An introduction to Bent Jørgensen's ideas

Gauss M. Cordeiro[a], Rodrigo Labouriau[b] and Denise A. Botter[c]

[a]*Universidade Federal de Pernambuco*
[b]*Aarhus University*
[c]*Universidade de São Paulo*

Abstract. We briefly expose some key aspects of the theory and use of dispersion models, for which Bent Jørgensen played a crucial role as a driving force and an inspiration source. Starting with the general notion of dispersion models, built using minimalistic mathematical assumptions, we specialize in two classes of families of distributions with different statistical flavors: exponential dispersion and proper dispersion models. The construction of dispersion models involves the solution of integral equations that are, in general, untractable. These difficulties disappear when more mathematical structure is assumed: it reduces to the calculation of a moment generating function or of a Riemann–Stieltjes integral for the exponential dispersion and the proper dispersion models, respectively. A new technique for constructing dispersion models based on characteristic functions is introduced turning the integral equations above into a tractable convolution equation and yielding examples of dispersion models that are neither proper dispersion nor exponential dispersion models. A corollary is that the cardinality of regular and non-regular dispersion models are both large.

Some selected applications are discussed including exponential families non-linear models (for which generalized linear models are particular cases) and several models for clustered and dependent data based on a latent Lévy process.

1 Introduction

The work of Bent Jørgensen is rich and deep. Working on the foundations of generalized linear models (GLMs) he gave the main steps in the construction of the theory of exponential dispersion models (EDMs), a class of parametric families of one-dimensional distributions. The necessity of expanding the EDMs to include further parametric families of distributions (*e.g.*, the von Mises and the simplex distributions) led to the construction of the class of proper dispersion models (PDMs) in a process which culminated with the development of the general theory of dispersion models. The class of dispersion models encompasses both the EDMs and the PDMs under the same umbrella. In this article, we will briefly expose some of the main aspects of the theoretical development referred above for which Bent Jørgensen played a crucial role being a driving force and an inspiration source.

The path from the construction of EDMs to the development of the general theory of dispersion models presents an increasing level of abstraction. While EDMs rely on relatively strong assumptions (inherited from exponential families of distributions), the general dispersion models are based on a minimalistic mathematical structure while keeping the desirable distributional and statistical properties. This theory-generation process was inductive in the sense that it represented a movement from several particular cases (disparate collections of distributions) to a general setup (large classes of families of probability measures with common properties). In this article, our exposition follows a reversed order, starting from the

---

general theory of dispersion models (Section 2.1) involving a weak mathematical structure and specializing in two different statistical scenarios: EDMs and PDMs (Sections 2.2 and 2.3, respectively).

We anticipate that when working in the general scenario, we meet some difficulties even for solving basic questions like finding a mathematical procedure to generate the families of probability involved there. For instance, the generation of dispersion models involves the solution of Fredholm integral equations that are known to be mathematically hard, and in many instances, it is difficult even to establish whether the equation has a solution or not. When introducing more mathematical structure, as expected, those difficulties gradually disappear. Interestingly, arguments of probabilistic and statistical nature (*e.g.*, a coincidence of an integral with the moment generating function, probabilistic properties of the characteristic function and exactness of the so-called $p^\star$ approximation) turn the hard mathematical problems tractable. The general integral equation becomes a convolution equation (solvable using standard methods of generalized functions or other methods of deconvolution) or involves a simple calculation of a Riemman–Stieltjes integral in some of the more specific scenarios.

We stress that the deductive path used in our exposition (starting from the general and moving to more particular cases) was only feasible because we knew the entire inductive path referred to above and then we essentially reversed the order of the construction. We believe that this choice of disposition of arguments is more comfortable to follow and enlighten the theoretical construction of the theory of dispersion models, but we invite the reader to imagine the reversed order while reading the text in order to get a feeling of the difficulties that she/he would face in constructing the general theory of EDMs. In our view, the most significant contribution that Bent Jørgensen gave to the theory of dispersion models was to envision the structure that we briefly present here.

Bent Jørgensen's professional trajectory was circular, starting in Denmark (Aarhus and Odense), he moved to several countries (England, Brazil and Canada) and eventually returned to Denmark (Aarhus-Foulum and finally Odense, where he was born). He obtained a master degree from Aarhus University (1979) with a thesis on the inverse Gaussian distribution, under the supervision of Ole Barndorff–Nielsen. His master thesis was later published in a book (Jørgensen, 1982). Bent Jørgensen received a Ph.D. title from the University of Southern Denmark (1987), also under the supervision of Ole Barndorff–Nielsen and a Doctor of Science degree from Aalborg University (1997). He also studied at the Imperial College, in London, from 1981 to 1983. His famous book "The Theory of Dispersion Models" (Jørgensen, 1997) was primarily derived from the results of his PhD thesis and his doctoral dissertation. A significant part of this work was published as lecture notes after teaching a course in the First School of Regression Models in São Paulo in 1989.

Bent Jørgensen came to the "Instituto de Matematica Pura e Aplicada" (IMPA, in Rio de Janeiro, Brazil) by the influence of Gauss Cordeiro who met him as a student at the Imperial College (London) in 1981. During a congress in York, in 1986, a few months before coming to Brazil, Bent expressed the desire to work in Brazil because of his Brazilian wife, Vera Botelho. Gauss Cordeiro took the steps to arrange that Bent came to IMPA since there he would have an environment suitable to do research without an overwhelming teaching and administrative load. He arrived in IMPA at the middle of the process of dissolution of the statistical group, but he fought to keep statistics alive there. From 1990 to 1992, Gauss Cordeiro stayed as a visiting professor at IMPA giving support to Bent. Together, they promoted an international event in asymptotic theory that attracted a large number of famous international statisticians.

Bent Jørgensen visited the main statistical groups in Brazil during his stay from 1986 to 1992 in the country. In particular, he worked intensively with the group of statisticians at the ESALQ-USP through Clarice Demétrio, with whom Bent kept a fruitful collaboration until

his last days. This collaboration, including some year latter Célestin C. Kokonendji and John P. Hinde (among others), generated two main research lines: one on models for counts, see Bonat et al. (2018), and one on the relations between the so called Taylor's law (according to which the variances of responses related to some natural phenomena tend to be proportional to a power of the mean) and the Tweedie models, see Jørgensen et al. (2011).

He recruited many students to work with him in IMPA (including Rodrigo Labouriau, José Raúl Martinez, among others). Bent attracted many international researchers in statistics as guests at IMPA including Ole Barndorff–Nielsen, Preben Blæsild, Michael Sørensen, Jesper Møller (from Aarhus University), Ib Skovgaard (University of Copenhagen), Stephen Laurizen, Søren Lundbye-Christensen (Aalborg University) and Gerard Letac (from Paul Sabatier University). The departure of Bent Jørgensen from IMPA to the University of British Columbia (UBC) in 1992 represented a significant loss for the Brazilian Statistics.

A group of statisticians was quickly gathered by Bent Jørgensen at the UBC (*e.g.*, Søren Lundbye-Christensen, as a recurrent visitor from Denmark, and Peter Song, among others). They worked in a new research line in which EDMs were used to represent a latent stochastic process governing the temporal development of a phenomena of interest, see Jørgensen, Labouriau and Lundbye-Christensen (1996a), Jørgensen et al. (1996b, 1996c), Jørgensen and Song (1997) and Section 3.3.1. In this period (around 1995–1997) Bent Jørgensen worked also with Rinaldo Artes (currently at Insper—Instituto de Ensino e Pesquisa, São Paulo) developing part of his thesis entitled "Extensions of generalized estimation equation theory to circular data and dispersion models", which was approved in 1997 at University of São Paulo, see also Artes and Jørgensen (2000).

Bent Jørgensen left UBC and came back to Denmark in a temporary position shared between the Aarhus University (at the Department for Theoretical Statistics) and the Biometry Research Unit at the Foulum Research Centre in 1996. In this period, he worked with Ole Barndorff–Nielsen (on simplex distributions and inferential separation techniques via concepts of sufficiency, ancillarity, and non-formation). The presence of Bent Jørgensen brought much life to the statistical discussion in the Foulum group of statisticians led at that time by Rodrigo Labouriau. He attracted Gordon Smyth who was working with dispersion parameter modeling and composite Poisson distributions, and latter Antonieta Peres (whom met him in IMPA and worked with him on state space models, but she died a few years later), see Botter, Jørgensen and Peres (2002), and Renjun Ma, who was working with models of repeated measures, see Ma and Jørgensen, Ma, Jørgensen and Willms (2007, 2009). The last of Bent Jørgensen's Ph.D. students was Wagner Hugo Bonat, who worked in multivariate extensions of dispersion models (currently at the Universidade Federal do Paraná) and Tweedie models, see Bonat and Jørgensen (2016) and Bonat and Kokonendjib (2017). All in all, Bent Jørgensen worked in many different academic environments, spread over several countries, and always gathered many collaborators around himself. In the following sections, we review and discuss some of his main contributions.

2 Dispersion models

2.1 Introduction

The notion of dispersion models can be seen as a generalization of the normal distribution as we expose below. Consider the density of a univariate normal distribution with expectation $\mu \in \mathbb{R}$ and variance $\tau \in \mathbb{R}_+$,

$$p(y; \mu, \tau) = (2\pi\tau)^{-1/2} \exp\left\{-\frac{1}{2\tau}(y-\mu)^2\right\} \quad \text{for } y \in C = \mathbb{R}.$$

Setting $d(y; \mu) = (y - \mu)^2$ and $a(y; \tau) = (2\pi\tau)^{-1/2}$, we can express the density above in the form

$$p(y; \mu, \tau) = a(y; \tau) \exp\left\{-\frac{1}{2\tau} d(y; \mu)\right\} \quad \text{for } y \in C, \tag{2.1}$$

where $d(y; \mu)$ is the squared Euclidean distance between the observation y and the location parameter μ. The idea of dispersion models (Jørgensen, 1997, p. 4) is to replace the squared distance d in (2.1) by another suitable function, called the unit deviance, measuring how far an observation y is from a central reference point μ of the distribution. This idea turned to be rather fruitful since it generates a rich class of families of distributions, the dispersion models (to be defined precisely below), including many classic continuous, discrete and mixed type distributions. Dispersion models are often used in applications since they are families of distributions of outstanding statistical models as generalized linear models, exponential family nonlinear models (see Section 3.1), generalized additive models, generalized linear mixed models, among others. Moreover, the dispersion models have some common mathematical properties that allow building an elegant and coherent theory of statistical inference and to construct models involving observations of well behaved stochastic processes (*e.g.*, processes with stationary and independent increments, *i.e.*, Lèvy processes, see Section 3.3.1). In this article, we study the notion of univariate dispersion models and refer the interested reader to Jørgensen and Lauritzen (2000) and Jørgensen (2013) for the multivariate case; see also Bonat and Jørgensen (2016).

2.1.1 *Basic definitions.* A dispersion model is a family of probability distributions parameterized by two parameters as follows. The starting point for the construction of the dispersion models is to define the concept of unit deviance. Let $S \subseteq \mathbb{R}$ be the set of the realizable values of the probability distributions contained in the family that we will construct (assumed to be the same for each element of the family). Denote the convex support of S (*i.e.*, the smallest interval containing S) by C and set $\Omega = \text{int}(C)$. Here Ω will be the parameter space of the referential parameter μ. A *unit deviance* is a function $d : C \times \Omega \to \mathbb{R}_+$ such that $d(\mu; \mu) = 0$ for all $\mu \in \Omega$, and $d(y; \mu) > 0$ for all $(y, \mu) \in C \times \Omega$ such that $y \neq \mu$. A unit deviance d is said to be *regular* when d is continuously twice differentiable in $C \times \Omega$ and $\partial^2 d(\mu; \mu)/\partial\mu^2 > 0$ for all μ in Ω.

A *dispersion model* generated by a unit deviance d is a parametric family of real probability measures with support contained in an interval $C \subseteq \mathbb{R}$ with density with respect to a suitable common dominating measure, υ, taking the form (2.1). The dominating measure υ is typically the Lebesgue measure on $\mathbb{R}$ or the counting measure. Here the function $a : C \times \mathbb{R}_+ \to \mathbb{R}_+$ is such that the integral of the density p is 1. The parameters μ and τ are called the *position parameter* and the *dispersion parameter*, respectively.

Classic examples of dispersion models are the normal ($d(y; \mu) = (y - \mu)^2$), gamma ($d(y; \mu) = 2\{y/\mu - \log(y/\mu) - 1\}$, for $y \in \mathbb{R}_+$), von Mises ($d(y; \mu) = 2\{1 - \cos(y - \mu)\}$, for $y \in [0, 2\pi)$), simplex ($d(y; \mu) = (y - \mu)^2/\{y(1 - y)\mu^2(1 - \mu)^2\}$) and Poisson ($d(y; \mu) = 2\{y\log(y/\mu) - y + \mu\}$) distributions, see (Jørgensen 1997, p. 13–23) for further examples.

An immediate consequence of the above definition is that the position parameter μ is the mode of the density (2.1) when the dispersion parameter τ is fixed, since $\log\{p(y; \mu, \tau)\}$ is proportional to $-d(y; \mu)$ and the function $d(y; \cdot)$ has by definition a minimum at $\mu = y$. A unit deviance can also be viewed as a generalization of the Kullback–Leibler information divergence (Jørgensen and Lauritzen, 2000).

We define for each regular unit deviance $d : C \times \Omega \to \mathbb{R}_+$ the associated *unit variance function* $V : \Omega \to \mathbb{R}_+$ given by $V(\mu) = 2/\{\partial^2 d(\mu; \mu)/\partial\mu^2\}$, for each $\mu \in \Omega$. The unit variance function plays an important rule in the theory of dispersion models since it expresses the dependency of the variance on the expectation under dispersion models and it characterizes

uniquely the elements of some important classes of these models. Moreover, the variance function is useful to characterize certain forms of convergence of dispersion models, *c.f.*, Jørgensen (1987a, 1987b, 1997).

Two major classes of dispersion models will be studied in details: the proper dispersion models (PDMs) and the exponential dispersion models (EDMs). A dispersion model with density (2.1) is said to be a *proper dispersion model* generated by a unit deviance $d : C \times \Omega \to \mathbb{R}_+$ when the function $a : C \times \mathbb{R}_+ \to \mathbb{R}_+$ factorizes as follows,

$$a(y; \tau) = a_0(\tau)b(y) \quad \text{for all } y \in C \text{ and } \mu \in \Omega,$$

for a suitable choice of the functions a_0 and b. A PDM is said to be a *regular proper dispersion model* when $C = \Omega$, the unit deviance d is regular and $b(\cdot) = V^{-1/2}(\cdot)$, where V is the unit variance function associated to the regular unit deviance d. There is essentially no loss of generality in assuming that $b(\cdot) = V^{-1/2}(\cdot)$ as we shall argue in Section 2.1.3.

A dispersion model generated by a unit deviance d is said to be a *reproductive exponential dispersion model* when the unit deviance has the form

$$d(y; \mu) = yf(\mu) + g(\mu) + h(y) \quad \text{for all } y \in C \text{ and } \mu \in \Omega, \tag{2.2}$$

for suitable functions f, g and h. The special form of the unit deviance in (2.2) will be absorbed in the density (2.1) to obtain (in Section 2.2, equation (2.6)) an alternative representation of the probability density of the reproductive EDM which is the standard used in most of the literature on the subject.

Examples of reproductive EDMs are the normal, gamma, inverse Gaussian and Poisson distributions. The von Mises, simplex, normal, gamma and inverse Gaussian distributions are classic examples of PDMs. Dispersion models that are neither a regular PDM nor a reproductive EDM "are still not well understood, mainly for lack of examples of this kind" and because methods for generating those models are currently non-existent (Jørgensen 1997, p. 8, last paragraph). However, we give below an example and show a general method for obtaining this type of dispersion models. There are only three PDMs that are also reproductive EDMs: the normal, the gamma and the inverse Gaussian distributions (Jørgensen 1997, Theorem 5.6, p. 188); see also Daniels (1980).

2.1.2 *Construction of dispersion models.*

We turn now to the problem of constructing dispersion models. Given a unit deviance d defined on $C \times \Omega$, for a given real interval C and $\Omega = \text{int}(C)$, a dispersion model can be obtained by finding a function $a : C \times \mathbb{R}_+ \to \mathbb{R}_+$ such that the integral of the right side of (2.1) is 1, that is, the function a is the solution of the following Fredholm integral equation of first kind

$$\int_C a(y; \tau) \exp\left\{-\frac{1}{2\tau} d(y; \mu)\right\} \upsilon(dy) = 1 \quad \text{for all } (\mu, \tau) \in \Omega \times \mathbb{R}_+. \tag{2.3}$$

The generation of a dispersion model involves then the construction of a real function from $C \times \Omega$ that is a unit deviance and finding a solution of the related integral equation (2.3). Not all unit deviances generate a dispersion model since this equation might not have a solution. Typically, solving this equation or even just establishing the existence and unicity for a solution is a hard problem. However, the integral equation (2.3) takes a simpler form for PDMs as we shall see in Section 2.3.2.

Regarding the construction of a unit deviance, note that any distance on $C \times \Omega$ or any increasing function of a distance on $C \times \Omega$ is a unit deviance. The following construction based on basic properties of characteristic functions yields unit deviances with a tractable related integral equation. Let P be a probability measure on $\mathbb{R}$ with characteristic function φ. If P is symmetric around zero, then $\varphi(t) \in \mathbb{R}$, for all $t \in \mathbb{R}$. Moreover, $\varphi(0) = 1$ and $|\varphi(t)| \leq 1$

for all $t \in \mathbb{R}$ (Lucaks 1970, p. 15). Assuming further that P is not a lattice distribution (*i.e.*, P is not concentrated on a set of the form $\{a + nh, n = 0, \pm 1, \pm 2, \dots \}$ for some $a, h \in \mathbb{R}$ and $h > 0$), then $|\varphi(t)| < 1$, for every $t \neq 0$ (Ushakov 1999, Theorem 1.1.3, p. 2). Therefore, defining $d : C \times \Omega \to \mathbb{R}_+$ by $d(y; \mu) = 1 - \varphi(y - \mu) = 1 - \varphi(\mu - y)$, for each $(y, \mu) \in C \times \Omega$ (the last equality follows from the symmetry of any characteristic function taking real values), yields a unit deviance. If we further require that the first two moments of P are finite, then d is twice differentiable and $\partial^2/\partial \mu^2 \, d(\mu; \mu) = 1 + m_2 > 0$ (where m_2 is the second central moment of P), so d is a regular unit deviance. For instance, the unit deviance given by $d(y; \mu) = 1 - \exp(-|y - \mu|)$, corresponding to the unit deviance constructed with the characteristic function of the Cauchy distribution, is not a regular unit deviance.

The integral equations (2.1) related to unit deviances constructed with characteristic functions as above, when $C = \mathbb{R}$, for a fixed $\tau = \tau_0$, becomes

$$1 = \int_{\mathbb{R}} a(y; \tau_0) \exp\left\{-\frac{1 - \varphi(\mu - y)}{2\tau_0}\right\} \upsilon(dy) = [a_{\tau_0} * K_{\tau_0}](\mu), \qquad (2.4)$$

for all $\mu \in \mathbb{R}$. Here the convolution operator, "$*$", refers to the convolution between functions. We want to solve the equation for $a_{\tau_0}(\cdot) = a(\cdot; \tau_0)$ where $K_{\tau_0}(\cdot) = \exp\{-1/(2\tau_0)[1 - \varphi(\cdot)]\}$ is the kernel of a convolution equation. It is remarkable that the kernel K_{τ_0} is a characteristic function of a probability measure as proved in the Corollary 1.3.4 in Ushakov (1999), page 8, and therefore it has a well defined Fourier transform. The solution of the convolution equation (2.4) does not depend on μ, which is a consequence of the Lemma 5.2 in Jørgensen (1997). A calculation involving the formalism of tempered distributions, see Rudin (1973), Chapters 7 and 9, and the delta Dirac distribution (in the sense of generalized functions), yields the general (formal weak) solution $a(y; \tau_0) = \mathcal{F}^{-1}[\delta(\cdot)/\mathcal{F}\{K_{\tau_0}\}(\cdot)] = 1/\{2\pi \int_{\mathbb{R}} K_{\tau_0}(z) dz\} \stackrel{\text{def}}{=} a_0(\tau_0)$. Here $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the Fourier transform and the inverse of the Fourier transform, respectively. Note that the function $a(\cdot; \tau) = a_0(\tau)$ can be factorized as a product of a function of the observations y (function taking constant values) and a function of the parameter τ; therefore, the model generated is a PDM. A calculation involving projections on the orthogonal complement of the $L^2(\mathbb{R})$ closure of the span$\{K(\cdot - \mu) : \mu \in \mathbb{R}\}$ allows to construct at least one DM that is neither a PDM nor an EDM, for each PDM generated as above. Let $\mathcal{P}_0$ be the class of Borel probability measures in $\mathbb{R}$ that are symmetric around zero and are not a lattice distribution. The discussion above allows us to claim that the cardinality of the regular dispersion models that are not a PDM and not a reproductive EDM is at least the same as the cardinality of $\mathcal{P}_0$. Moreover, the cardinality of those models that are regular is at least the cardinality of the set of elements of $\mathcal{P}_0$ that have the first two central moments finite. To our best knowledge, this result has never been exposed in the literature before. Further details and developments are exposed in Labouriau (2020).

2.1.3 *Some general properties of dispersion models.* Even though PDMs and reproductive EDMs have somewhat different distributional features, these two classes share some fundamental statistical properties, which are common to all regular dispersion models. We summarize these common properties below. First, we note that for any regular unit deviance d it holds that

$$\frac{\partial^2 d(\mu; \mu)}{\partial y^2} = \frac{\partial^2 d(\mu; \mu)}{\partial \mu^2} = -\frac{\partial^2 d(\mu; \mu)}{\partial \mu \, \partial y} \quad \text{for all } \mu \in \Omega, \qquad (2.5)$$

Jørgensen (1997), Lemma 1.1, p. 24. This general result has two immediate consequences: it gives alternative ways to calculate the unit variance function and it implies that the unit deviance behaves similarly to the unit deviance of the normal family near its minimum, μ_0, since it follows from (2.5) that

$$d(\mu_0 + x\delta; \mu_0 + m\delta) = \frac{\delta^2}{V(\mu_0)} (x - m)^2 + o(\delta^2).$$

This approximation sends us back to the initial idea of viewing dispersion models as a form of generalization of the normal distribution.

A useful characteristic of dispersion models is a duality property for (well behaved) transformations, as we explain below. Given a unit deviance d on $C \times \Omega$ and a one-to-one function $f : C \to C_f$ ($C_f \subseteq \mathbb{R}$), the function $d_f : C_f \times \text{int}(C_f) \to \mathbb{R}_+$ given by $d_f = d(f^{-1}(z); f^{-1}(\xi))$ for each $(z, \xi) \in C_f \times \text{int}(C_f)$ is also a unit deviance. Moreover, if a random variable Y follows a dispersion model with unit dispersion d and the function f is monotone and differentiable, then the distribution of a transformed random variable $Z = f(Y)$ is in a dispersion model with density (in the continuous case)

$$p_Z(z; \xi, \tau) = \frac{a\{f^{-1}(z), \tau\}}{|f'\{f^{-1}(z)\}|} \exp\left\{-\frac{1}{\tau} d_f(z; \xi)\right\} \quad \text{for all } z \in C_f.$$

In the discrete case, we eliminate the Jacobian $1/|f'\{f^{-1}(z)\}|$ of the above expression. The new dispersion model generated as above is called in the literature a *re-parametrization by a transformation*, in an abuse of nomenclature since the new dispersion model is actually not a re-parametrization of the original dispersion model. If the unit deviance d is regular and the transformation f is twice continuously differentiable with $|f'(y)| > 0$ for all $y \in \Omega$, then the unit deviance d_f is also regular and has the associated unit variance $V_f(\xi) = V\{f^{-1}(\xi)\}[f'\{f^{-1}(\xi)\}]^2$, where V is the unit variance associated to the unit deviance d. The transformation $f(y) = \int_{y_*}^{y} V^{-1/2}(v)\, dv$ (for a fixed y_*) has a constant variance function $V_*(\xi) = 1$ for all $\xi \in \text{int}(C_f)$. This transformation, called the *variance stabilizing transformation*, plays an important rule in the asymptotic theory of dispersion models.

The probability density of a regular dispersion model, p given in (2.1), can be well approximated by

$$q(y; \mu, \tau) = \{2\pi\tau V(y)\}^{-1/2} \exp\left\{-\frac{1}{2\tau} d(y; \mu)\right\},$$

in the sense that $p(y; \mu, \tau)/q(y; \mu, \tau) \to 1$ when $\tau \to 0$ for each $y \in C$ and $\mu \in \Omega$. This approximation is called the *saddlepoint approximation*. Clearly, this convergence is equivalent to $a(y, \tau)/\{2\pi\tau V(y)\}^{-1/2} \to 1$ when $\tau \to 0$. The saddlepoint approximation is often very accurate and it is useful because the function $a(\cdot, \cdot)$ in (2.1) is difficult to be calculated or numerically evaluated. Note that the integral of the functions $q(\cdot; \mu, \tau)$ are not necessarily 1, therefore we define the *renormalized saddlepoint approximation* by $q_0(\cdot; \mu, \tau) = q(\cdot; \mu, \tau) a_0(\mu, \tau)$, where $a_0(\mu, \tau) = 1/\int_C q(y; \mu, \tau) \nu(dy)$. The approximation $p(y; \mu, \tau) \sim q_0(y; \mu, \tau)$ (as $\tau \to 0$) is often more accurate than the original saddlepoint approximation.

2.2 Exponential dispersion models

An EDM is defined as a class of real distributions having a density with respect to a suitable dominating measure taking the form

$$p(y; \theta, \tau) = \exp\left\{\frac{1}{\tau}[y\theta - b(\theta)] + c(y; \tau)\right\}. \tag{2.6}$$

Here $b : \Theta \to \mathbb{R}$, and $c : C \times \mathbb{R}_+ \to \mathbb{R}$ are known appropriate functions, Θ is an open set in $\mathbb{R}$, $\theta \in \Theta$ and $\tau > 0$ are called the *canonical parameter* and the *dispersion parameter*, respectively. Typically, the dominating measure is the Lebesgue measure in $\mathbb{R}$, yielding continuous distributions, or the counting measure, generating discrete distributions. A *natural exponential family* is obtained when the dispersion parameter τ is kept fixed. The notion of EDMs was pioneered by Tweedie (1947), who studied several special cases and pointed out important structural properties. The theory of EDMs was systematically exposed in Jørgensen

(1987a, 1997) where several mathematical properties of EDMs were presented for the first time. The terminology "exponential dispersion model" reflects the exponential form of the density of those distributions and the important role played by the dispersion parameter τ.

The cumulant generating function (cgf) of a distribution with density given by (2.6) is

$$K(t; \theta, \tau) = \frac{1}{\tau}[b(\theta + \tau t) - b(\theta)], \tag{2.7}$$

which depends only on the function b, termed the *cumulant generator*. Setting $b(\theta) = \theta^2/2$, $b(\theta) = -\log(-\theta)$ and $b(\theta) = -(-2\theta)^{1/2}$ yield the normal, gamma and inverse Gaussian distributions, respectively. The discussion above implies that we can construct one EDM for any specified non-degenerate function b since this function uniquely determines the class of cgf of an EDM via (2.7) and the cgf uniquely determines the distribution (an instance of the Fourier inversion theorem for the characteristic functions). Therefore, there exist many EDMs.

Differentiating (2.7) yields the expectation of the distribution with density (2.6) given by $\mu \stackrel{\text{def}}{=} M(\theta) \stackrel{\text{def}}{=} \partial b(\theta)/\partial\theta$. The function M, associating each value of θ to the expectation of the corresponding distribution, is called the *mean value mapping*. By inverting the mean value mapping (we show below that M is indeed invertible) we obtain that $\theta = M^{-1}(\mu) = b'^{-1}(\mu) \stackrel{\text{def}}{=} q(\mu)$. The variance of the distributions given by (2.6) are then $\tau V(\mu)$, where the *variance function* $V(\cdot)$ has the following alternative forms, for each value of μ,

$$V(\mu) = \frac{\partial^2 b(\theta)}{\partial\theta^2} = \frac{\partial b'\{q(\mu)\}}{\partial\theta} = \frac{\partial M\{q(\mu)\}}{\partial\theta} = m'\{q(\mu)\} > 0.$$

Consequently, the mean value mapping μ is a strictly increasing function and the parameter $\theta = q(\mu) = \int V(\mu)^{-1} d\mu$ is a known one-to-one function of μ. For a given variance function $V(\cdot)$, we can easily obtain the inverse of the mean value mapping $q(\cdot)$ and then calculate $b(\theta) = \int q^{-1}(\theta) d\theta$ (for each θ) and the cgf given by (2.7). The elements of the EDMs are then uniquely determined by their variance functions $V(\cdot)$. Moreover, the variance functions play a key role for studying several of the structural properties of EDMs.

The EDMs with quadratic variance function given by $V(\mu) = a\mu^2 + b\mu + c$ form an important and well studied class of EDMs. Morris (1982) studied those EDMs in details and proved that this subclass is composed of only six distributions: the normal ($a = b = 0, c = 1$), the gamma ($a = 1, b = c = 0$), the Poisson ($a = c = 0, b = 1$), the binomial ($a = -1, b = 1, c = 0$), the negative binomial ($a = b = 1, c = 0$) and the generalized secant hyperbolic ($a = c = 1, b = 0$) distributions. Jørgensen (1997) discussed further the generalized secant hyperbolic by taking $b(\theta) = -\log\{\cos(\theta)\}$ and

$$c(y; \tau) = \log\left[\frac{2^{(1-2\tau)/\tau}}{\tau\Gamma(\tau^{-1})}\right] - \sum_{j=1}^{\infty} \log\left[1 + \frac{y^2}{(1 + 2j\tau)^2}\right],$$

where $\Gamma(\cdot)$ is the gamma function.

The rth cumulant of a distribution contained in an EDM with density given by (2.6) is $\kappa_r = \kappa_r(\theta, \tau) = \partial^r K(t; \theta, \tau)/\partial^r t|_{t=0}$. Therefore, the function $b(\theta)$ generates all cumulants of the distribution $\kappa_r = \tau^{r-1}\partial^r b(\theta)/\partial\theta^r$ for $r \geq 1$. A direct implication of that is that any distribution contained in an EDM has finite cumulants of all orders, which rules out the use of EDM for modeling situations where it is required that the distribution has very heavy tails.

Tweedie (1947) proved the normal convergence of the random variable $Z = (Y - \mu)/\sqrt{\tau}$, where the distribution of the random variable Y is contained in an EDM, using an expansion for its cgf. The cgf of Z follows from (2.7) as

$$K_Z(t; \theta, \tau) = -\mu\frac{t}{\sqrt{\tau}} + \frac{1}{\tau}[b(\theta + \sqrt{\tau}t) - b(\theta)].$$

By expanding $b(\theta + \sqrt{\tau}t)$ in a Taylor series around zero with the cumulants of Y as coefficients and collecting like terms, we obtain

$$K_Z(t; \theta, \tau) = V(\mu)t^2/2 + \sum_{k=3}^{\infty} \frac{\partial^k b(\theta)}{\partial^k \theta} \frac{\tau^{k/2-1} t^k}{k!}.$$

Based on this expansion, we conclude that

$$Z = (Y - \mu)/\sqrt{\tau} \xrightarrow{\text{D}} \text{N}(0, V(\mu)) \quad \text{when } \tau \to 0, \tag{2.8}$$

where $\xrightarrow{\text{D}}$ denotes convergence in distribution. Clearly, equation (2.8) is exact for the normal distribution since the derivatives of $b(\theta)$ of order greater than two vanish.

Equation (2.8) generalizes a number of known results on convergence to normality such as those for the gamma and inverse Gaussian distributions. In fact, under some regularity conditions, distributions contained in EDMs are approximate normally distributed for small values of τ. In this way, the standard asymptotic theory applies for small values of the dispersion parameter as well as for large sample sizes. The mathematical conditions for this result called *small dispersion asymptotics* were fully discussed in Jørgensen (1987b).

Let $Y_1, \ldots, Y_n$ be independent and identically distributed (iid) random variables with distribution contained in an EDM with mean μ and dispersion parameter τ. The form of the cgf in (2.7) implies that the distribution of the sample mean $\bar{Y} = \sum_{i=1}^{n} Y_i/n$ belongs to the same EDM as the distribution of the elements of the sample with mean μ and dispersion parameter τ/n. This result includes well-known convolution properties of the normal, gamma and inverse Gaussian distributions and implies that the distributions contained in EDMs are infinite divisible provided the dispersion parameter is allowed to take values arbitrarily close to zero.

2.2.1 *Tweedie models.* A important sub-class of the models defined by (2.6), called *Tweedie models* is obtained when $V(\mu) = \mu^p$, for $p \in (-\infty, 0] \cup [1, +\infty)$. A full discussion of these distributions with varying p was first addressesed by Tweedie (1947, 1984) and further discussed by Jørgensen (1987a). The support of the Tweedie models depends on the value of p. The normal, Poisson, gamma and inverse Gaussian distributions can be obtained from (2.6) for $V(\mu) = \mu^p$ and $p = 0, 1, 2$ and 3, respectively. Distributions generated by extreme stable distributions are determined when $p < 0$ with support on $\mathbb{R}$. Jørgensen (1987a) demonstrated that there exist no EDM with power variance functions for $0 < p < 1$. When $1 < p < 2$, we obtain the Gamma compound Poisson distributions which are interesting because they are continuous for $y > 0$ but have positive probability mass at zero. For $p > 2$ ($p \neq 3$), we obtain continuous distributions generated by positive stable distributions. When p increases to $+\infty$ the Tweedie models converge to extreme stable distributions.

We define the cumulant generator $b_p(\theta)$ for Tweedie models ($p \neq 1, 2$) by

$$b_p(\theta) = (2 - p)^{-1} [(1 - p)\theta]^{\frac{p-2}{p-1}}.$$

Further, $b_1(\theta) = \exp(\theta)$ and $b_2(\theta) = -\log(-\theta)$.

The unit deviance $d_p(\cdot; \cdot)$ of a Tweedie model follows by the following straightfoward calculation

$$d_p(y; \mu) = 2 \int_{\mu}^{y} \frac{(y - t)}{V(t)} \, dt = 2 \left\{ \frac{[\max(y, 0)]^{2-p}}{(1 - p)(2 - p)} - \frac{y\mu^{1-p}}{1 - p} + \frac{\mu^{2-p}}{2 - p} \right\}.$$

The remaining quantities in equations (2.6) can be evaluated numerically for the cases $1 < p < 2$ and $p > 2$ in power series following Jørgensen (1987a). See, also Bonat and Kokonendjib (2017).

2.2.2 *Saddlepoint approximations.* We adopt the notation $K_Y^{(j)}(t; \theta, \tau) = \partial^j K_Y(t; \theta, \tau) / \partial t^j$ for $j \geq 1$. The saddlepoint approximation for the density of Y takes the form

$$\pi_Y(y; \theta, \tau) \simeq \left[\frac{1}{2\pi K_Y^{(2)}(\hat{\lambda}; \theta, \tau)} \right]^{1/2} \exp[K_Y(\hat{\lambda}; \theta, \tau) - \hat{\lambda} y], \qquad (2.9)$$

where the saddlepoint $K_Y(\lambda; \theta, \tau) - y$ is found by solving the (usual nonlinear) equation $K_Y'(\hat{\lambda}; \theta, \tau) = y$.

This equation was derived by Daniels (1954) as an approximation for any density given its cgf, although it can be applied here to approximate the density of an EDM when τ is small. In fact, the rth cumulant of Y is of order $O(\tau^{r-1})$ (for $r \geq 1$) and then the density of Y with large precision parameter $1/\tau$ can be considered as the density of the sample average in large samples. By differentiating (2.7), we find $b'(\theta + \tau \hat{\lambda}) = y$ and then $\hat{\lambda} = [q(y) - \theta]/\tau$. By definition of the variance function, we have $K_Y^{(2)}(\hat{\lambda}; \theta, \tau) = \tau V(y)$. Inserting these quantities in the last density approximation, the saddlepoint approximation for the ED density can be expressed in a simple form

$$\pi_Y(y; \theta, \tau) \simeq \left[\frac{1}{2\pi \tau V(y)} \right]^{1/2} \exp\left[-\frac{1}{2\tau} d(y; \mu) \right], \qquad (2.10)$$

which holds when $\tau \to 0$. This approximation is exact independent of τ for the normal distribution since the third and fourth cumulants of Y vanish. Equation (2.10) is equivalent to the asymptotic result from (2.1): $\sqrt{\tau} a(y; \tau) \to [2\pi V(y)]^{1/2}$ when $\tau \to 0$.

The cumulative distribution function of Y follows approximately from (2.10) as

$$\Pi_Y(y; \theta, \tau) = P(Y \leq y) \simeq \int_{-\infty}^{y} \left[\frac{1}{2\pi \tau V(x)} \right]^{1/2} \exp\left[-\frac{1}{2\tau} d(x; \mu) \right] dx.$$

By integrating the last equation, Lugannani and Rice (1980) proved that

$$\Pi_Y(y; \theta, \tau) \simeq \left[\Phi\left(\frac{r}{\sqrt{\tau}} \right) + \sqrt{\tau} \phi\left(\frac{r}{\sqrt{\tau}} \right) \left(\frac{1}{r} - \frac{1}{u} \right) \right],$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution and the density functions of a standard normal distribution, respectively. The quantity $r = \text{sgn}(y - \mu)\sqrt{d(y; \mu)}$ is called the *deviance residual* and $u = \frac{V(y)^{1/2}}{2} \frac{\partial d(y; \mu)}{\partial y}$ is the *dual score residual*. This equation is very easy to be applied to compute probabilities for any EDM since it involves only the variance and the deviance functions.

We now move to the density of the sample average $\overline{Y} = \sum_{i=1}^{n} Y_i / n$ of iid random variables $Y_1, \ldots, Y_n$ having density (2.6) and cgf (2.7). The density function of $\overline{Y}$ follows from the Fourier inversion integral as

$$\pi_{\overline{Y}}(y; \theta, \tau) = \frac{1}{2\pi} \exp[-\mathrm{i} t y + n K_Y(\mathrm{i} t/n; \theta, \tau)],$$

where $\mathrm{i} = \sqrt{-1}$. This equation is suitable for Daniels' saddlepoint approximation. Setting $z = \mathrm{i} t/n$, the saddlepoint of $K_Y(z; \theta, \tau) - zy$ is $K_Y'(\hat{z}; \theta, \tau) = y$. Then, the density approximation of $\overline{Y}$ can be expressed as

$$\pi_{\overline{Y}}(y; \theta, \tau) = \left[\frac{n}{2\pi K_Y^{(2)}(\hat{z}; \theta, \tau)} \right]^{1/2} \exp\{n[K_Y(\hat{z}; \theta, \tau) - \hat{z} y]\},$$

which can provide good results in practice.

It is much more frequent in statistical applications to compute distribution functions than density functions. By integrating the last equation, the cumulative distribution function (cdf) of $\overline{Y}$ has the form

$$\Pi_{\overline{Y}}(y; \theta, \tau) \simeq \int_{-\infty}^{y} \left[\frac{n}{2\pi K_Y^{(2)}(t; \theta, \tau)} \right]^{1/2} \exp\{n[K_Y(t; \theta, \tau) - tx]\} \, dx,$$

where $t = t(x)$ is determined by $K_Y'(t; \theta, \tau) = x$. By transformation of variables and integration with respect to the saddlepoint variable t instead of x, we obtain $K_Y^{(2)}(t; \theta, \tau) \, dt = dx$ and then

$$\Pi_{\overline{Y}}(y; \theta, \tau) \simeq \int_{-\infty}^{t(y)} \left[\frac{n K_Y^{(2)}(t; \theta, \tau)}{2\pi} \right]^{1/2} \exp\{n[K_Y(t; \theta, \tau) - t K_{Y'}(t; \theta, \tau)]\} \, dt,$$

where $t(y)$ is found by solving $K_Y'(t(y); \theta, \tau) = y$. This integral for $\Pi_{\overline{Y}}(y; \theta, \tau)$ is much easier to compute than the previous one because it includes explicitly the saddlepoint function in the integrand. The saddlepoint approximation for the cdf of $\overline{Y}$ follows from Lugannani and Rice (1980) as

$$\Pi_{\overline{Y}}(y; \theta, \tau) \simeq \Phi[r(y)] + \phi[r(y)] \left[\frac{1}{r(y)} - \frac{1}{u(y)} \right], \qquad (2.11)$$

where

$$r(y) = \text{sgn}[t(y)]\{2n[yt(y) - K_Y(t(y); \theta, \tau)]\}^{1/2}$$

and

$$u(y) = t(y)[n K_Y^{(2)}(t(y); \theta, \tau)]^{1/2}.$$

Equation (2.11) provides highly accurate results for the probabilities associated with $\overline{Y}$.

2.3 Proper dispersion models

We consider now the notions of general dispersion models and general PDMs that will allow us to understand some statistical inferential aspects of the notion of PDMs introduced in Section 2.1. A *general dispersion model* is a family of real distributions parameterized by two parameters, $\theta \in \Theta \subseteq \mathbb{R}$ and $\lambda \in \Lambda \subseteq \mathbb{R}_+$, where Θ, Λ are intervals, Λ is unbounded to the right and the density (or Radon–Nikodym derivative) with respect to a common dominating real measure υ is of the form

$$p(y; \theta, \lambda) = a(y; \lambda) \exp\{\lambda t(y; \theta)\}, \quad \text{for all } y \in S, \theta \in \Theta \text{ and } \lambda \in \Lambda. \qquad (2.12)$$

Here S is the support of the dominating measure υ, and $a : S \times \Lambda \to \mathbb{R}_+$, and $t : S \times \Theta \to \mathbb{R}_+$ are suitable functions. When the function a factorizes as $a(y; \lambda) = a_0(\lambda)b(y)$, for all $y \in S$ and $\lambda \in \Lambda$, the family of distributions above is said to be a *general proper dispersion model* (general PDM). Note that when the function $-d$, used in (2.1), is a unit dispersion, $\lambda = 1/\tau$ and $\mu = \theta$, then the general dispersion model and the general PDM coincide with the dispersion models and the PDMs defined in Section 2.1. Moreover, (2.12) coincides with (2.1) defined with the unit deviance t instead of d. Apparently, the setup above is much more general than the situation considered in Section 2.1 since here we are not requiring the function t to be a unit deviance. However, we will argue that in order to obtain families of distributions with some desirable statistical inferential properties we will need to introduce some restrictions in the general definition given above that will render the present definition essentially equivalent to the setup discussed in Section 2.1. In this way, we are using an embedding of classes of families of distributions to obtain a better understanding of inferential properties of the statistical models based on PDMs.

2.3.1 *Some key properties.* The notion of yoke and yokable function defined below will allow us to connect the notions of PDMs (as defined in Section 2.1) and general dispersion models (defined above). Moreover, these notions will allow us to characterize the existence of maximum likelihood estimates (MLEs) for PDMs. Given an interval $\Omega \subseteq \mathbb{R}$, a function $t : \Omega \times \Omega \to \mathbb{R}$ is said to be a *yoke* if $\sup_{\theta \in \Omega} t(y; \theta) = t(y; y)$ for all $y \in \Omega$. When additionally, $t(y; y) = 0$, for all $y \in \Omega$, then the function t is called a *normed yoke*. If $d : \Omega \times \Omega \to \mathbb{R}_+$ is a unit deviance, then $-d$ is a normed yoke. On the other hand, if the function $t : \Omega \times \Omega \to \mathbb{R}$ is a yoke and we denote for each $y \in \Omega$ the supremum $\sup_{\theta \in \Omega} t(y; \theta)$ by $\hat{t}(y)$, then the function given by $\tilde{t}(y; \theta) = \hat{t}(y) - t(y; \theta)$ is a normed yoke.

A function $t : C \times C \to \mathbb{R}$, where C is a real interval, is said to be *yokable* when the following three conditions are satisfied: (i) $\sup_{\theta \in \Omega} t(y; \theta) < \infty$ for all $y \in C$; (ii) There exists an open interval $\Omega \subseteq C$ such that, for each $y \in \Omega$ the supremum $\hat{\theta}_y = \sup_{\theta \in \Omega} t(y; \theta)$ is unique; and (iii) The function $\hat{\theta} : \Omega \to \text{int}(\Theta)$ given by $\hat{\theta}(y) = \hat{\theta}_y$ (for each $y \in \Omega$) is a bijection.

If the function t used in (2.12) to define general PDMs is yokable, then we might define a PDM using the unit deviance $d : C \times \Omega \to \mathbb{R}$ given by $d(y; \mu) = 2[\hat{t}(y) - t\{y; \hat{\theta}(\mu)\}]$, for all $y \in C$ and $\mu \in \Omega$. In this case, the density of a general PDM takes the form

$$p(y; \theta, \lambda) = a_0(\lambda)b(y) \exp\left\{\lambda \hat{t}(y) - \frac{\lambda}{2} d(y; \mu)\right\}, \tag{2.13}$$

for all $y \in S$, $\mu \in \Theta$ and $\lambda \in \Lambda$. Here $a(y; \lambda) = a_0(\lambda) \exp\{\lambda \hat{t}(y)\}$. If we further assume that $\Omega = S = C$, that the function d is regular unit deviance, and that for all $y \in \Omega$, $b(y) = V^{-1/2}(y)$, then the family of distributions defined by (2.13) is a regular PDM as defined in Section 2.1.1. We argue below that it is natural from the statistical point of view to assume the function t in (2.12) to be yokable. First, the conditions (i) and (ii) of the definition of yokable function ensure the existence of at least a local maximum of the likelihood function for μ, obtained when fixing the index parameter λ. Note that this is a minimal necessary requirement for the existence and uniqueness of the maximum likelihood estimates. The condition (iii) implies that θ is a re-parametrization of the position parameter of the related proper dispersion model.

In order to elucidate some basic statistical properties of general dispersion models, let us consider the so called Barndorff–Nielsen's p^* formula (Barndorff-Nielsen 1983, 1988) for approximating the conditional distribution of the MLE (for a given statistical model) given an ancillary statistic (*i.e.*, a statistic carrying no information on the parameter of interest, see Jørgensen and Labouriau (2012), Chapter 2 for details). Here we consider the maximum likelihood estimation of the parameter μ in a statistical model defined by (2.13) when the parameter λ is kept fixed and the estimation is based on a single observation, $y \in S$. Taking a degenerate ancillary statistics (*e.g.*, a constant statistic) the p^* formula yields an approximation to the marginal distribution of the MLE. Since in these circumstances the MLE of μ is $\hat{\mu}(y) = y$ and the p^* approximation is defined by

$$p\big(y; \theta(\mu), \lambda\big) \sim p_0(y; \mu, \lambda), \tag{2.14}$$

where p_0 is the renormalized saddlepoint approximation corresponding to the unit deviance d, with $p_0(y; \mu, \lambda) = a_0(\mu, \lambda) V^{-1/2}(y) \exp\{-\frac{\lambda}{2} d(y; \mu)\}$, for all $y \in S$, $\mu \in \Theta$ and $\lambda \in \Lambda$. The Barndorff–Nielsen's formula is said to be *exact* if the two sides of (2.14) coincide for all y in S and all (μ, λ) in $\Omega \times \Lambda$.

A consequence of the saddlepoint approximation above is that standard dispersion models are asymptotically normal distributed for λ large. In this way, the Barndorff–Nielsen's formula may be viewed as a refinement of the normal approximation to the distribution of the

MLE. Furthermore, there is a strong result for regular PDMs stating that (assuming b continuous at $y = \mu$) the following three statements are equivalent: i) The Barndorff–Nielsen's formula is exact for all $\lambda \in \Lambda$; ii) the Barndorff–Nielsen's formula is asymptotically exact in the sense that the ratio p_0/p tends to 1 as $\lambda \to \infty$ for all y and μ in Ω; iii) The function $\hat{t}(y)$ is constant on Ω, and $b(y) \propto V^{-1/2}(y)$ (see Jørgensen (1997), Theorem 5.4 and Corollary 5.5). When these statements hold, the normalizing constant $a_0(\mu, \lambda)$ does not depend on μ, and satisfies $a_0(\mu, \lambda) \propto a(\lambda) \exp\{\lambda \hat{t}(\mu)\}$, and $a(\lambda) \sim \sqrt{\frac{\lambda}{2\pi}} \exp\{-\lambda \hat{t}(\mu)\}$ as $\lambda \to \infty$. In conclusion, the density in (2.13) defines a regular PDM when the Barndorff–Nielsen's formula is exact.

We discuss below some other properties of PDMs which will show some of the peculiarities of those families of distributions. For any fixed value of the position parameter μ, say $\mu = \mu_0$, the family given by (2.1) is an exponential family with canonical statistic $d(\cdot; \mu_0)$ and canonical parameter τ. This is enough to recognize that the general form of the density of a distribution belonging to the exponential family is obtained when setting the position parameter μ fixed.

Another property of PDMs is that when the dispersion parameter is fixed, say $\tau = \tau_0$, the unit deviance is a pivotal statistic for μ. That is, if Y is a random variable having density (2.13) with respect to υ, then the distribution of the random variable $d(Y, \mu)$ does not depend on μ. This property follows by observing that the integral $\int_C b(y) \exp\{-d(y, \mu)/(2\tau_0)\}\upsilon(dy) = 1/a_0(\tau_0)$ does not depend on the value of μ; it is easy then to prove that the moment generating functions of the random variables $T = T_\mu = d(Y, \mu)$, for $\mu \in \Omega$, are all equal and depend only on τ_0. This key result is the Lemma 5.2 in Jørgensen (1997).

2.3.2 *Construction of proper dispersion models.*

As discussed in Section 2.1.2, given a unit deviance an associated dispersion model can in principle be constructed by solving the integral equation (2.3), but this is in general a hard problem. However, that integral equation takes the following simpler form for PDMs

$$a_0(\tau) \int_C b(y) \exp\left\{-\frac{1}{2\tau} d(y; \mu)\right\} \upsilon(dy) = 1 \quad \text{for all } (\mu, \tau) \in \Omega \times \mathbb{R}_+, \tag{2.15}$$

which has a solution $a_0(\tau) = 1/\int_C b(y) \exp\{-d(y; \mu)/2\tau\}\upsilon(dy)$, provided that the integral involved exists. The integral $\int_C b(y) \exp\{-d(y; \mu)/2\tau\}\upsilon(dy)$ does not vanish (since the integrand is positive apart from a υ-null set) and does not depend on the parameter μ because of the second key property of PDMs discussed before.

Another useful technique for constructing PDMs involves the use of a transformation group, say G, acting freely and transitively on $\Omega = C = \mathbb{R}$; see Section 3.3 of Jørgensen and Labouriau (2012) for the basic definitions and a short account on transformation groups in statistical inference. Here we denote the action of G by $(g, y) \mapsto gy$ for $(g, y) \in G \times \Omega$. Let $t, b : \Omega \to \mathbb{R}_+$ where the function b is invariant by the action of G (i.e., $b(gy) = b(y)$ for all $g \in G$ and $y \in \Omega$) and t is an arbitrary function. Assume, moreover, that $\int_\Omega b(y) \exp\{\lambda t(g^{-1}y)\} dy < \infty$, for λ in an interval $\Lambda \subseteq \mathbb{R}_+$ that is unbounded from the right. Since G acts freely and transitively on Ω, there is a one-to-one correspondence between Ω and G. Assume further that the supremum $\hat{t} = \sup_{g \in G} t(g^{-1}y)$ is finite. Then, it is easy to see that $t(g^{-1}y)$ is yokable and we might define the unit deviance $d(y; \mu) = 2\{\hat{t} - t(\hat{g}_\mu^{-1}y)\}$, where g_μ^{-1} is the MLE of g when the observation is μ. The type of dispersion model constructed with this kind of unit deviance is called a *transformation dispersion model*.

Examples of the construction techniques based on transformation groups described above are the special PDMs called *location-dispersion model* for which the density function with respect to the dominating measure υ has the form $p(y; \mu, \tau) = c(\tau) \exp\{-d(y - \mu)/(2\tau)\}$,

where $C = \Omega = \mathbb{R}$ and d is a unit deviance; here the transformation group is the group of translations with action $y \mapsto (g + y)$. Another example is the von Mises distribution defined using the transformation group of rotations of the unit circle with action $y \mapsto$ module$\{2\pi; (g + y)\}$.

3 Applying dispersion and exponential dispersion regression

We review the Exponential Dispersion (ED) regression to extend the well-known GLMs, discuss some improved hypotheses tests and some models for clustered and dependent data based on latent Lévy processes.

3.1 Exponential family non-linear models

We consider models where the random variables $Y_1, \ldots, Y_n$ are assumed independent and each Y_i has a density or probability function of the form (2.1) with mean $\mu_i = E(Y_i)$ on a convenient support.

We define the ED regression by the random component (2.1) and the systematic component

$$g(\mu_i) = \eta_i = f(\boldsymbol{x}_i; \boldsymbol{\beta}), \tag{3.1}$$

where $g(\cdot)$ is a known one-to-one twice continuously differentiable link function, η_i and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^T$ denote the linear predictor and the $p \times 1$ vector of non-stochastic independent variables associated with the ith observation, respectively, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is a p-vector of unknown parameters, and $f(\cdot; \cdot)$ is a (possibly nonlinear) twice continuously differentiable function with respect to $\boldsymbol{\beta}$. Here a standard GLM is obtained when the function $f(\cdot; \cdot)$ is bi-linear, that is, when $f(\boldsymbol{x}_i; \boldsymbol{\beta}) = \boldsymbol{x}_i^T \boldsymbol{\beta}$.

The systematic component relates the explanatory variables $\boldsymbol{x}_i$ to the mean parameter μ_i of interest. The $n \times p$ matrix of derivatives of $\boldsymbol{\eta}$ with respect to $\boldsymbol{\beta}$, specified by $\widetilde{\mathbf{X}} = \widetilde{\mathbf{X}}(\boldsymbol{\beta}) = \partial \boldsymbol{\eta}/\partial \boldsymbol{\beta}$, is assumed to have rank p for all $\boldsymbol{\beta}$. We have $p + 1$ parameters to be estimated: the vector $\boldsymbol{\beta}$ and τ. The ED regression model has two important components: the ED class for the response variable and a possible nonlinear regression on a vector $\boldsymbol{\beta}$ by means of the link function. We assume that the standard regularity conditions for likelihood theory hold. The ED regression model was called *exponential family nonlinear* (EFNL) model by Cordeiro and Paula (1989), thus extending the well-known idea of the GLMs by allowing a nonlinear regression structure for the explanatory variables. Wei (1998) wrote an excellent book for the EFNL models.

Let $\boldsymbol{y} = (y_1, \ldots, y_n)^T$ be a vector of observations and $\ell = \ell(\boldsymbol{\beta}, \phi)$ be the total log-likelihood function for a given ED regression model expressed in terms of $\boldsymbol{\beta}$ and τ. A simple calculation shows that $E(\partial^2 \ell/\partial \tau \partial \boldsymbol{\beta}) = 0$, that is, the parameters $\boldsymbol{\beta}$ and τ are globally orthogonal. Let $\widehat{\boldsymbol{\beta}}$ and $\hat{\tau}$ be the MLEs of $\boldsymbol{\beta}$ and τ, respectively. Let $\mu_i = g^{-1}(\eta_i)$ be the inverse link function evaluated at the linear predictor. Given a data vector $\boldsymbol{y}$, the total deviance for the ED regression is defined as

$$D(\boldsymbol{y}; \boldsymbol{\mu}) = \sum_{i=1}^{n} d(y_i; \mu_i).$$

The vector $\widehat{\boldsymbol{\beta}}$ can be calculated by minimizing the total deviance $D(\boldsymbol{y}; \boldsymbol{\mu})$ with respect to this parameter vector. The MLE of $\boldsymbol{\beta}$ does not depend on the dispersion parameter τ. Let $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\eta}}$ and $\widehat{\boldsymbol{\mu}} = g^{-1}(\widehat{\boldsymbol{\eta}}) = (g^{-1}(\hat{\eta}_1), \ldots, g^{-1}(\hat{\eta}_n))^T$ be the MLEs of the vector of regression coefficients $\boldsymbol{\beta}$, the vector of linear predictors $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)^T$ and the vector of means $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T$, respectively. The Fisher information matrix for $\boldsymbol{\beta}$ is $K(\boldsymbol{\beta}) = \tau \widetilde{\mathbf{X}}^T \mathbf{W} \widetilde{\mathbf{X}}$, where $\mathbf{W} = \mathrm{diag}\{w_1, \ldots, w_n\}$ is a diagonal matrix with weights $w_i = V(\mu_i)^{-1}(\partial \mu_i/\partial \eta_i)^2$.

The algorithm to estimate $\boldsymbol{\beta}$ can be carried out by using iteratively re-weighted least squares (IRLLS)

$$\widehat{\boldsymbol{\beta}} = (\widehat{\widetilde{\mathbf{X}}}^T \widehat{\mathbf{W}} \widehat{\widetilde{\mathbf{X}}})^{-1} \widehat{\widetilde{\mathbf{X}}}^T \widehat{\mathbf{W}} \widehat{z},$$

where $\widehat{\widetilde{\mathbf{X}}}$ and $\widehat{\mathbf{W}}$ are the quantities $\widetilde{\mathbf{X}}$ and W evaluated at $\widehat{\boldsymbol{\beta}}$, $\widehat{z} = (\hat{z}_1, \ldots, \hat{z}_n)^T$ is the working vector with components $z_i = \eta_i + (y_i - \mu_i) \partial \eta_i / \partial \mu_i$ at $\widehat{\boldsymbol{\beta}}$. These nonlinear equations have the same form as the estimating equations for GLMs with a local model matrix $\widetilde{\mathbf{X}}$ instead of a known design matrix and can be solved by iterative methods. The IRLLS algorithm is easily implemented using some standard statistical software such as SAS or the GAMLSS script in **R** (R Development Core Team, 2007).

Some asymptotic results for the ED regression were obtained by Cordeiro and Paula (1989), Cordeiro and McCullagh (1991) and Simas and Cordeiro (2009), among others, which produce wider results than those for GLMs.

Estimation of the parameter τ is a more difficult problem than the estimation of $\boldsymbol{\beta}$ and the complexity depends entirely on the functional form of $c(y; \tau)$. The MLE $\widehat{\tau}$ is a function of the deviance of the model, namely $\widehat{\tau}$ is the solution of the following equation

$$\widehat{\tau}^2 \sum_{i=1}^{n} \frac{\partial c(y_i; \tau)}{\partial \tau} \bigg|_{\tau=\widehat{\tau}} = \left[\sum_{i=1}^{n} l(y_i; y_i) - \frac{D(\mathbf{y}; \boldsymbol{\mu})}{2} \right]. \tag{3.2}$$

Equation (3.2) requires in general the use of a nonlinear algorithm to compute numerically $\hat{\tau}$ except for normal and inverse Gaussian models. However, for some ED regressions, the form of $c(y; \tau)$ is complicated, and $\hat{\tau}$ could be difficult to compute from (3.2). In these cases, we can use a moment estimate of τ directly from $\widehat{\boldsymbol{\mu}}$ given by $\widetilde{\tau} = \frac{1}{n-p} \sum_{i=1}^{n} (y_i - \hat{\mu}_i)^2 V(\hat{\mu}_i)^{-1}$ on the grounds that the expected value of $D(\mathbf{y}; \boldsymbol{\mu})/\tau$ is approximately $n - p$ for a well fitted model.

If (2.6) is a two-parameter full exponential family with canonical parameters $1/\tau$ and θ/τ, the following decomposition holds

$$c(y; \tau) = \frac{1}{\tau} a(y) + d(\tau) + e(y) \tag{3.3}$$

and then explicit expressions for $\widehat{\tau}$ are possible. Clearly, equation (3.3) is valid for normal, gamma and inverse Gaussian distributions but does not hold for all ED distributions in (2.6).

The above results apply to all GLMs by setting $f(\widetilde{\mathbf{X}}; \boldsymbol{\beta}) = \mathbf{X}\boldsymbol{\beta}$ for a given design matrix $\mathbf{X}$. Several diagnostic measures for the ED regression are simple extensions of those measures for GLMs.

The Tweedie regression models are extensively used in several areas for non-negative right-skewed data and continuous data that allow zero observations. Bonat and Kokonend-jib (2017) proposed maximum likelihood, quasi-likelihood and pseudo-likelihood methods for estimation and inference of Tweedie regression models with unknown power parameter p in the variance function. The last two methods are fast and computationally simple because they employ the first two moments only and thus it does not require the function $c_p(y; \tau)$.

3.2 Improved tests

Bartlett and Bartlett-type corrections improve the large-sample χ^2 approximation to the null distribution of the likelihood ratio, score and gradient statistics, when the sample size is finite. For a detailed discussion, see Cordeiro and Cribari-Neto (2014). In several papers, like those from Ferrari, Cordeiro and Cribari-Neto (2001), Cordeiro, Paula and Botter (1994) and Medeiros, Ferrari and Lemonte (2017), for example, improved likelihood ratio, score and gradient tests, respectively, were obtained in the class of dispersion models.

3.3 Modelling clustered and dependent data

3.3.1 *Latent stochastic processes based models*. Dispersion and ED models were used to study data containing clustered and dependent observations (Jørgensen, Labouriau and Lundbye-Christensen, 1996a, Jørgensen et al. 1996b, 1996c, 1999, Jørgensen and Song 1997, 2007, Jørgensen and Tsao, 1999, Artes and Jørgensen, 2000, Botter, Jørgensen and Peres, 2002, Ma and Jørgensen, 2007, Ma, Jørgensen and Willms, 2009) in recent years. The common idea explored there is that the dependence in the data is modeled using a latent stochastic process, the observations being conditionally independent given the latent process. For example, in Jørgensen et al. (1996c) when modeling the number of hospital emergency visits and in Botter, Jørgensen and Peres (2002) when modeling the mortality by lung diseases a latent (unobservable) stochastic process represented a time varying morbidity. These stochastic process based constructions are possible if the distributions of the increments of processes are infinite divisible, see Jørgensen, Labouriau and Lundbye-Christensen (1996a) for details, which is the case for the EDMs that have the index set Λ equal to $\mathbb{R}_+$. The infinite divisibility condition is satisfied by the Tweedie exponential dispersion models.

3.3.2 *Estimating equations inference*. Zeger and Liang (1986) considered the generalized estimating equations (GEEs) to analyze longitudinal data based on quasi-likelihood methods. Liang and Zeger (1986) derived the GEEs from a different and slightly more limited context. The method derives from EDMs, but is essentially based on second-moment assumptions for the response. In both articles, the GEEs are derived without fully specifying the joint distribution. The regression coefficients are consistently estimated even when the correlation structure is misspecified. However, efficiency depends on the working correlation matrices proximity to the true one (Liang and Zeger, 1986). Liang and Zeger's method has been widely used in several areas dealing with non-Gaussian correlated data (Hardin and Hilbe, 2003).

Artes and Jørgensen (2000) extended the GEE method to the class of dispersion models to handle certain types of non-normal data such as angles and proportions that are not well accommodated by EDMs, and for which there are currently no good methods available for longitudinal data analysis.

Song and Tan (2000) proposed methods to directly model the marginal means of the longitudinal proportional responses using the simplex distribution that takes into account the fact that such responses are percentages restricted between zero and one and may as well have large dispersion.

4 Concluding remarks and future perspectives

As mentioned above, the Danish statistician Bent Jørgensen (April 15, 1954; November 19, 2015) made several vital contributions in the area of statistical modeling. He supervised many students in Denmark, Brazil, and Canada and developed a vast international scientific collaboration network. Bent worked on a combination of theoretical and applied topics, including exponential families, univariate and multivariate dispersion models, exponential dispersion models, proper dispersion models, Tweedie distributions, generalized estimating equations and other types of statistical models. Although Bent's domain was mainly theoretical statistics, he also made significant contributions in a wide range of applied fields such as insurance, meteorology, and marine ecology, among others. In this article, as colleagues and friends, we outlined some details of his carrier and reviewed some of his main contributions, especially, univariate dispersion models, exponential dispersion and proper dispersion models.

The work of Bent Jørgensen opened new research areas and inspired other researchers in the field. As a piece of evidence, we mention higher-order asymptotics (see Section 3.2), the

field of exponential family non-linear regressions (see Section 3.1) and the latent stochastic processes based models (see Section 3.3.1). We envision that further developments might appear by expanding the general theory of dispersion models to different multivariate and dependent observation contexts (already partially done). Defining general dispersion models via integral transforms other than the Laplace transform might allow incorporating heavy tail distributions and different types of stochastic processes not considered yet.

Acknowledgments

References

Artes, R. and Jørgensen, B. (2000). Longitudinal data estimating equations for dispersion models. *Scandinavian Journal of Statistics* **27**, 321–334. MR1777507 https://doi.org/10.1111/1467-9469.00192

Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–365. MR0712023 https://doi.org/10.1093/biomet/70.2.343

Barndorff-Nielsen, O. E. (1988). *Parametric Statistical Models and Likelihhod. Lecture Notes in Statistics* **50**. Berlin: Springer. MR0971982 https://doi.org/10.1007/978-1-4612-3934-5

Bonat, W. H. and Jørgensen, B. (2016). Multivariate covariance generalized linear models. *Journal of the Royal Statistical Society Series C Applied Statistics* **65**, 649–675. MR3564998 https://doi.org/10.1111/rssc.12145

Bonat, W. H., Jørgensen, B., Kokonendji, C. C., Hinde, J. and Demétrio, C. G. B. (2018). Extended Poisson-Tweedie: Properties and regression models for count data. *Statistical Modelling* **18**, 24–49. MR3746503 https://doi.org/10.1177/1471082X17715718

Bonat, W. H. and Kokonendjib, C. C. (2017). Flexible Tweedie regression models for continuous data. *Journal of Statistical Computation and Simulation* **87**, 2138–2152. MR3656096 https://doi.org/10.1080/00949655.2017.1318876

Botter, D. A., Jørgensen, B. and Peres, A. A. Q. (2002). A longitudinal study of mortality and air pollution for São Paulo, Brazil. *Journal of Exposure Analysis and Environmental Epidemiology* **12**, 335–343.

Cordeiro, G. M. and Cribari-Neto, F. (2014). *An Introduction to Bartlett Correction and Bias Reduction*. Berlin: Springer. MR3289997 https://doi.org/10.1007/978-3-642-55255-7

Cordeiro, G. M. and McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society, Series B, Methodological* **53**, 629–643. MR1125720

Cordeiro, G. M. and Paula, G. A. (1989). Improved likelihood ratio statistics for exponential family nonlinear models. *Biometrika* **76**, 93–100. MR0991426 https://doi.org/10.1093/biomet/76.1.93

Cordeiro, G. M., Paula, G. A. and Botter, D. A. (1994). Improved likelihood ratio tests for dispersion models. *International Statistical Review* **62**, 257–274.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics* **25**, 631–650. MR0066602 https://doi.org/10.1214/aoms/1177728652

Daniels, H. E. (1980). Exact sadlepoint approximation. *Biometrika* **67**, 59–63. MR0570505 https://doi.org/10.1093/biomet/67.1.59

Ferrari, S. L. P., Cordeiro, G. M. and Cribari-Neto, F. (2001). Higher-order asymptotic refinements for score tests in proper dispersion models. *Journal of Statistical Planning and Inference Applied Mathematical Modelling* **97**, 177–190. MR1851381 https://doi.org/10.1016/S0378-3758(00)00352-9

Hardin, J. W. and Hilbe, J. M. (2003). *Generalized Estimating Equations*. London: Chapman & Hall/CRC. MR2000388

Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution. Lecture Notes in Statistics* **9**. Berlin: Springer. MR0648107

Jørgensen, B. (1987a). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society, Series B, Methodological* **49**, 127–162. MR0905186

Jørgensen, B. (1987b). Small-dispersion asymptotics. *Brazilian Journal of Probability and Statistics* **1**, 59–90. MR0932188

Jørgensen, B. (1997). *The Theory of Dispersion Models*. London: Chapman & Hall Ltd. MR1462891

Jørgensen, B. (2013). Construction of multivariate dispersion models. *Brazilian Journal of Probability and Statistics* **27**, 285–309. MR3064725 https://doi.org/10.1214/11-BJPS171

Jørgensen, B., Demétrio, C. G. B., Kristensen, E., Banta, G. T., Petersen, H. C. and Delefosse, M. (2011). Bias-corrected Pearson estimating functions for Taylor's power law applied to benthic macrofauna data. *Statistics & Probability Letters* **81**, 749–758. MR2793740 https://doi.org/10.1016/j.spl.2011.01.005

Jørgensen, B. and Labouriau, R. (2012). Exponential families and theoretical inference. In *Monografias de Matematica* **52**, 196. Rio de Janeiro, Brazil: Instituto de Matematica Pura e Aplicada (IMPA). https://pure.au.dk/portal/files/51499534/Mon52.pdf. MR1183933

Jørgensen, B., Labouriau, R. and Lundbye-Christensen, S. (1996a). Linear growth curve analysis based on exponential dispersion models. *Journal of the Royal Statistics Society B* **58**, 573–592. MR1394368

Jørgensen, B. and Lauritzen, S. (2000). Multivariate dispersion models. *Journal of Multivariate Analysis* **74**, 267–281. MR1794004 https://doi.org/10.1006/jmva.1999.1885

Jørgensen, B., Lundbye-Christensen, S., Song, P. X.-K. and Sun, L. (1996b). Linear state-space models for multivariate longitudinal data of mixed types. *Canadian Journal of Statistics* **24**, 385–402. MR1421864 https://doi.org/10.2307/3315747

Jørgensen, B., Lundbye-Christensen, S., Song, P. X.-K. and Sun, L. I. (1999). A state space model for multivariate longitudinal count data. *Biometrika* **86**, 169–181. MR1688080 https://doi.org/10.1093/biomet/86.1.169

Jørgensen, B., Lundbye-Christensen, S., Song, X.-K. and Sun, L. (1996c). A longitudinal study of emergency room visits and air pollution for Prince George, British Columbia. *Statistics in Medicine* **15**, 823–836.

Jørgensen, B. and Song, P. X.-K. (1997). Stationary time series models with exponential dispersion model margins. *Journal of Applied Probability* **35**, 78–92. MR1622447 https://doi.org/10.1239/jap/1032192553

Jørgensen, B. and Song, P. X.-K. (2007). Stationary state space models for longitudinal data. *Canadian Journal of Statistics* **35**, 461–483. MR2381394 https://doi.org/10.1002/cjs.5550350401

Jørgensen, B. and Tsao, M. (1999). Linear dispersion models and longitudinal data analysis. *Statistics in Medicine* **18**, 2257–2270.

Labouriau, R. (2020). Construction and extension of dispersion models. Available at arXiv:2008.05448.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal analysis using generalized linear models. *Biometrika* **73**, 13–22. MR0836430 https://doi.org/10.1093/biomet/73.1.13

Lucaks, E. (1970). *Characteristic Functions*, 2nd rev. ed. MR0346874

Lugannani, R. and Rice, S. (1980). Saddle point approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability* **12**, 475–490. MR0569438 https://doi.org/10.2307/1426607

Ma, R. and Jørgensen, B. (2007). Nested generalized linear mixed models: An orthodox best linear unbiased predictor approach. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **69**, 625–641. MR2370072 https://doi.org/10.1111/j.1467-9868.2007.00603.x

Ma, R., Jørgensen, B. and Willms, J. D. (2009). Nested generalized linear mixed models: An orthodox best linear unbiased predictor approach. *Journal of the Royal Statistical Society, Series B, Statistical Methodology* **69**, 625–641. MR2370072 https://doi.org/10.1111/j.1467-9868.2007.00603.x

Medeiros, F. M. C., Ferrari, S. L. P. and Lemonte, A. J. (2017). Improved inference in dispersion models. *Applied Mathematical Modelling* **51**, 317–328. MR3694537 https://doi.org/10.1016/j.apm.2017.06.045

Morris, C. N. (1982). Natural exponential families with quadratic variance functions. *The Annals of Statistics* **10**, 65–80. MR0642719

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Rudin, W. (1973). *Functional Analysis. TMH Edition*. Limited, New Delhi: Tata McGraw-Hill Publishing Company. MR0365062

Simas, A. and Cordeiro, G. M. (2009). Adjusted Pearson residuals in exponential family nonlinear models. *Journal of Statistical Computation and Simulation* **79**, 411–425. MR2522359 https://doi.org/10.1080/00949650701829380

Song, P. X.-K. and Tan, M. (2000). Marginal models for longitudinal continuous proportional data. *Biometrics* **56**, 496–502.

Tweedie, M. C. K. (1947). Functions of a statistical variate with given means, with special reference to Laplacian distributions. *Proceedings of the Cambridge Philosophical Society* **49**, 41–49. MR0018381

Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. In *Statistics Applications and New Directions* (J. K. Ghosh and J. Roy, eds.), *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, 579–604. Indian Statistical Institute. MR0786162

Ushakov, N. G. (1999). *Selected Topics in Characteristic Functions*. Utrecht, The Netherlands: VSP. MR1745554 https://doi.org/10.1515/9783110935981

Wei, B.-C. (1998). *Exponential Family Nonlinear Models*. Berlin: Springer. MR1646023

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42**, 121–130.

G. M. Cordeiro
Departamento de Estatística
Universidade Federal de Pernambuco
Recife
Brazil
E-mail: gauss@de.ufpe.br

R. Labouriau
Department of Mathematics
Aarhus University
Ny Munkegade 118
DK-8000 Aarhus C
Denmark
E-mail: rodrigo.labouriau@math.au.dk

D. A. Botter
Instituto de Matemática e Estatística
Universidade de São Paulo
Rua do Matão 1010, 05508–090
São Paulo SP
Brazil
E-mail: botter@ime.usp.br