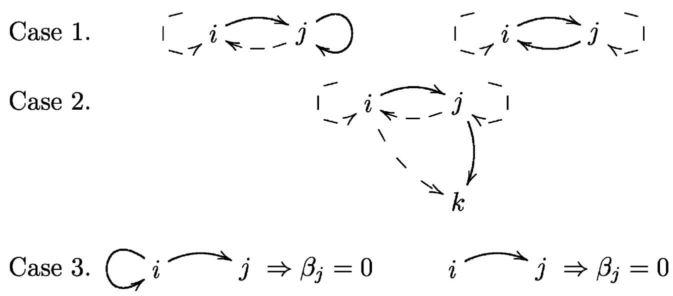

Fig. 1. Illustration of the lemma.

As a summary, the picture of the graph G is

1. a collection T_0 of vertices without any arrow and any loop.
2. q disjoint classes $T_1, \ldots, T_q$ without arrows between vertices of different classes, and with all possible arrows (including loops) inside a same class T_m.

Consider a fixed m in $\{1, \ldots, q\}$. For all i, j in T_m, by setting $k = i$ in (**??**), we obtain $\alpha_{ij}\alpha_{ji} = \alpha_{ji}\alpha_{ii}$. Thus, $\alpha_{ji} \neq 0$ implies $\alpha_{ij} = \alpha_{ii}$. By using (**??**), for $i \in T_m$ and for any $n \in \mathbb{N}^d$, the following result can be obtained:

$$\frac{L^{(n+e_i)}(\theta - \mathbf{1})}{L^{(n)}(\theta - \mathbf{1})} = \alpha_{ii}\left(\frac{\beta_i}{\alpha_{ii}} + \sum_{k \in T_m} n_k\right). \tag{12}$$

Recall that the number $p_m = \beta_i/\alpha_{ii}$ does not depend on i when i runs T_m (indeed $\alpha_{ij} = \alpha_{ii}$ and use (**??**)). Denote $a_k = \alpha_{kk}$ as above. The imitation of the proof of (**??**) and the formula (**??**) lead to:

$$\frac{L(\theta - \mathbf{1} + h)}{L(\theta - \mathbf{1})} = \left(1 - \sum_{k \in T_m} a_k h_k\right)^{-p_m}, \tag{13}$$

for any $h = \sum_{k \in T_m} h_k e_k$. This concludes the proof of theorem. $\square$

References

[1] S. Bar-Lev, P. Enis, Reproducibility and natural exponential families with power variance functions, Ann. Statist. 14 (1986) 1507–1522.
[2] K. Bobecka, J. Wesołowski, Multivariate Lukacs theorem, J. Multiv. Anal. 91 (2004) 143–160.
[3] A. Ferrari, G. Letac, J.-Y. Tourneret, Multivariate mixed Poisson distributions, in: Proceedings of 12th Conference on Signal Processing (EUSIPCO'04), Vienna, Austria, September 6–10, 2004, pp. 1067–1070.
[4] J. Goodman, Statistical Optics, Wiley, New York, 1985.
[5] J. Grandell, Mixed Poisson Processes, Chapman & Hall, London, 1997.
[6] P. Hougaard, M.-T.T. Lee, G.A. Whitmore, Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes, Biometrics 53 (1997) 1225–1238.
[7] B. Jorgensen, The Theory of Dispersion Models, Chapman & Hall, London, 1997.
[8] G. Letac, M. Mora, Natural exponential families with cubic variance functions, Ann. Statist. 18 (1990) 1–37.
[9] E. Lukacs, A characterization of the gamma distribution, Ann. Math. Statist. 26 (1955) 319–324.
[10] C.N. Morris, Natural exponential families with quadratic variance functions, Ann. Statist. 10 (1982) 65–80.