ISSN 1066-5307, Mathematical Methods of Statistics, 2017, Vol. 26, No. 3, pp. 159–175. © Allerton Press, Inc., 2017.

On the Mean Value Parametrization of Natural Exponential Families – a Revisited Review

S. K. Bar-Lev[1*] and C. C. Kokonendji[2]**

[1]*Dept. of Statist., Univ. of Haifa, Israel; NYU Shanghai, China*

[2]*Univ. Bourgogne Franche-Comté Besançon, France*

Received April 18, 2017; in final form, August 7, 2017

Abstract—It is well known that any natural exponential family (NEF) is characterized by its variance function on its mean domain, often much simpler than the corresponding generating probability measures. The mean value parametrization appeared to be crucial in some statistical theory, like in generalized linear models, exponential dispersion models and Bayesian framework. The main aim of the paper is to expose the mean value parametrization for possible statistical applications. The paper presents an overview of the mean value parametrization and of the characterization property of the variance function for NEF's. In particular it introduces the relationships existing between the NEF's generating measure, Laplace transform and variance function as well as some supplemental results concerning the mean value representation. Some classes of polynomial variance functions are revisited for illustration. The corresponding NEF's of such classes are generated by counting probabilities on the nonnegative integers and provide Poisson-overdispersed competitors to the homogeneous Poisson distribution.

Keywords: natural exponential families, exponential dispersion models, variance functions, polynomial variance functions, Poisson-overdispersed distribution.

2000 Mathematics Subject Classification: 60E10, 62E10.

DOI: 10.3103/S1066530717030012

1. INTRODUCTION

Natural exponential families (NEFs) on $\mathbb{R}$ (or on $\mathbb{R}^k$) play an important role both in probability and statistical applications. Most of the frequently used distributions are indeed belonging to such families. However, as will be described in the sequel, a huge number of NEFs have not been used in probabilistic or statistical modelling for a mainly one reason: they have not yet been revealed, although they could provide new models useful in applications.

There are various reasons why parametric models are employed for modelling. Some of them are obtained as limiting distributions of sample functionals (as the normal distribution and extreme value distributions). Others are obtained as a result of a probabilistic characterization (as stable and self decomposable distributions), but mainly some have been used as they are analytically tractable for statistical and probabilistic analysis as the Poisson and exponential distributions. Just for example, the Poisson distribution (or process) has been used for decades (if not centuries) to model counting characteristics of random phenomena (e.g., the number of insurance claims or an arrival process in queueing modelling aspects). The exponential distribution has been used, among other targets, to model survival data. The latter two distributions are manipulable and easily managed for obtaining relatively easy explicit relevant expressions.

For quite some time it is known that both distributions do not necessarily reflect real data analysis. The $M/M/k$ or $M/G/k$ (i.e., Poisson arrival, exponential service time M – or arbitrary service time distribution G, respectively – and k servers) policies that have been used in the early beginning of queueing modelling are now known not to really describe real data of queueing processes. As in practice

*E-mail: barlev@stat.haifa.ac.il

**E-mail: celestin.kokonendji@univ-fcomte.fr

the arrival process into a queueing system (or call centers) is rather much more complicated than that described by the homogeneous Poisson process, and likewise is the exponential service time of a server. Also, it is well known that Poisson distribution does not really reflect the number of claims into an insurance system or describing mortality rates (e.g., [22, 20, 14]), as real data are overdispersed and zero-inflated; c.f., [40] and the references cited therein. Moreover, the exponential distribution is hardly served in practice as a distribution for describing the survival lifetime to failure of electronic (or other) components (as had been used for several decades of the last century) and is often replaced by more suitable parametric distributions or by nonparametric modelling.

Nowadays, we are well aware of the fact that real data are more complex than just describing them by the above two distributions (or some other well used distributions). It seems to us that distributions such as the above have been frequently utilized for their tractability to obtain nice and explicit expressions and by that avoiding a computational complexity. In a way, such a use of 'light' distributions for inference appears like 'looking for the lost coin under the first light street'. Notwithstanding, with the availability of powerful mathematical software as MATHEMATICA, R, S and Matlab, it appears that intractable and more complex functional forms of competitive distributions is no longer a hindrance for computation and data analysis and can be used accordingly. A referee correctly argued that "neither numerical calculations nor lack of available software, never were a major obstacle, at least for the last several decades, in addressing well known statistical problems". Nonetheless, we trust that the complexity of the exact analytic forms of various distributions (as will be demonstrated in the sequel) is still often an obstacle in implementing such distributions. This is in spite of the availability of advanced mathematical software that could in principle cope with these distributions.

The main purpose of this study is to expose the mean value parametrization of NEFs and their associated variance functions for possible statistical applications. The latter parametrization is sometimes available in closed analytical form whereas the related Laplace transform is either unknown or rather has an intractable and cumbersome form. Moreover, the mean value parametrization provides and generates huge classes of NEFs with rather tractable variance functions that have not been known before. This will be exemplified by illustrative cases and examples, particularly in the Bayesian framework.

This paper is organized as follows. In Section 2 we review some background material on NEFs and exponential dispersion models (EDMs), their related variance functions. Section 3 is devoted to the mean value parametrization. In particular, it introduces the relationships existing between the NEF's generating measure, Laplace transform and variance function as well as some supplemental results concerning the mean value representation. A class of polynomial variance functions is introduced in Section 4 along with some of its subclasses associated with NEF's composed of counting distributions. Such subclasses can be smoothly utilized in Bayesian approach. Illustrations and applications are then provided. Some concluding remarks are presented in Section 5.

Most of the exposition of the paper is a collection of relevant results in the literature. We however present also some new results (particularly Proposition 1, Lemma 2 and Corollary 3) which aim at clarifying some indistinct related results.

2. NEF'S — A GENERAL DESCRIPTION

2.1. Preliminaries

The following preliminaries are mainly taken from [34]. Let μ be a non-Dirac positive Radon measure on $\mathbb{R}$, S_μ the support of μ, and C_μ the convex-hull of S_μ. The Laplace transform (LT) of μ is a mapping $L_\mu\colon \mathbb{R} \to (0,\infty]$ defined by

$$L_\mu(\theta) = \int_{\mathbb{R}} e^{\theta x}\mu(dx).$$

Let D_μ denote the effective domain of μ, i.e., $D_\mu = \{\theta \in \mathbb{R}\colon L_\mu(\theta) < \infty\}$, assume that $\Theta_\mu = \operatorname{int} D_\mu \neq \emptyset$ and let $k_\mu(\theta) = \log L_\mu(\theta)$ be the cumulant transform of L_μ. Also, let $\mathcal{M}(\mathbb{R})$ denote the set of positive measures μ on $\mathbb{R}$ not concentrated on one point such that $\Theta_\mu \neq \emptyset$. Then the NEF F generated by μ is defined by probabilities

$$F = F(\mu) = \big\{P(\theta, \mu(dx)) = \exp\{\theta x - k_\mu(\theta)\}\mu(dx),\ \theta \in \Theta_\mu\big\}. \tag{1}$$

The Mean value parametrization. The cumulant transform k_μ is strictly convex and real analytic on Θ_μ and

$$k'_\mu(\theta) = \int_{\mathbb{R}} x \exp\{\theta x - k_\mu(\theta)\}\, \mu(dx)$$

is the mean function of F. The open interval $M_F = k'_\mu(\Theta_\mu)$ is called the mean domain of F. Since the map $\theta \longmapsto k'_\mu(\theta)$ is one-to-one, its inverse function $\psi_\mu\colon M_F \longrightarrow \Theta_\mu$ is well defined. Hence the map $m \longmapsto P(m, F) = P(\psi_\mu(m), \mu)$ is one-to-one from M_F onto F and is called the mean domain parametrization of F.

The Variance Function. The variance corresponding to $P(m, F)$ is

$$V_F(m) = 1/\psi'_\mu(m) = k''_\mu(\theta). \tag{2}$$

The map $m \longmapsto V_F(m)$ from M_F into $\mathbb{R}^+$ is called the variance function (VF) of F. In fact a VF of an NEF F is a pair (V_F, M_F). It uniquely determines an NEF within the class of NEFs. Morris [37] characterized all NEFs having quadratic VFs (six families) and Letac and Mora [34] all NEFs with strict cubic VFs (six families). Bar-Lev [2] showed that any kth degree polynomial with nonnegative coefficients constitutes an infinitely divisible VF with mean domain $\mathbb{R}^+$. Tweedie [45], Bar-Lev and Enis [10] and Jørgensen [24] have considered, in different contexts, the class of NEFs with power variance functions (NEF-PVFs) of the form $V(m) = bm^\gamma$, for some $b > 0$, $\gamma \in \mathbb{R}$, and where $M = \mathbb{R}$ or $\mathbb{R}^+$ (depending on γ). This class is huge and contains the following families: normal (γ =0), Poisson ($\gamma = 1$), gamma (γ =2), inverse Gaussian ($\gamma = 3$), compound Poisson NEFs generated by all gamma distributions ($1 < \gamma < 2$), NEFs generated by extreme stable distributions with stable index belonging to the open interval $(0, 1)$ ($\gamma > 2$). All of the above NEFs are steep (see definition below) and $M = \operatorname{int} C = \mathbb{R}^+$ (c.f., [10]). The remaining subclass of NEF-PVFs with $\gamma < 0$ are NEFs generated by extreme stable distributions with stable index belonging to the interval $(1, 2)$. The latter subclass is composed of nonsteep NEFs with $M = \mathbb{R}^+$ and $\operatorname{int} C = \mathbb{R}$. All such NEF-PVFs are infinitely divisible. No NEFs-PVFs exist if $\gamma \in (0, 1)$. All of the NEFs with power VFs are often refereed to as Tweedie class.

A steep NEF. An NEF $F = F(\mu)$ is called steep if and only if $M_F = \operatorname{int} C_\mu$. An equivalent condition to steepness of F is presented in Theorem 2.1 in [34] and it states the following. Let $\Theta_\mu = (c, d)$, $-\infty \le c < d \le \infty$, then F is steep if and only if the two following conditions hold: (i) either $c \notin D_\mu$ or $c \in D_\mu$ and $\lim_{\theta \downarrow c} k'_\mu(\theta) = -\infty$; and (ii) either $d \notin D_\mu$ or $d \in D_\mu$ and $\lim_{\theta \uparrow d} k'_\mu(\theta) = \infty$. If F is regular, i.e., $D_\mu = \Theta_\mu$, then F is steep.

Bases of F. The measure μ generating $F = F(\mu)$ is called a basis. The basis of F is not unique. If μ and μ' belong to $\mathcal{M}(\mathbb{R})$, then $F(\mu) = F(\mu')$ if and only if there exists $(a, b) \in \mathbb{R}^2$ such that $\mu'(dx) = e^{ax+b}\mu(dx)$, implying that all bases of F generate the same VF (V_F, M_F).

The Jørgensen set of NEF and Exponential Dispersion Models (EDMs). EDMs have been studied thoroughly by Jørgensen [25, 26] and others, suggesting them to describe the error component in generalized linear models (GLIM). An EDM is related to an NEF as follows. The Jørgensen set $\Lambda = \Lambda_F$ associated with F is defined by

$$\Lambda = \big\{\lambda \in \mathbb{R}^+\colon \lambda k_\mu(\theta) \text{ is a cumulant transform of some measure } \mu_\lambda \text{ on } \mathbb{R}\big\}, \tag{3}$$

and is nonempty since by convolution it contains $\mathbb{N}$.

It is worthy of note the following equivalent statements:

$$F(\mu) \text{ is composed of infinitely divisible distributions} \Longleftrightarrow \\ \mu \text{ is infinitely divisible} \Longleftrightarrow \Lambda = \mathbb{R}^+. \tag{4}$$

For $\lambda \in \Lambda$, the NEF F_λ generated by μ_λ is

$$F_\lambda = F_\lambda(\mu_\lambda) = \big\{P(\theta, \lambda, \mu_\lambda(dx)) = \exp\{\theta x - \lambda k_\mu(\theta)\}\mu_\lambda(dx),\ \theta \in \Theta_\mu\big\}, \tag{5}$$

where the support of μ_λ may depend on λ.

For $\lambda \in \Lambda$, the mean function, the mean domain and VF of the NEF F_λ, denoted by m_λ, M_λ and V_λ, respectively, are given by

$$m_\lambda = \lambda\kappa'(\theta) = \lambda m, \qquad M_\lambda = \lambda M_F$$

and

$$V_\lambda(m) = \lambda\kappa''_\mu(\theta) = \lambda V_F(m), \quad m \in M_F, \qquad \text{or} \qquad V_\lambda(m_\lambda) = \lambda V_\lambda(m_\lambda/\lambda), \quad m_\lambda \in M_\lambda. \quad (6)$$

The set of probabilities

$$\mathbb{G} = \{P(\theta, \lambda, \mu_\lambda) \colon \theta \in \Theta_\mu, \lambda \in \Lambda\},$$

is called the EDM associated with F and is parameterized by $(\theta, \lambda) \in \Lambda \times \Theta_\nu$. The parameter $\sigma^2 = 1/\lambda$ is termed the dispersion parameter.

As can be seen, EDMs form a large class of models for data modelling and are used, among other applications, to describe the error component in generalized linear models, c.f., [39] and [24, 25].

The properties of infinite divisibility and steepness are important for the present study. Infinite divisibility ensures that $\Lambda = \mathbb{R}^+$ and thus EDM distributions resemble in GLIM the role of the normal distribution in simple regression models for describing the distribution of the error component. Steepness of F (or μ) ensures that the MLE of m exists with probability one and is given as the unique solution of the maximum likelihood equation $\hat{m} - \bar{X}_n = 0$ (c.f., [13], Chapter 9), where $\bar{X}_n$ is the sample mean based on n independent replicas $X_1, \ldots, X_n$ of X, where the distribution of X belongs to (1).

2.2. Some Statistical Applications of VFs Related to NEFs and EDMs

Various applications of VFs, NEFs and EDMs appeared in the statistical literature. Here we mention only a few of them.

1. *Statistical modelling*. Numerous works have dealt with the applications of VFs of NEFs and EDMs for statistical modelling by using GLIM. For example, the Tweedie class has been applied to actuarial problems related to insurance claims (cars and life insurance), c.f., [44, 16, 17, 46] and the references cited therein.

2. *Estimation*. EDMs with unknown mean and dispersion parameters were studied in [12] with respect to second-order minimax estimation of the mean. Bar-Lev, Bshouty and Landsman [9] studied the problem of the improvement of the sample mean in the second order minimax estimation sense for a mean belonging to an unrestricted mean. They solved such a problem for the class of NEFs whose VFs are regular at zero and at infinity. Such a class of VFs is huge and contains (among others): Polynomial VF's (e.g., quadratic VFs in the Morris class, cubic VFs in the Letac and Mora class and VFs in the Hinde—Demétrio class); VFs belonging to the Tweedie class with power VF's, VFs belonging to the Babel class (see [32, 33]) and many others.

3. *Prior distribution and information theory*. Jeffreys and Shtarkov distributions play an important role in universal coding and minimum description length inference, two central areas within the field of information theory. It was recently discovered that in some situations Shtarkov distributions exist while Jeffreys distributions do not. Some of these situations were considered in [8] in which they constructed numerous classes of infinitely divisible NEFs for which Shtarkov distributions exist and Jeffreys do not. The method used to obtain these general results was based on the VFs of such NEFs.

2.3. Relationships between the Generating Measure of an NEF and Its LT and VF

There are infinitely many $\mu \in \mathcal{M}(\mathbb{R})$ and such are the associated LT's, NEF's and VF's. Various cases of relationships exist among the triple: the generating measure μ of F, the LT L_μ and the associated VF (V_F, M_F). We outline only some of them with relevant examples.

(i) Cases in which the three of them are known and have nice explicit forms (e.g., NEF's with quadratic VF's including, among others, the normal, gamma, binomial, negative binomial and Poisson NEF's).

(ii) Cases in which the three are explicitly known but either one or two of them have intractable forms. For example, positive stable measures with stable index $0 < \rho < 1$ have LT's of the form $\exp(-(-\theta)^\rho), \theta < 0$, and VF's of the form

$$(V, M) = \big(\rho(1-\rho)(m/\rho)^{\frac{\rho-2}{\rho-1}}, \mathbb{R}^+\big),$$

whereas μ (except for the inverse Gaussian measure with $\rho = 1/2$) can be either represented as an infinite sum, or in terms of some transcendental function. Note that stable distributions have been obtained by a probabilistic characterization and also appear as limiting distributions for some stochastic processes; c.f., [35, 36].

(iii) Cases where both μ and its VF are explicitly specified whereas the respective LT is a solution of an implicit equation. An example of this type that will be considered in the sequel (see Example 1) is the Ressel distribution; c.f., [41, 34, 11, 3] (and various references cited therein). Its probability density function and VF are given by

$$\mu(dx) = \frac{x^{x-1}e^{-x}}{\Gamma(x+1)}(dx), \quad x > 0, \quad (V, M) = \big(m^2(1+m), \quad \mathbb{R}^+\big), \tag{7}$$

whereas its LT is the solution of an implicit equation which was first obtained by Prabhu [41], p. 73, to describe the LT of the busy period in an $M/G/1$ queue.

(iv) Cases in which only the VF is explicit. Such cases are important as they apparently reveal new NEF's that have not been used in statistical modelling. Examples of cases of this kind are (almost) all polynomials of degree ≥ 4 whose generating measures are concentrated on the set of nonnegative integers. These are well exemplified in Section 4.

2.4. Derivation of Cumulants via the VF

Needless to say that the mean value parametrization is of a more interest for statistical applications as the canonical parameter θ is a somewhat artificial parameter for any statistical usage. Nonetheless, as the present study focuses on some practical aspects, the mean value parametrization is useful whenever both integrals of $\psi_\mu(m)$ and $k_\mu(\psi_\mu(m))$ in (10) are expressible in terms of m. Note that if this is indeed the case, and whether or not μ is known, all moments of the corresponding NEF F are also nicely expressible in terms of m. This follows as the rth cumulant of F is $k^{(r)}(\theta) \doteq dk^r(\theta)/d\theta^r$, where $k^{(r)}$ denotes the rth derivative of k (here, when no ambiguity is caused, we suppress the dependence of k_μ, Θ_μ, V_F on μ and F, and so on). By following [7] and defining an operator L acting on V by $L(V) \equiv L_1(V) = VV'$ and $L_n(V) = L(L_{n-1}(V))$, $n \in \mathbb{N}$, with $L_0(V) = V$, the rth cumulant of F, expressed in terms of m, is given by

$$k_{r+2}(m) \doteq k^{(r+2)}(\psi_\mu(m)) = L_r(V(m)) \quad \text{for all} \quad r = 0, 1, \ldots \quad \text{and} \quad m \in M, \tag{8}$$

where $k_j = k_j(m)$ stands for the jth cumulant expressed in terms of m. Consequently, the skewness and kurtosis of F are easily obtained.

3. THE MEAN VALUE PARAMETRIZATION

As there are necessary and/or sufficient conditions for a function to be a LT there are also such conditions for a pair (V, M) to be a VF of an NEF (some of which will be introduced in the sequel), with or without knowing the corresponding measure μ.

Given an LT (or a VF) related to a measure μ, theoretically it is possible to reveal μ by using the inversion theorem for LT's. Practically, however, in most cases this mission seems to be impossible. Likewise is the situation with a given VF, where it is not possible to reveal the generating measure except for some few cases. We shall deal with this aspect, but first we present the mean value parametrization of an NEF.

Let (V_F, M_F) be a given VF of an NEF F generated by μ. Then as $m = k'_\mu(\theta)$ and $V_F(m) = 1/\psi'_\mu(m) = k''_\mu(\theta)$ (see (2)), the canonical parameter θ and the cumulant transform k_μ can be represented in terms of m as

$$\theta = \psi_\mu(m) = \int \frac{dm}{V_F(m)} + c_1, \qquad k_\mu(\psi_\mu(m)) = \int \frac{m}{V_F(m)}\, dm + c_2, \tag{9}$$

where one needs to determine the constants c_1 and c_2. The simplest way to do so is to assume that μ is a probability. Under this circumstance one chooses m_0 in the domain of the means M_F or at the boundary of M_F and writes

$$\psi_\mu(m) = \int_{m_0}^{m} \frac{dt}{V_F(t)}, \qquad k_\mu(\psi_\mu(m)) = \int_{m_0}^{m} \frac{t\, dt}{V_F(t)}. \tag{10}$$

Indeed, the following proposition shows how to cope with the determination of the constants c_1 and c_2 in (9) by just expressing the mean value parametrization of the NEF F as generated by a probability measure (and not by an arbitrary measure).

Proposition 1. *Suppose that (V_F, M_F) is the VF of the NEF F, where $M_F = (a, b)$ with $-\infty \leq a < b \leq \infty$. Then the mean value parametrization of $F = F(\mu)$ can be expressed in terms of probabilities as follows.*

1. *Let $m_0 \in (a, b)$, then*

$$P(m, F)(dx) = \exp\left(\int_{m_0}^{m} \frac{x - t}{V_F(t)}\, dt\right) P(m_0, F)(dx), \quad m \in M_F, \tag{11}$$

where

$$P(m_0, F)(dx) = \frac{\mu(dx)}{\exp\left[-(x\psi_\mu(m_0) - k_\mu(\psi_\mu(m_0)))\right]}.$$

2. *Suppose that $\int_m^b \frac{dt}{V_F(t)}$ and $\int_m^b \frac{tdt}{V_F(t)}$ both exist for $m \in M_F$. Then there exists a unique probability $\mu = P(b, F)$ such that*

$$P(m, F)(dx) = \exp -\left(\int_{m}^{b} \frac{x - t}{V_F(t)}\, dt\right) P(b, F)(dx), \quad m \in M_F \tag{12}$$

holds. Furthermore if $b = \infty$ then $\int_{\mathbb{R}} \max(-x, 0)\, \mu(dx) < \infty$ and $\int_{\mathbb{R}} x\, \mu(dx) = \infty$. If $b < \infty$, then F is not steep and $\int_{\mathbb{R}} xP(b, F)(dx) = b$.

Proof. 1. If $F = F(\mu)$ then, from $P(m, F) = \exp\left(x\psi_\mu(m) - k_\mu(\psi_\mu(m)\right)\mu(dx)$, we have

$$\mu(dx) = \exp\left[-(x\psi_\mu(m_0) - k_\mu(\psi_\mu(m_0))\right] P(m_0, F)(dx) \quad \text{and}$$
$$P(m, F) = \exp\left(x(\psi_\mu(m) - \psi_\mu(m_0)) - (k_\mu(\psi_\mu(m) - k_\mu(\psi_\mu(m_0))\right) P(m_0, F)(dx).$$

Finally we use the fact that

$$\frac{d}{dm}(\psi_\mu(m)-\psi_\mu(m_0))=\frac{1}{V_F(m)} \quad \text{and} \quad \frac{d}{dm}(k_\mu(\psi_\mu(m)-k_\mu(\psi_\mu(m_0))=\frac{m}{V_F(m)}$$

for getting (11).

2. Fix $m_0\in(a,b)$. Let $a<m_n<b$ be an increasing sequence such that $\lim_{n\to\infty}m_n=b$ and denote

$$g_n(x)=\int_{m_0}^{m_n}\frac{x-t}{V_F(t)}\,dt.$$

Then $P(m_n,F)=e^{g_n(x)}P(m_0,F)$ and there exists a subsequence n_k and a measure μ of mass ≤ 1 such that $P(m_{n_k},F)\to_{k\to\infty}\mu$ weakly. If $X\sim P(m_0,F)$ then since $\displaystyle\int_{m_0}^b\frac{dt}{V_F(t)}$ and $\displaystyle\int_{m_0}^b\frac{tdt}{V_F(t)}$ are finite, we can claim that for all $\epsilon>0$ there exists $M>0$ such that $\lim_k\Pr(g_{n_k}(X)>M)<1-\epsilon$. In other terms the convergence $P(m_{n_k},F)\to_{k\to\infty}\mu$ is tight and μ is a probability with

$$\mu(dx)=\exp(g_\infty(x))P(m_0,F)(dx)=\exp\left(\int_{m_0}^b\frac{x-t}{V_F(t)}\,dt\right)P(m_0,F)(dx),$$

which does not depend on the particular subsequence (n_k). This proves (12).

If $b=+\infty$ consider the sequence $n\mapsto\displaystyle\int_{-\infty}^0 xe^{g_n(x)}P(m_0,F)$. It is increasing since $g_{n+1}(x)-g_n(x)<0$ for $x<0$. Therefore its limit $L_-=\displaystyle\int_0^\infty xe^{g_\infty(x)}P(m_0,F)$ is finite.

Similarly one proves by monotone convergence that the limit $L_+=\displaystyle\int_0^\infty xe^{g_\infty(x)}P(m_0,F)$ of the sequence $n\mapsto\displaystyle\int_0^\infty xe^{g_n(x)}P(m_0,F)$ exists in $(0,\infty]$. For seeing that $L_+=\infty$, observe that $+\infty=b=\lim_n m_n=L_++L_-$.

If $b<\infty$ then the fact that $\int_m^b\frac{dt}{V_F(t)}$ is finite implies that F is not steep as

$$b=\int_{-R}xP(b,F)(dx)=\lim_n\int_{\mathbb{R}}xe^{g_n(x)}P(m_0,F),$$

an equality which is obtained by dominated convergence. □

We present now a few illustrative examples for the above proposition in which the measure μ is known, whereas in the next subsection we present examples in which the measure μ is unknown.

Example 1 (The Ressel NEF)**.** This example is essentially important in terms of mean value parametrization as its LT is not expressible and rather is obtained as a solution of implicit functional equation. Its probability density is given by (12). By the Stirling formula $\Gamma(x+1)\sim_{x\to\infty}\sqrt{2\pi}x^{x+\frac{1}{2}}e^{-x}$ and therefore

$$m_0=\int_0^\infty x\mu(dx)=\infty,$$

implying that $m_0=\infty$ is a limit point of M_F. Consequently, using the second part of the proposition we have

$$\psi_\mu(m)=-\int_m^\infty\frac{dt}{V_F(t)}=-\frac{1}{m}+\log\frac{1+m}{m}$$

and

$$k_\mu(\psi_\mu(m))=-\int_m^\infty\frac{tdt}{V_F(t)}=-\log\frac{1+m}{m},$$

implying that both $\psi_\mu(m)$ and $k_\mu(\psi_\mu(m))$ can be written explicitly as a function of m. For more details see [3].

The following two examples (Poisson and exponential) are obvious as their triples: μ, LT and VF are well known. They are presented just for the sake of illustration of the proposition.

Example 2 (The Poisson NEF). The well known and typical generating measure of the Poisson NEF is $\mu(dx) = \sum_{n=0}^{\infty} \frac{1}{n!}\delta_n$, where δ_n is the Dirac mass on n, with $\psi_\mu(m) = \log m$ and $k_\mu(m) = m$. It has a mass e and it is not a probability, but $\mu_1 = \mu/e$ is. If you apply the above proposition to $m_0 = 1$, one gets $\psi_{\mu_1}(m) = \log m$, $k_{\mu_1}(m) = m - 1$ and

$$P(m, F)(dx) = e^{x \log m - m + 1}\mu_1(dx) = e^{x \log m - m}\mu(dx).$$

Example 3 (The exponential NEF). The typical generating measure of the exponential NEF is $\mu(dx) = \mathbf{1}_{(0,\infty)}(x)dx$. It is unbounded and has $\psi_\mu(m) = -1/m$ and $k_\mu(m) = \log m$. Choose $m_0 = 1$ then $\mu_1(dx) = e^{-x}\mu(dx)$, $\psi_{\mu_1}(m) = 1 - 1/m$, $k_{\mu_1}(m) = \log m$ and thus

$$P(m, F)(dx) = \exp\Big\{x - \frac{x}{m} - \log m\Big\}\mu_1(dx) = \exp\Big\{ - \frac{x}{m} - \log m\Big\}\mu(dx).$$

4. POLYNOMIAL VF'S

Various necessary and/or sufficient conditions for a pair (V, M) to be a VF of an NEF have appeared in the literature. See for example: [37, 38, 24, 25, 34, 32, 5, 6, 29].

We shall however focus here on sufficient conditions leading to polynomial VF's which have a simple functional structure as a polynomial can be. Their corresponding cumulants are also polynomials and can be easily computed by (8).

The main significant question related to polynomial VFs is the following. Given an nth degree polynomial

$$P_n(m) = \sum_{i=0}^{n} a_i m^i,$$

under which conditions on the polynomial coefficients a_i, $i = 1, \ldots, n$, can P_n serve as a VF of an NEF, i.e., does there exist a measure μ which generates an NEF and for which P_n (along with a corresponding mean parameter space M) is a VF? Moreover, if (P_n, M) is a VF, can we then reveal the forms of

$$\theta = \psi_\mu(m), \quad k_\mu(\psi_\mu(m)), \quad L_\mu(\theta) \quad \text{or the generating measure } \mu? \tag{13}$$

The answer to this question is equivocal as occasionally none of these functionals, or only some of them, are available explicitly. The cases of quadratic, cubic or power VF's are indeed exceptional. (Just note that even for the Ressel NEF having a cubic VF, the corresponding LT is a solution of an implicit functional equation).

However, for $n = 0, 1, 2$, Morris [37] characterized all polynomial VFs and revealed six NEFs (normal, Poisson, binomial, negative binomial, gamma and hyperbolic cosine) for which the corresponding θ, k_μ and μ are explicitly expressed. For $n = 3$, Letac and Mora [34] characterized all cubic VFs and revealed six additional families (Takács, strict arcsine, large arcsine Ressel and inverse Gaussian). Also, the special polynomial $P_n(m) = a_n m^n$, $n = 3, 4, \ldots$, is a special case of power VFs belonging to the Tweedie class. Apart from the latter three cases of quadratic, cubic and power VFs, it seems that for $n \geq 4$ the generating measure μ of a polynomial VF cannot be evaluated explicitly and neither are L_μ or $k_\mu(\psi_\mu(m))$. Nonetheless, it would be of interest to know which polynomials P_n, $n \geq 4$, are VFs of NEFs (even without being able to delineate their corresponding generating measures). Indeed, and as already indicated, various necessary and/or sufficient conditions have been imposed on the coefficients a_i's for showing whether or not P_n is a VF. Worthy noted on this subject are the works [4, 7].

However, one of the most relevant important result was given by Bar-Lev [2] who provided a sufficient condition for a pair (V, M) to be a VF; a condition which has a significant relevance to the construction of polynomial VF's. Bar-Lev showed that if $V : \mathbb{R}^+ \to \mathbb{R}^+$ is an absolutely monotone mapping satisfying

$$\lim_{m \to 0+} \int_m^{m_0} \frac{dt}{V(t)} = \infty, \quad \text{where} \quad m_0 \in \mathbb{R}^+, \tag{14}$$

then there exists an infinitely divisible NEF with VF $(V, \mathbb{R}^+)$. As a simple consequence of this result it readily follows that any polynomial $V = P_n$ with nonnegative coefficients

$$V(m) = \sum_{i=1}^{q} a_i m^i, \quad m \in \mathbb{R}^+, \quad q \in \mathbb{N}, \quad a_i \geq 0, \quad \sum_{i=1}^{q} a_i > 0, \tag{15}$$

constitutes a VF of an infinitely divisible NEF with mean domain $M = \mathbb{R}^+$ (for a more rigorous treatment of this result see Theorem 3.2 and Corollary 3.3 in [34]). Indeed, four of the six NEFs in the Morris class and all of the six NEFs in the Letac and Mora Class have VFs of the form (15). Also note that as any polynomial VF of the form (15) is infinitely divisible it can be used to generate an EDM with dispersion parameter space $\Lambda = \mathbb{R}^+$ and VF of the form

$$V_\lambda(m) = \lambda\left(\sum_{i=1}^{q} a_i \left(\frac{m}{\lambda}\right)^i\right), \quad m_\lambda \in \mathbb{R}^+;$$

c.f., (4) and (6).

Also, any polynomial VF of the form (15) represents a steep NEF (i.e., $M = \operatorname{int} C_\mu = \mathbb{R}^+$) as the following lemma shows.

Lemma 2. *Let $(V, \mathbb{R}^+)$ be a given VF of the form (15) of an NEF with generating measure μ. Then,*

(i) *as $m \to 0+$ then $\theta \to -\infty$ implying that the constant c_1 in (9) is zero,*

(ii) *either $0 \in D_\mu$ or $0 \notin D_\mu$ and $\theta = \psi_\mu(m) \to -\infty$ as $m \to \infty$.*

Proof. (i) The VF V in (15) is positive. Without loss of generality assume that all zeroes (other than the origin) of V in absolute value are larger than some $\rho > 0$. Then the Laurent series of $1/V(t)$ converges absolutely for $0 \leq t \leq \rho$ and has the form

$$\theta = \psi_\mu(m) = \int_\rho^m \frac{1}{t^n}(a_{-n} + a_{-n+1}t + a_{-n+2}t^2 + \cdots)\, dt,$$

for some $n \in \mathbb{N}$. The Laurent series is bounded away from zero and therefore the convergence or divergence of $\psi_\mu(m)$ is equivalent to that of $\int_\rho^m t^{-q}\, dt$, which is $-\infty$, implying that c_1 in (9) is zero.

(ii) Since $V(m) \sim a_q m^q$ as $m \to \infty$ with $q \geq 1$ and $a_q > 0$ without loss of generality, one has as previously $\theta = \psi_\mu(m) \sim -q a_q \int_\rho^m t^{-q-1}\, dt$ as $m \to \infty$ which leads to the desired result. This would then imply by Theorem 2.1 of Letac and Mora [34] that the corresponding NEF is steep. □

The class of polynomial VF's (15) is huge and contains, among others,

- The class of quadratic VF's (except for the binomial and Normal NEF's) characterized by Morris [37], like Poisson and negative binomial distributions.
- The class of cubic VF's characterized by Letac and Mora [34], like strict arcsine, large arcsine, Abel or generalized Poisson, and Takács or generalized negative binomial distributions.
- The class of positive integer power VF's in the form $(\alpha m^\ell, \mathbb{R}^+)$, $\alpha > 0$, $\ell \in \mathbb{N}$; c.f., [45, 10, 24].

- The HD (Hinde and Demétrio) class of VF's of the form

$$(V, M) = (m + m^{\ell}, \mathbb{R}^+), \quad \ell = 2, 3, \dots, \tag{16}$$

corresponding to NEF's of counting distributions supported on the set $\mathbb{N}_0$ of the nonnegative integers and including negative binomial ($\ell = 2$) and strict arcsine ($\ell = 3$) distributions.

- The LM (Letac and Mora) class of VF's of the form

$$(V, M) = \left(m \prod_{i=1}^{\ell} \left(1 + \frac{m}{p_i}\right), \mathbb{R}^+\right), \quad p_i > 0, \quad i = 1, \dots, \ell, \quad \ell \in \mathbb{N}, \tag{17}$$

which correspond to NEF's generated by counting measures on $\mathbb{N}_0$.

- The ABM (Awad, Bar-Lev and Makov) class of VF's of the form

$$(V, M) = \left(m \left(1 + \frac{m}{p}\right)^r, \mathbb{R}^+\right), \quad p > 0, \quad r = 1, 2, 3, \dots, \tag{18}$$

constituting a subclass of the LM class, including Poisson ($r = 0$ with $p = 1$), negative binomial ($r = 1$) and Abel or generalized Poisson ($r = 2$) distributions.

We elaborate on the latter three classes in more detail in subsequent sections.

4.1. The LM Class

Letac and Mora [34] presented the class of polynomial VF's given by (17) which correspond to NEF's supported on $\mathbb{N}_0$. This class includes four (of the six) NEF's with quadratic VF's and also four (of the six) with cubic VF's. Due to the factorization of (17), both primitives in (9) for $\theta = \psi_\mu(m)$ and $k_\mu(\psi_\mu(m))$ can be computed explicitly (with some unpleasant forms) but not their inverse functions (e.g., $m = \psi_\mu^{-1}(\theta)$). This means that the likelihood function of a random sample taken from an NEF with VF as in (17) can be explicitly expressed in terms of the mean parameter m, an important feature in a Bayesian framework.

However, the main importance of the LM class comes from a frequentist approach as their Proposition 4.4 [34] enables to express the generating measure μ in terms of m. Indeed, let $\mu(dx) = \sum_{n=0}^{\infty} \mu_n \delta_n(dx)$ be a generating measure of an NEF with VF of the form (17) and let $G(m) = m \exp(-\psi(m)$ (here we suppress for simplicity the dependence of ψ on μ). Then, roughly writing, this proposition states that μ_0 and μ_n, $n \geq 1$, are given by

$$\begin{cases} \mu_0 = \exp\left(k(\psi(m))\right)|_{m=0}, \\ \mu_n = \dfrac{1}{n!}\left[\left(\dfrac{d}{dm}\right)^{n-1} \exp\left(k(\psi(m)) \times k'(\psi(m)) \times (G(m))^n\right)\right]\Big|_{m=0}. \end{cases} \tag{19}$$

We illustrate (19) with two examples taken from [34].

Example 4 (The Poisson NEF). Here,

$$(V, M) = (m, R^+), \quad \psi(m) = \log m, \quad k(\psi(m) = m, \quad G(m) \equiv 1,$$

so that

$$\mu_0 = 1, \quad \mu_n = \frac{1}{n!}, \quad n \geq 1, \quad \text{and} \quad \mu(dx) = \sum_{n=0}^{\infty} \frac{1}{n!} \delta_n(dx).$$

Example 5 (The negative binomial NEF). Here, for $p > 0$,

$$(V, M) = \left(m\left(1 + \frac{m}{p}\right), R^+\right), \quad \psi(m) = \log\frac{m}{m+p}, \quad k(\psi(m) = p\log(m+p),$$
$$G(m) = m\exp\left[-\log\left(\frac{m}{m+p}\right)\right],$$

and thus

$$\mu_0 = \mu_1 = p^p, \quad \mu_n = \frac{1}{n!}p^p\prod_{i=1}^{n-1}(i+p), \quad n \geq 2.$$

Other explicit expressions for the generating measure μ of NEF's with quadratic and cubic VF's were derived in [34] using their Proposition 4.4 (the derivation of some of them, however, required some rather cumbersome calculations).

Explicit derivations of (19) for polynomial VF's of degree $n \geq 4$ do not seem to be feasible and numerical computations based on mathematical software will be then required.

4.2. The HD Class

The HD class was presented by Hinde and Demétrio [21] for over dispersed models and characterized by Kokonendji et al. [26]. They considered the class of VF's of the form:

$$V(m) = m + m^p, \quad p \in \{0\} \cup [1, \infty),$$

where $m > -1$ if $p = 0$ and $m > 0$ if $p \geq 1$. The corresponding support S_p is $S_0 = \{-1\} \cup \mathbb{N}_0$, $S_1 = 2\mathbb{N}_0$ and $S_p = p\mathbb{N}_0 \cup \mathbb{N}_0$ for $p > 1$. Accordingly, such VF's are associated with NEF's supported on $\mathbb{N}_0$ only for $p = 2, 3, \dots$, in which case they are polynomial of the form (15). Such a polynomial version of the HD class has been used by Kokonendji et al. [27] to analyze overdispersed and zero-inflated count data. They employed the data used in Ridout et al. [43], originally modelled by the Poisson and negative binomial distributions, and showed that their HD modelling performs better for such data. Further theoretical and data analysis of the HD class can be found in [28] and [30].

We shall now illustrate (9) with respect to the HD class given in (16). As by Lemma 2, $c_1 = 0$, $\psi_\mu(m)$ can be obtained explicitly as

$$\theta = \psi_\mu(m) = \int\frac{dm}{m + m^\ell} + c_1 = \frac{1}{\ell - 1}\log\frac{1}{1 + m^{-(\ell-1)}} + c_1, \quad \ell \in \mathbb{N}.$$

On the other hand the integral for $k_\mu(\psi_\mu(m))$ in (9), although can be computed explicitly for each $\ell \in \mathbb{N}$, does not have a closed form. Nonetheless, the corresponding cumulant LT can be expressed in terms of the transcendental Gaussian hypergeometric series as

$$e^\theta \times_2 F_1\left(\frac{1}{\ell-1}, \frac{1}{\ell-1}; \frac{\ell}{\ell-1}; e^{\theta(\ell-1)}\right), \quad \theta < 0, \quad \ell = 2, 3, \dots,$$

(see [30] and [23], pp. 17–19).

Summarizing the HD class in terms of the question posed in (13), we have: $\theta = \psi_\mu(m)$ has a closed form whereas $k(\psi_\mu(m))$ does not (though it can be explicitly expressed for each $\ell \geq 2$); the LT can be expressed as a transcendental function (and thus has not practical appealing); the measure μ can be computed only numerically by using (19).

4.3. The ABM Class

The Lee—Carter model [31] and variants thereof (e.g., [42]) is a largely acceptable method of mortality forecasting. Awad, Bar-Lev and Makov [1] have dealt with predicting mortality rates by embedding the Lee—Carter model within a Bayesian framework. They used the ABM class of counting distributions as alternatives to the Poisson counts of events (deaths) under the Lee—Carter modeling for mortality forecast. They demonstratively obtained an improved prediction of mortality, especially on the sparse of elderly groups, for real data from three countries (Ireland, the US and Ukraine).

More specifically, the Lee—Carter model was originally designed to forecast age-specific mortality rates with the following specification:

$$\log m_{xt} = \alpha_x + \beta_x k_t + \varepsilon_{xt},$$

where the logarithm of the age and time-specific mortality rate m_{xt} is decomposed into an overall age profile, α_x, averaged over the entire period under consideration, and age-specific changes in mortality β_x. The subscripts x and t denote age and time, respectively. The β_x parameter describes which rates decline rabidly and which rates decline slowly in response to changes in the time-specific effect k_t. The error term ε_{xt} is assumed to be distributed with mean 0 and variance σ_ε^2 reflecting particular age-specific historical influences not captured by the model. The age and time-specific mortality rate m_{xt} should be calculated as (D_{xt}/E_{xt}), while D_{xt} denote the number of deaths in a population at age $x = 1, 2, \ldots, P$ and time $t = 1, 2, \ldots, T$ and E_{xt} the matching exposure to the risk of death. The general consensus in actuarial modelling, the Poisson distribution is selected as the Poisson response model with the response variable equal to the number of deaths (c.f., [42]). Namely

$$D_{xt} \sim \text{Poisson}(\mu_{xt}), \quad \mu_{xt} = E_{xt} m_{xt}.$$

Such a Poissonian assumption does not seem to fit various kinds of data as zero-inflated and overdispersed data. Delwarde, Denuit and Partrat [15] demonstrated that it is possible to take into account the overdispersion present in the mortality data by estimating the parameter in a negative binomial regression model. Haberman and Renshaw [19] made comparison between recent model enhancements including binomial model, and which these enhancements address the deficiencies that have been identified of some of the models. Awad, Bar-Lev and Makov [1] have assumed that $D_{xt} \sim ABM(p, r)(\mu_{xt})$, where the latter expression stands for an ABM distribution with parameters p and r with mean μ_{xt} depending on x and t. Under further prior distributions imposed on α_x, β_x and k_t they found an improved prediction of mortality, especially on the sparse of elderly groups, for real data from the above mentioned three countries.

Particularly, since Awad, Bar-Lev and Makov [1] considered a Bayesian approach it was unnecessary to compute the generating measures μ of the NEF's related to the ABM class of VF's in (18), as the likelihood function of m is the only relevant component in such an approach. Nevertheless, for other statistical purposes, they used the R package to numerically compute (19) for various values of $r \geq 3$ and p and then stopped computing μ_n in (19) after getting that the corresponding cumulative probability function is lager then 0.999.

For the sake of theoretical completeness we shall now apply Proposition 1 to represent the probabilities of an NEF belonging to the ABM class in (18) in a more general form.

Corollary 3. *Let F be an NEF with VF* (18) *belonging to the ABM class. Then the mean value parametrization of F is given by the probabilities*

$$P(m, F)(dx) = \exp\Big(x\big(\log m - R_r(m)\big) + \frac{p^r}{(n-1)(m+p)^{r-1}} - H(1)\Big)\mu(dx), \quad m > 0, \quad p > 0, \tag{20}$$

where

$$R_r(m) = \log(m+p) - \sum_{k=1}^{r-1} \frac{p^k}{k(m+p)^k}, \quad r \in \mathbb{N}, \tag{21}$$

and

$$H(1)=\sum_{i=1}^{\infty}\frac{1}{i!}\times\frac{1}{i}\left[\left(\frac{d}{dm}\right)^{i-1}e^{iR_r(m)}\right]_{m=0}.$$

Proof. By using (18) it is easily shown that

$$\theta=\psi_\mu(m)=\log\frac{m}{p+m}+\sum_{i=1}^{r-1}\frac{1}{i}\frac{p^i}{(p+m)^i}+c_1,\quad r\geq 1,\quad \text{where}\quad \sum_{i=1}^{0}=0, \tag{22}$$

and

$$k_\mu(\psi_\mu(m)=-\frac{p^r}{(r-1)(m+p)^{r-1}}+c_2.$$

Alternatively, we write

$$\begin{aligned}
k_\mu(\psi_\mu(m))&=-\int_m^\infty\frac{dt}{(\frac{t}{p}+1)^r}=-p\int_{\frac{m}{p}+1}^\infty\frac{ds}{s^r}=-\frac{p^r}{(r-1)(m+p)^{r-1}}\\
\psi_\mu(m)&=-\int_m^\infty\frac{dt}{t(\frac{t}{p}+1)^r}=\int_{\frac{m}{p}+1}^\infty\frac{ds}{(1-s)s^r}\\
&=\int_{\frac{m}{p}+1}^\infty\left(\left(\frac{1}{s}+\frac{1}{1-s}\right)+\frac{1}{s^r}+\cdots+\frac{1}{s^2}\right)ds\\
&=\log m-R_r(m).
\end{aligned}$$

Note that, for $r=0$ and $r=1$, the expressions for $k_\mu(\psi_\mu(m))$ and $\psi_\mu(m)$ are given explicitly in Examples 4 and 5, respectively. Denote $w=e^\theta=e^{\psi_\mu(m)}$ and $h(w)=k'(\theta)=m$. Since $w=e^{\psi_\mu(m)}=me^{-R_r(m)}$, we have

$$w=h(w)e^{-R_r(h(w))}.$$

The Lagrange formula states that if $h(w)=wg(h(w))$ then

$$h(w)=\sum_{i=1}^{\infty}\frac{w^i}{i!}\left[\left(\frac{d}{dm}\right)^{i-1}(g(m))^i\right]_{m=0}.$$

We apply the Lagrange formula to $m=h(w)=k'_\mu(\theta)$ and to $g(m)=e^{R_r(m)}$ and denote

$$H(w)=\int_0^w h(z)\frac{dz}{z}.$$

Then clearly

$$k_\mu(\theta)=-C+H(e^\theta).$$

Hence, for μ to be a probability, C should satisfy

$$C=H(1)=\int_0^1\frac{h(z)}{z}\,dz=\sum_{i=1}^{\infty}\frac{1}{i!}\times\frac{1}{i}\left[\left(\frac{d}{dm}\right)^{i-1}e^{iR_r(m)}\right]_{m=0},$$

and we need to show by Part 2 of Proposition 1 that $\mu=P(\infty,F)$. To prove this we need to show that $m=h(1)=\infty$ or that the only solution $h(1)$ of $1=h(1)e^{-R_r(h(1))}$ is $h(1)=\infty$. For this purpose denote $u=p/(p+h(1))\in(0,1]$, so that the equation $1=h(1)e^{-R_r(h(1))}$ becomes

$$\exp\left(u+\frac{u^2}{2}+\cdots+\frac{u^{r-1}}{n-1}\right)=\frac{1-2u}{1-u}. \tag{23}$$

On $[0, 1]$, the left-hand side of (23) is increasing and the right-hand side is decreasing. Consequently, the only solution of (23) is $u = 0$ and therefore $h(1) = \infty$. Hence, by Part 2 of Proposition 1 it follows that

$$P(m, F)(dx) = \exp\left(x\big(\log m - R_r(m)\big) + \frac{p^r}{(n-1)(m+p)^{r-1}} - H(1)\right)\mu(dx),$$

where μ is the probability generating F with infinite mean (just note that $b = \infty$ in the notation of Proposition 1). □

Summarizing the ABM class in terms of the question raised in (13): We have both $\theta = \psi_\mu(m)$ and $k(\psi_\mu(m))$ in closed forms and thus the likelihood function of m is nicely expressed; the LT is not expressible for $r \geq 3$; the measure μ can be computed only numerically by using (19) or by (20).

5. CONCLUDING REMARKS

The main aim of this paper has been to expose the mean value parametrization of NEF's (or EDM's) for possible implementation in both probability modelling and statistical inference. Particularly, this parametrization is of great worth when the corresponding LT and/or the generating measures are not known explicitly (as has been demonstrated for the ABM class or the Ressel NEF for which the LT is a solution of an implicit functional equation).

The existence of polynomial VF's of NEF's in the form (15) provides a huge class of families of distributions (either discrete or continuous), the majority of which have not been known (and thus have not been utilized in applications). Moreover, by using (8), polynomial VF's easily allow the computation of all cumulants and moments (in terms of polynomial forms) as well as all measures of skewness and kurtosis of the associated NEF's.

The special case, but rather general though, of (15) in the form (17) presented by Letac and Mora [34] sets ample examples of discrete distributions supported on $\mathbb{N}_0$ for which the generating measures can in general be computed numerically by using available efficient software as MATHEMATICA, S, R, MAPLE and the like. Many of these count NEF's distributions can be used as competitors to the well used Poisson distribution for describing overdispersed or zero-flatted data. Nonetheless, the form of the generating measure is not needed in the Bayesian framework (as has been indicated for the ABM family).

We trust that this revisited presentation of the mean value parametrization of NEF's will lead to the exposure of new classes of VF's (particularly of the polynomial type) which will be used for probabilistic and statistical modelling (frequentist or Bayesian) as are the HD and ABM classes. New special classes of VF's in the form (17) for which both $\theta = \psi(m)$ and $k(\psi(m))$ have nicely closed forms can be definitely obtained and used accordingly.

More specifically, consider the LM class in which the VF's are products of linear functions whose roots are trivially specified. This entails that the integrands of both $\theta = \psi(m)$ and $k(\psi(m))$ are proper rational functions (where the corresponding numerator in both is 1), implying that one can use the partial fraction method (c.f., [18], Section 2.10, pp. 2.101–2.104) to specifically (cumbersomely though) calculate $\theta = \psi(m)$ and $k(\psi(m))$. For example one might consider a two-parameter subclass of the LM class with VF of the form

$$V(m) = m(1+m)\Big(1 + \frac{m}{p}\Big)^l, \quad \text{where} \quad p > 0, \quad l \in \mathbb{N}, \tag{24}$$

for which it can be shown by the partial fraction method (computational details are omitted for brevity) that

$$\begin{aligned} \theta(m) &= \int \frac{1}{m(1+m)(1+\frac{m}{p})^l}\,dm \\ &= \log m + (1-p)^l \log(m+p) \sum_{i=0}^{l-1} (-1)^{i+1} \binom{l}{i} p^i - \Big(\frac{p}{p-1}\Big)^l \log(m+1) \\ &\quad + \sum_{i=1}^{l-1} \frac{1}{i} \frac{A_i}{(m+p)^i (p-1)^{l-i}} + c, \end{aligned} \tag{25}$$

where

$$A_i = \sum_{j=1}^{l-1} (-1)^{j+l} \binom{l-i}{j-i} p^j \quad \text{and} \quad \sum_{i=1}^{0} = 0,$$

and

$$k(\psi(m)) = \Big(\frac{p}{p-1}\Big)^l \log \frac{m+1}{m+p} + p^l \sum_{i=2}^{l} \frac{1}{(i-1)(p-1)^{l-i+1}(m+p)^{i-1}} + c, \tag{26}$$

where

$$\sum_{i=2}^{1} = 0.$$

The latter expressions for $\theta = \psi(m)$ and $k(\psi(m))$ resemble those in (22) of the ABM class (a subclass of the LM class) which has been implemented, Bayesian-wise) under the Lee–Carter modelling for mortality forecast and demonstratively obtained an improved prediction of mortality, especially on the sparse of elderly groups, for real data from three countries.

We confide that the subclass of VF's in (24) with $\theta = \psi(m)$ and $k(\psi(m))$ as in (25) and (26) might be challenging when implemented similarly under the Lee–Carter modelling to perhaps yielding an improved forecast for various age groups. In either case, this subclass provides a new class of counting distributions whose probabilities can be computed numerically by using (19) and an appropriate software.

ACKNOWLEDGMENTS

We thank a referee for helpful comments which improved the presentation of the paper. Part of this work was done while the second author was a visitor at the University of Haifa and supported by the Zimmerman foundation. The paper was completed while S. K. Bar-Lev was a visiting professor at NYU Shanghai, China, during the Spring semester of 2017. We sincerely thank Gérard Letac for illuminating and interesting discussion and comments.

REFERENCES

1. Y. Awad, S. K. Bar-Lev, and U. Makov, *A New Class of Counting Distributions Embedded in the Lee–Carter Model for Mortality Projections: A Bayesian Approach*, Techn. Rep. (Actuarial Research Center, Univ. of Haifa, Israel, 2016) (in press).
2. S. K. Bar-Lev, "Discussion on paper by B. Jørgensen 'Exponential dispersion models'," J. Roy. Statist. Soc., Ser. B **49**, 153–154 (1987).
3. S. K. Bar-Lev, B. Boukai, and I. Kleiner, "The Kendal–Ressel Exponential Dispersion Model: Some Statistical Aspects and Estimation", Int. J. Statist. and Probab. **5** (3), 32–41 (2016).
4. S. K. Bar-Lev and D. Bshouty, "A Class of Infinitely Divisible Variance Functions with an Application to the Polynomial Case", Probab. and Statist. Lett. **10**, 377–379 (1990).
5. S. K. Bar-Lev, D. Bshouty, and P. Enis, "Variance Functions with Meromorphic Means", Ann. Probab. **19**, 1349–1366 (1991).
6. S. K. Bar-Lev, D. Bshouty, and P. Enis, "On Polynomial Variance Functions", Probab. Theory and Rel. Fields **94**, 69–82 (1991).
7. S. K. Bar-Lev, D. Bshouty, P. Enis, and A. Y. Ohayon, "Compositions and Products of Infinitely Divisible Variance Functions", Scandinavian J. Statist. **19**, 83–89 (1992).
8. S. K. Bar-Lev, D. Bshouty, P. Grünwald, and P. Harremoës, "Jeffreys versus Shtarkov Distributions Associated with Some Natural Exponential Families", Statist. Methodology **7**, 638–643 (2010).
9. S. K. Bar-Lev, D. Bshouty, and Z. Landsman, "Second Order Minimax Estimation of the Mean", J. Statist. Planning and Inference **140**, 3282–3294 (2010).
10. S. K. Bar-Lev and P. Enis, "Reproducibility and Natural Exponential Families with Power Variance Functions", Ann. Statist. **14**, 1507–1522 (1986).
11. S. K. Bar-Lev and G. Letac, "Increasing Hazard Rate of Mixtures for Natural Exponential Families", Advances in Appl. Probab. **44**, 373–390 (2009).

12. S. K. Bar-Lev and Z. Landsman, “Exponential Dispersion Models: Second-Order Minimax Estimation of the Mean for Unknown Dispersion Parameter”, J. Statist. Planning and Inference **136**, 3837–3851 (2006).
13. O. E. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory* (Wiley, New York, 1978).
14. J. Castillo and M. Pérez-Casany, “Overdispersed and Underdispersed Poisson Generalizations”, J. Statist. Planning and Inference **134**, 486–500 (2005).
15. A. Delwarde, M. Denuit, and C. Partrat, “Negative Binomial Version of the Lee–Carter Model for Mortality Forecasting”, Appl. Stochastic Models in Business and Industry **23** (5), 385–401 (2007).
16. P. K. Dunn and G. K. Smith, “Series Evaluation of Tweedie Exponential Dispersion Densities”, Statistics and Computing **15**, 267–280 (2005).
17. P. K. Dunn and G. K. Smith, “Evaluation of Tweedie Exponential Dispersion Model Densities by Fourier Inversion”, Statistics and Computing **18**, 73–86 (2008).
18. I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (seventh edition) (Elsevier and Academic Press, New York, 2007).
19. S. Haberman and A. Renshaw, “A Comparative Study of Parametric Mortality Projection Models”, Insurance: Mathematics and Economics **48** (1), 35–55 (2011).
20. D. B. Hall and K. S. Berenhaut, “Score Tests for Heterogeneity and Overdispersion in Zero-Inflated Poisson and Binomial Regression Models”, Canadian J. Statist. **30**, 1–16 (2002).
21. J. Hinde and C. G. B. Demétrio, “Overdispersion: Models and Estimation”, Comput. Statist. and Data Analysis **27**, 151–170 (1998).
22. P. Hougaard, K.-L. T. Lee, and G. A. Whitmore, “Analysis of Overdispersed Count Data by Mixtures of Poisson Variables and Poisson Processes”, Biometrics **53**, 1225–1238 (1997).
23. N. L. Johnson, S. Kotz, and A. W. Kemp, *Univariate Discrete Distributions*, 2nd ed. (Wiley, New York, 1992).
24. B. Jørgensen, “Exponential Dispersion Models (with Discussion)”, J. Roy. Statist. Soc., Ser. B **49**, 127–162 (1987).
25. B. Jørgensen, *The Theory of Exponential Dispersion Models*, in *Monographs on Statistics and Probability* (Chapman and Hall, London, 1997), Vol. 76.
26. C. C. Kokonendji, S. Dossou-Gbété, and C. G. B. Demétrio, “Some Discrete Exponential Dispersion Models: Poisson–Tweedie and Hinde–Demétrio Classes”, Statist. and Operations Research Transactions **28**, 201–214 (2004).
27. C. C. Kokonendji, C. G. B. Demétrio, and S. S. Zocchi, “On Hinde–Demétrio Regression Models for Overdispersed Count Data”, Statist. Methodology **4**, 277–291, (2007).
28. C. C. Kokonendji and M. Khoudar, “On Strict Arcsine Distribution”, Commun. Statist. – Theory and Methods **33**, 993–1006 (2004).
29. C. C. Kokonendji and M. Khoudar, “On Lévy Measures for Infinitely Divisible Natural Exponential Families”, Statist. and Probab. Lett. **76**, 1364–1368 (2006).
30. C. C. Kokonendji and D. Malouche, “A Property of Count Distributions in the Hinde–Demétrio Family”, Commun. Statist. – Theory and Methods **37**, 1823–1834 (2008).
31. R. D. Lee and L. Carter, “Modeling and Forecasting the Time Series of US Mortality”, J. Amer. Statist. Assoc. **87**, 659–671 (1992).
32. G. Letac, *Lectures on Natural Exponential Families and Their Variance Functions*, in *Monografias de Matemática* (Insituto de Matemática Pura Eplicada, Rio de Janeiro, 1992), Vol. 50.
33. G. Letac, *Associated Natural Exponential Families and Elliptic Functions*, in *The Fascination of Probability and Statistics and their Applications*, Ed. by M. Podolskij, R. Stelzer, S. Thorbjornsen, and A. E. D. Veraat (Springer, 2016).
34. G. Letac and M. Mora, “Natural Real Exponential Families with Cubic Variance Functions”, Ann. Statist. **18**, 1–37 (1990).
35. E. Lukacs, *Characteristic Functions*, 2nd ed. (Hafner, New York, 1970).
36. E. Lukacs, *Developments in Characteristic Functions Theory* (Griffin, London, 1983).
37. C. N. Morris, “Natural Exponential Families with Quadratic Variance Functions”, Ann. Statist. **10**, 65–80 (1982).
38. C. N. Morris, “Natural Exponential Families with Quadratic Variance Functions: Statistical Theory”, Ann. Statist. **11**, 515–529 (1983).
39. J. A. Nelder and R. W. M. Wedderburn, “Generalized Linear Models”, J. Roy. Statist. Soc., Ser. A **135**, 370–384 (1972).
40. R. S. Pimentel, M. Niewiadomska-Bugaj, and J.-C. Wang, “Association of Zero-Inflated Continuous Variables”, Statist. and Probab. Lett. **96**, 61–67 (2015).
41. N. U. Prabhu, *Queues and Inventories: A Study of Basic Stochastic Processes* (Wiley, New York, 1965).

42. A. E. Renshaw and S. Haberman, “A Cohort-Based Extension to the Lee–Carter Model for Mortality Reduction Factors”, Insurance: Mathematics and Economics **38**, 556–570 (2006).
43. M. Ridout, C. G. B. Demétrio, and J. Hinde, “Models for Count Data with Many Zeros”, in *Proc. XIXth Intern. Biometrics Conf., Cape Town, 1998*, pp. 179–192.
44. G. K. Smyth and B. Jørgensen, “Fitting Tweedie’s Compound Poisson Model to Insurance Claims Data: Dispersion Modelling”, Astin Bulletin **32**, 143–157 (2002).
45. M. C. K. Tweedie, “An Index Which Distinguishes between Some Important Exponential Families”, in *Statistics: Applications and New Directions, Proc. Indian Statist. Inst. Golden Jubilee International Conference*, Ed. by J. K. Ghosh and J. Roy (Indian Statist. Inst., Calcutta, 1984), pp. 579–604.
46. V. Vinogradov, R. B. Paris, and O. Yanushkevichiene, “New Properties for Members of the Power-Variance Family. I”, Lithuanian Math. J. **52**, 444–461 (2012).
47. V. Vinogradov, R. B. Paris, and O. Yanushkevichiene, “New Properties for Members of the Power-Variance Family. II”, Lithuanian Math. J. **53**, 103–120 (2013).