NEW EXPONENTIAL DISPERSION MODELS FOR COUNT DATA: THE ABM AND LM CLASSES

Shaul K. Bar-Lev[1,*] and Ad Ridder[2]

Abstract. In their fundamental paper on cubic variance functions (VFs), Letac and Mora (*The Annals of Statistics*, 1990) presented a systematic, rigorous and comprehensive study of natural exponential families (NEFs) on the real line, their characterization through their VFs and mean value parameterization. They presented a section that for some reason has been left unnoticed. This section deals with the construction of VFs associated with NEFs of counting distributions on the set of non-negative integers and allows to find the corresponding generating measures. As EDMs are based on NEFs, we introduce in this paper two new classes of EDMs based on their results. For these classes, which are associated with simple VFs, we derive their mean value parameterization and their associated generating measures. We also prove that they have some desirable properties. Both classes are shown to be overdispersed and zero inflated in ascending order, making them as competitive statistical models for those in use in both, statistical and actuarial modeling. To our best knowledge, the classes of counting distributions we present in this paper, have not been introduced or discussed before in the literature. To show that our classes can serve as competitive statistical models for those in use (*e.g.*, Poisson, Negative binomial), we include a numerical example of real data. In this example, we compare the performance of our classes with relevant competitive models.

1. Introduction and background

Natural exponential families (NEFs) and exponential dispersion models (EDMs) on $\mathbb{R}$ play an important role both in probability and statistical applications. Most of the frequently used distributions are indeed belonging to such models. However, a huge number of NEFs (or EDMs) have not been used in probabilistic or statistical modelling for two main reasons: they have not been revealed or do not have explicit functional forms (even not via power series expansions). This, despite the fact that they could have provided significant and new models useful in statistical applications. Indeed, the main purpose of this paper is to expose the statistical research community to various classes of such NEFs. A thorough discussion on this observation is presented in [5].

One of the most neglected reference representing the above situation is the fundamental paper [31] on NEFs which provides a thorough description and analytic properties of such families along with their mean value

[1] Faculty of Industrial Engineering and Technology Management, Holon Institute of Technology, Holon, Israel.

[2] School of Business and Economics, Vrije University of Amsterdam, Amsterdam, The Netherlands.

* Corresponding author: barlev@stat.haifa.ac.il

parameterization. In spite of the fact that their article received many citations, a major and important part of the article was somehow abandoned without being noticed. This part refers to the section dealing with the construction of NEFs of counting distributions on the set of nonnegative integers $\mathbb{N}_0$. These families are represented by either polynomial variance functions (VFs) or other nice forms. Moreover, in their Proposition 4.4 they explicitly present a formula that allows to compute, at least numerically, the counting measure μ which generates the appropriate NEF in terms of its mean m, for further details see also [5]. Such a formula requires (except for a few limited special cases) some rather cumbersome numerical calculations of the n-th derivative of product of functions depending on the mean m which are needed for calculating the mass of μ at the point n.

In our opinion, one of the reasons why this formula as well as Proposition 4.4 of [31], were not used is that in the nineties of the last century (when the [31] article was just published) is related to the fact that there were no powerful mathematical programs that would allow the complex and cumbersome calculations of the mass of μ on the nonnegative integers. Fortunately, nowadays, the situation has changed and existing powerful computing software are available and might be used to calculate. However, despite the nowadays availability of existing powerful software, it is still intricate or even not possible to carry out the probability calculation of the relevant NEFs in their general settings. It is therefore necessary to locate special cases of NEFs complying Proposition 4.4 of [31] for which the software application is possible. And indeed, our aim in this paper is to achieve this goal and introduce two classes of NEFs, which through further mathematical improvements, allow the calculation of the appropriate count probabilities of these subclasses of NEFs. We need to point out here that locating such classes is not as simple as it seems, and requires great care and thoughts in choosing them. To our best knowledge, the classes of counting distributions we present in this paper have not been introduced or discussed before in the literature. A fact that will lead to exposure of numerous counting NEFs (as well as EDMs) that can serve as competitive statistical models for those in use (*e.g.*, Poisson, Negative binomial) in both, statistical and actuarial modeling.

For this we need to present some preliminaries. As is well known, and as will seen in the sequel, EDMs are based on NEFs. Hence, we first need to present some basic properties of NEFs and VFs, mean value parameterization, and then EDMs.

Let μ be a positive Radon measure on $\mathbb{R}$ with convex support C_μ. Consider the set

$$D_\mu \doteq \left\{ \theta \in \mathbb{R} : L_\mu(\theta) \doteq \int_{\mathbb{R}} \exp(\theta x)\mu(\mathrm{d}x) < \infty \right\}, \tag{1.1}$$

and assume that $\Theta_\mu \doteq \mathrm{int}\, D_\mu$ is nonempty. Then, the NEF $\mathcal{F}(\mu)$ generated by μ is defined by the set of probability distributions

$$\mathcal{F}(\mu) \doteq \left\{ F\big(\theta, \mu(\mathrm{d}x)\big) = \exp\big(\theta x - k_\mu(\theta)\big)\mu(\mathrm{d}x) : \theta \in \Theta_\mu \right\}, \tag{1.2}$$

where $k_\mu(\theta) \doteq \log L_\mu(\theta)$ is the cumulant transform of μ; k_μ is strictly convex and real analytic on Θ_μ. Moreover, $k'_\mu(\theta)$ and $k''_\mu(\theta)$, $\theta \in \Theta_\mu$, are the respective mean and variance corresponding to $F(\theta, \mu)$, and the open interval $M_\mu \doteq k'_\mu(\Theta_\mu)$ is called the mean domain of $\mathcal{F}(\mu)$.

An important observation is that measure μ is not unique for $\mathcal{F}(\mu)$. Let $\mathcal{M}$ be the set of Radon measures ν on $\mathbb{R}$ for which $L_\nu(\theta) < \infty$ on domain Θ_ν. Consider two measures $\mu, \mu^* \in \mathcal{M}$, and suppose that μ^* is an exponential shift of μ; *i.e.*, $\mu^*(\mathrm{d}x) = e^{a+bx}\mu(\mathrm{d}x)$ for some real a, b. Then a simple calculation shows that $\mathcal{F}(\mu) = \mathcal{F}(\mu^*)$. This holds also reversely, if $\mathcal{F}(\mu) = \mathcal{F}(\mu^*)$ for two measures $\mu, \mu^* \in \mathcal{M}$, then one is an exponential shift of the other. Consequently, we may denote the NEF by $\mathcal{F} = \mathcal{F}(\mu)$ and its the mean domain $M = M_\mu$ to stress that these do not depend on μ.

Since the function $k'_\mu : \Theta_\mu \to M$ is one-to-one, its inverse function $\big(k'_\mu\big)^{-1} : M \to \Theta_\mu$ is well defined. When we compute the variance $V_\mu(\theta) \doteq k''_\mu(\theta)$ of the distribution $F(\theta, \mu)$ as a function of the mean $m \in M$, *i.e.*,

$$V_\mu(m) = k''_\mu\big((k'_\mu)^{-1}(m)\big),$$

it also does not depend on μ, and we denote it by $V(m)$. The map $m \mapsto V(m)$ from M into $\mathbb{R}^+$ is called the variance function (VF) of $\mathcal{F}$. In fact, a VF of an NEF $\mathcal{F}$ is a pair (V, M) which uniquely determines the NEF within the class of NEFs, see [31, 32]. It is important to emphasize that a VF is a transform, not of a particular distribution, but rather of a family $\mathcal{F}$ in the sense that if two VFs (V_1, M_1) and (V_2, M_2) of two NEFs $\mathcal{F}_1$ and $\mathcal{F}_2$, respectively, satisfy $V_1 = V_2$ on $M_1 \cap M_2 \neq \emptyset$, then $\mathcal{F}_1 = \mathcal{F}_2$. This would imply that given a VF (V, M), the mean domain M is the largest open interval on which V is positive real analytic.

Suppose that we would denote the function $\left(k'_\mu\right)^{-1}(\cdot)$ by $\psi_\mu(\cdot)$. Then we get by differentiating

$$\psi'_\mu(m) = \frac{1}{k''_\mu\left((k'_\mu)^{-1}(m)\right)} = \frac{1}{V(m)}.$$

Similarly, when we would define $\phi_\mu(\cdot) \doteq k_\mu\left((k'_\mu)^{-1}(\cdot)\right)$, we get by differentiating

$$\phi'_\mu(m) = \frac{m}{V(m)}.$$

Again, we find these derivatives to be not dependent on the specific distribution, but only on the VF. Remarkebly, then their antiderivatives do not depend either on μ. Thus, rather defining the functions $\psi_\mu(\cdot)$ and $\phi_\mu(\cdot)$ as above for specific measure μ, we define these as functions on the mean domain M of NEF $\mathcal{F}$ as (arbitrary) primitives of $1/V(m)$ and $m/V(m)$, respectively, i.e.,

$$\psi(m) \doteq \int \frac{\mathrm{d}m}{V(m)}, \tag{1.3}$$

and

$$\phi(m) \doteq \int \frac{m\,\mathrm{d}m}{V(m)}. \tag{1.4}$$

As a result, consider that a VF (V, M) of an NEF $\mathcal{F} = \mathcal{F}(\mu)$ is given, and suppose that we choose any two primitives $\psi(m)$ and $\phi(m)$ of $1/V(m)$ and $m/V(m)$, respectively. Then there exists a positive Radon measure $\mu^* \in \mathcal{M}$ such that

$$\phi(m) = \log \int_{\mathbb{R}} \exp(\psi(m)x)\mu^*(\mathrm{d}x), \quad m \in M,$$

and

$$\mathcal{F} = \mathcal{F}(\mu^*) = \left\{ F\big(m, \mu^*(\mathrm{d}x)\big) = \exp\big(x\psi(m) - \phi(m)\big)\mu^*(\mathrm{d}x) : m \in M \right\}. \tag{1.5}$$

The reparameterization of $\mathcal{F}$ in (1.5) is called the mean value parameterization of $\mathcal{F}$, see Proposition 2.3 of [31]. Accordingly, an NEF has two natural presentations: one is parameterized by canonical parameter θ and is given in (1.2) and the second by the mean parameter m, and is given in (1.5). However, as far as statistical applications are concerned, the rather more important presentation is the mean value parameterization (as θ is just an artificial parameter – the argument of the corresponding Laplace transform).

We now present the definitions of steep NEFs and EDMs:

- **Steep NEFs**: An NEF $\mathcal{F}(\mu)$ is called steep $\Leftrightarrow$ its cumulant transform $k_\mu(\theta)$ is essentially smooth convex function on D_μ (defined in (1.1)) $\Leftrightarrow$ $M = \operatorname{int} C_\mu$, cf. [8, 31]. We shall refer to this definition in the sequel.

– **EDMs**: Let $\mathcal{F} = \mathcal{F}(\mu)$ be an NEF generated by μ with Laplace and cumulant transforms L_μ and k_μ, respectively. Denote

$$\Lambda = \left\{ p \in \mathbb{R}^+ : L_\mu^p \text{ is the Laplace transform of a positive Radon measure } \mu_p \right\},$$

then Λ is nonempty due to convolution, and it is called the Jorgensen set (or the dispersion parameter space in the terminology of EDMs). It has been shown that $\Lambda = \mathbb{R}^+$ iff μ (and thus all members of $\mathcal{F}(\mu)$) is infinitely divisible. If $p \in \Lambda$, the cumulant function of μ_p is

$$k_{\mu_p}(\theta) = \log L_\mu^p(\theta) = pk_\mu(\theta).$$

Hence, the NEF generated by μ_p is the set of probability distributions

$$\mathcal{F}_p = \mathcal{F}(\mu_p) \doteq \left\{ F\big(\theta, \mu_p(\mathrm{d}x)\big) = \exp\big(\theta x - pk_\mu(\theta)\big)\mu_p(\mathrm{d}x) : \theta \in \Theta_{\mu_p} = \Theta_\mu \right\}.$$

Furthermore, the mean parameterization goes similarly as above. Denote the VF of $\mathcal{F}_p$ by (V_p, M_p), given by

$$V_p(m) = k_{\mu_p}''\big((k_{\mu_p}')^{-1}(m)\big) = pk_\mu''\big((k_\mu')^{-1}(m/p)\big) = pV(m/p), \tag{1.6}$$

and $M_p = pM_\mu$; denote primitives of $1/V_p(m)$ and $m/V_p(m)$ by $\psi_p(m)$ and $\phi_p(m)$, respectively. We choose these to satisfy

$$\psi_p(m) = \int \frac{\mathrm{d}m}{V_p(m)} = \int \frac{\mathrm{d}m}{pV(m/p)} = \int \frac{\mathrm{d}(m/p)}{V(m/p)} = \psi(m/p), \tag{1.7}$$

and

$$\phi_p(m) = \int \frac{m\mathrm{d}m}{V_p(m)} = \int \frac{m\mathrm{d}m}{pV(m/p)} = p\int \frac{(m/p)\mathrm{d}(m/p)}{V(m/p)} = p\phi(m/p). \tag{1.8}$$

Then, there is a positive Radon measure μ_p^* such that

$$\begin{aligned} \mathcal{F}_p = \mathcal{F}(\mu_p^*) &= \left\{ F\big(m, \mu_p^*(\mathrm{d}x)\big) = \exp\big(x\psi_p(m) - \phi_p(m)\big)\mu_p^*(\mathrm{d}x) : m \in M_p \right\} \\ &= \left\{ F\big(m, \mu_p^*(\mathrm{d}x)\big) = \exp\big(x\psi(m/p) - p\phi(m/p)\big)\mu_p^*(\mathrm{d}x) : m \in pM \right\}. \end{aligned} \tag{1.9}$$

The set of NEFs

$$\cup_{p \in \Lambda} \mathcal{F}_p$$

was termed by [23] the EDM corresponding to μ. In particular if $\Lambda = \mathbb{R}^+$ (*i.e.*, μ is infinitely divisible) then EDMs are used to describe the error component in generalized linear models.

Many types of VFs of NEFs have been presented and discussed in the literature, for a thorough survey see [5]. Associated with our study are VF's having a polynomial structure, for which all of the respective cumulants and moments are also polynomials. In quite generality, [3] and Corollary 3.3 of [31] showed that any r-th degree

polynomial of the form

$$V(m) = \sum_{i=1}^{r} a_i m^i, \quad m \in \mathbb{R}^+, r \in \mathbb{N}, \tag{1.10}$$

where $a_i \geq 0, i = 1, \ldots, r$, and $\sum_{i=1}^{r} a_i > 0$, is a VF of an infinitely divisible NEF. Special cases are quadratic VFs [32], and the six strictly cubic VFs [31]. Related to our study are classes of EDM's with variance functions that are power functions, or sums of power functions such as in the Tweedie class with unit variance function $V(m) = \alpha m^\gamma$ [4, 23, 24, 36], the Hinde-Demétrio class with unit variance $V(m) = m + m^\gamma$ [22, 26, 27, 29], Poisson-Tweedie mixture class with variance $V(m) = m + \alpha m^\gamma$ [25, 26], and Poisson-exponential-Tweedie models which have the VF $V(m) = m + m^2 + \alpha m^\gamma$ [1]. All these variance functions allow for non-integer powers $\gamma > 0$.

Recall that for a given VF (V, M), $\psi(m)$ and $\phi(m)$ are primitives of $1/V(m)$ and $m/V(m)$, see (1.3) and (1.4), respectively. Accordingly, if V is of the general form (1.10) it is not possible to explicitly express $\psi(m)$ and $\phi(m)$, in which case the mean value parameterization (1.9) is useless for any practical consideration. If, however, for some special cases of the a_i's coefficients, it can be calculated nicely and explicitly then so can be the corresponding likelihood function based on an appropriate random sample. This fact has a tremendous significance in statistical inference.

After this long introduction we arrive at the crux of the paper. [31] proved a proposition that (a) characterizes the VF (V, M) of NEFs that are generated by counting measures on $\mathbb{N}_0$, and (b) gives an expression for the generating measures of such NEFs.

Proposition 1.1 ([31], Prop. 4.4). *Let $\mathcal{F} = \mathcal{F}(\mu)$ be an NEF on $\mathbb{R}$ with VF (V, M).*

(a). $\mathcal{F}$ is concentrated on $\mathbb{N}_0$ with $\mu_0 > 0$ and $\mu_1 > 0$ if and only if (i) $M = (0, b)$ for some $0 < b \leq \infty$; (ii) there exists a real analytic function ϕ' on M such that $\phi'(m) = m/V(m)$ on M, and such that

$$\lim_{m \to 0} \phi'(m) = 1, \tag{1.11}$$

(b). In this case, let ϕ be a primitive of ϕ', choose a primitives $\psi(m)$ of $1/V(m)$, satisfying

$$\lim_{m \to 0} m \exp\left(-\psi(m)\right) = 1, \tag{1.12}$$

and define,

$$G(m) \doteq m \exp\left(-\psi(m)\right), \quad m \in M.$$

Then the NEF $\mathcal{F}$ of (1.5) is generated by the measure $\mu^* \in \mathcal{M}$ whose entries are computed by

$$\begin{cases} \mu_0^* = \exp\left(\phi(m)\right)\big|_{m=0}, \\ \mu_n^* = \frac{1}{n!}\left(\frac{d}{dm}\right)^{n-1}\left(\left(\exp(\phi(m))\right) \times \phi'(m) \times \left(G(m)\right)^n\right)\big|_{m=0}, \quad n = 1, 2, \ldots. \end{cases} \tag{1.13}$$

Condition (1.11) is necessary for the NEF to be concentrated on the nonnegative integers, however, it leaves infinitely many choices for the function ϕ which are all equal upto an additive constant. For our purposes it is most convenient to impose the condition

$$\lim_{m \to 0} \phi(m) = 0. \tag{1.14}$$

This condition gives $\mu_0^* = 1$, which is convenient in our analysis of the zero-inflation properties that we discuss in the sequel. Secondly, we remark that condition (1.12) could be relaxed. We state this as a lemma.

Lemma 1.2. *Proposition 1.1 holds for* (1.12) *relaxed to*

$$\lim_{m \to 0} m \exp\big(-\psi(m)\big) = c,$$

where $c > 0$ is any positive constant.

Proof. Denote the ψ-primitive that satisfies this condition by $\psi_c(m)$. Clearly, $\psi_c(m) = \psi_1(m) - \log c$, where $\psi_1(m)$ is the primitive of Proposition 1.1. When we implement the computations (1.13) with ψ_c in stead of ψ_1, the resulting measure $\widetilde{\mu}$ is an exponential shift of the μ^* of (1.13). $\qquad\square$

Returning to the concept of EDM, it is now obvious that the generating measures $\mu_p^*, p \in \Lambda$ are computed similarly as in Proposition 1.1.

Corollary 1.3. *Let $\{\mathcal{F}_p, p \in \Lambda\}$ be the EDM originating from the NEF $\mathcal{F} = \mathcal{F}(\mu)$ of Proposition 1.1. For any $p \in \Lambda$, a generating measure μ_p^* for the NEF* (1.9) *is obtained by the computations* (1.13) *in which the primitives are $\psi_p(m) = \psi(m/p)$, and $\phi_p(m) = p\phi(m/p)$.*

Proof. Because (i) the mean domain of the NEF $\mathcal{F}_p$ is $M_p = pM$, and (ii) $\phi_p'(0) = \phi'(0) = 1$, the NEF $\mathcal{F}_p$ is concentrated on $\mathbb{N}_0$. Furthermore,

$$\lim_{m \to 0} m \exp\big(-\psi_p(m)\big) = p \lim_{m \to 0}(m/p) \exp\big(-\psi(m/p)\big) = p. \tag{1.15}$$

Now apply Lemma 1.2. $\qquad\square$

Two classes of variance functions that satisfy the conditions for the corresponding NEFs to be concentrated on the nonnegative integers, were mentioned in [31]. These are the polynomial function

$$V(m) = m \prod_{i=1}^{r} \left(1 + \frac{m}{p_i}\right), \quad M = \mathbb{R}^+, \tag{1.16}$$

and the rational function

$$V(m) = \frac{m}{\prod_{i=1}^{r} \left(1 - \frac{m}{p_i}\right)}, \quad M = \big(0, \min(p_1, \ldots, p_r)\big), \tag{1.17}$$

where $p_i > 0, i = 1, 2, \ldots$, and $r \in \mathbb{N}$.

A few simple cases ($r \leq 2$) of these classes have been considered in the statistical literature, for instance, the VF

$$V(m) = m\left(1 + \frac{m}{p}\right)^2,$$

which results in the the Abel distribution, also known as generalized Poisson [13, 14].

For $r \geq 3$, the corresponding μ_n's in (1.13), and thus also the NEF probabilities in (1.5), cannot be presented neither in closed and explicit forms nor in terms of infinite sum (or some transcendental functions). They can be derived only through numerical calculations by either mathematical software as Mathematica, or Maple or by writing appropriate computer programs in 4th generation languages as Matlab, R and Python. This explains our statement above that many NEFs (at least with polynomial VF structure and degree $r \geq 3$) have not been used for statistical modeling or applications for the mere fact that they have not been known before and

thus not been considered and investigated. Therefore in this paper we intend to correct to a certain extent the 'injustice' caused to these discrete NEFs. Notice however an important point. When we refer to (1.5) in a Bayesian framework and when (1.5) serves as a prior distribution then the μ_n's calculation becomes superfluous and redundant when calculating the posterior distribution, as one can choose arbitrarily any two primitives $\psi(m)$ and $\phi(m)$. We shall further relate to this point in Section 2.

In particular, we present in the sequel two subclasses: one of the form (1.16), and one of the form (1.17). For convenience we say classes although they are subsets of (1.16) and (1.17). These classes of VFs were chosen because of the relative simplicity of the calculations of $\psi(m)$ and $\phi(m)$ for which explicit expressions are available. The two VF classes are

$$V_p(m) = m\Big(1 + \frac{m}{p}\Big)^r, \quad M = \mathbb{R}^+; \tag{1.18}$$

$$V_p(m) = \frac{m}{\big(1 - \frac{m}{p}\big)^r}, \quad M = (0, p), \tag{1.19}$$

where $p > 0$ and $r \in \mathbb{N}_0$. Later we will coin each class a name and discuss its properties. However, at this point, we will notice a very important fact. Both classes are of the form (1.6) representing VFs of EDMs. Consequently, their corresponding probabilities belong to the realm of EDMs.

The paper is organized as follows. In Section 2 we will discuss further important aspects related to the practical implementation of Proposition 1.1. In Section 3 we elaborate the two classes presented in (1.18) and (1.19). For each class we derive expressions for $\psi(m)$ and $\phi(m)$ which fulfills the premises of Proposition 1.1. We then describe some of their properties. In particular it will be shown that the corresponding NEFs' distributions are overdispersed and zero-inflated in ascending order in r. A numerical example of real data, presented in Section 4, compares the performances of our two classes to other well used discrete distributions. This example demonstrates the superiority of the members of these classes for larger power r of the polynomial VF, vis-a-vis all other distributions. Section 5 is devoted to some concluding remarks.

2. Further aspects and analysis and presentation of the two classes

As stated above, our goal is to locate classes of VFs, subclasses of (1.16) and (1.17), for which we can derive explicitly and relatively simple expressions both for the $\psi(m)$ and $\phi(m)$ functions. Our first class has variance function (see (1.18))

$$V_p(m) = m\Big(1 + \frac{m}{p}\Big)^r,$$

which belongs to the realm of (1.16). The special cases $r = 0$, $r = 1$ and $r = 2$ correspond, respectively, to the Poisson, negative binomial and Abel (or generalized Poisson) NEF's. The class in (1.18) is called the ABM class, as it was first presented by [2] in a Bayesian framework. Further details regarding such a Bayesian framework for the ABM class can be found in [5]. The second class does not have a polynomial structure. Its variance function has the form (see (1.19))

$$V_p(m) = \frac{m}{\big(1 - \frac{m}{p}\big)^r}.$$

We call this class the LM class, as being a subclass of (1.17) which was presented in [31].

Before we proceed to discuss the two classes separately in the subsections below, we will present a number of general comments regarding these classes (as well as any other classes too).

1. **Steepness**: The NEFs corresponding to the two classes of VFs are concentrated on $\mathbb{N}_0$, thus their convex support is $C = [0, \infty)$. The ABM class belongs to steep NEFs as its mean domain $M = \mathbb{R}^+$ coincides with $int C$. In contrast, the LM class is nonsteep as the corresponding mean domain $M = (0, p)$ is a proper subset of $(0, \infty)$.

2. **Infinitely divisibility and EDMs**: Both classes constitute infinitely divisible NEFs as they are subsets of (1.10) and thus the dispersion parameter space $\Lambda = \mathbb{R}^+$ (*i.e.*, they are VFs for all $p \in \mathbb{R}^+$). Thus, as indicated above, they establish EDMs.

3. **The form of Θ**: We notice that the set Θ is the image of $\mathbb{R}^+$ for the ABM clas, and the image of $(0, p)$ of the LM class by the map $m \mapsto \theta = \psi(m)$. Thus, it has the form $(-\infty, q)$, for some $q \in \mathbb{R}$. Obviously, the calculation of the inverse function $m \mapsto \theta = \psi(m)$ cannot be done in an elementary way for $r > 2$ (and sometimes also not for $r = 2$).

   If μ is bounded then one can impose conditions on μ to be a probability. The question arises, therefore, when μ is bounded. The following lemma provides an answer for the ABM class, it goes similarly for the LM class.

   Lemma 2.1. *The generating μ of the NEF $\mathcal{F}$ is bounded iff $q \geq 0$.*

   Proof. $\impliedby$: If $q > 0$ then $k(0) < \infty$. Recall that $e^{k(0)}$ is the total mass of μ. If $q = 0$, then for $\theta < 0$ we assume that $\lim_{\theta \to -\infty} k(\theta) = 0$ and write

   $$k(\theta) = \int_{-\infty}^{\theta} k'(t)\mathrm{d}t = \int_0^{k'(\theta)} k'(\psi(s))\psi'(s)\mathrm{d}s = \int_0^{k'(\theta)} \frac{s}{V(s)}\mathrm{d}s$$
   $$= \int_0^{k'(\theta)} \frac{\mathrm{d}s}{(1 + \frac{s}{p})^r},$$

   where in the last equality we used VF corresponding to the ABM class. Since $\lim_{\theta \to 0} k'(\theta) = \infty$ we can claim that

   $$\lim_{\theta \to 0} k(\theta) = \int_0^{\infty} \frac{\mathrm{d}s}{(1 + \frac{s}{p})^r} = \frac{p}{r-1}, r \geq 2.$$

   This shows that when $q = 0$, the total mass of μ is $e^{p/(r-1)}$. If μ is normalized to make it a probability then $\lim_{\theta \to -\infty} k(\theta) = 0$ is no longer fulfilled after such a normalization.
   $\implies$: If $q < 0$ the measure μ is unbounded since 0 does not belong to the closure of Θ. $\qquad\square$

4. **Overdispersion**: Recall that in statistics, overdispersion is the presence of greater variability in a data set than would be expected based on a given statistical model. For instance, the Poisson NEF which is commonly used in practice to model count data (*e.g.*, number of insurance claims; number of customers arriving into a queueing system). The theoretical mean and variance for the Poisson model are equal. On the other hand, in a large number of empirical data sets, the sample variance is considerably larger than the sample mean. Consequently, researchers have tried to model such data sets by families of distributions, such as the negative binomial and the generalized Poisson -Abel) distributions, for which the variance is larger than the mean. The statistical literature is full of articles on this subject, but we refrain from citing them for the sake of brevity.

   Consider the polynomial VF in (1.16) and denote explicitly its degree; *i.e.*,

   $$V_{r+1}(m) \doteq m \prod_{i=1}^{r} \left(1 + \frac{m}{p_i}\right).$$

Then trivially we have that the larger the degree of the polynomial, the larger is V_r, *i.e.*,

$$m = V_1(m) < V_2(m) < \cdots .$$

Firstly, the latter property indicates that all of the associated NEFs distributions are overdispersed with respect to the Poisson distribution, and secondly, there is an ascending order in r of such an overdispersion. Similarly, this overdispersion property trivially holds also for the second class given in (1.17). As the ABM, and the LM are subclasses of (1.17) or(1.16), they share the same overdispersion property. Moreover, one can simply realize that for any degree r one has $V_r^{ABM} < V_r^{LM}$, *i.e.*, the LM class is more overdispersed than the ABM one.

3. Some analysis of the ABM and LM classes

In the following two subsections we will discuss the two classes in two aspects. One is to find explicit expressions for $\psi(m)$ and $\phi(m)$ functions that satisfy the conditions (1.11), (1.12), and (1.14), and for which Proposition 1.1 is applicable. The second aspect is to show that the distributions of the relevant NEFs are zero-inflated with respect to the Poisson NEF and among themselves in an ascending order. Recall that a zero-inflated model is a statistical model based on a zero-inflated probability distribution, *i.e.* a distribution that allows for frequent zero-valued observations. In various insurance data the probability of the event of no claims during the insured period is rather large and the Poisson model does not fit. Various other models have been suggested in the realm of zero-inflated models in which the probability of zero is larger than the probability of nonzero. Such zero-inflated distributions are naturally overdispersed relative to the Poisson distribution. On this subject, too, the statistical literature is full of relevant articles, but we refrain from quoting them for reasons of brevity.

In each subsection we provide two propositions. One relates to the computations of the $\psi_p(m)$, $\phi_p(m)$, and $G_p(m)$ functions fulfilling the conditions (1.11), (1.14), and (1.15); the second proposition relates to the zero-inflated property.

3.1. The ABM class

The ABM class has been first introduced by [2] for implementing mortality projections in actuarial science. In this respect, the Lee-Carter model [30], and variants thereof (*e.g.* [34]) is a largely acceptable method of mortality forecasting. [2] 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 and showed that members of the ABM class predicts better than the Poisson the mortality rates of elderly age people. This has been demonstrated for national data of the US, Ireland and Ukraine. Since the Bayesian approach was involved, it was not relevant there to calculate neither the constants of integration for the primitives ψ and ϕ, nor the μ_n^*'s in (1.13), as these constants and mass points are cancelled out while computing the appropriate posterior distribution, for further details see [5]. They also did not demonstrate how the general expressions are obtained for ψ and ϕ. Therefore, we will provide the appropriate proof.

Proposition 3.1. *Consider the ABM class with the variance function given in (1.18) for $r \geq 1$. The corresponding $\psi_p(m)$, $\phi_p(m)$ and $G_p(m)$ functions fulfilling the conditions (1.11), (1.14), and (1.15), have the forms*

$$\psi_p(m) = \log \frac{m}{m+p} + \sum_{j=1}^{r-1} \frac{1}{j}\Big(\frac{p^j}{(m+p)^j} - 1 \Big); \tag{3.1}$$

$$\phi_p(m) = \begin{cases} p \log \frac{m+p}{p}, & (r = 1); \\ \frac{p}{r-1}\left(1 - \left(\frac{p}{m+p}\right)^{r-1}\right), & (r \geq 2); \end{cases} \tag{3.2}$$

$$G_p(m) = (m + p) \exp\left(\sum_{j=1}^{r-1} \frac{1}{j}\left(\frac{p^j}{(m+p)^j} - 1\right)\right). \tag{3.3}$$

We exclude the trivial case $r = 0$ (Poisson). As usual, an empty sum $\sum_{j=1}^{0} \cdot = 0$ in (3.1) and (3.3) in the $r = 1$ case.

Proof. Using the EDM properties (1.7) and (1.8), it suffices to determine the $\psi(m), \phi(m)$ and $G(m)$ functions of the originating ABM NEF with variance function $V(m) = m(1 + m)^r$. Consider

$$\int \frac{1}{V(m)}\, \mathrm{d}m = \int \frac{1}{m(1 + m)^r}\, \mathrm{d}m,$$

and apply polynomial division,

$$\frac{1}{m(1+m)^r} - \frac{1}{m} = -\frac{1}{m}\left(1 - \left(\frac{1}{1+m}\right)^r\right)$$
$$= -\frac{1}{m}\left(1 - \frac{1}{1+m}\right)\left(1 + \frac{1}{1+m} + \left(\frac{1}{1+m}\right)^2 + \cdots + \left(\frac{1}{1+m}\right)^{r-1}\right)$$
$$= -\frac{1}{1+m}\sum_{j=0}^{r-1}\left(\frac{1}{1+m}\right)^j = -\sum_{j=1}^{r}\frac{1}{(1+m)^j}.$$

Hence

$$\psi(m) = \int \frac{1}{m(1+m)^r}\, \mathrm{d}m = \int \left(\frac{1}{m} - \sum_{j=1}^{r}\frac{1}{(1+m)^j}\right) \mathrm{d}m$$

$$= \log m - \log(1 + m) + \sum_{j=2}^{r}\frac{1}{j-1}\frac{1}{(1+m)^{j-1}} + c = \log\frac{m}{1+m} + \sum_{j=1}^{r-1}\frac{1}{j}\frac{1}{(1+m)^j} + c,$$

where c is the integration constant. By the EDM property (1.7) we get

$$\psi_p(m) = \psi(m/p) = \log\frac{m}{m+p} + \sum_{j=1}^{r-1}\frac{1}{j}\frac{p^j}{(m+p)^j} + c.$$

Substituting $G_p(m) = m \exp\left(-\psi_p(m)\right)$, and solving $G_p(0) = p$ (see Cor. 1.3), gives

$$G_p(0) = \lim_{m \to 0}(m+p)\exp\left(-\sum_{j=1}^{r-1}\frac{1}{j}\frac{p^j}{(m+p)^j} - c\right) = p \exp\left(-\sum_{j=1}^{r-1}\frac{1}{j} - c\right) = p$$

$$\Leftrightarrow \quad c = -\sum_{j=1}^{r-1}\frac{1}{j},$$

resulting in the forms (3.1) and (3.3). Recall that ψ_p maps the mean domain M into the space Θ of the natural parameter θ. From (3.1) we note that it is not possible to express the inverse map (m as a function of θ), implying that the corresponding Laplace transform cannot be explicitly expressed as a function of θ. This is the situation that will prevail in the other classes of VFs under consideration.

Now, to find $\phi_p(m)$ fulfilling (1.14), we again solve first for

$$\phi(m) = \int \frac{m}{V(m)}\, \mathrm{d}m = \int \frac{1}{(1+m)^r}\, \mathrm{d}m = \begin{cases} \log(1+m) + d, & (r = 1); \\ -\frac{1}{r-1}\frac{1}{(1+m)^{r-1}} + d, & (r \geq 2), \end{cases}$$

with integration constant d. Applying EDM property $\phi_p(m) = p\phi(m/p)$, and the condition $\phi_p(0) = 0$, we obtain

$$d = \begin{cases} 0, & (r = 1); \\ \frac{p}{r-1}, & (r \geq 2). \end{cases}$$

Consequently, the primitive ϕ_p to be used in Proposition 1.1 has the form (3.2). □

Now we go back to discussing the ABM class in the context of zero-inflated distributions. The probability mass at zero, using $(\mu_p^*)_0 = 1$, is

$$P_r(0; p, m) \doteq (\mu_p^*)_0 \exp\big(-\phi_{p,r}(m)\big) = \exp\big(-\phi_{p,r}(m)\big), \tag{3.4}$$

$r = 1, 2, \ldots$, where $\phi_{p,r}(m)$ denotes the $\phi_p(m)$ function when the VF has degree $r + 1$. Note that the probability at 0 of the Poisson NEF with $r = 0$, is e^{-m}. We present the following proposition according to which the probability at 0 is an increasing function in r. the associated distributions become more and more zero-inflated, a feature that enables the ABM class to serve as statistical model for zero-inflated data.

Proposition 3.2. *The zero-mass probability $P_r(0; p, m)$ in the ABM class is increasing in $r \geq 0$.*

Proof. The zero probabiities are by using (3.4) and (1.8), for any $r, p, m > 0$,

$$P_r(0; p, m) = \exp\big(-\phi_{p,r}(m)\big) = \exp\big(-p\phi_{1,r}(m/p)\big).$$

Thus, for showing that $P_r(0; p, m)$ is increasing, it suffices to take $p = 1$; *i.e.*, to prove that $\phi_{1,r}(m)$ is decreasing, where $\phi_{1,r}(m)$ is given in (3.2) with $p = 1$. First, we consider $r > 1$ for which

$$\phi_{1,r}(m) = \frac{1}{r-1}\Big(1 - \big(\frac{1}{1+m}\big)^{r-1}\Big).$$

Define functions $\{f_x(s) : (0, \infty) \to \mathbb{R}, x \in (0, 1)\}$ by

$$f_x(s) = \frac{1}{s}(1 - x^s).$$

We shall argue that $f_x(s)$ is decreasing (in $s > 0$) for any $x \in (0, 1)$. The derivative

$$\frac{\mathrm{d}}{\mathrm{d}s} f_x(s) = -\frac{1}{s^2}(1 - x^s) - \frac{1}{s}x^s \log x = -\frac{1}{s}\Big(\frac{1}{s} + \frac{1}{s}\big(x^s + sx^s \log x\big)\Big).$$

A simple calculus shows that $x^s + sx^s \log x > -1$ for $s > 0$ for any $x \in (0, 1)$ (for instance by determining its minimum). Thus, $f_x(s)$ is decreasing (as function of $s > 0$), and consequently, $\phi_{1,r}(m)$ is decreasing for $r > 1$, and $P_r(0; p, m)$ is increasing for $r \geq 2$ for any values of $p > 0$ and $m > 0$.

To complete the proof we show (i) $P_0(0; p, m) < P_1(0; p, m)$, and (ii) $P_1(0; p, m) < P_2(0; p, m)$.

(i). $P_0(0; p, m) = e^{-m}$ is the zero-probability of the Poisson distribution with mean m. From (3.2) we see $P_1(0; p, m) = \big(p/(m+p)\big)^p$. Thus

$$P_0(0; p, m) < P_1(0; p, m) \iff e^{-m} < \Big(\frac{p}{m+p}\Big)^p \iff e^{-m}\Big(\frac{m+p}{p}\Big)^p < 1.$$

The latter inequality follows easily by taking the logarithm, and by noting that $\log(1+x) < x$ for $x > 0$.

(ii). It suffices to consider $p = 1$ to conclude

$$P_1(0; p, m) < P_2(0; p, m) \iff P_1(0; 1, m) < P_2(0; 1, m) \iff \phi_{1,1}(m) > \phi_{1,2}(m)$$
$$\iff \log(m+1) > 1 - \Big(\frac{1}{m+1}\Big)^2.$$

Equivalently, $\log x > 1 - 1/x^2$ for $x > 1$, which is again a simple calculus exercise.

$\square$

3.2. The LM class

The LM class is given by VFs of the form (1.19). Recall that the corresponding class of NEFs when $r \geq 1$ is non-steep with mean domain $(0, p)$, support $\mathbb{N}_0$ and convex support $[0, \infty)$. When $r = 0$ the corresponding VF is the Poisson one. [11] considered a special case $V(m) = m/(1 - \frac{m}{p})$ on the mean domain $(0, p)$ and compute explicitly a measure μ such that $\mathcal{F} = \mathcal{F}(\mu)$ is an NEF supported on $\mathbb{N}_0$.

We will compute the primitives ψ_p and ϕ_p that fulfill the conditions (1.11), (1.14), and (1.15) in much the same way as we did in Section 3.1 for the ABM class. In Appendix B we will explore a second way to find these functions, namely by following more closely the proof of Proposition 4.4 of [31], which is based on using the Lagrange formula, and express the μ_n^*'s by means of Hermite polynomials.

Proposition 3.3. *Consider the LM class with the variance function given in (1.19) for $r \geq 0$. The corresponding $\psi_p(m)$, $\phi_p(m)$, and $G_p(m)$ functions fulfilling the conditions (1.11), (1.14), and (1.15), have the forms*

$$\psi_p(m) = \log \frac{m}{p} + \sum_{i=1}^{r}(-1)^i \frac{1}{i}\binom{r}{i}\Big(\frac{m}{p}\Big)^i; \tag{3.5}$$

$$\phi_p(m) = \frac{p}{r+1}\Big(1 - \big(1 - \frac{m}{p}\big)^{r+1}\Big); \tag{3.6}$$

$$G_p(m) = p \exp\Big(-\sum_{i=1}^{r}(-1)^i \frac{1}{i}\binom{r}{i}\Big(\frac{m}{p}\Big)^i\Big) \tag{3.7}$$

Proof. We determine the $\psi(m), \phi(m)$ and $G(m)$ functions of the originating LM NEF with variance function $V(m) = m/(1-m)^r$. Namely,

$$\int \frac{1}{V(m)}\,\mathrm{d}m = \int \frac{(1-m)^r}{m}\,\mathrm{d}m = \int \frac{1}{m}\sum_{i=0}^{r}\binom{r}{i}(-m)^i\,\mathrm{d}m$$
$$= \int \frac{1}{m}\,\mathrm{d}m + \sum_{i=1}^{r}(-1)^i\binom{r}{i}\int m^{i-1}\,\mathrm{d}m = \log m + \sum_{i=1}^{r}(-1)^i\frac{1}{i}\binom{r}{i}m^i + c,$$

which gives by (1.7)

$$\psi_p(m) = \log\frac{m}{p} + \sum_{i=1}^{r}(-1)^i\frac{1}{i}\binom{r}{i}\left(\frac{m}{p}\right)^i + c.$$

Substituting $G_p(m) = m\exp\left(-\psi(m)\right)$, and solving $G_p(0) = p$ (see Cor. 1.3), gives

$$G_p(0) = \lim_{m\to 0} p\exp\left(-\sum_{i=1}^{r}(-1)^i\frac{1}{i}\binom{r}{i}\left(\frac{m}{p}\right)^i - c\right) = p \quad\Leftrightarrow\quad c = 0,$$

resulting in the forms (3.5) and (3.7).

For $\phi(m)$,

$$\int\frac{m}{V(m)}\,\mathrm{d}m = \int(1-m)^r\,\mathrm{d}m = -\frac{1}{r+1}(1-m)^{r+1} + d.$$

Applying EDM property $\phi_p(m) = p\phi(m/p)$, and the condition $\phi_p(0) = 0$, we obtain

$$d = \frac{p}{r+1}.$$

Consequently, the primitive ϕ_p has the form (3.6). $\qquad\square$

We now examine the zero-inflated property. We use similar notations as in the ABM case and denote by $P_r(0; p, m)$ the zero-mass probability of the distribution associated with the LM NEF obtained by VF $m/(1-\frac{m}{p})^r$.

Proposition 3.4. *The zero-mass probability $P_r(0; p, m)$ in the LM class is increasing in $r \geq 0$.*

Proof. Similar as in the proof of Proposition 3.2 it suffices to show that $\phi_{1,r}(m)$ is decreasing, where $\phi_{1,r}(m)$ is given in (3.6) with $p = 1$:

$$\phi_{1,r}(m) = \frac{1}{r+1}\left(1 - (1-m)^{r+1}\right), \quad 0 < m < 1;\ r = 0, 1, \ldots.$$

Define functions $\{f_x(s) : (0,\infty) \to \mathbb{R}, x \in (0,1)\}$ by

$$f_x(s) = \frac{1}{s}(1 - x^s).$$

In the proof of Proposition 3.2 we showed that $f_x(s)$ is decreasing (in $s > 0$) for any $x \in (0,1)$. Specifically, we get that $\phi_{1,r}(m)$ is decreasing for $r \geq 0$, and $P_r(0; p, m)$ is increasing for $r \geq 0$ for any values of $p > 0$ and $m \in (0, p)$.

$\qquad\square$

Remark 3.5. For the LM class we assumed that r is a natural number. However, all results obtained for this class are also correct for any real number $r \geq 1$ as the LM class of VFs can be shown to fulfill the premises of [31], Proposition 4.4. Consequently, the finite sum in (3.5) could be replaced by sum of entire series using the binomial series of Newton instead of the binomial formula of Pascal. Note that the proof of Proposition 3.4 used already any $r \geq 1$.

However, we focus in this work only on classes for which we can obtain relatively simple expressions for both $\psi_p(m)$ and $\phi_p(m)$ in the form of finite sums and the like and not in sums of entire series.

TABLE 1. Data set of insurance claims.

Value	0	1	2	3	4	5
Frequency	3719	232	38	7	3	1

TABLE 2. Descriptive statistics.

Number of observations	4000
Mean	0.0865000
Cariance	0.122548
Skewness	5.31602
Kurtosis	41.0067
Fraction zeros	0.929750
Index of dispersion	1.41674

4. A NUMERICAL EXAMPLE

In this section we show that our classes are very well suited for fitting small counting data. Consider the well-known 6 data sets of automobile insurance claims per policy over a fixed period of time that have been studied in [21]. They fitted Poisson (P) and the negative Binomial distributions (NB), Since then, many models have been developed for fitting one or more of these data sets [9, 12, 15, 17–20, 26, 28, 35, 37]. It is not the purpose of this paper to give a full description of all the 6 data sets, of all these fitting models, and of a full comparison with our ABM and LM models. For the complete picture, we refer to the ancillary file of [7]. Here, we consider the data set of insurance claims in Zaire in 1974.
The descriptive statistics of the data show over-dispersion (index of dispersion, defined as the variance divided by the mean, is larger than one), zero-inflation (fraction of zeros is more than 90%), and relative large skewness and kurtosis (see Tab. 2).

The following two-parameter models and their probability mass functions have been considered for fitting. We leave out models with one parameter, with three or more parameters, and regression models.

(a). PIG (Poisson-inverse Gaussian distribution) in [37],

$$p_n = \int_0^\infty \frac{e^{-x} x^n}{n!} f(x; \beta, \mu) \, \mathrm{d}x, \ n = 0, 1, \ldots,$$

where $f(x; \beta, \mu)$ is the inverse Gaussion pdf of the form

$$f(x; \beta, \mu) = \frac{\mu}{\sqrt{2\pi\beta x^3}} e^{-\frac{(x-\mu)^2}{2\beta x}}, \ x > 0,$$

with positive parameters β and μ.

(b). NLD (new logarithmic distribution) in [19],

$$p_n = \frac{\log(1 - \alpha\theta^n) - \log(1 - \alpha\theta^{n+1})}{\log(1 - \alpha)} \ n = 0, 1, \ldots,$$

with parameters $\alpha < 1$ ($\alpha \neq 0$), and $0 < \theta < 1$.

(c). PLB (Poisson-Lindley-Beta prime distribution) in [20],

$$p_n = \frac{\alpha(1+\alpha)\Gamma(\alpha+\beta)\Gamma(\beta+n)}{\Gamma(\beta)\Gamma(\alpha+\beta+n+3)} \left((\beta+n)(2+n)+\alpha+2\right) n = 0, 1, \ldots,$$

with positive parameters α, β.

(d). GDP (a new geometric discrete Pareto distribution) in [9],

$$p_n = \frac{q^n}{(n+1)^\alpha} - \frac{q^{n+1}}{(n+2)^\alpha}, \; n = 0, 1, \ldots,$$

with parameters $0 < q \leq 1$ and $\alpha \geq 0$.

(e). BTD (Bell-Touchard discrete distribution) in [12],

$$p_n = \frac{1}{n!} e^{\theta(1-e^\alpha)} \alpha^n T_n(\theta), \; n = 0, 1, \ldots,$$

with positive parameters α, θ, and where the $T_n(\cdot)$ are Touchard polynomials; *i.e.*,

$$T_n(\theta) = e^{-\theta} \sum_{k=0}^{\infty} \frac{k^n \theta^k}{k!}$$

We compare the fitted distributions of the models given above with the distributions of our ABM and LM classes for a range of powers r. The performances of the fitted distributions are computed, using the parameters that are reported in the cited references. The mean parameter m in our models is estimated by the sample mean of the data, while the dispersion parameter p is computed by maximum likelihood estimation. The resulting values for $r = 1, \ldots, 10$ are shown in Table 3.

The computation of the probabilities of the NEF (1.9) is done by a numerical computer program. Given parameters m and p of the variance function $V_p(m)$, the functions ψ_p and ϕ_p follow from Proposition 3.1 (ABM class) and Proposition 3.3 (LM class), the measure $(\mu_p^*)_n$ is computed numerically by solving the derivatives in (1.13). In Appendix A we give the main computational ingredients of this procedure, but we refer to [7] for more technical details. Moreover, the code is available at https://github.com/adriaanridder/abm.

TABLE 3. The estimated dispersion parameter p for the ABM and LM classes.

r	ABM	LM
1	0.216600	0.277098
2	0.459964	0.520502
3	0.704120	0.764666
4	0.948471	1.009018
5	1.192899	1.253448
6	1.437365	1.497914
7	1.681853	1.742403
8	1.926354	1.986905
9	2.170867	2.231417
10	2.415385	2.475934

TABLE 4. Performance measures of the fitting models. ABM and LM for $r \in \{1, \ldots, 10\}$ that gave the highest p-value.

Model	χ^2	df	p-value	RMSE	KL
PIG	0.543789	2	0.761935	1.760396	1.9866e-04
NLD	2.312184	2	0.314714	2.235378	3.1047e-04
PLB	0.370556	2	0.830873	1.488770	2.0715e-04
GDP	0.383445	2	0.825536	1.303073	1.7734e-04
BTD	9.251567	2	0.009796	4.056789	8.6835e-04
ABM $(r = 10)$	0.444362	2	0.800770	0.726298	1.5896e-04
LM $(r = 4)$	0.382901	2	0.825760	1.044667	1.6972e-04

Let $x_0, x_1, \ldots, x_K$ be the data set of counts as presented in Table 1; thus $K = 5$, and x_k is the observed number of value k. Let $N = \sum_{k=0}^{K} x_k$ be the total number of observations. The empirical probability mass function is

$$p_k^{(\text{emp})} = \frac{x_k}{N}, \ k = 0, \ldots, K.$$

The probabilities of the distribution of a fitting model are denoted by p_k^{mod}. The performance of a fitting model is expressed through the following measures.

– χ^2 value; taken into account sufficiently large expected numbers in the categories. In this way, the data set of Table 1 has 5 categories $0, \ldots, 3, \geq 4$. Doing this as well for the fitting models, the χ^2 statistic is

$$\chi^2 = \sum_{k=0}^{3} \frac{\left(x_k - N p_k^{(\text{mod})}\right)^2}{N p_k^{(\text{mod})}} + \frac{\left(\sum_{k \geq 4} x_k - N \sum_{k \geq 4} p_k^{(\text{mod})}\right)^2}{N \sum_{k \geq 4} p_k^{(\text{mod})}}.$$

– p-value of the χ^2 quantile; taken into account the number of categories and the number of parameters that are estimated from the data. For the data set of Table 1 partitioned in 5 categories and 2 parameters, we get $5 - 2 - 1 = 2$ degrees of freedom.

– Root mean squared error (RMSE):

$$\sqrt{\frac{1}{K+1} \sum_{k=0}^{K} \left(x_k - N p_k^{\text{mod}}\right)^2}.$$

– Kullback-Leibler divergence (KL):

$$\sum_{k=0}^{K} p_k^{(emp)} \log \frac{p_k^{(emp)}}{p_k^{(mod)}}.$$

When we consider the p-value criterion (see Tab. 4), several distributions are compatible, including the ABM and LM models. However, for the RMSE and Kullback-Leibler criteria, the ABM and LM models show a major improvement. In [7] we give an overview of comparisons of these models for many more data sets having various different statistical properties. The overall picture is that our models show competitive, or best performances in all cases, while the other models perform sometimes good sometimes bad.

It might be of interest to present the performances of the investigated ABM and LM classes for $r = 1, \ldots, 10$. We show these in figures.
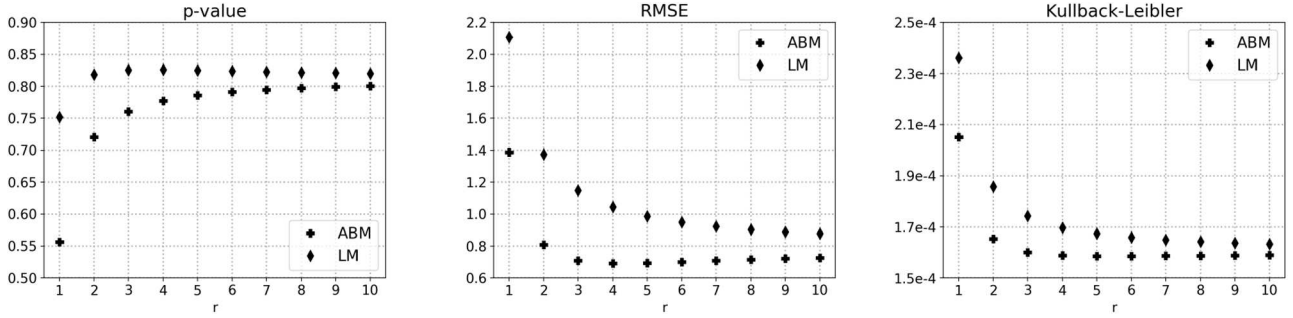
FIGURE 1. Performances of the fitted models of the ABM and LM classes for r upto 10.

From these figures we make a few observations. Firstly, the LM class performs better than the ABM class for the p-value performance criterion, but not for the other criteria. Secondly, when the degree r of the variance function increases, the performances of ABM and LM become more and more equal. Also, note that performance criteria do not alsway show a monotone behaviour.

4.1. Poisson-Tweedie class

A corresponding approach of modelling counting distributions by exploiting the concept of EDM's with power variance functions, has been developed in [10, 25, 26] by considering the Poisson-Tweedie class. We shall sketch the distribution, compute the fit to the data set of Table 1, and compare with our models.

The Poisson-Tweedie distribution consists of a Tweedie distribution mixture of Poisson probabilities [10, 26]. We choose the Tweedie distribution from the EDM with mean $m > 0$, $p > 0$ the dispersion parameter, and $\gamma \geq 1$ the power of the Tweedie variance function. However, we exclude $\gamma = 1$ (Poisson distribution), $1 < \gamma < 2$ (compound Poisson), and $\gamma = 2$ (Gamma). Then, the variance satisfies the power function

$$V(m) = m + p^{1-\gamma}m^\gamma = m\Big(1 + \big(\frac{m}{p}\big)^{\gamma-1}\Big).$$

Furthermore, its cumulant is [26],

$$\kappa(\theta) = \frac{p(\gamma-1)^\alpha}{\gamma-2}\Big((-t)^\alpha - \big(1 - t - e^\theta\big)^\alpha\Big),\ \theta \leq -t,$$

where $\alpha \in (0,1)$ is the stability parameter of a positive α-stable distribution, satisfying $(\gamma - 1)(1 - \alpha) = 1$. Parameter $t \leq 0$ is the natural parameter of the Tweedie distribution which becomes by reparameterization with respect to the mean

$$t = -\frac{1}{\gamma-1}\big(\frac{m}{p}\big)^{1-\gamma}.$$

From the cumulant one derives a recursion for the associated probabilities $P(n) = P_\gamma(m, p)$ [16],

$$P(0) = \exp\Big(\frac{p(\gamma-1)^\alpha}{\gamma-2}\big((-t)^\alpha - (1-t)^\alpha\big)\Big),$$

TABLE 5. Performance measures of ABM, and LM models in comparison with the Poisson-Tweedie.

Model	χ^2	df	p-value	RMSE	KL
ABM($r = 10$)	0.444362	2	0.800770	0.726298	1.5896e-04
ABM($r = 20$)	0.428988	2	0.806950	0.754472	1.5961e-04
LM($r = 4$)	0.382901	2	0.825760	1.044667	1.6972e-04
PT($\gamma = 2.683$)	0.437404	2	0.803561	1.051929	1.6452e-04

and for $n \geq 0$,

$$P(n+1) = \frac{p(\gamma-1)^{\alpha-1}(1-t)^{\alpha-1}}{n+1} \sum_{j=0}^{n} \frac{\Gamma(j+1-\alpha)}{\Gamma(1-\alpha)j!} \left(\frac{1}{1-t}\right)^j P(n-j).$$

Given the data of Table 1, we executed similar experiments as we did for our models. We considered a range of γ-values, and given a γ-value, the mean parameter m is estimated by the sample mean of the data, while the dispersion parameter p is computed by maximum likelihood estimation. From all these fitted models the highest p-value was obtained for

$$m = 0.0865, \ p = 0.1432, \ \gamma = 2.683.$$

The performance of this fitted model is shown in Table 5, where we repeat our best models. The p-values of the fitted ABM models were increasing with the power r, thus we decided to run the ABM model for powers r upto $r = 20$.

The conclusion is that the Poisson-Tweedie is a competitve model to our models, but it performs slightly worse in all criteria.

5. CONCLUDING REMARKS

1. In this paper we analysed in depth two classes of EDMs of distributions supported on the set of non-negative integers, suggested in literature. Specifically, we derive exact formulations of their mean value parametrization, and we provide computational results. This enables us to execute data modelling which we show by fitting oiur models to a real data set.

2. The classes of EDMs introduced in this paper can be used, for example, as competitors and alternatives to the Poisson or negative binomial NEFs for modeling count data in various actuarial aspects and insurance claims. This has been indeed demonstrated in the numerical section. However, based on our experience in the insurance and actuarial industry, we have noticed that professionals are very concerned about using new (both discrete and continuous) distributions to estimate and evaluate various relevant parameters as the insurance risk factor. So in another paper of ours [6] we considered, just for the sake of demonstration, the problem of computing the insurance risk factor

$$\ell(x) = \mathbb{P}\Big(\sum_{k=0}^{N} Y_k > x\Big),$$

for large values of x, where N is a discrete random variable, counting the number of claims during a fixed period of time, and the Y_i's are the respective independent claim sizes. The conventional actuarial literature is full with models in which N has either Poisson or negative binomial distributions whereas the Y_i's have a common gamma or inverse Gaussian or even positive stable distribution. [6] used 'unconventional'

NEF distributions for N by taking the Abel, strict arcsine and Takács NEFs, *i.e.*, NEFs having cubic VFs characterized by [31]. For data of a Swedish claims at a car insurance company they considered all combinations of the distributions of N and the Y_i's mentioned above and demonstrated that the best fit for such data is obtained for the pair (arcsine, positive stable).

3. Consequently, we trust that the ABM and and LM classes (as well as other similar classes) are going to play a significant role as a 'new generation' of counting distributions and to have a 'prosperous future' in applications to actuarial science data as well as to other statistical data. Indeed, in [7] the present authors conducted a project in which more than 20 sets of count data from the statistical literature were collected. Such data were modeled by some conventional discrete distributions and were compared to count probabilities belonging to the ABM and LM classes. And so, as we expected, all of the latter count probabilities, and with respect to various of metrics or goodness-of-fit tests, have shown superiority and provided a much better fit for each of these data sets.

4. One last remark. Researchers may avoid using the LM class as it is non-steep. However, another important class of NEFs having power VFs of the form $(V, M) = (\alpha m^\gamma, \mathbb{R}^+), \alpha > 0, \gamma < 0$ (which belong to the Tweedie scale) is also non-steep. Indeed, for the latter class $M = \mathbb{R}^+$ whereas its convex support $C = \mathbb{R}$. This class though is frequently used in various applications.

Appendix A. ABM computations

In this section we briefly sketch how the measure μ^* of the ABM distribution is computed in our code. An extended version can be found in [7], and the code is available at https://github.com/adriaanridder/abm. Consider the variance function of the EDM, $V_p(m) = m(1 + m/p)^r$, for the cases $r \geq 2$. The required functions ψ_p, ϕ_p, and G_p are given in Proposition 3.1, and the measure μ^* is given in Proposition 1.1 to be

$$\mu_n^* = \frac{1}{n!}\Big(\frac{\mathrm{d}}{\mathrm{d}m}\Big)^{n-1} e^{H_n(m)}\Big|_{m=0},$$

where

$$\begin{aligned} H_n(m) &= \phi_p(m) + \log \phi_p'(m) + n \log G_p(m) \\ &= \phi_p(m) + \log \phi_p'(m) - n\big(\psi_p(m) - \log(m)\big). \end{aligned}$$

Apply the chain rule for μ_n^*:

$$\Big(\frac{\mathrm{d}}{\mathrm{d}m}\Big)^{n-1} e^{H_n(m)} = \sum \frac{(n-1)!}{k_1! k_2! \cdots k_{n-1}!} \Big(\big(\frac{\mathrm{d}}{\mathrm{d}x}\big)^k e^x\Big)_{x=H_n(m)} \prod_{j=1}^{n-1} \Big(\frac{H_n^{(j)}(m)}{j!}\Big)^{k_j},$$

where the sum is over all nonnegative integer solutions of the Diophantine equation $\sum_{j=1}^{n-1} j k_j = n - 1$, and where $k = \sum_{j=1}^{n-1} k_j$. This leads to

$$\mu_n^* = \frac{1}{n} e^{H_n(0)} \sum \prod_{j=1}^{n-1} \frac{\big(H_n^{(j)}(0)/j!\big)^{k_j}}{k_j!}.$$

Using the expressions of ψ_p and ϕ_p as presented in (3.1) and (3.2), we can write the H_n function in the form

$$H_n(m) = q_0 + \sum_{i=1}^{r-1} q_i (m + p)^{-i} + q_r \log(m + p).$$

Thus the derivatives of H_n are sums of derivatives of $(m + p)^{-i}$ and $\log(m + p)$.

The computation of the LM class measure goes similarly.

Appendix B. LM miscellaneous

We now present a second way to compute the μ_n's for the NEFs corresponding to the LM class by means of Hermite polynomials, a way suggested to us by Gérard Letac (a personal communication).

Proposition B.1. *Let $\mathcal{F}$ be the NEF corresponding to the VF $m/(1 - \frac{m}{p})^r, 0 < m < p$. Then $\mathcal{F}$ is generated by*

$$\delta_0 + \sum_{k=1}^{\infty} \frac{p^k}{k!} \nu^{*k},$$

where ν^{*k} is the k-th fold convolution of the positive measure ν on $\mathbb{N}$ given by

$$\nu(n) = \frac{1}{n!n} \left[\left(\frac{\mathrm{d}}{\mathrm{d}m} \right)^{n-1} e^{nP(m)} \right]_{m=0}$$

and

$$P(m) = -\sum_{k=1}^{\infty} \frac{(-r)_k}{k!k} m^k,$$

where $(-r)_k$ is the Pochhammer symbol

$$(-r)_k = -r(-r + 1)(-r + 2) \cdots (-r + k - 1).$$

Proof. It sufficers to consider $p = 1$. For this case we have

$$d\theta = \frac{\mathrm{d}m}{V_{\mathcal{F}}(m)} = (1 - m)^r \frac{\mathrm{d}m}{m} = \frac{\mathrm{d}m}{m} + \sum_{k=1}^{\infty} \frac{(-r)_k}{k!} m^{k-1} \mathrm{d}m$$

Thus $\theta = \log m - P(m)$ which by denoting $w = e^\theta$ we get $m = we^{P(m)}$. Now apply the Lagrange formula which states that if $h(w) = wg(h(w))$ then

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

When applying this formula to $m = h(w) = k'_\mu(\theta)$ and $g(m) = e^{P(m)}$ we get

$$k'_\mu(\theta) = \sum_{n=1}^{\infty} \frac{w^n}{n!} \left[(\frac{\mathrm{d}}{\mathrm{d}m})^{n-1} e^{nP(m)} \right]_{m=0}.$$

Since $d\theta = dw/w$ we obtain

$$k_\mu(\theta) = \sum_{n=1}^{\infty} \frac{w^n}{n!n} \left[(\frac{\mathrm{d}}{\mathrm{d}m})^{n-1} e^{nP(m)} \right]_{m=0} = \sum_{n=1}^{\infty} \nu(n) w^n,$$

and the remainder of the proof is standard. □

Example B.2. For $r = 1$, $P(m) = m$ and $\nu(n) = n^{n-2}/n!$.

Example B.3. For $r = 2$, $P(m) = 2m - m^2/2$ but the computation of

$$\left[(\frac{\mathrm{d}}{\mathrm{d}m})^{n-1} e^{n(2m - m^2/2)} \right]_{m=0}$$

is more delicate. For such a computation with use the formula for Hermite polynomials, see] [33, p.130], by which

$$e^{2xt - t^2} = \sum_{k=0}^{\infty} H_k(x) \frac{t^k}{k!}.$$

Setting $x = \sqrt{2n}$ and $t = \sqrt{n/2}m$ yields

$$e^{n(2m - m^2/2)} = \sum_{k=0}^{\infty} H_k(\sqrt{2n}) \left(\frac{n}{2}\right)^{k/2} \frac{m^k}{k!}.$$

By employing the Taylor formula it follows that

$$\left[(\frac{\mathrm{d}}{\mathrm{d}m})^{n-1} e^{n(2m - m^2/2)} \right]_{m=0} = H_{n-1}(\sqrt{2n}) \left(\frac{n}{2}\right)^{(n-1)/2}$$

and thus

$$\nu(n) = \frac{1}{n!n} H_{n-1}(\sqrt{2n}) \left(\frac{n}{2}\right)^{(n-1)/2}.$$

References

[1] R. Abid, C.C. Kokonendji and A. Masmoudi, On Poisson-exponential-Tweedie models for ultra-overdispersed count data. To appear in: *AStA Adv. Stat. Anal.* (2020). 10.1007/s10182-020-00375-4.

[2] Y. Awad, S.K. Bar-Lev and U. Makov, A new class counting distributions embedded in the Lee-Carter model for mortality projections: a Bayesian approach. A technical report No. 146. Actuarial Research Center, University of Haifa, Israel (2016).

[3] S.K. Bar-Lev, Discussion on paper by B. Jørgensen, "Exponential dispersion models". *J. Roy. Stat. Soc. Ser. B* **49** (1987) 153–154.

[4] S.K. Bar-Lev and P. Enis, Reproducibility and natural exponential families with power variance functions. *Ann. Stat.* 14 (1986) 1507–1522.

[5] S.K. Bar-Lev and C.C. Kokonendji, On the mean value parameterization of natural exponential families – a revisited review. *Math. Methods Stat.* **26** (2017) 159–175.

[6] S.K. Bar-Lev and A. Ridder, Monte Carlo methods for insurance risk computation. *Int. J. Stat. Probab.* **8** (2019) 54–74.

[7] S.K. Bar-Lev and A. Ridder, Exponential dispersion models for overdispersed zero-inflated count data. Preprint arXiv: 2003.13854v1 (2020).

[8] O. Barndorff-Nielsen, Information and Exponential Families in Statistical Theory. Wiley, New York (1978).

[9] D. Bhati and H.S. Bakouch, A new infinitely divisible discrete distribution with applications to count data modelling. *Commun. Stat. Theory Methods* **48** (2019) 1401–1416.

[10] W.H. Bonat, B. Jørgensen, C.C. Kokonendji, J. Hinde and C.G.B. Demétrio, Extended Poisson–Tweedie: Properties and regression models for count data. *Stat. Model.* **18** (2018) 24–49.

[11] W. Bryc and M. Ismail, Approximation operators, q-exponential, and free exponential families. Preprint arXiv:math/0512224 (2005).

[12] F. Castellares, A.J. Lemonte and G. Moreno-Arenas, On the two-parameter Bell–Touchard discrete distribution. *Commun. Stat. Theory Methods* **49** (2020) 4834–4852.

[13] P.C. Consul, Generalized Poisson Distributions: Properties and Applications. Marcel Dekker, New York (1989).

[14] P.C. Consul and F. Famoye, Lagrangian Probability Distributions. Birkhäuser, Boston, Basel, Berlin (2006).

[15] M. Denuit, A new distribution of Poisson-type for the number of claims. *ASTIN Bull.* **27** (1997) 229–242.

[16] A.H. El-Shaarawi, R. Zhu and H. Joe, Modelling species abundance using the Poisson-Tweedie family. *Environmetrics* **22** (2011) 152–164.

[17] Y. Gençtürk and A. Yiğiter, Modelling claim number using a new mixture model: negative binomial gamma distribution. *J. Stat. Comput. Simul.* **86** (2016) 1829–1839.

[18] E. Gómez-Déniz and E. Calderin-Ojeda, The discrete Lindley distribution: properties and applications. *J. Stat. Comput. Simul.* **81** (2011) 1405–1416.

[19] E. Gómez-Déniz, J.M. Sarabia and E. Calderin-Ojeda, A new discrete distribution with actuarial applications. *Insurance: Math. Econ.* **48** (2011) 406–412.

[20] E. Gómez-Déniz, A. Hernández-Bastida and M.P.A. Fernández-Sánchez, A suitable discrete distribution for modelling automobile claim frequencies. *Bull. Malays. Math. Sci. Soc.* **39** (2016) 633–647.

[21] A. Gossiaux and J. Lemaire, Methodes d'ajustement de distributions de sinistres. *Bull. Assoc. Swiss Actuar.* **81** (1981) 87–95.

[22] J. Hinde and C.G.B. Demétrio, Overdispersion: models and estimation. *Comput. Stat. Data Anal.* **27** (1998) 151–170.

[23] B. Jørgensen, Exponential dispersion models (with discussion). *J. Roy. Stat. Soc. B* **49** (1987) 127–162.

[24] B. Jørgensen, The Theory of Exponential Dispersion Models. Vol. 76 of *Monographs on Statistics and Probability*. Chapman and Hall, London (1997).

[25] B. Jørgensen and C.C. Kokonendji, Discrete dispersion models and their Tweedie asymptotics. *AStA Adv. Stat. Anal.* **100** (2016) 43–78.

[26] C.C. Kokonendji, S. Dossou-Gbété and C.G.B. Demétrio, Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes. *Stat. Oper. Res. Trans.* **28** (2004) 201–214.

[27] C.C. Kokonendji, C.G.B. Demétrio and S.S. Zocchi, On Hinde–Demétrio regression models for overdispered count data. *Stat. Methodol.* **4** (2007) 277–291.

[28] C.C. Kokonendji and M. Khoudar, On strict arcsine distribution. *Commun. Stat. - Theory Methods* **33** (2004) 993–1006.

[29] C.C. Kokonendji and D. Malouche, A property of count distributions in the Hinde-Demétrio family. *Commun. Stat. Theory Methods* **37** (2008) 1823–1834.

[30] R.D. Lee and L. Carter, Modelling and forecasting the time series of US mortality. *J. Am. Stat. Assoc.* **87** (1992) 659–671.

[31] G. Letac and M. Mora, Natural real exponential families with cubic variance functions. *Ann. Stat.* **18** (1990) 1–37.

[32] C.N. Morris, Natural exponential families with quadratic variance functions. *Ann. Stat.* **10** (1982) 65–80.

[33] E.D. Rainville, Special Functions. The Macmillan Company, New York (1960).

[34] A.E. Renshaw and S. Haberman, A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Math. Econ.* **38** (2006) 556–570.

[35] M. Ruohonen, On a model for the claim number process. *ASTIN Bull.* **18** (1988) 57–68.

[36] M.C.K. Tweedie, An index which distinguishes between some important exponential families. Statistics: Applications and New Directions. Edited by J. K. Ghosh and J. Roy. *Proc. Indian Institute Golden Jubilee lnternat Conf.* Indian Statist. Inst., Calcutta (1984) 579–604

[37] G. Willmot, The Poisson-inverse Gaussian distribution as an alternative to the negative binomial. *Scand. Actuar. J.* **1987** (1987) 113–127.