J. R. Statist. Soc. B (1987)
49, *No.* 2, *pp.* 127–162

Exponential Dispersion Models

By BENT JØRGENSEN†

Odense University, Denmark

[*Read before the the* Royal Statistial Society, *at a meeting organized by the* Research Section *on Wednesday, December 10th, 1986*, Professor A. F. M. Smith *in the Chair*]

SUMMARY

We study general properties of the class of exponential dispersion models, which is the multivariate generalization of the error distribution of Nelder and Wedderburn's (1972) generalized linear models. Since any given moment generating function generates an exponential dispersion model, there exists a multitude of exponential dispersion models, and some new examples are introduced. General results on convolution and asymptotic normality of exponential dispersion models are presented. Asymptotic theory is discussed, including a new small-dispersion asymptotic framework, which extends the domain of application of large-sample theory. Procedures for constructing new exponential dispersion models for correlated data are introduced, including models for longitudinal data and variance components. The results of the paper unify and generalize standard results for distributions such as the Poisson, the binomial, the negative binomial, the normal, the gamma, and the inverse Gaussian distributions.

Keywords: ASYMPTOTIC THEORY; COMBINATIONS; COMPOUND DISTRIBUTIONS; CONVOLUTION; EXPONENTIAL FAMILIES; GENERALIZED LINEAR MODELS; LONGITUDINAL DATA; MIXTURES; POWER VARIANCE FUNCTIONS; SMALL-DISPERSION ASYMPTOTICS; STABLE DISTRIBUTION; VARIANCE COMPONENTS; VARIANCE FUNCTIONS

1. INTRODUCTION

The increasingly powerful computational tools available to the statistician allow him to handle increasingly complex models. However, there remains a need for models based on simple, yet general, ideas. Thus, the success of Nelder and Wedderburn's (1972) generalized linear models relies to some extent on the balance they achieve between simplicity and generality, computationally as well as conceptually, and on the fact that they include some important standard statistical models as special cases, specifically linear normal models and log-linear models for contingency tables.

In the present paper we study the error distribution of generalized linear models, which in its multivariate form is

$$f(y;\,\lambda,\,\theta)=a(\lambda,\,y)e^{\lambda\{y^{\mathrm{T}}\theta-\kappa(\theta)\}},\ y\in\mathbb{R}^k, \tag{1.1}$$

where a and κ are given functions, θ varies in a subset of $\mathbb{R}^k$ and λ varies in a subset of $\mathbb{R}_+$. In order to distinguish between the random and the systematic part of a generalized linear model we call (1.1) an *exponential dispersion model*, a terminology that reflects the partly exponential form of (1.1) and the important role played by the *dispersion parameter* $\sigma^2=1/\lambda$. A generalized linear model is obtained if $y_1,\ldots,y_n$ are independent one-dimensional variables, such that y_i is distributed according to (1.1) with parameters λ and $\theta_i=h(\eta_i)$, where h is called the link function, and $(\eta_1,\ldots,\eta_n)^{\mathrm{T}}=X\beta$, where β is a $p\times 1$ vector parameter and X is an $n\times p$ matrix.

Implicitly, the main theme of the paper is thus generalized linear models, but because the partly linear systematic form employed in generalized linear models is not necessary for the theory considered here, we emphasize properties and examples of exponential dispersion

†*Present address*: Department of Mathematics, Odense University, Campusvej 55, DK-5230 Odense M, Denmark.

 0035–9246/87/49127 $2.00

models. Current research, including the book of McCullagh and Nelder (1983), has concentrated on a few standard univariate error distributions, such as the Poisson, binomial, normal, gamma and inverse Gaussian distributions. The multivariate form (1.1) has been suggested previously by various authors, but no systematic theory is present, see for example McCullagh (1980, 1983), Barndorff-Nielsen and Blaesild (1983) and Jørgensen (1983). Our results unify and generalize standard results for the distributions mentioned above. Many of the ideas in this paper may also be found in the works of M. C. K. Tweedie (cf. Tweedie, 1947, 1984), dating back almost 40 years, although his ideas have remained largely unnoticed. Tweedie (1947) suggested (1.1) in the one-dimensional case, and (1.1) was introduced again by Nelder and Wedderburn (1972) without reference to Tweedie's work.

Exponential dispersion models have certain analogies with location and scale models, with the role of location and scale being played by respectively the expectation $\mu = E(y)$ and the dispersion parameter σ^2. Exponential dispersion models thus constitute a useful generalization of exponential families (strictly speaking of linear exponential families, cf. Section 2), and our results go beyond exponential family theory. As a parallel to the way a location and scale model is generated from a given density function, we show that any given moment generating function generates an exponential dispersion model in a unique way. The theory of exponential dispersion models is thus to some extent the study of the class of tractable moment generating functions, and our strategy is to be able to apply the theory and to make inferences without explicitly inverting the moment generating function.

An important advantage of exponential dispersion models is the elegant asymptotic theory, which generalizes the analysis of deviance for generalized linear models, giving rise to asymptotic versions of the familiar t-, F-, and χ^2-tests from linear normal theory. Expanding the domain of application of standard large-sample theory, we develop a new asymptotic framework, called small-dispersion asymptotics, which applies for λ large.

Based on techniques of combining and mixing exponential dispersion models, we present methods for constructing models for correlated data, such as longitudinal data or variance component problems. These models resemble, but do not include, the standard normal-theory models usually employed in these fields.

In Section 2 we study elementary properties of exponential dispersion models, including the variance function and discrete exponential dispersion models. In Section 3 we consider exponential dispersion models with power variance functions. In Section 4 we consider asymptotic theory. In Section 5 we consider combinations and mixtures and discuss longitudinal data and variance component models.

2. PROPERTIES OF EXPONENTIAL DISPERSION MODELS

We now summarize the basic properties of continuous and discrete exponential dispersion models. For basic results and terminology concerning exponential families used in the sequel, the reader is referred to Barndorff-Nielsen (1978a). The mathematical details of the techniques discussed in Section 2.1 are given in the paper Jørgensen (1986a).

2.1. *Generating New Exponential Dispersion Models*

Let Q be a given distribution on $\mathbb{R}^k$. The *linear exponential family* (i.e. the exponential family with canonical statistic x) *generated by* Q has probability density function

$$\frac{dQ_{1,\theta}}{dQ}(x) = e^{x^{\mathrm{T}}\theta - \kappa(\theta)}, \quad x \in \mathbb{R}^k, \tag{2.1}$$

with respect to Q. Here $\kappa(\theta) = \log \int \exp(x^{\mathrm{T}}\theta) dQ(x)$ is the cumulant generating function of Q, and θ varies in the set

$$\Theta = \{\theta \in \mathbb{R}^k \colon \kappa(\theta) < \infty\}, \tag{2.2}$$

the effective domain for κ. Note that a linear exponential family is sometimes called a natural exponential family (cf. Morris, 1982).

Let Λ be the set of $\lambda > 0$ such that $\lambda\kappa(\cdot)$ is the cumulant generating function of some distribution Q_λ, say. The linear exponential family generated by Q_λ has probability density function

$$\frac{dQ_{\lambda,\theta}}{dQ_\lambda}(x) = e^{x^{\mathrm{T}}\theta - \lambda\kappa(\theta)}, \; x \in \mathbb{R}^k, \tag{2.3}$$

with respect to Q_λ. Transforming to $y = x/\lambda$, we obtain an exponential dispersion model of the form

$$\frac{dP_{\lambda,\theta}}{dP_\lambda}(y) = e^{\lambda\{y^{\mathrm{T}}\theta - \kappa(\theta)\}}, \; y \in \mathbb{R}^k, \tag{2.4}$$

where P_λ is Q_λ transformed by $y = x/\lambda$, and (λ, θ) varies in $\Lambda \times \Theta$.

Since any distribution with a moment generating function generates an exponential dispersion model in the above way, there exists an abundance of exponential dispersion models. A number of new examples are introduced below and in Section 3. The following example illustrates the technique of generating a new exponential dispersion model.

Example 2.1. According to Feller (1971, p. 437), the distribution with probability density function

$$g(x; \lambda) = \lambda x^{-1} I_\lambda(x) e^{-x}, \; x > 0, \tag{2.5}$$

has moment generating function

$$M(s; \lambda) = [1 - s - \{(1-s)^2 - 1\}^{1/2}]^\lambda, \; s \leqslant 0,$$

for $\lambda > 0$, where I_λ denotes the modified Bessel function of the first kind with index λ. For integer values of λ, (2.5) is the probability density of the first passage time through λ for a randomized random walk. The exponential dispersion model generated by (2.5) is

$$f(y; \lambda, \theta) = \lambda y^{-1} I_\lambda(\lambda y) \exp\lambda[y\theta - \log\{-\theta - (\theta^2 - 1)^{\frac{1}{2}}\}], \; y > 0,$$

with

$$\Theta = (-\infty, -1] \quad \text{and} \quad \Lambda = \mathbb{R}_+. \qquad \square$$

Any given exponential dispersion model $\mathscr{P}$ may be generated from any given member of $\mathscr{P}$ via (2.4). The model (2.4) is uniquely determined by the distribution Q as the exponential dispersion model that contains Q. Similarly, any given linear exponential family generates an exponential dispersion model in a unique way via (2.1).

As in Example 2.1, P_λ may be dominated by Lebesgue measure for any λ in Λ. In this case, $P_{\lambda,\theta}$ has a density of the form (1.1) with respect to Lebesgue measure on $\mathbb{R}^k$, and we speak of a *continuous exponential dispersion model*. The other main type is the discrete exponential dispersion models, which we consider in Section 2.4. There exist intermediate cases, where generally the dominating measure for (1.1) depends on λ, cf. Jørgensen (1986a). Such models are of little practical relevance, but the properties derived in the following apply for an arbitrary exponential dispersion model.

For a given exponential dispersion model (1.1), (λ, θ) varies in the set $\Lambda \times \Theta$, where Θ is defined by (2.2) and Λ is the set of $\lambda > 0$ such that $\lambda\{\kappa(\cdot + \theta_0) - \kappa(\theta_0)\}$ is a cumulant generating function, θ_0 being an arbitrary member of Θ. Generalizing exponential family terminology, we call θ the *canonical parameter* and Θ the *canonical parameter domain* of (1.1). The function κ is called the *cumulant generator* of (1.1). We call λ the *index parameter*, and Λ is called the *index set* for (1.1). We have $\Lambda = \mathbb{R}_+$ if and only if any member of (1.1) is infinitely divisible, and (1.1) is infinitely divisible either for none or for all values of the parameter (λ, θ), cf. Jørgensen (1986a).

For any given value of λ, (1.1) is a linear exponential family, which we denote by $\mathscr{P}_\lambda$. If the exponential representation (1.1) for $\mathscr{P}_\lambda$ is minimal for one value of λ, then it is minimal for all values of λ in Λ (Jørgensen, 1986a). Minimality, which we assume throughout the rest of the paper, is equivalent to P_λ having affine support $\mathbb{R}^k$ for all values of λ in Λ.

2.2. *Moments, Cumulants, and the Variance Function*

Let y be a random variable with distribution (1.1). The moment generating function of y is

$$M(s;\,\lambda,\,\theta) = e^{\lambda\{\kappa(\theta+s/\lambda)-\kappa(\theta)\}}. \tag{2.6}$$

Hence, the cumulant for y of order $i=(i_1,\,\ldots,\,i_k)$ is

$$\kappa_i(\lambda,\,\theta) = D^i\kappa(\theta)\lambda^{1-i_1-\ldots-i_k},\ \theta\in\mathrm{int}\Theta,$$

where D is the differential operator, and $\mathrm{int}\Theta$ denotes the interior of Θ. In particular, the mean value of y is $\mu=\tau(\theta)$, where $\tau(\theta)=\partial\kappa/\partial\theta$ is called the *mean value mapping*. From exponential family theory (Barndorff-Nielsen, 1978a, p. 121) is follows that τ is a one-to-one both ways continuously differentiable mapping of $\mathrm{int}\Theta$ onto $\Omega=\tau(\mathrm{int}\Theta)$. Ignoring the boundary of Θ, we may thus parametrize (1.1) by $(\mu,\,\sigma^2)$, where $\sigma^2=1/\lambda$ is the dispersion parameter, and μ and σ^2 are variation independent. In the following, $ED(\mu,\,\sigma^2)$ denotes the distribution (1.1), using the parametrization (μ,σ^2). The form of the cumulants and the cumulant transform $\log M(s;\,\lambda,\,\theta)$ was given for $k=1$ by Tweedie (1947).

Since τ is one-to-one, we may write the variance matrix of y in the form $\sigma^2V(\mu)$, where

$$V(\mu) = \left.\frac{\partial^2\kappa}{\partial\theta^{\mathrm{T}}\partial\theta}\right|_{\theta=\tau^{-1}(\mu)}$$

is called the *variance function* of (1.1). The matrix $V(\mu)$ is positive-definite for any μ in Ω, cf. Barndorff-Nielsen (1978a, p. 114). For a given exponential dispersion model (1.1), the variance function depends on the parametrization used, but any two parametrizations give proportional variance functions. For the purpose of the next theorem we shall identify V with its corresponding equivalence class under multiplication. For a linear exponential family $\mathscr{F}$ given by (2.1) we call V the variance function of $\mathscr{F}$.

Theorem 1. A linear exponential family is characterized within the class of all linear exponential families by its variance function. An exponential dispersion model is characterized within the class of all exponential dispersion models by its variance function.

Proof. For the linear exponential family (2.1), κ and the inverse of the function τ satisfy the differential equations

$$\frac{\partial\tau^{-1}(\mu)}{\partial\mu^{\mathrm{T}}} = V(\mu)^{-1} \tag{2.7}$$

and

$$\frac{\partial\kappa}{\partial\theta} = \tau(\theta), \tag{2.8}$$

where $V(\mu)^{-1}$ is the inverse of the matrix $V(\mu)$. Given V, we may thus obtain κ by solving (2.7), inverting τ^{-1} and solving (2.8), and from κ we may obtain the moment generating function $M(\cdot;\,1,\,\theta)$ of (2.1), where M is defined by (2.6). It is easily shown that the class of distributions corresponding to $M(\cdot;\,1,\,\theta)$ does not depend on the initial conditions for (2.7) and (2.8). Hence we have shown that a linear exponential famly is characterized by its variance function. The corresponding result for exponential dispersion models now follows from the theory in Section 2.1.

Theorem 1 was shown for linear exponential families with $k = 1$ by Morris (1982), but the essence of the proof goes back to Tweedie (1947).

2.3. *Convolution and Asymptotic Normality*

Exponential dispersion models enjoy a remarkable convolution property, which is somewhat analogous to the convolution property of the normal distribution. Suppose that $y_1, \ldots, y_n$ are independent and $y_i \sim ED(\mu, \sigma^2/w_i)$, where $w_i/\sigma^2 \in \Lambda$, $i = 1, 2, \ldots, n$. From the form of the moment generating function (2.6) it follows that

$$\frac{1}{w_.} \sum_{i=1}^{n} w_i y_i \sim ED(\mu, \sigma^2/w_.), \tag{2.9}$$

where $w_. = \Sigma w_i$. In particular, for $y_1, \ldots, y_n$ independent and identically distributed $ED(\mu, \sigma^2)$ we have

$$\bar{y}_. \sim ED(\mu, \sigma^2/n), \tag{2.10}$$

where $\bar{y}_.$ denotes the average of $y_1, \ldots, y_n$. The result (2.9) includes as special cases the well-known convolution properties of the normal, gamma and inverse Gaussian distributions, except that the normal distribution is closed under arbitrary convolutions, whereas (2.9) requires $y_1, \ldots, y_n$ to have identical means. In the case $k = 1$, (2.9) was first obtained by Tweedie (1947), see also McCullagh (1983).

Using the central limit theorem, we get from (2.9) that for $y \sim ED(\mu, \sigma^2)$,

$$\sqrt{\lambda}(y - \mu) \rightarrow N_k(0, V(\mu)) \quad \text{for} \quad \lambda \rightarrow \infty, \tag{2.11}$$

where N_k denotes the k-variate normal distribution and the convergence is in distribution. A more formal proof of (2.11) proceeds via expansion of the moment generating function of the variable $\sqrt{\lambda}(y - \mu)$. The result (2.11), which for $k = 1$ was given by Tweedie (1947), unifies and generalizes a number of known results on convergence to normality, for example for the gamma and inverse Gaussian distributions.

2.4. *Discrete Exponential Dispersion Models*

Jørgensen (1986a) showed that there exist no exponential dispersion models with support on the integer lattice $\mathbb{Z}^k$. More precisely, the support of an exponential dispersion model generated by a distribution with support on a subset of $\mathbb{Z}^k$ is contained in $c_\lambda + \lambda^{-1}\mathbb{Z}^k$ for some vector c_λ. With a support that depends on the value of λ, such a distribution is not practically relevant.

However, many standard discrete distributions have the form

$$f(x; \lambda, \theta) = a(\lambda, x) e^{x^{\mathrm{T}}\theta - \lambda\kappa(\theta)}, \; x \in \mathbb{Z}^k. \tag{2.12}$$

We call (2.12) a *discrete exponential dispersion model*, and we write $x \sim ED^*(\lambda, \theta)$. For example, the multinomial distribution corresponding to λ trials with $k + 1$ possible outcomes with probabilities $\mu_1, \ldots, \mu_k$, $1 - \Sigma\mu_i$, the sum being from 1 to k, is a discrete exponential dispersion model with

$$f(x; \lambda, \theta) = \left[\overset{\lambda}{x_1 \ldots x_k} \right] \exp\left\{ \sum_{i=1}^{k} x_i\theta_i - \lambda \log\left(1 + \sum_{i=1}^{k} e^{\theta_i} \right) \right\}, \tag{2.13}$$

where $x_i \geqslant 0, \Sigma x_i \leqslant \lambda$ and $\theta_i = \log\{\mu_i/(1 - \Sigma\mu_i)\}$. We denote (2.13) by $Bi_k(\lambda, \mu_1, \ldots, \mu_k)$. Similarly, the negative multinomial distribution of order k with index parameter λ and probability parameters $\exp(\theta_1), \ldots, \exp(\theta_k)$ is a discrete exponential dispersion model with

$$f(x; \lambda, \theta) = \frac{\Gamma(\lambda + x_1 + \cdots + x_k)}{\Gamma(1 + x_1) \ldots \Gamma(1 + x_k)\Gamma(\lambda)} \exp\left\{ \sum_{i=1}^{k} x_i\theta_i + \lambda \log\left(1 - \sum_{i=1}^{k} e^{\theta_i} \right) \right\} \tag{2.14}$$

where $x_i \geqslant 0$ for $i = 1, \ldots, k$. We denote (2.14) by $Nb_k(\lambda, \theta_1, \ldots, \theta_k)$. Further examples of discrete exponential dispersion models are the Hermite distribution (Kemp and Kemp, 1965), and the bivariate Poisson distribution (Johnson and Kotz, 1969, p. 298).

Since $x \sim ED^*(\lambda, \theta)$ implies $x/\lambda \sim ED(\mu, \sigma^2)$, properties of discrete exponential dispersion models are easily derived from general properties of exponential dispersion models, and in the following we use notations such as $\mu = \tau(\theta) = \partial\kappa/\partial\theta$ and $\sigma^2 = \lambda^{-1}$ corresponding to $ED(\mu, \sigma^2)$. In particular, the mean and variance of x are

$$m = E_{\lambda,\theta}(x) = \lambda\mu, \ \mathrm{Var}_{\lambda,\theta}(x) = \lambda V(\mu).$$

For λ known, (2.12) is a linear exponential family with variance function $\lambda V(m/\lambda)$. From (2.9) we obtain

$$ED^*(\lambda_1, \theta) * ED^*(\lambda_2, \theta) = ED^*(\lambda_1 + \lambda_2, \theta), \tag{2.15}$$

where the second $*$ denotes convolution. Formula (2.15) includes the convolution properties of the multinomial and the negative multinomial distributions as special cases.

The Poisson distribution is a linear exponential family, and as such it formally generates a discrete exponential dispersion model via (2.3), with probability function

$$f(x; \lambda, \theta) = \frac{\lambda^x}{x!} e^{x\theta - \lambda e^\theta}, \ x = 0, 1, \ldots. \tag{2.16}$$

Since (2.16) is a Poisson distribution $Po(m)$ with mean $m = \lambda e^\theta$, it is not a proper generalization of the Poisson distribution. But (2.16) shows that the standard convolution property of the Poisson distribution is formally a special case of (2.15). Similarly, the asymptotic normality of the Poisson distribution for large values of the mean is a special case of (2.11), because (2.11) applies to $y = x/\lambda$ for any discrete exponential dispersion model $x \sim ED^*(\lambda, \theta)$. Jørgensen (1986a) showed that, under mild regularity conditions, any positive univariate discrete exponential dispersion model $ED^*(\lambda, \theta)$ converges to the Poisson distribution $Po(m)$ for λ large and $m = E_{\lambda,\theta}(x)$ fixed. Special cases of this result are the convergence of the binomial and the negative binomial distributions to the Poisson distribution. The Poisson distribution thus occupies a rather unique position among the discrete exponential dispersion models.

3. POWER VARIANCE FUNCTIONS

3.1. *Finding all Exponential Dispersion Models with Power Variance Functions*

As shown in Section 2.2, an exponential dispersion model is characterized by its variance function $V(\mu)$. An important class of exponential dispersion models corresponds to power variance functions

$$V(\mu) = \mu^p. \tag{3.1}$$

A number of authors have studied power variance functions, independently of the present author, for example Morris (1981), Tweedie (1984), Hougaard (1986), Bar-Lev and Enis (1985). The most complete study is by Tweedie (1984), but as shown below, his claim that (3.1) does not correspond to an exponential family for $p < 0$ is not correct. With this exception, the following results about power variance functions may all be found in his paper.

By solving (2.7) and (2.8) based on the variance function (3.1), we obtain the following cumulant generator κ_α (if it exists as such)

$$\kappa_{-\infty}(\theta) = e^\theta \qquad (p = 1), \tag{3.2}$$

$$\kappa_0(\theta) = -\log(-\theta) \qquad (p = 2), \tag{3.3}$$

$$\kappa_\alpha(\theta) = \frac{\alpha - 1}{\alpha}\left(\frac{\theta}{\alpha - 1}\right)^\alpha \qquad (p \neq 1, 2), \tag{3.4}$$

where

$$\alpha = \frac{p-2}{p-1}. \tag{3.5}$$

The case (3.2) ($p = 1$) corresponds to the Poisson distribution (2.16), and (3.3) ($p = 2$) corresponds to the gamma distribution. For $0 < \alpha < 1$ and $1 < \alpha \leqslant 2$, (3.4) is the cumulant generating function of an extreme stable distribution with index α (cf. Eaton, Morris and Rubin, 1971), and hence these extreme stable distributions generate exponential dispersion models with variance function (3.1) with $p \leqslant 0$ ($1 < \alpha \leqslant 2$) or $p > 2$ ($0 < \alpha < 1$). In particular $p = 0$ ($\alpha = 2$) and $p = 3$ ($\alpha = \frac{1}{2}$) correspond to respectively the normal and the inverse Gaussian distributions. For $p > 2$ the distributions are positive and for $p \leqslant 0$ they have support on $\mathbb{R}$. In Section 5.1 we show that the case $1 < p < 2$ ($\alpha < 0$) corresponds to certain compound Poisson distributions, see also Tweedie (1984). The following theorem shows that the remaining cases ($0 < p < 1$) do not correspond to exponential dispersion models. Table 1 summarizes all exponential dispersion models with power variance functions.

Theorem 2. Power variance functions with power $0 < p < 1$ do not correspond to exponential dispersion models.

Proof. We assume that (3.4) is the cumulant generator of an exponential dispersion model, and show that this leads to a contradiction. Let p and α be related by (3.5), so that $0 < p < 1$ corresponds to $\alpha > 2$. The solution to (2.7) corresponding to (3.1) is

$$\tau_\alpha(\theta) = \left(\frac{\theta}{\alpha - 1}\right)^{\alpha - 1}. \tag{3.6}$$

Since $V(\mu) > 0$ for $\mu \in \Omega$, and because $V(0) = 0$ in the present case, we get $\Omega = \mathbb{R}_+$. By (3.6) this implies that $\operatorname{int}\Theta = \mathbb{R}_+$. Since $\kappa_\alpha(0) < \infty$ in the present case, we conclude that $\theta = [0, \infty)$, so $\exp\{\kappa_\alpha(\cdot)\}$ is a moment generating function. However, this moment generating function corresponds to a stable distribution with index $\alpha > 2$, and such stable distributions do not exist (Feller, 1971, p. 170). By a stable distribution with index α, we here mean a distribution P for which $x_1 + \cdots + x_n$ and $n^{1/\alpha}x_1$ have identical distributions, if $x_1, \ldots, x_n$ are independent and all have distribution P. Hence we have reached a contradiction, and the proof is complete.

The compound Poisson distributions ($1 < p < 2$) are interesting, because they are continuous for $y > 0$, but allow zero observations. Data of this form are a common problem in many areas, for example meteorology (daily rainfall, Coe and Stern, 1984) or insurance (yearly claims by individual insurance holders). If p is known, the model is easily fitted in *GLIM*. An example is given in Section 4.4.

As the form of the variance function (3.1) suggests, the corresponding exponential dispersion model is closed under scale transformations, a property which is important when modelling positive continuous data. Thus, if $ED^{(p)}(\mu, \sigma^2)$ denotes the exponential dispersion model corresponding to (3.1), then $y \sim ED^{(p)}(\mu, \sigma^2)$ implies

$$cy \sim ED^{(p)}(c\mu, c^{2-p}\sigma^2), \tag{3.7}$$

for any positive constant c.

Special cases of (3.7) are the scale transformation properties of the normal, the gamma and the inverse Gaussian distributions, and (3.7) includes the scale transformation property of the extreme stable distributions as a limiting case. For the Poisson distribution, (3.7) applies to the distribution of $y = x/\lambda$ for x distributed according to (2.16), but the resulting formula for the Poisson distribution is void.

It is well-known that the stable distributions, the gamma distribution and the Poisson distribution etc. are infinitely divisible, and in fact, all the exponential dispersion models in Table 1 are infinitely divisible. For $p \neq 2$, this may be seen as a consequence of (3.7), because (3.7) implies $\Lambda = \mathbb{R}_+$.

TABLE 1
Exponential dispersion models with power variance functions

Distribution	p	α	*Support*	Ω	Θ
Generated by extreme stable distributions	$p<0$	$1<\alpha<2$	$\mathbb{R}$	$\mathbb{R}_+$	$[0, \infty)$
Normal distribution	$p=0$	$\alpha=2$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$
Not exponential dispersion models	$0<p<1$	$\alpha>2$	—	$\mathbb{R}_+$	$[0, \infty)$
Poisson distribution	$p=1$	$\alpha=-\infty$	$\mathbb{N}_0$	$\mathbb{R}_+$	$\mathbb{R}$
Compound Poisson distributions	$1<p<2$	$\alpha<0$	$\mathbb{R}_+^*$	$\mathbb{R}_+$	$\mathbb{R}_-$
Gamma distribution	$p=2$	$\alpha=0$	$\mathbb{R}_+$	$\mathbb{R}_+$	$\mathbb{R}_+$
Generated by positive stable distributions	$2<p<3$	$0<\alpha<\frac{1}{2}$	$\mathbb{R}_+$	$\mathbb{R}_+$	$(-\infty, 0]$
Inverse Gaussian distribution	$p=3$	$\alpha=\frac{1}{2}$	$\mathbb{R}_+$	$\mathbb{R}_+$	$(-\infty, 0]$
Generated by positive stable distributions	$p>3$	$\frac{1}{2}<\alpha<1$	$\mathbb{R}_+$	$\mathbb{R}_+$	$(-\infty, 0]$

*Continuous for $y>0$, with an atom at $y=0$.

3.2. *Reproductive Exponential Families*

Three of the most important exponential dispersion models with power variance functions, the normal, the inverse Gaussian and the gamma distribution, are also *reproductive exponential families* (Barndorff-Nielsen and Blaesild, 1983, 1984). Here we follow the terminology of the second paper, and speak of reproductive, rather than strongly reproductive, exponential families.

Reproductive exponential families are exponential families which satisfy a certain convolution property, similar to (2.10). By Theorem (3.2) of Barndorff-Nielsen and Blaesild (1983), a reproductive exponential family with canonical statistic of the form $(y, t(y))$, with $t(y)$ one-dimensional, is of the form

$$f(y;\, \lambda,\, \theta)=b(y)e^{\lambda\{y^{\mathrm{T}}\theta-\kappa(\theta)-t(y)\}-G(\lambda)},$$

where $t(y)=\sup_{\theta}\{y^{\mathrm{T}}\theta-\kappa(\theta)\}$. Blaesild and Jensen (1985) showed that the above three distributions are the only univariate reproductive exponential families of this form. Table 2 summarizes these distributions, complementing Table 1.

In higher dimensions, the number of reproductive exponential families proliferate, and Barndorff-Nielsen and Blaesild (1983, 1986) showed how to combine reproductive exponential families to obtain new reproductive exponential families. A similar technique for combining exponential dispersion models is introduced in Section 5.

4. ASYMPTOTIC THEORY

4.1. *Basic Results*

We now consider inference on the parameters of an exponential dispersion model. Although there exist exact results in certain cases, the only viable general approach (at the moment) is via asymptotic theory. We develop what has come to be known as *analysis of deviance* techniques, which are fairly direct analogues of techniques from analysis of variance for normal

TABLE 2
Reproductive exponential families

Name	*Symbol*	$t(y)$	$b(y)$	$\kappa(\theta)$	$e^{-G(\lambda)}$
Normal	$N(\mu, \sigma^2)$	$\frac{1}{2}y^2$	1	$\frac{1}{2}\theta^2$	$(\lambda/2\pi)^{1/2}$
Gamma	$\Gamma(\lambda, \mu)$	$-\log y$	y^{-1}	$-\log(-\theta)$	$\lambda^{\lambda}/\Gamma(\lambda)$
Inverse Gaussian	$IG(\mu, \sigma^2)$	$\frac{1}{2}y^{-1}$	$y^{-3/2}$	$-(-2\theta)^{1/2}$	$(\lambda/2\pi)^{1/2}$

models. The analysis of deviance for generalized linear models dates back to Nelder and Wedderburn (1972) and Baker and Nelder (1978).

We assume that the k-dimensional vector y consists of independent component vectors $y_1, \ldots, y_n$ with distributions $y_i \sim ED_i(\mu_i, \sigma^2/w_i)$, where $w_1, \ldots, w_n$ are known weights, and $ED_1, \ldots, ED_n$ are suitable exponential dispersion models. The joint distribution of $y_1, \ldots, y_n$ is then an exponential dispersion model, which we denote $y \sim ED(\mu, \sigma^2)$. Note that k is effectively proportional to the sample size. We consider inference on μ, with σ^2 either known, or an unknown nuisance parameter.

We consider three nested hypotheses:

H_0: $\mu \in \Omega$ (the saturated hypothesis)

H_1: $\mu = \mu(\beta)$

H_2: $\mu = \mu(\beta)$ and $\beta = \beta(\gamma)$,

of dimensions k, p_1 and p_2, respectively, where $k > p_1 > p_2$.

The *deviance* for μ, for the observation y, is defined by

$$D(y;\, \mu) = 2[\sup_{\theta \in \Theta}\{y^{\mathrm{T}}\theta - \kappa(\theta)\} - \{y^{\mathrm{T}}\tau^{-1}(\mu) - \kappa(\tau^{-1}(\mu))\}].$$

We define the deviance for the hypothesis H_1 by $D(y;\, \hat{\beta}) = D(y;\, \mu(\hat{\beta}))$, where $\hat{\beta}$ is the maximum likelihood estimate for β under H_1, and similarly, $D(y;\, \hat{\gamma}) = D(y;\, \mu(\beta(\hat{\gamma})))$ denotes the deviance for H_2. Note that the maximum likelihood estimate $\hat{\beta}$ does not depend on the value of σ^2, and that $\hat{\beta}$ may be obtained by minimizing the deviance $D(y;\, \mu(\cdot))$. The statistic $\lambda D(y;\, \hat{\beta})$ is the log-likelihood ratio test for H_1 under H_0, and $\lambda\{D(y;\, \hat{\gamma}) - D(y;\, \hat{\beta})\}$ is the log-likelihood ratio test for H_2 under H_1. We let $\lambda i(\beta) = \lambda X(\beta)^{\mathrm{T}} V(\mu(\beta))^{-1} X(\beta)$ denote the expected information matrix for β, where $X(\beta) = \partial\mu/\partial\beta^{\mathrm{T}}$ is the local design matrix.

We distinguish between two types of asymptotic results namely *large-sample results*, which apply for n tending to infinity, and *small-dispersion results*, which apply for λ tending to infinity. The argument for small-dispersion asymptotics is based on formula (2.10), by which we may, for λ large and known, think of y as the sufficient statistic for a large sample from an exponential dispersion model. This implies that the standard large-sample results apply for λ large, as well as for n large. Under circumstances to be discussed below, we thus have the following approximate distribution results

$$\sqrt{(n\lambda)}(\hat{\beta} - \beta) \sim N_{p_1}(0,\, ni(\beta)^{-1}) \tag{4.1}$$

$$\lambda D(y;\, \hat{\beta}) \sim \chi^2(k - p_1) \tag{4.2}$$

$$\lambda\{D(y;\, \hat{\gamma}) - D(y;\, \hat{\beta})\} \sim \chi^2(p_1 - p_2), \tag{4.3}$$

where $\chi^2(d)$ is the chi-squared distribution on d degrees of freedom.

All three statements (4.1) – (4.3) hold in the small-dispersion case, provided H_1 and H_2 are sufficiently smooth. However, (4.2) does not apply in the large-sample case, because the dimension of the saturated hypothesis H_0 depends on the sample size n. The results (4.1) and (4.3) hold for n large under the usual regularity conditions. In particular, the matrix $n^{-1}i(\beta)$ must converge to a positive-definite limit and for (4.3), the same must hold for the information matrix for γ. We do not go into details about regularity conditions here, and the reader is referred to Fahrmeier and Kaufmann (1985) and McCullagh (1983) for detailed accounts of large-sample results. Jørgensen (1986b) generalizes the idea of small-dispersion asymptotics to certain distributions outside the family of exponential dispersion models.

The above argument for small-dispersion asymptotics based on (2.10) may also be used in connection with the saddle-point approximation (Barndorff-Nielsen and Cox, 1979) for the distribution of y, because for λ large we may think of y as the sample average based on a large sample from an exponential dispersion model. If $\tilde{\mu}$ denotes the maximum likelihood estimate of μ under H_0, we thus obtain the approximation for the density of $ED(\mu, \sigma^2)$

$$f(y;\, \mu, \sigma^2) \simeq \{(2\pi\sigma^2)^k |V(\tilde{\mu})|\}^{-1/2} \mathrm{e}^{-\lambda D(y;\, \mu)/2}, \tag{4.4}$$

where, for steep models, we have $\tilde{\mu} = y$ (cf. Jørgensen (1986a) for the definition of steepness for exponential dispersion models). All the models considered in the paper are steep, except for the models $ED^{(p)}(\mu, \sigma^2)$ with power variance function, for $p < 0$. The saddle-point aproximation applies to continuous and discrete exponential dispersion models. In the latter case, (4.4) applies to the distribution of $y = x/\lambda$, where $x \sim ED^*(\lambda, \theta)$.

4.2 *Inference, λ known*

For λ known, formulae (4.1), (4.2) and (4.3) give the small-dispersion asymptotic distribution of respectively the maximum likelihood estimate, the deviance goodness-of-fit statistic for H_1, and the likelihood ratio test for H_2 under H_1. For a discrete exponential dispersion model $x \sim ED^*(\lambda, \theta)$, these results apply for the distribution of $y = x/\lambda$, and so may be used for inference on μ for λ known. As a special case we obtain the standard asymptotic results for contingency tables based on multinomial sampling, because a large value of λ corresponds to large expected cell frequencies $m = \lambda\mu$. In this case, λ is the number of trials, and is hence effectively the sample size. For Poisson sampling, the small-dispersion results are also equivalent to the standard results, because by (2.16), λ large means large expected cell frequencies.

As mentioned above, (4.2) does not hold in the large-sample case. Hence, (4.2) should in general be used with care, and often an inspection of residuals is preferable to a formal goodness-of-fit test based on (4.2).

4.3. *Inference, λ Unknown*

We first discuss small-dispersion results, and later discuss more general results.

An estimate of σ^2 under H_1 may be obtained from the asymptotic distribution in (4.2), namely

$$\tilde{\sigma}^2 = D(y;\, \hat{\beta})/(k - p_1). \tag{4.5}$$

Note that for $\Lambda \neq \mathbb{R}_+$ we may have to modify $\tilde{\sigma}^2$ in order to obtain an estimate of λ which belongs to Λ. Similar remarks apply to the estimates of σ^2 considered below.

By inserting $\tilde{\sigma}^2$ for λ^{-1} in (4.1) and using the asymptotic independence of (4.1) and (4.2) under H_1 for λ large, we may obtain asymptotic confidence intervals for the components of β based on the t-distribution. Similarly, (4.2) and (4.3) are asymptotically independent for large λ, and hence a test for H_2 may be based on convergence in distribution (under H_2) of

$$\tilde{F} = \frac{\{D(y;\, \hat{\gamma}) - D(y;\, \hat{\beta})\}/(p_1 - p_2)}{\tilde{\sigma}^2} \to F(p_1 - p_2,\, k - p_1) \quad \text{as} \quad \lambda \to \infty, \tag{4.6}$$

where $F(d_1, d_2)$ is the F-distribution on (d_1, d_2) degrees of freedom.

Some known results may be obtained as special cases of small-dispersion asymptotics, for example asymptotic results for nonlinear normal models for σ^2 small (Johansen, 1983, Bickel and Doksum, 1981), and results for the gamma distribution (McCullagh and Nelder, 1983, p. 157).

The F- and t-tests defined above show a very direct analogy with normal-theory analysis of variance methods, with the deviance taking the place of the residual sum of squares. However, the estimate $\tilde{\sigma}^2$ in (4.5) is in general not consistent for σ^2 in the large-sample case. In order to obtain results which apply in both the large-sample case and the small-dispersion case, we now develop two alternative estimates.

The first estimate, which was suggested by Wedderburn (1972) and McCullagh (1983), is $\bar{\sigma}^2 = X^2/(k - p_1)$, where $X^2 = (y - \hat{\mu})^{\mathrm{T}} V(\hat{\mu})^{-1}(y - \hat{\mu})$ is the generalized Pearson statistic, and $\hat{\mu} = \mu(\hat{\beta})$. The second is based on the modified profile likelihood for σ^2 (Barndorff-Nielsen, 1983), $L^0(\sigma^2) = \sigma^{p_1} f(y;\, \hat{\mu}, \sigma^2)$, where $f(y;\, \mu, \sigma^2)$ is the density of the distribution $ED(\mu, \sigma^2)$. The estimate $\hat{\sigma}_0^2$ is the value of σ^2 that maximizes $L^0(\sigma^2)$. Like $\tilde{\sigma}^2$, these two estimates depend on the hypothesis for μ under consideration, and the versions above are for the hypothesis H_1.

From the saddle-point approximation (4.4), it follows that $\hat{\sigma}_0^2$ is equivalent to $\tilde{\sigma}^2$ for λ large, and similarly $\bar{\sigma}^2$ is equivalent to $\tilde{\sigma}^2$ for λ large, because $D(y; \hat{\beta})$ and X^2 are asymptotically equivalent for λ large. Both $\hat{\sigma}_0^2$ and $\bar{\sigma}^2$ are consistent estimators of σ^2 in the large-sample case. Hence, we shall replace $\tilde{\sigma}^2$ by either $\bar{\sigma}^2$ or $\hat{\sigma}_0^2$ in the F- and t-tests above. Let $\bar{F}$ and $\hat{F}_0$ denote the modified versions of the F-statistic in (4.6), with $\tilde{\sigma}^2$ replaced by respectively $\bar{\sigma}^2$ or $\hat{\sigma}_0^2$. Then, by the asymptotic equivalence of $\bar{\sigma}^2$ or $\hat{\sigma}_0^2$ with $\tilde{\sigma}^2$, the limit distribution in (4.6) applies to $\hat{F}_0$ and $\bar{F}$ in the small-dispersion case. In the large-sample case, the consistency of $\hat{\sigma}_0^2$, combined with (4.3), imply that $\hat{F}_0$ converges in distribution (under H_2) to $\chi^2(p_1 - p_2)/(p_1 - p_2)$ (and similarly for $\bar{F}$), in agreement with the limit of $F(p_1 - p_2, k - p_1)$ for $k - p_1$ large. Hence, $\hat{F}_0$ and $\bar{F}$ are approximately distributed as $F(p_1 - p_2, k - p_1)$ both for λ large and n large, indicating that this approximation applies under a wide range of conditions. Similar arguments apply to the modified versions of the t-test defined above.

For linear normal models, the three versions of the F- and t-tests are identical, and are exactly F- or t-distributed, respectively. For certain nested classification models for the inverse Gaussian distribution, the F-statistic in (4.6) is exactly F-distributed (Folks and Chhikara (1978), Tweedie (1957)). These seem to be the only cases in which $\tilde{F}$ is exactly F-distributed.

In general the estimate $\hat{\sigma}^2$, being based on the modified profile likelihood, is preferable to the moment estimate $\bar{\sigma}^2$. To illustrate the difference between $\bar{\sigma}^2$ and $\hat{\sigma}_0^2$, note that for the inverse Gaussian distribution, the saddlepoint approximation (4.4) gives the exact distribution of y. Hence $\tilde{\sigma}^2 = \hat{\sigma}_0^2$ in this case, so in the cases where $\tilde{F}$ in (4.6) is exactly F-distributed for the inverse Gaussian distribution, the same is the case for $\hat{F}_0$, but not for $\bar{F}$.

For a discrete exponential dispersion model with λ unknown, we note that the transformation from x to $y = x/\lambda$ involves the unknown parameter λ. Hence, the above results for the case λ unkown do not apply for inference in discrete exponential dispersion models, and we shall not discuss this problem any further here.

4.4. *Amazonian Peasants' Data*

The following data analysis illustrates the asymptotic results developed above.

In a study of the conditions of migrants in the Amazonian area of Brazil, Botelho (1986) followed 210 families from two settlements. We analyse the amount of money (Cr$) spent by a family (per year) on hiring outside labour power, divided by the number of working members of the family (w). The resulting variable is called y, and the variable w is used as weight in the analysis.

As a model, we take a generalized linear model with log-link and power variance function. Taking the weight w into account, the model is thus $y_i \sim ED^{(p)}(\mu_i, \sigma^2/w_i)$ for the ith family, where $\mu_i = \exp(x_i^{\mathrm{T}}\beta)$, with x_i a vector of explanatory variables.

In Section 5 we show that the model $ED^{(p)}(\mu, \sigma^2)$ with $1 < p < 2$ is a compound Poisson distribution. In particular, the model has a positive probability for a zero observation, a feature which is clearly present in the data, with 148 of the 210 families having no use of outside labour power. In the analysis we used the value $p = 1.75$. This value was chosen fairly arbitrarily in the interval $1 < p < 2$, but with an aim of obtaining a nearly symmetric distribution of the residuals from the fit.

We first eliminated a number of x-variables from the model, based on t-tests, using the estimate $\bar{\sigma}^2$ for σ^2. Table 3 shows parameter estimates for the final model, with standard errors based on $\bar{\sigma}^2$. We used *GLIM* for the computations.

The estimates in Table 3 show that family labour power has a significant negative effect on the use of outside labour power. Thus, families with many working members tend to hire less outside labour power per working family member. Similarly, gross income has a positive effect, and wages received has a negative effect on the use of outside labour power. The family size (number of working and non-working members) seems to have opposite effects in the two settlements, although the effects are hardly significant.

For the qualitative variable debt/no debt, there seems to be a significant difference between

TABLE 3
Parameter estimates for the Amazonian peasants' data

Variable	*Coefficient*	*s.e.*
Gross income	0.0030	0.0004
Wages received	−0.022	0.004
Family labour power	−0.0060	0.0007
Family size (settlement 1)	0.17	0.08
Family size (settlement 2)	−0.21	0.18
No debt (settlement 1)	−5.3	0.3
No debt (settlement 2)	−5.7	0.3
Debt (settlement 1)	−27	12
Debt (settlement 2)	−6.9	0.8
$\bar{\sigma}^2$	1199 (d.f. = 201)	
$\tilde{\sigma}^2$	990	

the two settlements, in that families in settlement 1 with debt hardly use any outside labour power at all, compared with other families in either settlement. For the quantitative variables, the mean of the variable was subtracted in each group, and hence the conclusion holds for an "average" family in the settlement. This difference between settlements indicates a more dramatic economic differentiation (in terms of the use of outside labour power) in settlement 1, which consists of spontaneous settlers, compared with settlement 2, which was sponsored by the federal government. An overall comparison of the two settlements based on an F-test ($\bar{F} = 3.73$, d.f. = 3, 201, $P \simeq 1\%$) shows that there is a real difference between the settlements. For comparison, note that $\tilde{F} = 4.52$, which is rather different, although in the present case $\tilde{F}$ would lead to the same conclusion.

5. COMBINATIONS AND MIXTURES

5.1. *Definition and Examples*

In this section we define two techniques for constructing new exponential dispersion models, called respectively combinations and mixtures. In a combination, we combine a conditional and a marginal distribution, both exponential dispersion models, to obtain a higher-dimensional exponential dispersion model. A mixture is obtained if we marginalize to the first component of a combination. We use these techniques to obtain new models for two important practical problems involving correlated data, namely longitudinal data and variance component models.

If $y \sim ED(\mu, \sigma^2)$ denotes a given exponential dispersion model, we use the notation $x \sim ED^*(\lambda, \theta)$ in the following to denote the distribution of $x = \lambda y$. In particular, $ED^*(\lambda, \theta)$ may denote a discrete exponential dispersion model, as in Section 2.4.

Let ED_1 and ED_2 denote two exponential dispersion models, and let κ_i, τ_i, Λ_i and k_i, $i = 1, 2$, denote the cumulant generator, mean value mapping, index set and dimension of ED_1 and ED_2, respectively. In the ED^*-notation introduced above, suppose that

$$x_1 | x_2 \sim ED_1^*(\lambda r + x_2^{\mathrm{T}} q, \theta_1),\ x_2 \sim ED_2^*(\lambda, \theta_2'), \tag{5.1}$$

where $r \in \mathbb{R}$ and $q \in \mathbb{R}^{k_2}$ are constants such that $\lambda r + x_2^{\mathrm{T}} q \in \Lambda_1 \cup \{0\}$. Here we interpret $ED_1^*(0, \theta_1)$ as a degenerate distribution at zero. The joint probability density function of $(y_1, y_2) = (x_1/\lambda, x_2/\lambda)$ is of the form

$$f(y_1, y_2; \lambda, \theta_1, \theta_2) = a(\lambda, y_1, y_2)\exp[\lambda\{y_1^{\mathrm{T}}\theta_1 + y_2^{\mathrm{T}}\theta_2 - \kappa_{12}(\theta_1, \theta_2)\}], \tag{5.2}$$

where

$$\theta_2 = \theta'_2 - q\kappa_1(\theta_1)$$

and

$$\kappa_{12}(\theta_1, \theta_2) = r\kappa_1(\theta_1) + \kappa_2(\theta_2 + q\kappa_1(\theta_1)). \tag{5.3}$$

The distribution (5.2) is an exponential dispersion model of dimension $k_1 + k_2$, called the *combination* of ED_1 and ED_2, and we denote (5.2) by

$$(y_1, y_2) \sim ED_1 \times ED_2(\mu_1, \mu_2, \sigma^2).$$

Here $\mu_2 = \tau_2(\theta_2')$ and $\mu_1 = \tau_1(\theta_1)(r + \mu_2^T q)$ are the expected values of y_2 and y_1, respectively, and $\sigma^2 = 1/\lambda$. The two components ED_1 and ED_2 are called respectively the *kernel distribution* and the *mixing distribution* of the combination. The marginal distribution of y_1 is called the *mixture distribution* associated with the combination (5.2), and is denoted by

$$y_1 \sim ED_1 \wedge ED_2(\mu_1, \mu_2, \sigma^2). \tag{5.4}$$

By (5.2), the mixture distribution (5.4) is an exponential dispersion model for θ_2 known. If the same exponential dispersion model is obtained for any value of θ_2, the mixture, and the corresponding combination, are called *simple*.

The distribution of (x_1, x_2) is denoted by

$$(x_1, x_2) \sim ED_1^* \times ED_2^*(\lambda, \theta_1, \theta_2). \tag{5.5}$$

If ED_1^* and ED_2^* are discrete exponential dispersion models, (5.5) is a discrete exponential dispersion model, and is called the *discrete combination* of ED_1^* and ED_2^*. Similarly, if ED_1^* is a discrete exponential dispersion model, then x_1 follows a discrete exponential dispersion model, called a *discrete mixture*, which we denote by $x_1 \sim ED_1^* \wedge ED_2(\lambda, \theta_1, \theta_2)$.

The combination (5.2) and the mixture (5.4) etc. depend on the constants r and q. The simplest case is obtained for $r = 0$ and $q = (1, \ldots, 1)^T$, with the components of x_2 positive, and unless otherwise stated, we assume this to be the case.

The above techniques generalize and unify certain well-known combination, mixing and compounding techniques, and many relations between standard distributions correspond to combinations and mixtures. A number of examples are summarized in Table 4, using the notations of Section 2.4 and 3.2. In the table, $Bi(\lambda, \mu)$ and $Nb(\lambda, \theta)$ denote respectively the binomial and the negative binomial distributions. The symbol $x \sim l(\lambda, \theta)$ denotes the discrete exponential dispersion model generated by the logarithmic distribution, cf. Jørgensen (1986a). Note that if $ED_2^*(\lambda, \theta)$ is a discrete exponential dispersion model, we write $ED_1 \wedge ED_2^*$ instead of $ED_1 \wedge ED_2$ etc., to avoid using two notations for the standard discrete distributions. For example, by $\Gamma \wedge Po$ we mean $ED^{(2)} \wedge ED^{(1)}$, in the notation of Section 3.

TABLE 4
Examples of mixtures and combinations

Symbol	*Name or reference*
$N \wedge IG$	Generalized hyperbolic distribution,
$N \wedge \Gamma$	Barndorff-Nielsen (1978b)
$\Gamma \wedge Po$	Compound Poisson distribution
$Po \wedge \Gamma = Nb$	Negative binomial distribution
$Po \wedge IG$	Holla (1966), Sankaran (1968), Sichel (1971).
$Po \wedge Po$	Neyman Type A, Johnson and Kotz (1969, p. 216)
$Bi_k \times Po = Po^{k+1}$	Poisson trick
$Bi \times Bi_k = Bi_{k+1}$	Barndorff-Nielsen (1978a, p. 26)
$l \wedge Po = Nb$	Barndorff-Nielsen (1978a, p. 127)
$Bi \wedge Nb = Nb$	Cacoullos and Papageorgiou (1981)
$\Gamma \wedge Nb = \Gamma$	Barndorff-Nielsen (1978a, p. 202)

Consider the combination $ED^{(p_1)} \times ED^{(p_2)}(\mu_1, \mu_2, \sigma^2)$ of two exponential dispersion models with power variance functions. By (3.7) we may in this case express the definition (5.1) directly in terms of y_1 and y_2 as

$$y_1 \mid y_2 \sim ED^{(p_1)}(\mu'_1(r + y_2 q), \sigma^2(r + y_2 q)^{1-p_1}) \tag{5.6}$$

$$y_2 \sim ED^{(p_2)}(\mu_2, \sigma^2), \tag{5.7}$$

where $\mu'_1 = \tau_1(\theta_1)$. The definition requires $p_2 \geqslant 1$ (making y_2 positive) and $r, q \geqslant 0$.

Example 5.1. The combination $\Gamma \times Po$ (taking $p_1 = 2$ and $p_2 = 1$ in (5.6) and (5.7)), with $r = 0$, and q arbitrary, corresponds to the cumulant generator

$$\kappa_{12}(\theta_1, \theta_2) = e^{\theta_2}(-\theta_1)^{-q}. \tag{5.8}$$

This is a bivariate distribution of mixed type, and the mixture $\Gamma \wedge Po$, which is a compound Poisson distribution, is positive and continuous with a positive probability mass at zero. The corresponding cumulant generator $\kappa_{12}(\cdot, \theta_2)$ essentially corresponds to (3.4) with $\alpha = -q$. We have thus obtained the exponential dispersion models with power variance function, with power $p = (q+2)/(q+1)$ in the interval $1 < p < 2$. The form of (5.8) shows that this mixture is simple. The special case $q = 1$ corresponds to the noncentral chi-squared distribution with zero degrees of freedom, cf. Siegel (1985) and references therein. □

Several factorizations related to the multinomial, Poisson and negative multinomial distribution correspond to discrete combinations of the form (5.5). For example, the so-called "Poisson trick", which relates a multinomial distribution with a set of independent Poisson distributions, corresponds to the combination $Bi_k \times Po$. A second example is

Example 5.2. Let x_1 be a one-dimensional and x_2 a k-dimensional vector, and assume

$$x_1 \mid x_2 \sim \mathrm{Bi}(n + x_2^{\mathrm{T}} q, \mu), \quad x_2 \sim \mathrm{Bi}_k(n, \mu_1, \ldots, \mu_k),$$

where $q = (-1, \ldots, -1)^{\mathrm{T}}$, defining a discrete combination with $r = 1$. This results in a multinomial distribution, $(x_1, x_2) \sim \mathrm{Bi}_{k+1}(n, \mu', \mu_1, \mu_2, \ldots, \mu_k)$, where $\mu' = \mu(1 - \mu_1 - \ldots - \mu_k)$. The corresponding discrete mixture is the binomial distribution $\mathrm{Bi}(n, \mu')$. □

Further examples of mixtures and combinations, some of which are indicated in Table 4, may be found in Barndoff-Nielsen (1978a) and Cacoullos and Papageorgiou (1981).

5.2 *Longitudinal data*

Combinations and mixtures may be employed to construct new probability models for correlated data. In the present section we introduce a model for longitudinal data, and analyse a set of data.

In physiological research, subjects doing physical exercise were instructed to express their RPE (rate of perceived exertion) on a scale from 6 to 20. In an experiment involving sustained exercise, the subjects were asked to state their RPE at 6, 12, 20 and 40 minutes after the start of the exercise period, or until exhaustion occurred. The experiment involved 6 female and 7 male subjects, see Pedersen and Froberg (1985) for further details.

We shall use an autoregressive model of order 1 for these data, with gamma conditional distributions. Letting $y_1, \ldots, y_n$ denote the RPEs for a given subject at times $1, \ldots, n$, we build up a combination iteratively from (5.6) and (5.7) with $p_1 = p_2 = 2$ and $r = 0$, which gives the model

$$y_1 \sim ED^{(2)}(\mu_1, \sigma^2),\ y_i \mid y_{i-1} \sim ED^{(2)}(\mu'_i y_{i-1} q, \sigma^2(y_{i-1} q)^{-1}), \quad i = 2, \ldots, n.$$

The restriction of the data to the interval [6, 20] makes it difficult to find a suitable model, and the gamma model is at best a rough approximation. However, if the coefficient of variation, σ, is small, the gamma distribution is nearly normal, and is concentrated in an interval around the mean.

The mean values $\mu_i = E(y_i)$, $i = 1, \ldots, n$, satisfy $\mu_i = \mu_i' q \mu_{i-1}$, $i = 2, \ldots, n$. Assuming a log-linear model for $\mu_1, \ldots, \mu_n$, $\log \mu_i = z_i^{\mathrm{T}} \beta$, $i = 1, \ldots, n$, we get the following model for the conditional mean values,

$$\log \mu_1 = z_1^{\mathrm{T}} \beta \tag{5.9}$$

$$\log(\mu_i' q y_{i-1}) = \log y_{i-1} + (z_i - z_{i-1})^{\mathrm{T}} \beta, \quad i = 2, \ldots, n. \tag{5.10}$$

Taking q as known, the model has the same likelihood as a generalized linear model with gamma errors, log link, offset vector $(0, \log y_1, \ldots, \log y_{n-1})^{\mathrm{T}}$, weights $1, qy_1, \ldots, qy_{n-1}$ and design matrix with rows z_1^{T} and $(z_i - z_{i-1})^{\mathrm{T}}$, $i = 2, \ldots, n$. Assuming that the subjects are independent, we obtain the full model for the data. The model may be fitted using *GLIM*, but the standard errors for the estimates produced by GLIM are not quite correct, because the weights and the offset are stochastic.

To estimate q, we fit separate models for the marginal distribution of y_1 and the conditional distribution of $y_2, \ldots, y_n$ given y_1. We then use the ratio of the dispersion parameter estimates for the two models to estimate q. By (5.10), the conditional distribution of $y_2, \ldots, y_n$ given y_1 does not involve the constant vector, which we assume is in the model. Hence we base the inference on the slope parameters on this conditional distribution. Table 5 shows an analysis of deviance table for the slope of the regression on log-time (LOG(T)). The models are written in *GLIM* notation, with factors sex and subject.

TABLE 5
Analysis of deviance for RPE data (slopes)

Model	*Deviance*	*d.f.*	$\tilde{\sigma}^2/\tilde{q}$	$\tilde{F}$	*d.f.*
SUBJECT.LOG(T)	1.317	20	0.0659		
SEX.LOG(T)	1.772	31	0.0572	0.63	11/20
LOG(T)	1.877	32	0.0587	1.84	1/31

Since the estimate for σ^2 is small (cf. Table 6), we use small-dispersion asymptotics. The F-tests in Table 5 indicate no differences between subjects within each sex, or between sexes. Hence we accept the model with a common slope for all the subjects.

By (5.9), the marginal distribution of y_1 depends on both the slope and the intercept for the regression, but we take the slope as known and equal to the estimate from the final model from the above analysis. Table 6 shows the analysis of deviance for this model. Assuming identical intercepts for the subjects within each sex, the F-test shows a difference between the sexes, significant at the 2.5% level.

Table 7 shows the estimates for the final model. According to this model, log (RPE) increases linearly with the logarithm of time, and with the same progression (slope) for both sexes, but with intercept depending on sex.

TABLE 6
Analysis of deviance for RPE data (intercepts)

Model	*Deviance*	*d.f*	$\tilde{\sigma}^2$	$\tilde{F}$	*d.f.*
SEX	0.2050	11	0.0186		
%GM	0.3665	12	0.305	8.67	1/11

TABLE 7
Parameter estimates for RPE data

Parameter	*Estimate*	*s.e.*
Slope	0.0509	0.0052
Intercept (women)	2.260	0.056
Intercept (men)	2.485	0.052
$\tilde{\sigma}^2 = 0.0186$	$\tilde{q} = 0.318$	

5.3 *Variance component models*

Using the idea of mixtures of exponential dispersion models, we now introduce a class of variance component models. Let ED_1 be a given exponential dispersion model, and let ED_1^k denote the exponential dispersion model corresponding to the joint distribution of k independent variables $u_1, \ldots, u_k$, where $u_j \sim ED_1(\mu_j, \sigma^2)$, $j = 1, \ldots, k$. Let $y_1, \ldots, y_n$ be independent vectors distributed as

$$y_i \sim ED_1^k \wedge ED_2(\mu_{i1}, \ldots, \mu_{ik}, \mu_{k+1}, \sigma^2),\ i = 1, \ldots, n \tag{5.11}$$

where $ED_2(\mu_{k+1}, \sigma^2)$ is a second exponential dispersion model such that the mixture (5.11) exists for suitable r and q.

Since the components of the vector y_i are correlated, we have a variance component model with blocks $y_1, \ldots, y_n$, and a between-block and a within-block variance component. By iterating the process we may specify any number of components for nested structures. For example, the model $y_i \sim (ED_1^k \wedge ED_2)^l \wedge ED_3$, $i = 1, \ldots, n$, where ED_3 is a third exponential dispersion model, corresponds to three variance components.

To examine the correlation structure of (5.11), we consider the simplest case where both ED_1 and ED_2 are one-dimensional. Let $z_i \sim ED_2(\mu_{k+1}, \sigma^2)$ represent the mixing distribution of the mixture (5.11). The conditional mean and variance of y_{ij} given z_i are

$$E(y_{ij}|z_i) = \mu_{ij}(r + qz_i)/g_{k+1},\ \mathrm{Var}(y_{ij}|z_i) = \sigma^2(r + qz_i)V_1(\mu_{ij}/g_{k+1}),$$

where $g_{k+1} = r + q\mu_{k+1}$, and V_i denotes the variance function of ED_i, $i = 1, 2$. The model thus resembles the variance component model suggested by McCullagh and Nelder (1983, p. 225). The cumulant generator of y_i is

$$\kappa_{12}(\theta_1, \ldots, \theta_k, \theta_{k+1}) = r\sum_{j=1}^{k}\kappa_1(\theta_j) + \kappa_2\left(\theta_{k+1} + q\sum_{j=1}^{k}\kappa_1(\theta_j)\right),$$

regarded as a function of $\theta_1, \ldots, \theta_k$ with θ_{k+1} fixed. Differentiating twice with respect to the θ's we obtain

$$\mathrm{Var}(y_{ij}) = \sigma^2\{V_1(\mu_{ij}/g_{k+1})g_{k+1} + V_2(\mu_{k+1})(\mu_{ij}q/g_{k+1})^2\} \tag{5.12}$$

$$\mathrm{Cov}(y_{ij}, y_{il}) = \sigma^2 V_2(\mu_{k+1})\mu_{ij}\mu_{il}(q/g_{k+1})^2,\quad j \neq l. \tag{5.13}$$

The factor $\mu_{ij}\mu_{il}$ thus determines the sign of the correlation between y_{ij} and y_{il}, and in particular, the correlation is positive for $\mu_{ij} = \mu_{il}$.

To represent crossed structures, we proceed in a slightly different way. Let $x_1, \ldots, x_l, z_1, \ldots, z_m$ be independent and distributed as $x_i \sim ED_1(\mu_1, \sigma^2)$, $i = 1, \ldots, l$ and $z_j \sim ED_2(\mu_2, \sigma^2)$, $j = 1, 2, \ldots, m$. Suppose that y_{ijk}, $i = 1, \ldots, l$, $j = 1, \ldots, m$, $k = 1, \ldots, t_{ij}$ are conditionally independent given the x's and z's. Assume that

$$\lambda y_{ijk}|\lambda(x_1, \ldots, x_l, z_1, \ldots, z_m) \sim ED_3^*(\lambda r + q_i x_i + p_j z_j, \theta_{ijk}), \tag{5.14}$$

where $r, q_1, \ldots, q_l, p_1, \ldots, p_m$ are constants. The joint marginal distribution of the y_{ijk}'s is a mixture distribution of a sort which is slightly more general than (5.4). This distribution represents a two-way model in which each of the effects i and j are random. By combining structures such as (5.11) and (5.14) it is possible to represent a considerable variety of variance component models.

For discrete data, we may obtain variance component models by using discrete mixture distributions instead of (5.11) and (5.14) etc. For example, Manton, Stallard and Woodbury (1981) suggested the model $Po^n \wedge \Gamma$, which is the negative multinomial distribution, as a model for variance components in contingency tables.

Let us consider the special case of (5.11) where both ED_1 and ED_2 are gamma distributions. In this case (5.12) and (5.13) turn into

$$\mathrm{Var}(y_{ij}) = \sigma_0^2 \mu_{ij}^2 \tag{5.15}$$

$$\mathrm{Cov}(y_{ij}, y_{il}) = \sigma_0^2 \mu_{ij} \mu_{il} \rho, \; j \neq l, \tag{5.16}$$

where $\sigma_0^2 = \sigma^2/\rho$ and $\rho = q\mu_{k+1}/(1 + q\mu_{k+1})$ is the correlation between y_{ij} and y_{il}. The variance function (5.15) for y_{ij} is the same as for a gamma distribution but the marginal distribution of y_{ij} is not a gamma distribution.

To illustrate this model, we apply it to the eelworm data of Cochran and Cox (1957, p. 46). The data come from a randomized block design with $n = 4$ blocks and $k = 12$ plots in each block. The treatments are four soil fumigants (single and double dose) and control (4 replications in each block). The data are the number of eelworm cysts in a sample of 400 grams of soil from each plot.

In the notation of (5.15) and (5.16), $i = 1, \ldots, 4$ denotes blocks and $j = 1, \ldots, 12$ denotes plots. We need to estimate the treatment means $\mu_1, \ldots, \mu_9$, the dispersion parameter σ_0^2 and the intra-block correlation ρ, where $0 \leqslant \rho < 1$. For fixed ρ, the treatment effects may be estimated by iterative weighted least squares (see for example Jørgensen, 1983), with inverse weight matrix given by (5.15) and (5.16).

From (5.15) and (5.16) we get

$$\mathrm{Var}\left(\sum_{j=1}^{k} y_{ij}/\mu_{ij}\right) = \sigma^2\{k + \rho k(k-1)\}. \tag{5.17}$$

Replacing the left-hand side of (5.17) by the estimate

$$\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{k} y_{ij}/\hat{\mu}_{ij} - k\right)^2$$

we obtain an estimate for ρ by solving (5.17) for given fitted values $\hat{\mu}_{ij}$. By updating the estimate for ρ in each cycle of the weighted least squares iteration, we obtain final estimates for the parameters. For the present model, however, the procedure is somewhat simpler, because for any value of ρ, the estimates of the treatment means are given by the sample treatment averages. The estimates are given in Table 8.

TABLE 8
Parameter estimates for eelworm data

Treatment	*Estimate*	*s.e.*
Control	366	95
1CK	223	74
1CN	267	88
1CM	358	118
1CS	232	77
2CK	281	93
2CN	316	105
2CM	310	103
2CS	219	73
$\bar{\sigma}_0^2 = 0.280$	$\bar{\rho} = 0.187$	d.f. = 39

This method of estimation does not produce a standard error for the estimate $\bar{\rho}$, but the value $\bar{\rho} = 0.187$ indicates a positive intra-block correlation. This corresponds to the significant block effect found in the analysis of variance (cf. Cochran and Cox, 1957, p. 57).

6. CONCLUSIONS

The paper was opened by a call for simplicity in statistical models, and we have tried to make a case for exponential dispersion models in this respect. For example, the analogy with the normal distributions gives a familiar ring to many of the results. Since the continuous and discrete exponential dispersion models include a comprehensive range of standard models, we have obtained an important unification and generalization of some previously unrelated results. However, if the idea is pushed to its limits, complications arise, such as the non-existence of proper exponential dispersion models for discrete data. The problems with the models for correlated data also illustrate this point.

The main problem with exponential dispersion models for correlated data is that the correlation structure is fixed, in the sense that the correlation depends on μ and σ^2 through μ only. If a more flexible correlation structure is needed, we move outside the exponential dispersion model framework, and this generally leads to complications. A general way to handle extra parameters in an exponential dispersion model is Nelder and Pregibon's *extended quasi-likelihood* (cf. Nelder, 1985). This method uses the saddle-point approximation (4.4) as an approximate likelihood, but the properties of this approach are not yet fully developed.

ACKNOWLEDGEMENTS

I am grateful to Alex Whitmore and Chris Small for helpful discussions during the preparation of the paper, and to the referees for constructive comments on earlier versions of the paper.

REFERENCES

Baker, R. J. and Nelder, J. A. (1978) *The GLIM System, Release 3*, Oxford: Numerical Algorithms Group.

Bar-Lev, S. K. and Enis, P. (1985) Reproducability and natural exponential families with power variance functions. Research Report 89, Department of Statistics, State University of New York at Buffalo.

Barndorff-Nielsen, O. (1978a) *Information and Exponential Families in Statistical Theory*. Chichester: Wiley.

Barndorff-Nielsen, O. (1978b) Hyperbolic distributions and distributions on hyperbolae. *Scand. J. Statist.*, **5**, 151–157.

Barndorff-Nielsen, O. (1983) On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.

Barndorff-Nielsen, O. and Blaesild, P. (1983) Reproductive exponential families. *Ann. Statist.*, **11**, 770–782.

Barndorff-Nielsen, O. and Blaesild, P. (1984). Reproductive models. In *Encyclopedia of Statistical Sciences* (S. Kotz and N. L. Johnson eds.) New York: Wiley (to appear).

Barndorff-Nielsen, O. Blaesild, P. (1986). Combination of reproductive models. To appear in *Annals of Statistics*.

Barndorff-Nielsen, O. and Cox, D. R. (1979). Edgeworth and saddlepoint approximations with statistical applications. *J. R. Statist. Soc.* **B**, **41**, 279–312.

Bickel, P. J. and Doksum, K. A. (1981) An analysis of transformations revisited. *J. Amer. Statist. Assoc.*, **76**, 296–311.

Blaesild, P. and Jensen, J. L. (1985) Saddlepoint formulas for reproductive exponential models. *Scand. J. Statist.*, **12**, 193–202.

Botelho, V. L. (1986) Social differentiation among peasants in the Brazilian Amazon frontier. Ph.D. thesis, University of London (in preparation).

Cacoullos, T. and Papageorgiou, H. (1981) On bivariate discrete distributions generated by compounding. In *Statistical Distributions in Scientific Work*, Vol. 4 (C. Taillie, G. P. Patil and B. A. Baldessari, eds.), pp. 197–212. Dordrecht-Holland: D. Reidel.

Cochran, W. G. and Cox, G. M. (1957). *Experimental designs* (2nd ed.). New York: Wiley.

Coe, R. and Stern, R. D. (1984) A model fitting analysis of daily rainfall data. *J. R. Statist. Soc.* **A, 147**, 1–34.

Eaton, M. L., Morris, C. and Rubin, H. (1971). On extreme stable laws and some applications. *J. Appl. Prob.*, **8**, 794–801.

Fahrmeir, L. and Kaufmann, H. (1985) Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.*, **13**, 342–368.

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, Vol. II (2nd ed.). New York: Wiley.

Folks, J. L. and Chhikara, R. S. (1978) The inverse Gaussian distribution and its statistical application – a review. *J. R. Statist. Soc.* **B**, **40**, 263–289.

Holla, M. S. (1966) On a Poisson-inverse Gaussian distribution. *Metrika*, **11**, 115–121.

Hougaard, P. (1986) Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, **73**, 387–396.
Johansen, S. (1983) Some topics in regression. *Scand. J. Statist.*, **10**, 161–194.
Johnson, N. L. and Kotz, S. (1969) *Discrete Distributions*. Boston: Houghton Mifflin.
Jørgensen, B. (1983) Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models. *Biometrika*, **70**, 19–28.
Jørgensen, B. (1986a) Some properties of exponential dispersion models. *Scand. J. Statist.*, **13**, 187–198.
Jørgensen, B. (1986b) Small-dispersion asymptotics (in preparation).
Kemp, C. D. and Kemp, A. W. (1965) Some properties of the Hermite distribution. *Biometrika*, **52**, 381–394.
Manton, K. G., Stallard, E. and Woodbury, M. A. (1981) A variance components approach to categorical data models with heterogeneous cell populations: Analysis of spatial gradients in lung cancer motality rates in North Carolina counties. *Biometrics*, **37**, 259–269.
McCullagh, P. (1980) Regression models for ordinal data. *J. R. Statist. Soc.* B, **42**, 109–142.
McCullagh, P. (1983) Quasi-likelihood functions. *Ann. Statist.*, **11**, 59–67.
McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models*. London: Chapman and Hall.
Morris, C. N. (1981) Models for positive data with good convolution properties. Memo no. 8949, Rand Corporation, Santa Monica, California.
Morris, C. N. (1982) Natural exponential families with quadratic variance functions. *Ann. Statist.*, **10**, 65–80.
Nelder, J. A. (1985) Quasi-likelihood and *GLIM*. In *Generalized Linear Models* (R. Gilchrist, B. Francis and J. Whittaker eds.), pp. 120–127. Lecture Notes in Statistics 32. Berlin: Springer-Verlag.
Nelder, J. A. and Wedderburn, R. W. M. (1972) Generalized linear models. *J. R. Statist. Soc.* A, **135**, 370–384.
Pedersen, P. K. and Froberg, K. (1985). Heart rate, work load, and perceived exertion in men and women during incremental and sustained exercise. In *Ergonomics International 85* (J. D. Brown, K. Coombes and M. A. Sinclair, eds.), pp. 571–573. London and Philadelphia: Taylor & Francis.
Sankharan, M. (1968) Mixtures by the inverse Gaussian distribution. *Sankhyā*, **30**, 455–458.
Sichel, H. S. (1971 On a family of discrete distributions particularly suited to represent long-tailed frequency data. In *Proceedings of the Third Symposium on Mathematical Statistics* (N. F. Laubscher, ed.) Pretoria: CSIR.
Siegel, A. F. (1985) Modelling data containing exact zeroes using zero degrees of freedom. *J. R. Statist. Soc.* B, **47**, 267–271.
Tweedie, M. C. K. (1947) Functions of a statistical variate with given means, with special reference to Laplacian distributions. *Proc. Cambridge Phil. Soc.*, **49**, 41–49.
Tweedie, M. C. K. (1957) Statistical properties of Inverse Gaussian distributions, I. *Ann. Math. Statist.*, **28**, 362-372.
Tweedie, M. C. K. (1984) An index which distinguishes between some important exponential families. In *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*. (Eds. J. K. Ghosh and J. Roy), pp. 579–604. Calcutta: Indian Statistical Institute.
Wedderburn, R. W. M. (1974) Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, **61**, 439–447.

DISCUSSION OF DR JØRGENSEN'S PAPER

Dr Robert Gilchrist (Polytechnic of North London): It is a great pleasure to propose the vote of thanks to Bent Jørgensen for tonight's paper. This paper is in the best traditions of the Society: theoretical but, not neglecting practical applications. New distributions are introduced which should prove of value to data analysts, particularly those who work in the generalised linear model (glm) framework. Clearly, the paper will be of particular interest to theoreticians but it should certainly pay the practical statistician to follow the ideas through.

It is, of course, true that the basic concept of what the author calls an exponential dispersion model was outlined some forty years ago by Tweedie, but his results have remained little-known to the new breed of glm devotees until very recently. And, when a good idea has lain dormant for so long, it is a pleasure to see it being promulgated, and indeed a pleasure to see Mr Tweedie here present tonight.

One of the delights of the paper for me is its treatment of the case where data are considered to have a variance which is a power p of the mean. The possibilities of using this family within a glm framework have been apparent for a decade or so but the validity of such a model has caused some concern. Indeed, when I looked at this some years ago my initial feelings on the matter were that, having developed a general form for the deviance D,

$$D(y, \mu, p) = -2y(\mu^{1-p} - y^{1-p})/(1-p) + 2(\mu^{2-p} - y^{2-p})/(2-p),$$

then an *ad hoc* assumption was necessary in order to allow for the different scales introduced by each value of p. This was put on a firmer foundation by Nelder and Pregibon (cf. Nelder, 1985), who noted