A Review of Dispersion Models Generated by Tilting

Bent Jørgensen*
University of Southern Denmark, Odense, Denmark - bentj@stat.sdu.dk

Abstract

We review recent results on dispersion models generated by a tilting operator acting on a generating function, similar to exponential dispersion models, which are generated by exponential tilting acting on the cumulant generating function. Other cases include extreme, geometric and factorial dispersion models. These models are important for modelling non-normal response variables in generalized linear regression models, but here we concentrate on their distributional properties. A key ingredient for each type of dispersion model is the associated dispersion function, which is a useful characterization and convergence tool, much like the variance function for exponential dispersion models. In particular, dispersion models corresponding to power dispersion functions appear as scaling or dilation limits, similar to the Tweedie convergence theorem for exponential dispersion models.

Keywords: dispersion function; exponential tilting; generating function; reproductive dispersion model.

1. Introduction

Dispersion models, in the sense of Jørgensen (1997), are motivated as being two-parameter position-dispersion models, similar to location-scale families. Such models are important for modelling non-normal response variables in generalized linear regression models. The two main types of dispersion models discussed by Jørgensen (1997) are exponential and proper dispersion models. We have recently introduced three other types of dispersion models analogous to exponential dispersion models, although they are not dispersion models in the sense of Jørgensen (1997). The new models may nevertheless be characterized as two-parameter position-dispersion models, and appear to have applications in the construction of new types of generalized linear models.

Each of the three new types of dispersion models apply to a special type of data, namely extremes (Jørgensen et al., 2010), geometric sums (Jørgensen and Kokonendji, 2011) and count data (Jørgensen and Kokonendji, 2015), whereas exponential dispersion models apply mainly to continuous data. Their construction, however, allows parallels to be drawn between the four types, in some cases showing parallels between well-known distributions, while in other cases producing new results.

To each type of dispersion model there is an associated dispersion function, which is a useful characterization and convergence tool, similar to the variance function for exponential dispersion models. In particular, power dispersion functions characterize important classes of dispersion models, which appear as scaling or dilation limits, similar to the Tweedie class of exponential dispersion models. The four types of dispersion models are summarized in Table 1.

Type	Compounding	Dilation	C-function	Dispersion
Exponential	Convolution	Scaling	$\kappa(t)$	$Var(X)$
Geometric	Geometric sum	Scaling	$1-e^{-\kappa(t)}$	$Var(X)-E^2(X)$
Factorial	Convolution	Thinning	$\kappa(\log(1+t))$	$Var(X)-E(X)$
Extreme	Minimum	Scaling	$-\log[1-F(t)]$	

Table 1: The four main types of dispersion models generated by tilting.

2. Generating position and dispersion

We begin by considering the cumulant generating function (CGF) κ defined for a given random variable X by

$$\kappa(t) = \kappa(t;X) = \log \mathrm{E}\left(e^{tX}\right),$$

assumed to be finite for t in a non-degenerate interval Θ. There are two basic operations that generate location and dispersion, respectively. The exponential tilting operation maps κ into the new CGF defined by

$$t \longmapsto \kappa(t+\theta) - \kappa(\theta) \text{ for } t \in \Theta - \theta, \tag{1}$$

which is the operation that generates a natural exponential family. The convolution/division operation maps κ into $\lambda\kappa$ for some or all $\lambda > 0$, the latter case implying infinite divisibility. For integer λ this mapping corresponds to convolution of i.i.d. random variables. By combining exponential tilting and convolution/division, we obtain the class of additive exponential dispersion models defined by the following family of CGFs

$$t \longmapsto \lambda\left[\kappa(t+\theta) - \kappa(\theta)\right]. \tag{2}$$

In addition, we consider location and scale transformations, characterized by

$$\kappa(t; aX+b) = \kappa(at;X) + bt.$$

Standardization by means of location and scaling is, as we know, a crucial device for stating the central limit theorem. Here, we apply scaling by $1/\lambda$ to (2) to obtain the family of reproductive exponential dispersion models, yielding the two-parameter family of CGFs

$$t \longmapsto \lambda\left[\kappa(t/\lambda+\theta) - \kappa(\theta)\right]$$

There are many stochastic operations that correspond to applying certain transforms to κ, see e.g. the review by Abate and Whitt (1996). We are interested in applying a function that goes through the origin with slope 1, either to κ directly, or to the argument of κ. Table 1 shows two examples of such modified CGFs, or C-functions, namely $C(t) = 1-e^{-\kappa(t)}$, which generates geometric dispersion models, and $C(t) = \kappa(\log(1+t))$, which yields the factorial cumulant generating function used for discrete distributions. The derivatives of C are the C-cumulants, the first being simply the mean: $E(X) = \dot{C}(0;X)$, where dots denote derivatives. The second factorial cumulant $S(X) = \ddot{C}(0;X)$ is called the dispersion. As Table 1 shows, the dispersion tends to involve both the mean and variance of X. It plays a role similar to the variance, but can often take both positive and negative values. The case of extreme dispersion models in Table 1 will be discussed below.

We now define the dispersion function v for a given C-function by

$$v(\mu) = \ddot{C} \circ \dot{C}^{-1}(\mu), \tag{3}$$

where $\mu = \dot{C}(\theta)$ is the mean, restricted to an interval where $\dot{C}$ is monotone. We note, in particular, that $v(\mu)$ is positive/negative if $\dot{C}$ is increasing/decreasing. The dispersion function can be inverted as follows:

$$\dot{C}^{-1}(\mu) = \int_{\mu_0}^{\mu} \frac{1}{v(z)}\,dz, \tag{4}$$

where μ_0 corresponds to an arbitrary additive constant for $\dot{C}^{-1}(\mu)$. We now invert to obtain $\dot{C}$, and integrate once more to obtain C, except that the arbitrary additive constant from (4) yields the family of C-functions given by

$$t \longmapsto C(t+\theta) - C(\theta), \tag{5}$$

similar to the exponential tilting used in (1) to generate a natural exponential family. We shall hence call (5) a tilting family. Just like a natural exponential family is characterized by its variance function, our construction shows that a tilting family is characterized by its dispersion function v.

Further development of this construction requires us to give meaning to the convolution/division operator that maps C into λC for some or all $\lambda > 0$. Similarly, we must define scaling and translation operators $\cdot$ and $\oplus$, respectively, satisfying

$$C(t; a \cdot X \oplus b) = C(at; X) + bt$$

for some or all $a, b > 0$. We may then consider the dispersion function $\gamma v(\mu)$, which gives rise to the reproductive dispersion model

$$t \longmapsto \gamma \left[C(t/\gamma + \theta) - C(\theta) \right], \tag{6}$$

where $\gamma = 1/\lambda > 0$ is a dispersion parameter. The corresponding mean is $\mu = \dot{C}(\theta)$ and the dispersion is $\gamma v(\mu)$. This concludes our construction of a dispersion model based on C, with position and dispersion parameters.

3. **Examples**

We now consider two examples of dispersion models, namely factorial dispersion models, which illustrate the construction outlined above, and extreme dispersion models, where κ is replaced by the negative log survival function.

Factorial dispersion models are based on the factorial cumulant generating function defined by

$$C(t) = \kappa(\log(1 + t)),$$

which is particularly useful for handling discrete distributions. The factorial tilting operation is defined by the factorial cumulant generating function

$$t \longmapsto C(t + \theta) - C(\theta),$$

similar to the exponential tilting operation (1). The convolution/division operator corresponding to λC for some $\lambda > 0$ is the same as for exponential tilting.

The dispersion of a random variable X is defined by

$$S(X) = Var(X) - E(X), \tag{7}$$

which is a measure of over/underdispersion relative to the Poisson distribution, making it a useful tool for discrete distributions.

In order to proceed, we consider the dilation operator $a \cdot X$ defined by

$$C(t; a \cdot X) = C(at; X),$$

for those $a > 0$ such that the right-hand side is a factorial cumulant generating function. This turns out to be the binomial thinning operator for $0 < a \leq 1$, but may also exist for certain values of $a > 1$.

Letting v denote the dispersion function (3), the reproductive factorial dispersion model (6) has dispersion function $\gamma v(\mu)$. In view of (7), the corresponding variance has the form $Var(X) = \mu + \gamma v(\mu)$, which is a convenient representation of the variance for discrete factorial dispersion models. An example is the negative binomial distribution, which corresponds to $v(\mu) = \mu^2$, whereas the dispersion function $v(\mu) = \mu$ corresponds to the Neyman Type A distribution. Table 2 summarizes the main examples of power dispersion functions $v(\mu) = \mu^p$ for the four different types of dispersion models considered here. The convergence theorem for power dispersion functions for factorial dispersion models correspond to limiting distributions that are Poisson-Tweedie mixtures.

Type	$p = 0$	$p = 1$	$p = 2$	$p = 3$
Exponential	Normal	Poisson	Gamma	Inverse Gaussian
Geometric	Laplace	Geometric Poisson	Geometric gamma	Geometric Mittag-Leffler
Factorial	Hermite	Neyman Type A	Negative binomial	Poisson-inverse Gaussian
Extreme	Rayleigh	Gumbel	Uniform	Weibull

Table 2: Examples of dispersion models with power dispersion functions.

We now turn to the case of extreme dispersion models, for which the integrated hazard function

$$H(t) = -\log\left[1 - F(t)\right] \tag{8}$$

replaces the cumulant generating function in the construction, where F denotes the cumulative distribution function. In this case, convolution is replaced by the minimum operation, and the tilting operation corresponds to translation. The first and second cumulants corresponding to (8) are $h(0)$ and $\dot{h}(0)$, where $h = \dot{H}$ denotes the hazard function. The dispersion function is defined by $v(\mu) = \dot{h} \circ h^{-1}(\mu)$. Many well-known distribution families have very simple dispersion functions. For example, the logistic distribution has integrated hazard function $H(t) = \log\left(1 + e^y\right)$, corresponding to $v(\mu) = \mu\left(1 - \mu\right)$ for $\mu \in (0, 1)$. A second example is the power dispersion function proportional to μ^p, which corresponds to the class of generalized extreme-value distributions. In this case, convergence to power dispersion functions correspond to the classical convergence theorem of extremes to generalized extreme-value distributions.

A curious aspect of the theory is that a given functional form of the dispersion function suggests certain unexpected parallels between distributions in the different classes, as illustrated by the examples in Table 2. In particular, the second column in the table corresponds to distributions that appear in central limit-type convergence results, giving rise to normal, Laplace, Hermite and Rayleigh distributions, respectively, for the four types of dispersion models.

4. Conclusions

We have considered the construction of new dispersion models using different versions of the cumulant generating function and the dispersion function as our main tools. The models are designed for the purpose of generalized linear regression modelling of non-normal data, where it is important to take both position and dispersion into account. One of the difficulties of working with generating functions is that likelihoods and probabilities are hard to calculate, but we expect that regression models may be fitted using quasi-likelihood methods, following e.g. the suggestion by Jørgensen et al. (2010) in the case of extremes.

Based on the general approach presented here, it seems likely that there may exist other types of dispersion models with a similar structure. One possible example is free probability, where Bryc (2009) introduced so-called free exponential families, and studied an analogue of quadratic variance functions. The multivariate case also deserves further attention, because of the need to model correlated data, including for example longitudinal and spatio-temporal data. Multivariate dispersion models have been discussed in general by Jørgensen (2013) and in the special case of exponential dispersion models by Jørgensen and Martínez (2013).

References

Abate, J., & Whitt, W. 1996. An operational calculus for probability distributions via Laplace transforms. Advances in Applied Probability, 28, 75–113.

Bryc, W. 2009. Free exponential families as kernel families. Demonstr. Math., XLII, 657–672.

Jørgensen, B. 1997. The Theory of Dispersion Models. London: Chapman & Hall.

Jørgensen, B. 2013. Construction of multivariate dispersion models. Brazilian J. Probab. Statist., 27, 285–309.

Jørgensen, B., & Kokonendji, C. C. 2011. Dispersion models for geometric sums. Brazilian J. Probab. Statist., 25, 263–293.

Jørgensen, B., & Kokonendji, C. C. 2015. Discrete dispersion models and their Tweedie asymptotics. AStA Advances in Statistical Analysis, submitted.

Jørgensen, B., & Martínez, J. R. 2013. Multivariate exponential dispersion models. Pages 73–98 of: Kollo, T. (Ed.), Multivariate Statistics: Theory and Applications. Proceedings of the IX Tartu Conference on Multivariate Statistics & XX International Workshop on Matrices and Statistics. Singapore: World Scientific.

Jørgensen, B., Goegebeur, Y., & Martínez, J. R. 2010. Dispersion models for extremes. Extremes, 13, 399–437.