International Statistical Review (2005), **73,** 1, 35–58, Printed in Wales by Cambrian Printers

Mixed Poisson Distributions

Dimitris Karlis[1] and Evdokia Xekalaki[2]

[1]*Department of Statistics, Athens University of Economics and Business, Greece*
[2]*Department of Statistics, Athens University of Economics and Business, Greece and Department of Statistics, Trinity College, Dublin, Ireland*

Summary

Mixed Poisson distributions have been used in a wide range of scientific fields for modeling non-homogeneous populations. This paper aims at reviewing the existing literature on Poisson mixtures by bringing together a great number of properties, while, at the same time, providing tangential information on general mixtures. A selective presentation of some of the most prominent members of the family of Poisson mixtures is made.

Key words: Mixtures; Discrete distributions; Distribution theory; Mixing distribution; Overdispersion.

1 Introduction

Mixtures of distributions have been widely used for modeling observed situations whose various characteristics as reflected by the data differ from those that would be anticipated under the simple component distribution. In actuarial applications, for example, observed data on the number of claims often exhibit a variance that noticeably exceeds their mean. Hence, assuming a Poisson form (or any other form that would imply equality of the mean to the variance) for the claim frequency distribution is not appropriate in such cases.

In general, the assumption of a specific form of distribution $F(\cdot|\lambda),\ \lambda \in \Theta$ for the parent distribution of a data set imposes a certain mean-to-variance relationship, which in practice may be severely violated. More general families of distributions, well known as mixtures, are usually considered as alternative models that offer more flexibility. These are superimpositions of simpler component distributions depending on a parameter, itself being a random variable with some distribution. Mixed Poisson distributions, in particular, have been widely used as claim frequency distributions.

Definition 1. A probability distribution is said to be a mixture distribution if its distribution function $F(\cdot)$ can be written in the form

$$F(\cdot) = \int_{\Theta} F(\cdot|\lambda) dG(\lambda),$$

where $F(\cdot|\lambda)$ denotes the distribution function of the component densities considered to be indexed by a parameter λ with distribution function $G(\lambda),\ \lambda \in \Theta$.

The above definition can also be expressed in terms of probability density functions, thus

$$f(x) = \int_{\Theta} f(x|\lambda) g_{\lambda}(\lambda) d\lambda.$$

In the sequel, the above mixture is denoted as $f(x|\lambda) \underset{\lambda}{\wedge} g(\lambda)$. The density $g(\cdot)$ is referred to as the *mixing density*. The mixing distribution can be continuous, discrete or a distribution with positive probability at a finite number of points, i.e. a finite step distribution. In the sequel, a mixture with a finite step mixing distribution is termed a *k-finite step mixture* of $F(\cdot|\lambda)$, where k is the number of points with positive probabilities in the mixing distribution.

Mixture models cover several distinct fields of the statistical science. Their broad acceptance as adequate models to describe diverse situations is evident from the plethora of their applications in the statistical literature. Titterington *et al.* (1985) provide a review of work in the area of applications of mixture models up to 1985. In recent years, the number of applications increased mainly because of the availability of high speed computer resources, which removed any obstacles to apply such methods. Thus, mixture models have found applications in fields as diverse as data modeling, discriminant analysis, cluster analysis, outlier-robustness studies, ANOVA models, kernel density estimation, latent structure models, empirical Bayes estimation, Bayesian statistics, random variate generation, approximation of the distribution of some statistic and others.

Interesting reviews for finite mixture models are given in the books by Everitt & Hand (1981), Titterington *et al.* (1985), McLachlan & Basford (1988), Lindsay (1995), B öhning (1999), McLachlan & Peel (2000). Also, literature reviews on special topics related to mixtures can be found in the papers by Gupta & Huang (1981) (on ranking and selection topics), Redner & Walker (1984) (on the EM algorithm for finite mixtures) and Titterington (1990); an update of some topics covered in the book by Titterington *et al.* (1985).

The aim of this paper is to bring together results concerning Poisson mixtures that have been dispersed in diverse scientific fields. An attempt is made to include a variety of related topics, though some topics are not treated in detail, due to, mainly, space limitations. In particular, in section 2, some general properties for mixture models are provided, while in section 3, a detailed description of the properties of mixed Poisson distributions is given. Some new results are also provided. Section 4 contains a brief description of various mixed Poisson distributions and some interrelationships pertaining to their derivation. A brief review of bivariate mixed Poisson distributions is given in section 5. Finally, several aspects of the mixed Poisson models in connection with applications are discussed in section 6.

2 Some Properties of Mixture Models

Mixture models have interesting properties. In this section, some of their properties that are used in the sequel are briefly presented. Some definitions, notation and terminology are also provided.

It can easily be shown that the following associative property holds for mixtures provided that there are no dependencies between the parameters of the distributions considered.

$$[f(x|\lambda) \underset{\lambda}{\wedge} g(\lambda|\mu)] \underset{\mu}{\wedge} h(\mu) \text{ is equivalent to } f(x|\lambda) \underset{\lambda}{\wedge} [g(\lambda|\mu) \underset{\mu}{\wedge} h(\mu)].$$

The proof is based on the definition of mixtures and the possibility of interchanging the order of integration or summation.

Regardless of the form of $f(\cdot|\lambda)$, the expected value of the function $h(X)$ is obtained as

$$E[h(X)] = \int_{\Theta} E_{x|\lambda}[h(X)]\, g(\lambda) d\lambda, \tag{1}$$

with the subscript in the expectation denoting that the expectation is taken with respect to the conditional distribution of X. Integration is replaced by summation in the case of a discrete mixing distribution. It follows that

$$E[X] = E[E_{x|\lambda}(X)] \text{ and } V[X] = V[E_{x|\lambda}(X)] + E[V_{x|\lambda}(X)], \tag{2}$$

i.e., the variance of the X in the mixed distribution is the sum of the variance of its conditional mean and the mean of its conditional variance. Relationship (2) indicates that the variance of the mixture model is always greater than that of the simple component model and this explains the use of the term "overdispersion models" used for mixture models.

Another interesting property that relates mixtures to products of random variables has been given by Sibuya (1979). The proposition that follows shows that mixing with respect to a scale parameter is equivalent to obtaining the distribution of the product of two random variables—the distribution of one being of the same form as that of the simple component model, but with unit scale parameter, and the distribution of the other being the same as the mixing distribution.

PROPOSITION 1 *(Sibuya, 1979). Suppose that the conditional density of X is given by $f(\cdot|\lambda)$, where λ is a scale parameter. Assume that λ is itself a random variable with density function $g(\cdot|\varphi)$ for some parameter φ (symbolically $\lambda \sim g(\cdot|\varphi)$). Then, the unconditional density of the random variable X is the same as the density of the random variable $Z = X_1 X_2$, where $X_1 \sim f(\cdot|1)$ and $X_2 \sim g(\cdot|\varphi)$.*

The above proposition justifies the dual derivation of certain distributions as mixtures and as distributions of products of two random variables. The Beta distribution and the t distribution are typical examples.

Definition 2. Consider a random variable S that can be represented as

$$S = X_1 + X_2 + \cdots + X_N,$$

where $N, X_1, X_2, \ldots$ are mutually independent, non-negative, integer valued random variables with the variables $X_1, X_2, \ldots$ distributed identically with density function f. Let the distribution of N be defined by a probability function p. Then, S is said to have a *compound p distribution* with density denoted by $p \vee f$. The distribution defined by the density f is referred to as *the summand distribution*, as it is the distribution of the summands X_i. Some authors use the term *generalized p distribution*.

Definition 3. A compound (or generalized) distribution is termed a *compound Poisson distribution* if the distribution of N is the Poisson distribution.

The following proposition connects mixture models to compound distributions in the discrete case.

PROPOSITION 2 *(Gurland, 1957). If a probability function g has a probability generating function of the form $[\phi(t,a)]^n$, where $\phi(t,a)$ is some probability generating function independent of n, the model $f \vee g$ is equivalent to the model $g(x|n) \underset{n}{\wedge} f(n)$.*

PROPOSITION 3 *(Gurland, 1957). It holds that*

$$\left[f(x|\lambda) \underset{\lambda}{\wedge} g(\lambda) \right] \vee h \quad \textit{is equivalent to} \quad [f \vee h] \underset{\lambda}{\wedge} g(\lambda).$$

Definition 4. A distribution with probability function p is said to be the *convolution of the distributions with probability functions f and g* denoted by $(f * g)$ if

$$p(x) = \sum_{n=0}^{x} f(x-n) g(n).$$

The convolution is the distribution of the sum $Y = X + Z$, where X follows a distribution with probability function f and Z follows a distribution with probability function g, respectively. In the case of continuous random variables X or Z we replace summation by integration.

PROPOSITION 4. *The models* $[f(x|\lambda)^* g(y|\mu)] \underset{\lambda}{\wedge} h(\lambda)$ *and* $[f(x|\lambda) \underset{\lambda}{\wedge} h(\lambda)]^* g(y|\mu)$ *are equivalent provided that the density function* $g(\cdot|\mu)$ *does not depend on* λ.

The above results constitute merely a few of the properties of general mixture models. In the sequel, we use them for deriving related results. De Vylder (1989) provided some other relationships between mixtures and compound distributions. (See also Douglas, 1980).

Definition 5. A random variable X follows a mixed Poisson distribution with mixing distribution having probability density function g if its probability function is given by

$$P(X = x) = P(x) = \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!} g(\lambda) d\lambda, \qquad x = 0, 1, \ldots \tag{3}$$

In terms of the probability generating function, $H(s)$, of X, (3) can be written in the form

$$H(s) = \int_0^\infty e^{\lambda(s-1)} g(\lambda) d\lambda.$$

Note that the right hand side of this equation is $M_\lambda(s-1)$, the moment generating function of the mixing distribution evaluated at $s-1$. This, immediately, implies that the probability generating function of the mixed Poisson distribution uniquely determines the mixing distribution through its moment generating function.

In the sequel, the mixed Poisson distribution with mixing distribution the distribution with density function g is denoted by the $MP(g)$, while its probability function is denoted by $P(x)$. Note that λ is not necessarily a continuous random variable. It can be discrete or it can take a finite number of values. The latter case gives rise to finite Poisson mixtures.

3 Properties of Mixed Poisson Distributions

Historically, the derivation of mixed Poisson distributions goes back to 1920 when Greenwood & Yule considered the negative binomial distribution as a mixture of a Poisson distribution with a Gamma mixing distribution. Depending on the choice of the mixing distribution, various mixed Poisson distributions can be constructed. (For an early review on mixed Poisson distributions, see Haight (1967)). Since then, a large number of mixed Poisson distributions has appeared in the literature. However, only a few of them have been used in practice, the main reason being that often their form is complicated.

Let X be a random variable whose distribution is a mixed Poisson distribution. Then, the following results hold (e.g. Neuts & Ramalhoto (1984) and Willmot (1990)).

(a). $P(X \le x) = \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!} G(\lambda) d\lambda$ and

(b). $P(X > x) = \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!} [1 - G(\lambda)] d\lambda$,

where $G(\lambda) = \int_0^\lambda g(x)dx$ is the distribution function of the parameter λ.

These results, although interesting as directly relating the distribution functions of mixtures to those of their mixing distributions, have had only limited practical use mainly due to the complexity of the form of $G(\cdot)$.

3.1 Comparison to the Simple Poisson Distribution

Let $P(x)$ be the probability function of a mixed Poisson distribution as given by (3) and $P(x|m)$ be the probability function of a simple Poisson distribution with the same mean, say m. Then, as shown by Feller (1943),

(i) $P(0) \geq P(0|m)$, i.e., the probability of observing a zero value is always higher under a mixed Poisson distribution than under a simple Poisson distribution with the same mean
and
(ii) $P(1)/P(0) \leq P(1|m)/P(0|m) = m$, i.e., the ratio of the probability of 1 to that of 0 is less than the mean for every mixed Poisson distribution.

Shaked (1980) showed that the function $P(x) - P(x|m)$ has exactly two sign changes of the form + - +, implying that a mixed Poisson distribution gives a higher probability to the event $\{X = 0\}$, and has a longer right tail. This is known as the *two crossings theorem* in the econometrics literature. This result can be used to test if a mixed Poisson distribution is adequate for describing a dataset. A similar result holds for other mixtures too.

Shaked (1980) also showed that for every convex function $c(\cdot)$ it holds that

$$\sum c(x)P(x) \geq \sum c(x)P(x|m).$$

(For mixtures of continuous densities summation is replaced by integration.) For example, for $c(x) = (x - m)^2$, the property that the variance of the mixed Poisson is greater than the variance of the simple Poisson is obtained. Multivariate extensions of this result are given in Schweder (1982). Another generalization can be found in Lynch (1988). Recently, Denuit *et al.* (2001) examined the s-convexity of Poisson mixtures and its application to actuarial topics. Stochastic comparisons of the simple Poisson distribution and mixed Poisson distributions are discussed in Misra *et al.* (2003). Roos (2003) described approximations of mixed Poisson distributions by simple Poisson distributions.

3.2 The Moments of a Mixed Poisson Distribution

As implied by (1), the moments of any mixed distribution can be obtained by weighting the moments of the simple component models with the mixing distribution as the weighting mechanism. In the discrete case, this holds true for the probability generating function too, for $h(X) = t^X$. For the probability generating function $Q(t)$ of the *MP*(g) distribution, in particular, we obtain

$$Q(t) = E\left[t^x\right] = \int_0^\infty \exp[\lambda(t-1)]g(\lambda)d\lambda. \tag{4}$$

From (4), it is clear that the factorial moments of the mixed Poisson distribution are the same as the moments of the mixing distribution about the origin. Thus, one may express the moments about the origin of the mixed Poisson distribution in terms of those of the mixing distribution. So, $E(X) = E(\lambda)$ and $E(X^2) = E(\lambda^2) + E(\lambda)$. In general,

$$E(X^r) = \sum_{j=1}^{r} S(r, j)E(\lambda^j), \quad r = 1, 2, \ldots,$$

where $S(n, k)$ denotes the Stirling numbers of the second kind (see e.g. Johnson *et al.*, 1992, chapter 1). In particular, we have for the variance of the mixed Poisson distribution that

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = E(\lambda^2) + E(\lambda) - [E(\lambda)]^2 = E(\lambda) + \mathrm{Var}(\lambda). \tag{5}$$

From the above result, it becomes obvious that the variance of a mixed Poisson distribution is always greater than the variance of a simple Poisson distribution with the same mean. Molenaar &

Van Zwet (1966) gave necessary and sufficient conditions for distributions to have this property. (See also Schweder (1982), Cox (1983) and Gelfand & Dalal (1990)). Further, from (5), it follows that the variance of the mixed Poisson distribution can be decomposed into two components: the first component can be attributed to randomness and is the variance of a simple Poisson distribution with mean equal to $E(\lambda)$ and the second component is the variance imposed by the mixing distribution. It is interesting to note that this scheme is similar to that used in the well known analysis of variance methods (ANOVA) for normal models. Moreover, (5) implies that the variance-to-mean ratio for a mixed Poisson distribution is always greater than 1, which is the value corresponding to the Poisson distribution. This property has been used to test the null hypothesis that the data come from a Poisson distribution versus the alternative hypothesis that the data come from a mixed Poisson distribution.

Carriere (1993) used the relationship between the moments of the mixed Poisson to those of the mixing distribution implied by (4) to construct a test of the hypothesis that a mixed Poisson distribution fits a dataset.

3.3 The Convolution of Two Mixed Poisson Random Variates

Definition 6. A distribution with probability density function f is said to be reproducible if the sum of two independent random variables X_1 and X_2, each with probability density function f, follows a distribution with probability density function of the same form as f but with possibly different parameters.

Some well known examples of reproducible distributions are the normal distribution, the exponential distribution and the Poisson distribution, among others.

An interesting result in this area is given by the proposition that follows:

PROPOSITION 5 *(Feller, 1943). The sum of two mixed Poisson variables (MP(f)) has an MP(f) distribution if the distribution defined by f is itself reproducible.*

The proof of Proposition 5, as given by Feller (1943), is based on the fact that the convolution of two mixed Poisson distributions is itself a mixed Poisson distribution with mixing density the convolution of the two mixing densities. It also follows as a consequence of (4), since the reproducibility of the distribution defined by f implies that the convolution of the mixing densities is also defined by f. Therefore, the mixed Poisson distribution is an $MP(f)$ distribution.

Next proposition is also a consequence of (4).

PROPOSITION 6 *(Willmot & Sundt, 1989). The convolution of an $MP(f)$ and a Poisson distribution with parameter λ is an $MP(g)$ distribution with $g(x) = f(x-\lambda)$, $x \geq \lambda$, (i.e. with g being a shifted version of f).*

As an example, consider the convolution of an $MP(f)$ and a Poisson distribution. Since the Poisson distribution can be regarded as an $MP(D_\lambda)$ distribution where D_λ is the density of a degenerate distribution at the point λ, the resulting distribution is a mixed Poisson distribution. The mixing density, in this case, is the convolution of the density f with the density D_λ, which results in a shifted version of f. (In particular, it is the distribution of the random variable $Y = X + \lambda$ with the density of X being f.) The Delaporte distribution (see e.g. Ruohonen, 1988) is the distribution of the convolution of a Poisson distribution with a negative binomial distribution. Willmot & Sundt (1989) showed that the Delaporte distribution is a mixed Poisson distribution with a shifted Gamma as mixing distribution.

3.4 Identifiability

The term identifiability of mixtures refers to the ability of identifying the mixing distribution of a given mixed distribution. Identifying a mixture is important since it ensures that the mixing distribution characterizes the mixed distribution.

Definition 7. Mixtures of the probability function $f(x|\theta)$ are *identifiable* if and only if $\int f(x|\theta)h_1(\theta)d\theta = \int f(x|\theta)h_2(\theta)d\theta$ implies that $h_1(\theta) = h_2(\theta)$ for all the values of θ. (In the case of discrete mixtures, integration is replaced by summation).

In our case, $f(x|\theta)$ is the probability function of a Poisson distribution. Mixtures of the Poisson distribution (finite or not) were shown to be identifiable by Feller (1943), who pointed out that the probability generating function of a mixed Poisson distribution is the Laplace transform of the mixing distribution. This means that every mixed Poisson distribution corresponds to one and only one mixing distribution. Teicher (1961) showed that mixtures on n of distributions with a probability generating function of the form $(h(t))^n$ are identifiable. The Poisson distribution belongs to this family, which also contains the normal, the gamma and the binomial distributions among others. The identifiability of Poisson mixtures has also been examined by Ord (1972), Xekalaki (1981), Xekalaki & Panaretos (1983), Lindsay & Roeder (1993) and Sapatinas (1995). The property of identifiability is important as only in this case is it meaningful to estimate the mixing distribution. Related material can be found in Barndorff-Nielsen (1965), Tallis (1969) and Yakowitz & Spragins (1969). Teicher (1963) and Al-Hussaini & El-Dab (1981) have also discussed the identifiability of finite mixtures.

3.5 Modality and Shape Properties

Holgate (1970) showed that an $MP(g)$ distribution is unimodal if g is unimodal. Note that he used the term unimodal to refer to distributions with one mode or with several modes at successive points. So, the unimodality of a mixed Poisson distribution depends on the unimodality of its mixing distribution. This result holds only if g is absolutely continuous. It is not true for discrete mixing distributions. For example, the Neyman distribution, which is a mixed Poisson distribution with a Poisson mixing distribution, is known to be multimodal (e.g. Douglas, 1980) even though the Poisson distribution is unimodal. Bertin & Theodorescu (1995) extended Holgate's (1970) results to the case of not absolutely continuous mixing distributions. The modality of general mixture models has been considered by Al-Zaid (1989) and Kemperman (1991).

The shape of the probability function of an $MP(g)$ distribution exhibits a resemblance to that of the probability density function of the mixing distribution when its parameters are appropriately adjusted. Lynch (1988) showed that mixing carries the form of the mixing distribution over to the resulting mixed distribution. This fact, has been used long before its formal proof for approximating the probability function of some mixed Poisson distributions. For example, Best & Gipps (1974) proposed the use of the cumulative distribution function of a Gamma distribution as an approximation to the cumulative distribution function of the negative binomial distribution. The resemblance is much greater for larger values of the mean. If the mean is small, there is a high probability of observing a 0 value, i.e. $P(0)$ is large. This is not true for many continuous densities, and thus the approximation is poor. Cassie (1964) discussed the use of the lognormal distribution instead of the Poisson-Lognormal distribution. Adell & de la Cal (1993), studied under fairly general assumptions, the order of convergence of a mixed Poisson distribution to its mixing distribution.

Willmot (1990) examined the asymptotic tail behaviour of some mixed Poisson distributions in the case of continuous mixing distributions. He showed that the tails of some mixed Poisson distributions look like the tails of their mixing distributions and proposed their approximation at the tails by their mixing distributions that are of a more tractable form. A similar result was derived by Perline (1998). Remillard & Theodorescu (2000) showed that certain Poisson mixtures are Paretian, in the sense that $\lim_{x\to\infty} x^a P(X > x) = \text{const} > 0$, for some $\alpha > 0$.

An interesting result implied by the fact that the probability generating function of a $MP(g)$ distribution specifies the moment generating function of g is provided by the following proposition.

PROPOSITION 7 *(Grandell, 1997). For two mixed Poisson distributions, say $MP(g_1)$ and $MP(g_2)$, we have that $MP(g_1) \to MP(g_2)$ if and only if $g_1 \to g_2$ where $\to$ denotes convergence in distribution.*

The implication of the above result is that a limiting case of a Poisson mixture is uniquely determined by a limiting case of the mixing distribution. For example, since the Beta distribution tends to the Gamma distribution for certain parameter values, the negative binomial distribution can be regarded as a limiting case of the Poisson-Beta distribution.

Chen (1995), working with finite mixture models (including Poisson mixtures), studied the estimation of the mixing distribution on the basis of its rate of convergence. He showed that, when the number n of support points is not known a priori, the best possible rate of convergence is $n^{-1/4}$ and can only be achieved by minimum distance estimators.

Hall (1979), Pfeifer (1987) and Khaledi & Shaked (2003) studied the distance between a mixture model and its associated simple component model.

3.6 Infinite Divisibility and Compound Poisson Distributions

A random variable X is said to have an *infinitely divisible distribution* if its characteristic function $\varphi(t)$ can be written in the form $\varphi(t) = [\psi_n(t)]^n$, where $\psi_n(t)$ are characteristic functions for any $n \geq 1$. In other words, a distribution is infinitely divisible if it can be written as the distribution of the sum of an arbitrary number n of independently and identically distributed random variables. The simple Poisson distribution is a typical example since the sum of n independent Poisson variables is itself a Poisson variable.

Two noteworthy results linking Poisson mixtures to compound Poisson models through infinite divisibility are provided by the next two propositions.

PROPOSITION 8 *(Maceda, 1948). If, in a Poisson mixture, the mixing distribution is infinitely divisible, the resulting mixture distribution is infinitely divisible, too.*

PROPOSITION 9 *(Feller, 1968; Ospina & Gerber, 1987). Any discrete infinitely divisible distribution can arise as a compound Poisson distribution.*

Combining the above two results implies that a mixed Poisson distribution that is infinitely divisible can also be represented as a compound Poisson distribution. Well known examples are the negative binomial distribution (Quenouille, 1949), the Poisson-inverse Gaussian distribution (Sichel, 1975), and the generalized Waring distribution (Xekalaki, 1983b). It is worth pointing out that, for the two first cases, the form of the summand distribution is known, while, for the latter case, the form of the summand distribution has not been derived in a closed form.

Note that a compound Poisson distribution has a probability generating function $G(z)$ of the form

$$G(z) = \exp\left[\lambda(Q(z) - 1)\right],$$

where $Q(z) = \sum_{r=0}^{\infty} q_r z^r$ is the probability generating function of the summand distribution $\{q_r, r = 0, 1, 2, \ldots\}$. Note that q_0 is arbitrary and hence λ is arbitrary. Obviously, the identification of the summand distribution allows the compound Poisson representation of a mixed Poisson distribution. Solving the above equation we find that the probability generating function of the summand is

$$Q(z) = \frac{\ln G(z)}{\lambda} + 1. \tag{6}$$

This result may also be obtained as a limiting form of the zero-truncated mixed Poisson distribution as shown by Kemp (1967). (See also Johnson *et al.*, 1992, p. 352). Therefore, the probability function of the summands can be obtained by successive differentiation of (6). In practice, this does not always lead to a closed form for the probability function of the summand distribution. Of course, in all the cases, one is able to calculate numerically the probability function, using Panjer's recursive scheme. In the case where the summand distribution is discrete with probability function, say $f(x)$, the probability function of the corresponding compound Poisson distribution, say $g(x)$, can be obtained recursively via the formula:

$$g(x) = \sum_{y=1}^{x} \frac{\lambda y}{x} f(y) g(x-y) \tag{7}$$

with

$$g(0) = \exp\left[-\lambda + \lambda f(0)\right]. \tag{8}$$

From (7), one is able to derive the probability function of a compound Poisson distribution from the form of the summand distribution. For the converse result, relations (7) and (8) can be solved for $f(x)$ yielding

$$f(0) = \frac{\ln g(0) + \lambda}{\lambda}, \; f(1) = \frac{g(1)}{\lambda g(0)} \text{ and } f(x) = \frac{g(x)}{\lambda g(0)} - \frac{1}{x g(0)} \sum_{y=1}^{x-1} y f(y) g(x-y), \; x = 2, 3, \ldots$$

Willmot (1986) proposed choosing the value of λ by imposing the condition $Q(0) = 0$. However, one may verify that the successive ratios of the form $f(x+1)/f(x)$ do not depend on λ, apart from the ratio $f(1)/f(0)$. It should be noted that the above scheme applies only to the case of discrete summand distributions and that it cannot be used for estimation purposes. Note that making use of the empirical relative frequencies to estimate $g(x)$ may lead to unacceptable values for the probability function of the summand, i.e. negative values or values greater than 1. Remedies of this fact as well as properties of this kind of estimate are reported in Buchmann & Grubel (2003).

3.7 Posterior Moments of λ

A useful result concerning the posterior expectation of the random variable λ is given by the following proposition.

PROPOSITION 10 *(Willmot & Sundt, 1989). Suppose that X follows an MP(g) distribution. Then, the posterior expectation $E(\lambda^r | X = x)$ is given by:*

$$E\left(\lambda^r | X = x\right) = \frac{P(x+r)}{P(x)} (x+1) \ldots (x+r),$$

where $P(x)$ is the probability function of an MP(g) distribution.

Note that the above results may be extended to the case of negative r whenever $(x+r) > 0$. This enables one to find, for example, posterior expectations of the form $E\left(\lambda^{-r} | X = x\right)$.

Johnson (1957) showed that the posterior first moment of λ is linear if and only if the mixing distribution is the Gamma distribution. Johnson (1967) generalized this result to show that the form of the first posterior moment of λ determines the mixing distribution. Nichols & Tsokos (1972) derived more general formulae for a variety of distributions. The results of Cressie (1982) are also pertinent. Sapatinas (1995) gave the special forms of the posterior expectation of λ for other mixtures of power series family distributions.

It is interesting to note that since the posterior expectation of λ is expressed through the ratio $P(x+1)/P(x)$, it can characterize any specific member of the family of Poisson mixtures. (See also

Papageorgiou & Wesolowski, 1997). Ord (1967) showed that, for some basic discrete distributions, the ratio $(x+1)P(X=x+1)/P(X=x)$ can provide useful information concerning the distributional form of the population from which the data come. The practical value of this result, however, is limited by the fact that different mixed Poisson distributions can lead to very similar graphs for the quantity $\left(x, \frac{(x+1)P(X=x+1)}{P(X=x)}\right)$, thus making identification very difficult.

Bhattacharya (1967) showed the following result in the context of accident theory if the mixing distribution is a Gamma distribution: Selecting at the beginning of a time period individuals with no accidents in an immediately preceding time period reduces the expected number of accidents. In particular, if X and Y are the numbers of accidents in the first and the second period, respectively, then $E(Y)/E(Y|X=0) \geq 1$. This result is also valid for the Poisson-confluent hypergeometric series distribution (Bhattacharya, 1966). In fact, it can be shown to hold for every mixed Poisson distribution.

PROPOSITION 11. *For any mixing density $g(\lambda)$, it holds that $E(Y)/E(Y|X=0) \geq 1$.*

Proof. From Proposition 10, $E(Y|X=x) = (x+1)P(X=x+1)/P(X=x)$. Setting $x=0$, we obtain $E(Y|X=0) = P(1)/P(0) \leq m$ (see section 3.1), where m is the mean of the mixed Poisson distribution and, hence, the mean of the unconditional distribution of Y.

Haight (1965) derived the distribution of the number of accidents in a given time period given the removal of persons with more than n accidents in an immediately preceding period for the case of a negative binomial accident distribution.

PROPOSITION 12. *For any mixing density $g(\lambda)$, $\lambda > 0$ it holds that*

$$E(\lambda|X \leq x) = \frac{\sum_{n=1}^{x+1} nP(n)}{\sum_{n=0}^{x} P(n)},$$

where $P(x)$ is the probability function of an $MP(g)$ random variable.

Proof. The posterior density of λ conditional on $X \leq x$ is given by

$$g(\lambda|X \leq x) = \frac{g(\lambda)\sum_{n=0}^{x} \frac{e^{-\lambda}\lambda^n}{n!}}{P(X \leq x)}.$$

Hence, the posterior expectation conditional on $X \leq x$ is given by

$$E(\lambda|X \leq x) = \frac{\int_0^\infty \lambda g(\lambda)\sum_{n=0}^{x} \frac{e^{-\lambda}\lambda^n}{n!} d\lambda}{P(X \leq x)} = \frac{\int_0^\infty g(\lambda)\sum_{n=0}^{x}(n+1)P(n+1|\lambda)d\lambda}{P(X \leq x)},$$

where $P(x|\lambda)$ denotes the Poisson distribution with parameter λ. Integrating by parts, we obtain

$$E(\lambda|X \leq x) = \frac{(x+1)P(X \leq x+1) - \sum_{n=0}^{x} P(X \leq n)}{P(X \leq x)}.$$

Since $\sum_{n=0}^{x} P(X \leq n) = \sum_{n=0}^{x}(x-n+1)P(n)$, the above expression becomes

$$E(\lambda|X \leq x) = \frac{(x+1)\sum_{n=0}^{x+1} P(n) - \sum_{n=0}^{x}(x-n+1)P(n)}{P(X \leq x)} = \frac{\sum_{n=1}^{x+1} nP(n)}{P(X \leq x)} = \frac{\sum_{n=1}^{x+1} nP(n)}{\sum_{n=0}^{x} P(n)}.$$

This completes the proof.

An interesting special case arises when $x=0$, leading to a specialisation of the result of Proposition 10 regarding the posterior mean of λ given that $X=0$.

3.8 Numerical Approximation for the Probability Function of a Mixed Poisson Distribution

The analytic calculation of the probabilities of Poisson mixtures is often quite involved and resorting to numerical calculation becomes necessary. However, even in the case of powerful recursive schemes, direct calculation of initial values is required. In the sequel, the devices used by some numerical methods for the efficient calculation of the probabilities are outlined.

a) Taylor expansions

One of the methods, given by Ong (1995), is based on a Taylor expansion of a special function of a gamma variable.

PROPOSITION 13 *(Ong, 1995). Let $g(\lambda)$ be the probability density function of the mixing distribution of a mixed Poisson distribution. If $g(\lambda)$ has a finite n-th derivative at the point k, the probability function $P(k)$ of the mixed Poisson distribution has the formal expansion:*

$$P(X = k) \simeq g(k) + \frac{1}{k} \sum_{y=2}^{n} \frac{\mu_y h^{(y)}(k)}{y!},$$

where $h(k) = kg(k)$, $h^{(i)}(k)$ denotes the i-th derivative of $h(k)$ with respect to k and μ_y is the y-th moment about the mean of a gamma random variable with scale parameter equal to 1 and shape parameter equal to k.

The above approximation has some disadvantages. The first is that we cannot obtain $P(0)$. On the other hand, evaluating the derivatives of the mixing distribution (provided that they exist), is a very tedious task.

b) The Probability Function of the Mixed Poisson Distribution as an Infinite Series Involving the Moments of the Mixing Distribution

An alternative useful method makes use of a formula linking the probability function of a mixed Poisson distribution to the moments of the mixing distribution as indicated by the next proposition.

PROPOSITION 14. *Provided that the moments of the mixing distribution in a mixed Poisson model exist, the probability function of the mixture distribution can be written as*

$$P(X = x) = \frac{1}{x!} \sum_{r=0}^{\infty} \frac{(-1)^r}{r!} \mu_{x+r}(\lambda),$$

where $\mu_r(\lambda)$ is the r-th moment of λ about the origin.

The proof of this result is a consequence of the fact that

$$\begin{aligned} P(X = x) &= \frac{1}{x!} \int_0^\infty \exp(-\lambda)\, \lambda^x g(\lambda) d\lambda = \int_0^\infty \left(\sum_{r=0}^{\infty} \frac{(-\lambda)^r}{r!} \right) \frac{\lambda^x}{x!} g(\lambda) d\lambda \\ &= \sum_{r=0}^{\infty} \int_0^\infty \frac{\lambda^{x+r}(-1)^r}{x!r!} g(\lambda) d\lambda = \sum_{r=0}^{\infty} \frac{(-1)^r}{x!r!} \int_0^\infty \lambda^{x+r} g(\lambda) d\lambda. \end{aligned}$$

Results of similar nature have been obtained by Katti (1966), for compound distributions, by Sivaganesan & Berger (1993), for mixtures of the binomial distribution, and by Karlis & Xekalaki (2000), for mixtures of the negative binomial distribution.

c) Gauss–Laguerre Polynomials

Other methods utilize representations of the probability function of the mixture in terms of adjusted Laguerre polynomials or Gauss–Laguerre polynomials. Following Press *et al.* (1992), some integrals can be approximated using certain weight functions of the integrand evaluated at certain points, thus

$$\int_0^\infty e^{-x} x^\alpha f(x) dx \simeq \sum_{j=1}^{n} w_j f(x_j). \tag{9}$$

Here, n is the number of points used for the approximation and w_j and x_j, $j = 1, \dots, n$, are, respectively, the Gauss–Laguerre weights and abscissas calculated using the methods and routines described by Press *et al.* (1992). Clearly, the probability function of any mixed Poisson distribution can be approximated using the above formula. It is interesting that using (9), the probability function of a mixed Poisson distribution is calculated as a finite mixture of the mixing distribution. Note that this approach has been used for fitting a mixed effect Poisson regression model by Aitkin (1996).

d) Recursive Relations for Mixed Poisson Distributions

Approximating all the probabilities of a probability function is not a good strategy, mainly because of the computational complexity. Willmot (1993) showed that, for several mixed Poisson distributions a recursive formula can be obtained. Specifically, provided that the mixing density $g(\cdot)$, satisfies the relationship

$$\frac{d \ln g(\lambda)}{d\lambda} = \frac{\sum_{i=0}^{k} s_i \lambda^i}{\sum_{i=0}^{k} w_i \lambda^i}, \lambda \in (0, +\infty)$$

for some constants $s_i, w_i, i = 0, 1, \dots, k,\ k > 0$, the probability function $P(x)$ of the *MP*(g) satisfies the following recursive formula

$$\sum_{n=-1}^{k} \{\varphi_n + m w_{n+1}\} (m+n)^{(n)} P(m+n) = 0,$$

where $a^{(b)} = a(a+1)\dots(a+b+1)$ and $\phi_n = s_n + (n+1)w_{n+1} + w_n$ with $\phi_{-1} = 0$. Appropriate modifications have been suggested by Willmot (1993) for different supports of λ. Note that this iterative scheme requires calculation of the first k probabilities only. Ong (1995) derived a method of using this iterative scheme that requires no exact evaluation of any probability. The idea is to start from a point n at the tail of the distribution setting arbitrarily $P(n) = 1$ and $P(n+1) = 0$. Then, by using the above recurrence, one may obtain the values $P(n-1),\ P(n-2), \dots, P(0)$. Rescaling so that the obtained series sums to 1, leads to the probability function. It is useful to start with a value of n, for which the true value of $P(n)$ is negligible. It should be noted at this point that the recursion described above is unstable and should therefore be used with caution.

The recursive scheme defined above led to the increase of the applicability of several mixed Poisson distributions. Earlier, the difficulties in evaluating the probabilities prevented the researchers to use many of the mixed Poisson distributions. Wang & Panjer (1993) argued the recurrence relations might be quite unstable, mainly because of the negative coefficients of some of the probabilities and proposed using as starting points those points where the instability occurs (i.e., the points with negative coefficients in the recurrence representation). They also provided several examples.

3.9 Simulation Based Reconstruction

An approximate way to construct the probability function for any mixed Poisson distribution is via simulation according to the following scheme:

Step 1. Generate λ from the mixing distribution.
Step 2. Generate X from the *Poisson* λ distribution.

Hence, if a very large number of values is generated, an approximation of the probability function can be obtained. Note that the speed of the simulation depends on the speed of generating a random variate from the mixing distribution. As the number of replications increases, the approximate probability function tends to the true probability function. Note that generating values from a mixed Poisson distribution is possible even when the exact form of its probability function is not known.

3.10 Weighting a Mixed Poisson Distribution

Very often, the distribution of recorded observations produced by a certain model may differ from that under the hypothesized model, mainly because of the recording mechanism that may employ unequal probabilities of recording the observations. The observed distribution will thus have a density of the form

$$f_w(x) = \frac{w(x)f(x)}{E[w(x)]},$$

where $f(x)$ is the original density anticipated under the hypothesized model and $w(x)$ is a function proportional to the probability with which an observation *n* is recorded. These models were introduced by Rao (1965) and are known as *weighted models*. When the *weight function* $w(x)$ is equal to x, these models are known as *size biased models*. The observed distribution is then termed *size biased distribution* and is defined by the density

$$f_w(x) = \frac{xf(x)}{E(x)}. \tag{10}$$

In an ecological context, Patil & Rao (1978), referring to the bias induced by the recording mechanism, used the term visibility bias and discussed various forms of $w(x)$ and their effect on the original distribution. Also, Patil *et al.* (1986) provided several results for discrete forms of distributions. It is interesting to note the following result concerning mixed Poisson distributions.

PROPOSITION 15. *A size biased MP(g) distribution can be obtained as a mixture of a size biased Poisson distribution with mixing distribution defined by the size biased version of the original mixing density* g.

Proof. The size biased version of the mixing distribution, say $g^*(\lambda)$ has density function

$$g^*(\lambda) = \frac{\lambda g(\lambda)}{E(\lambda)}, \lambda > 0. \tag{11}$$

Also, the size biased Poisson distribution has probability function of the form

$$f^*(x|\lambda) = \frac{x}{\lambda}\frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}, \quad x, 1, 2, \ldots \tag{12}$$

Then, using (11) and (12), the mixed size biased Poisson distribution with mixing distribution g^* has probability function

$$f(x) = \int_0^\infty f^*(x)g^*(\lambda)d\lambda = \int_0^\infty \frac{x}{\lambda}\frac{e^{-\lambda}\lambda^x}{x!}\frac{\lambda}{E(\lambda)}g(\lambda)d\lambda = \frac{xP(x)}{E(\lambda)} = \frac{xP(x)}{E(x)},$$

which is (from (10)) the size biased version of an *MP*(g) distribution.

For example, the size biased version of the negative binomial distribution can be obtained as a mixed size biased Poisson distribution with a size biased Gamma distribution as the mixing distribution. Relationship (11) has been widely used in the context of empirical Bayesian estimation with mixed Poisson distributions (see e.g. Sundt, 1999). Seshadri (1991) has shown that mixtures of the size biased Poisson distribution are in fact 2-finite mixtures with components the simple Poisson distribution and the size biased Poisson distribution.

3.11 Compound Mixed Poisson Distributions

In actuarial applications, the distribution of the number of claims can as mentioned earlier, often be assumed to be of a mixed Poisson form since it is very common that the population under investigation is not homogeneous. Depending on the distribution of the claim size, the total amount paid by the insurance company follows a compound mixed Poisson distribution. The probability density function of the total claimed amount is usually hard to derive and to compute. So, the use of recursive relations becomes necessary. One of the best known mixed Poisson distributions, the negative binomial distribution, has been treated in the fundamental paper of Panjer (1981) as it is the only member of the family of mixed Poison distributions with linear first order recurrence relations. Compound mixed Poisson distributions are discussed in detail in a series of papers, by Sundt (1999), Hesselager (1994, 1996), Wang & Sobrero (1994), Willmot (1993) and Grandell (1997) among others. The idea is to construct recurrence schemes based on the recurrence relation for the probabilities of the mixed Poisson distribution. In these papers, several examples are given for many members of the family of mixed Poisson distributions. Recurrence relationships for the moments of compound mixed Poisson distributions are given by Chadjiconstantinidis & Antzoulakos (2002).

3.12 Mixed Poisson Distributions Arising from the Mixed Poisson Process

The derivation of the Poisson process was based on the assumptions of a constant infinitesimal risk over the entire period of observation and of independence between any two events. These assumptions are not always realistic. Arbous & Kerrich (1951) proposed the so-called contagious model by assuming that, after the first event, the infinitesimal risk is reduced and remains constant for the remaining period of observation i.e., each event results in a change in the infinitesimal risk. More formally, the infinitesimal risk $k_m(t)$ depends on both the number m of previously occurred events and on time t. A well known example of such a process is the Pólya process for which $k_m(t) = \frac{a+m}{b+t}$. The resulting distribution is the negative binomial distribution. McFadden (1965) described a more general process the so-called mixed Poisson process. (See also Willmot & Sundt (1989) and Grandell (1997)). Any $MP(g)$ distribution can be shown to arise from a contagion model if the infinitesimal risk is defined to be the quantity

$$k_m(t) = \frac{\int_0^\infty \lambda^{m+1} e^{-\lambda t} g(\lambda) d\lambda}{\int_0^\infty \lambda^m e^{-\lambda t} g(\lambda) d\lambda}.$$

For $t = 1$, the above relation simplifies to

$$k_m(1) = \frac{(m+1)P(m+1)}{P(m)}, \tag{13}$$

where $P(m)$ is the probability function of the $MP(g)$ distribution as given in (3). This, however, has unfortunate implications in practice as observing a data set, which can be described adequately by a particular mixed Poisson distribution, one is not in a position to know which model led to it: a mixed model or a contagion model? At least two models can result in the same mixed Poisson distribution. (See Cane (1977) and Xekalaki (1983a) for a discussion on this problem).

As seen before, any mixed Poisson distribution can be obtained via a mixed Poisson process defined by the infinitesimal risk given in (13). In the notation adopted earlier, such a model can be represented by

$$\text{Poisson}(t\lambda) \underset{\lambda}{\wedge} g(\lambda), \tag{model 1}$$

where t is the period of observation. In general, such models lead to mixed Poisson distributions that differ from those obtained from the model

$$\text{Poisson}(\lambda) \underset{\lambda}{\wedge} g(\lambda). \tag{model 2}$$

Our main interest is on the second model. However, in some circumstances, it is of interest to consider the first model. If the observed time period is assumed to be of unit length, the two models are identical. On the other hand, it is often interesting to consider *model 1* and examine how this more general model relates to *model 2*, which is most commonly used in practice. Note also that several mixed Poisson regression models utilize *model 1*. In such cases, the Poisson parameter λ is treated as a regressor depending on a series of covariates, while t is a random variable having its own probability function, termed the *overdispersion parameter*. For more details on mixed Poisson regression models, one can refer to Lawless (1987), Dean *et al.* (1989), Xue & Deddens (1992), McNeney & Petkau (1994), Wang *et al.* (1996) and Chen & Ahn (1996). Moreover, it is worth mentioning that mixed Poisson regression models allow for different mean to variance relationships offering a wide range of different models for real applications (see e.g., Hinde & Demetrio, 1998).

Using Definition 6 for reproducible distributions, one can see that if the mixing distribution is reproducible, a rescaling of the random variable does not affect the distributional form of the resulting mixed Poisson distribution, but it does affect the parameters. In this case, the probabilities of the mixed Poisson distribution are easily obtainable. The gamma and the inverse Gaussian distributions are some well known examples of reproducible distributions commonly used as mixing distributions.

4 Some Mixed Poisson Distributions

In this section, several mixed Poisson distributions considered in the literature are presented (Table 1). Most of them have been of limited use in the applied statistical literature, mainly due to the complexity of their form. Numerical techniques are in some cases necessary for the evaluation of their probability (density) functions combined with some recursive scheme facilitated greatly by Willmot's (1993) fundamental method. Obtaining moment estimates for their parameters however, presents no difficulty due to the property of Poisson mixtures linking the moments of the mixing and mixed distributions discussed in section 3.2. (For more information about the distributions in Table 1, see Karlis (1998)).

Some other miscellaneous mixed Poisson distributions, not included in Table 1, have also been considered in the literature, such as the distributions proposed by Burrel & Cane (1982), Willmot (1986, 1993), Ghitany & Al-Awadhi (2001) and Gupta & Ong (2004). Albrecht (1984) described several mixed Poisson distributions based on the Mellin and Laplace transforms of their mixing distributions. Devroye (1993) described some mixed Poisson distributions related to the stable law. Gerber (1991) and Wahlin & Paris (1999) described a mixing density that has the gamma and the inverse Gaussian as special cases. Finally, Ong (1996) described a mixing density which is based on the modified Bessel function. The polylogarithmic distribution described by Kemp (1995) can also be considered as a mixed Poisson distribution. Discrete mixing distributions have also been used (see Johnson *et al.*, 1992).

5 Multivariate Mixed Poisson Distributions

This section aims at providing a brief description of multivariate mixed Poisson distributions. To simplify the presentation, most of the results are given for the bivariate case, but their generalization to more dimensions is obvious.

Definition 8. The *bivariate Poisson* (hereafter *BP*) *distribution* is defined as the bivariate discrete distribution with joint probability function given by

$$P(x, y; \lambda_1, \lambda_2, \lambda_3) = \frac{e^{-\lambda_1}\lambda_1^x}{x!}\frac{e^{-\lambda_2}\lambda_2^y}{y!}e^{-\lambda_3}\sum_{i=0}^{\min(x,y)}\binom{x}{i}\binom{y}{i}i!\left\{\frac{\lambda_3}{\lambda_1\lambda_2}\right\}^i, \quad x, y = 0, 1, \ldots . \quad (14)$$

We will denote the BP distribution with parameters $\lambda_1, \lambda_2, \lambda_3$ as $BP(\lambda_1, \lambda_2, \lambda_3)$.

If $\lambda_3 = 0$, the joint probability function given in (14) reduces to the product of two univariate Poisson probability functions, thus implying independence of the associated variables. The joint

Table 1

Some mixed Poisson distributions.

Mixed Poisson Distribution	Mixing Distribution	A Key Reference
Negative Binomial	Gamma	Greenwood & Yule (1920)
Geometric	Exponential	Johnson *et al.* (1992)
Poisson-Linear Exponential Family	Linear Exponential Family	Sankaran (1969)
Poisson–Lindley	Lindley	Sankaran (1970)
Poisson-Linear Exponential	Linear Exponential	Kling & Goovaerts (1993)
Poisson-Lognormal	Lognormal	Bulmer (1974)
Poisson-Confluent Hypergeometric Series	Confluent Hypergeometric Series	Bhattacharya (1966)
Poisson-Generalized Inverse Gaussian	Generalized Inverse Gaussian	Sichel (1974)
Sichel	Inverse Gaussian	Sichel (1975)
Poisson-Inverse Gamma	Inverse Gamma	Willmot (1993)
Poisson-Truncated Normal	Truncated Normal	Patil (1964)
Generalized Waring	Gamma Product Ratio	Irwin (1975)
Simple Waring	Exponential $\wedge$ Beta	Pielou (1962)
Yule	Beta with Specific Parameter Values	Simon (1955)
Poisson-Generalized Pareto	Generalized Pareto	Kempton (1975)
Poisson-Beta I	Beta Type I	Holla & Bhattacharya (1965)
Poisson-Beta II	Beta Type II	Gurland (1958)
Poisson-Truncated Beta II	Truncated Beta Type II	Willmot (1986)
Poisson-Uniform	Uniform	Bhattacharya (1966)
Poisson-Truncated Gamma	Truncated Gamma	Willmot (1993)
Poisson-Generalized Gamma	Generalized Gamma	Albrecht (1984)
Dellaporte	Shifted Gamma	Ruohonen (1988)
Poisson-Modified Bessel of the 3rd Kind	Modified Bessel of the 3rd Kind	Ong & Muthaloo (1995)
Poisson–Pareto	Pareto	Willmot (1993)
Poisson-Shifted Pareto	Shifted Pareto	Willmot (1993)
Poisson–Pearson Family	Pearson's Family of Distributions	Albrecht (1982)
Poisson-Log-Student	Log-Student	Gaver & O'Muircheartaigh (1987)
Poisson-Power Function	Power Function Distribution	Rai (1971)
Poisson–Lomax	Lomax	Al-Awadhi & Ghitany (2001)
Poisson-Power Variance	Power Variance Family	Hougaard *et al.* (1997)
Neyman	Poisson	Douglas (1980)
	Other Discrete Distributions	Johnson *et al.* (1992)

probability function given in (14) is quite complicated and an iterative scheme is necessary for efficient calculation of the probabilities (see Kocherlakota & Kocherlakota, 1992). The marginal distributions of the BP distribution are simple Poisson distributions, while the conditional distributions are convolutions of a Poisson distribution and a binomial distribution. For more details on the BP distribution, the interested reader may be referred to Kocherlakota & Kocherlakota (1992). The multivariate Poisson distribution can be defined analogously. A thorough treatment of this distribution can be found in Johnson *et al.* (1997).

In the literature, the term *bivariate (multivariate) mixed Poisson distributions* has been used for distributions derived from Poisson distributions by several very different ways of mixing. The majority of mixed BP distributions can be placed in two main categories.

Definition 9. A distribution is said to be a *mixed BP distribution of the first kind (MBP1(g))* if it arises by assuming that the parameters of the *BP* distribution are proportional to some random variable a with density $g(a)$.

More formally, an $MBP1(g)$ distribution arises as

$$BP(a\lambda, a\lambda_2, a\lambda_3)_1 \underset{a}{\wedge} g(a).$$

The joint probability function of an $MBP1(g)$ distribution is given by

$$P_g(x, y) = \int_a P(x, y; a\lambda_1, a\lambda_2, a\lambda_3) g(a) da, \qquad (15)$$

where g is the density of the mixing distribution. If g is discrete or finite, the integral in (15) is replaced by a sum. Generalization to more variables is obvious.

Definition 10. A distribution is termed *mixed BP distribution of the second kind (MBP2(g))* if it arises by assuming that the parameters of the *BP* distribution are themselves random variables having a joint distribution with density function $g(\lambda_1, \lambda_2, \lambda_3)$. The joint probability function of an $MBP2(g)$ distribution is given by

$$P_g(x, y) = \int_{\lambda_1} \int_{\lambda_2} \int_{\lambda_3} P(x, y) g(\lambda_1, \lambda_2, \lambda_3) d\lambda_3 d\lambda_2 d\lambda_1. \tag{16}$$

Note that, in (16), the joint distribution of λ' may give positive mass only to a finite number of points. In this case, the integral may be substituted by a finite sum, thus leading to a finite *BP* mixture of the second kind. Variants of these models can be obtained if some of the parameters are considered as fixed.

5.1 Mixed Bivariate Poisson Distributions of the 1st Kind

Properties of $MBP1(g)$ distributions are given by Kocherlakota (1988), including expressions for their probability function and moment relationships.

The covariance of the $MBP1(g)$ distribution is

$$\text{Cov}(X, Y) = (\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3)\sigma^2 + \lambda_3 \mu,$$

where μ and σ^2 are the mean and the variance of α respectively. It is worth noting that, in case of a simple *BP* distribution, the covariance would be $\alpha\lambda_3$. Hence, the covariance can be decomposed into two factors, one due to the mixing distribution and one due to the assumed *BP* distribution. Note also that, if $\sigma^2 = 0$ (as in the case of a degenerate mixing distribution), the *MBP*1 distribution reduces to the *BP* distribution. The correlation of an $MBP1(g)$ distribution is always positive.

Another interesting result refers to the generalized variance ratio between an *MBP*1 distribution and a *BP* distribution with fixed marginal means. It can be seen that

$$\text{GVR} = \frac{|V_{MBP2}|}{|V_{BP}|} = 1 + \frac{\sigma^2(\lambda_1 + \lambda_2)}{1 - \frac{\lambda_3}{(\lambda_1+\lambda_3)(\lambda_2+\lambda_3)}} = 1 + \frac{\sigma^2(\lambda_1 + \lambda_2)}{1 - \rho_I} \geq 1.$$

So, provided that the mixing distribution is not degenerate, mixing increases the generalized variance. This is in analogy to the increase of the variance in the univariate case. Note also that the generalized variance increases as the variance of the mixing distribution increases.

One can see that the marginal distributions are mixed Poisson distributions and, in particular, of the form $P(a(\lambda_2 + \lambda_3)) \underset{a}{\wedge} g(a)$. The conditional distributions are of a much more complicated form. They are convolutions of a binomial distribution and some other form of distribution. It has been proved that the conditional expectation is linear only when the mixing distribution is a Gamma distribution or a degenerate distribution (Johnson, 1967).

Members of this family are the bivariate negative binomial distribution studied by Edwards & Gurland (1961) and Subrahmaniam (1966) in its general form, and a reduced form of it examined by Bates & Neyman (1952a,b) and Mellinger *et al.* (1965). Ashford & Hunt (1974) used a multivariate extension of this model to describe the patient-doctor contacts in a hospital. Bivariate Neyman type A distributions have been considered by Kocherlakota & Kocherlakota (1992), while the bivariate Poisson-generalized inverse Gaussian distribution has been used by Stein *et al.* (1987), Stein & Juritz (1987), Kocherlakota (1988) and Kocherlakota & Kocherlakota (1992). Some other miscellaneous members have been proposed by Bhattacharya (1966). Munkin & Trivedi (1999) described multivariate mixed Poisson regression models based on this type of mixing and a gamma mixing distribution. Gurmu & Elder (2000) used an extended gamma density as a mixing distribution.

5.2 Mixed Bivariate Poisson Distributions of the 2nd kind

This type of mixture modeling imposes a multivariate mixing distribution on the parameter vector $(\lambda_1, \lambda_2, \lambda_3)$ of the *BP* distribution, with some probability density function $g(\lambda_1, \lambda_2, \lambda_3)$. Mixtures of this type are often very complicated. In practice, the majority of the models that have been proposed are in two dimensional contexts assuming $\lambda_3 = 0$ (see e.g. Xekalaki, 1984a,b). Xekalaki (1986) defined and studied the multivariate generalized Waring distribution, a multivariate mixed Poisson model of this type. The literature is sparse in mixtures of this type in their more general form (i.e. with λ_3 not a constant). Stein & Juritz (1987) described briefly bivariate models giving a few properties of them.

The form of the joint probability function of a multivariate mixed Poisson distribution is usually complicated and no general recurrence relationships are available in the *n*-dimensional case. In the bivariate case, denoting by $g_{ij}, i, j = 1, 2, 3$ the joint bivariate marginal distribution of the mixing density for any pair of the parameters of the bivariate Poisson distribution, one can see that the marginal distributions of an *MBP*2 distribution are of the form:

$$X \sim \left[\text{Poisson}(\lambda_1 + \lambda_3) \underset{\lambda}{\wedge} g_{13}(\lambda_1, \lambda_3) \right],$$

where $g_{13}(\lambda_1, \lambda_3)$ denotes the joint density of λ_1 and λ_3. The marginal distribution of Y can be derived similarly. Obviously, the marginal distributions are also mixed Poisson distributions.

For the case of constant λ_3, the covariance can be written in the form

$$\text{Cov}(X, Y) = \lambda_3 + \text{Cov}(\lambda_1, \lambda_2)$$

and can, therefore, be decomposed into two parts, one due to the covariance imposed by the *BP* distribution for each individual and one due to the mixing distribution. An important consequence of the above formula is that if $\text{Cov}(\lambda_1, \lambda_2) < 0$ then, the covariance of X and Y can be negative as well. Recall that, for type I mixtures, we can have only positive covariance.

Stein & Juritz (1987) derived the correlation coefficient in the case where $\lambda_3 = 0$, i.e., when two independent Poisson distributions have been considered. They termed such a process *the conceal process*.

If λ_3 is fixed, one can show that the Generalized Variance is again always greater than the Generalized Variance of a simple *BP* distribution. In particular,

$$\text{GVR} = \frac{|V_{MBP2}|}{|V_{BP}|} = 1 + \frac{\lambda_3\,(V_1 + V_2 - 2\text{Cov}(\lambda_1, \lambda_2)) + V_1 V_2 - [\text{Cov}(\lambda_1, \lambda_2)]^2 + V_1 c_2 + V_2 c_1}{(c_1 + c_2)\,\lambda_3 + c_1 c_2} \geq 1,$$

where $E(\lambda_1) = c_1$, $E(\lambda_2) = c_2$ and $V_i = \text{Var}(\lambda_i)$, $i = 1, 2$.

The exact form of the conditional distribution is not obtainable in general. The same is true for the conditional expectation.

Mixtures of *BP* distributions of type II are rather complicated and it is not surprising that their use in practice has been limited. The case where $\lambda_3 = 0$ has been given attention mainly because it can induce negative correlation between counts, an interesting fact not possible under other simpler standard bivariate discrete distributions. Steyn (1976) proposed the use of a bivariate normal distribution as the mixing distribution. Some years later, Aitchinson & Ho (1989) proposed the use of the bivariate lognormal distribution instead of the simple bivariate normal distribution. For a recent application of this distribution, see Chib & Winkelmann (2001). Stein & Juritz (1987) proposed the use of the bivariate inverse Gaussian distribution as the mixing distribution. Their model assumes two independent Poisson distributions with parameters jointly distributed according to the bivariate inverse Gaussian distribution. Barndorff-Nielsen *et al.* (1992) extended this model to the case of multivariate Poisson-multivariate generalized inverse Gaussian distributions. Xekalaki (1984b) demonstrated various models leading to the bivariate generalized Waring distribution. The model has the form

$$BP(\lambda_1, \lambda_2, 0) \underset{\lambda_1,\lambda_2}{\wedge} [\text{Gamma}(\nu, r), \text{Gamma}(\nu, k)] \underset{\nu}{\wedge} \text{Beta}(a, b).$$

In the context of an *MBP* model, Goutis & Galbraith (1996) assumed that the vector of parameters $(\lambda_1, \lambda_2, \lambda_3)$ follows a Wishart distribution. They derived the resulting *MBP*2 distribution as well as some properties of this distribution. Stein & Juritz (1987) considered the derivation of a mixed *BP* distribution via a trivariate reduction. Nelson (1985) derived a mixed *BP* model which makes use of both of the mixing procedures described above. His model was of the form:

$$BP(p_1\lambda,\ p_2\lambda, 0) \underset{\lambda}{\wedge} \text{Gamma}(a, b) \underset{(p_1, p_2)}{\wedge} \text{Dirichlet}.$$

Tables 2 and 3 present several bivariate and multivariate mixed Poisson models considered in the literature.

6 Discussion

Several results related to mixed Poisson distributions were brought together in this paper with the aim of highlighting their application potential and the theoretical as well as practical value of the relationships among the members of this vast family of discrete distributions.

An aspect common to several of the members of the family of mixed Poisson models worth drawing one's attention is the latency of the genesis of the mixing distribution. The study of this latency reveals interesting properties of the mixed Poisson distributions and leads to models giving rise to them. The Poisson parameter λ can be regarded as a random variable linked to μ, ν—random variables themselves with respective densities g_1, g_2—either multiplicatively ($\lambda = \mu\nu$) or additively ($\lambda = \mu + \nu$). The generalized Waring distribution is a typical example of the multiplicative model examined by Xekalaki (1983a). In the case of mixtures under the additive model, the mixing density is the convolution of g_1 and g_2. The Delaporte distribution may arise from the additive model. Other interesting examples of this kind are given in Barndorff-Nielsen *et al.* (1992), where the convolution of a Gamma with a generalized inverse Gaussian distribution leads to another generalized inverse Gaussian distribution. It should be noted that, in the multiplicative model, the distributions of μ and ν are not identifiable in general. This difficulty however, can be overcome in the case of the generalized Waring distribution (Xekalaki, 1984a,b).

Another interesting point is the connection of mixed Poisson models to Bayesian methods. From the Bayesian perspective, the mixing distribution corresponds to the *prior distribution*, while the resulting mixed distribution is termed as the *predictive distribution*. Classical Bayesian analysis for Poisson models is confined to the conjugate case of a gamma prior distribution, leading to a negative binomial predictive distribution. Hierarchical Bayes models allow for imposing hyperparameters to the parameters of the gamma density so as to remove subjectivity (George *et al.*, 1993). The form of the resulting predictive distributions is more complicated. Markov Chain Monte Carlo (MCMC) methodologies have enhanced the plausibility of different models using different priors. For example, Maiti (1998) used a log-normal prior distribution, while the posterior distribution was derived via sampling based methods. From the empirical Bayes perspective, Gaver & O'Muircheartaigh (1987) considered gamma and log-Student types of prior distributions. Interestingly, as shown in section 3.8, the posterior moments of the parameter can be derived easily for a wide range of prior distributions. Moreover, the generalized inverse Gaussian distribution is also conjugate for the parameter of the Poisson distribution (Willmot & Sundt, 1989). Note that, the differences between various mixed Poisson distributions do not seem to have been examined. From the Bayesian point

Table 2.
Some bivariate and multivariate mixed Poisson distributions of the first kind.

Mixing Density	Resulting Distribution	A Key Reference	Comments
Gamma	Bivariate Negative Binomial	Edwards & Gurland (1961)	$\lambda_3 = 0$
		Subrahmaniam (1966)	$\lambda_3 \neq 0$
		Munkin & Trivedi (1999)	$\lambda_3 = 0$, includes covariates
	Multivariate Negative Binomial	Ashford & Hunt (1974)	$\lambda_3 = 0$
Generalized Inverse Gaussian		Stein *et al.* (1987)	$\lambda_3 \neq 0$
Confluent Hypergeometric		Bhattacharya (1966)	$\lambda_3 = 0$
Extended Gamma		Gurmu & Elder (2000)	$\lambda_3 = 0$, includes covariates

Table 3.
Some bivariate and multivariate mixed Poisson distributions of the second kind.

Mixing Density	Resulting Distribution	A Key Reference	Comments
Bivariate Normal		Steyn (1976)	$\lambda_3 = 0$, lacks physical interpretation
Bivariate Lognormal		Munkin & Trivedi (1999)	$\lambda_3 = 0$, includes covariates
Multivariate Lognormal		Aitchison & Ho (1989)	$\lambda_3 = 0$
		Chin & Winkelmann (2001)	$\lambda_3 = 0$, includes covariates
Gamma Mixed by a Beta II	Bivariate Generalized Waring Distribution	Xekalaki (1984a)	$\lambda_3 = 0$
Multivariate Gamma with Independent Components Mixed by a Univariate Beta II	Multivariate Generalized Waring Distribution	Xekalaki (1986)	$\lambda_3 = 0$
Wishart		Goutis & Galbraith (1996)	This is the only case where λ_3 is treated as random variable

of view, examining the effect of using different prior distributions can be considered as equivalent to examining the robustness of the prior distributions.

References

Adell, J. & de la Cal, J. (1993). On the Uniform Convergence of Normalised Poisson Mixtures to their Mixing Distribution. *Statistics and Probability Letters*, **18**, 227–232.

Aitchinson, J. & Ho, C.H. (1989). The Multivariate Poisson-Log Normal Distribution. *Biometrika*, **75**, 621–629.

Aitkin, M. (1996). A General Maximum Likelihood Analysis of Overdispersion in Generalized Linear Models. *Statistics and Computing*, **6**, 251–262.

Albrecht, P. (1982). On Some Statistical Methods Connected with the Mixed Poisson Process. *Scandinavian Actuarial Journal*, **9**, 1–14.

Albrecht, P. (1984). Laplace Transforms, Mellin Transforms and Mixed Poisson Processes. *Scandinavian Actuarial Journal*, **11**, 58–64.

Al-Awadhi, S.A. & Ghitany, M.E. (2001). Statistical Properties of Poisson–Lomax Distribution and its Application to Repeated Accidents Data. *Journal of Applied Statistical Science*, **10**, 365–372.

Al-Hussaini, E.K. & El-Dab, A.K. (1981). On the Identifiability of Finite Mixtures of Distributions. *IEEE Transactions on Information Theory*, **27**, 664–668.

Al-Zaid, A.A. (1989). On the Unimodality of Mixtures. *Pakistan Journal of Statistics*, **5**, 205–209.

Arbous, A.G. & Kerrich, J.E. (1951). Accident Statistics and the Concept of Accident-Proneness. *Biometrics*, **7**, 340–432.

Ashford J.R. & Hunt R.G. (1974). The Distribution of Doctor-Patient Contacts in the National Health Service. *Journal of the Royal Statistical Society* A, **137**, 347–383.

Barndorff-Nielsen, O.E. (1965). Identifiability of Mixtures of Exponential Families. *Journal of Mathematical Analysis and Applications*, **12**, 115–121.

Barndorff-Nielsen, O.E., Blaesild, P. & Seshardi, V. (1992). Multivariate Distributions with Generalized Inverse Gaussian Marginals and Associated Poisson Mixtures. *Canadian Journal of Statistics*, **20**, 109–120.

Bates, G. & Neyman, J. (1952a). Contributions to the Theory of Accident Proneness Part II: True Or False Contagion? *University of California Publications in Statistics*, pp. 255–275.

Bates, G. & Neyman, J. (1952b). Contributions to the Theory of Accident Proneness Part I: An Optimistic Model of the Correlation Between Light and Severe Accidents. *University of California Publications in Statistics*, pp. 215–253.

Bertin, E. & Theodorescu, R. (1995). Preserving Unimodality by Mixing. *Statistics and Probability Letters*, **25**, 281–288.

Best, A. & Gipps, B. (1974). An Improved Gamma Approximation to the Negative Binomial. *Technometrics*, **16**, 621–624.

Bhattacharya, S.K. (1966). Confluent Hypergeometric Distributions of Discrete and Continuous Type with Application to Accident Proneness. *Bulletin of the Calcutta Statistical Association*, **15**, 20–31.

Bhattacharya, S.K. (1967). A Result in Accident Proneness. *Biometrika*, **54**, 324–325.

Böhning, D. (1999). *Computer Assisted Analysis of Mixtures (C.A.M.AN)*. New York: Marcel Dekker Inc.

Buchmann, B. & Grubel, R. (2003). Decompounding: an Estimation Problem for Poisson Random Sums. *Annals of Statistics*, **31**, 1054–1074.

Bulmer, M.G. (1974). On Fitting the Poisson Lognormal Distribution to Species Abundance Data. *Biometrics*, **30**, 101–110.

Burrel, Q. & Cane, V. (1982). The Analysis of Library Data. *Journal of the Royal Statistical Society* A, **145**, 439–471.

Cane, V. (1977). A Class of Non-Identifiable Stochastic Models. *Journal of Applied Probability*, **14**, 475–482.

Carriere, J. (1993). Nonparametric Tests for Mixed Poisson Distributions. *Insurance Mathematics and Economics*, **12**, 3–8.

Cassie, M. (1964). Frequency Distributions Models in the Ecology of Plankton and Other Organisms. *Journal of Animal Ecology*, **31**, 65–92.

Chadjiconstantinidis, S & Antzoulakos, D.L. (2002). Moments of Compound Mixed Poisson Distributions. *Scandinavian Actuarial Journal*, (3) 138–161.

Chen, J. (1995). Optimal Rate of Convergence for Finite Mixture Models. *Annals of Statistics*, **23**, 221–233.

Chen, J. & Ahn, H. (1996). Fitting Mixed Poisson Regression Models Using Quasi-Likelihood Methods. *Biometrical Journal*, **38**, 81–96.

Chib, S. & Winkelmann, R. (2001). Markov Chain Monte Carlo Analysis of Correlated Count Data. *Journal of Business and Economic Statistics*, **19**, 428–435.

Cox, D. (1983). Some Remarks on Overdispersion. *Biometrika*, **70**, 269–274.

Cressie, N. (1982). A Useful Empirical Bayes Identity. *Annals of Statistics*, **10**, 625–629.

De Vylder, F. (1989). Compound and Mixed Distributions. *Insurance Mathematics and Economics*, **8**, 57–62.

Dean, C.B., Lawless, J. & Willmot, G.E. (1989). A Mixed Poisson-Inverse Gaussian Regression Model. *Canadian Journal of Statistics*, **17**, 171–182.

Denuit, M., Lefevre, C. & Shaked, M. (2001). Stochastic Convexity of the Poisson Mixture Model. *Methodology and Computing in Applied Probability*, **2**, 231–254.

Devroye, L. (1993). A Triptych of Discrete Distributions Related to the Stable Law. *Statistics and Probability Letters*, **18**, 349–351.

Douglas, J.B. (1980). *Analysis with Standard Contagious Distributions*. Statistical Distributions in Scientific Work Series 4. Fairland, MD, USA: International Cooperative Publishing House.

Edwards, C.B. & Gurland, J. (1961). A Class of Distributions Applicable to Accidents. *Journal of the American Statistical Association*, **56**, 503–517.

Everitt, B.S. & Hand, D.J. (1981). *Finite Mixtures Distributions*. Chapman and Hall.

Feller, W. (1943). On a Generalized Class of Contagious Distributions. *Annals of Mathematical Statistics*, **14**, 389–400.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*. Vol I, 3rd Edition. New York: John Wiley and Sons.
Gaver, D. & O'Muircheartaigh, I.G. (1987). Robust Empirical Bayes Analyses of Event Rates. *Technometrics*, **29**, 1–15.
Gelfand, A. & Dalal, S. (1990). A Note on Overdispersed Exponential Families. *Biometrika*, **77**, 55–64.
George, E., Makov, U. & Smith, A.F.M. (1993). Conjugate Likelihood Distributions. *Scandinavian Journal of Statistics*, **20**, 147–156.
Gerber, H.U. (1991). From the Generalized Gamma to the Generalized Negative Binomial Distribution. *Insurance Mathematics and Economics*, **10**, 303–309.
Ghitany, M.E. & Al-Awadhi, M.E. (2001). A Unified Approach to Some Mixed Poisson Distributions. *Tamsui Oxford Journal of Mathematical Sciences*, **17**, 147–161.
Goutis, C. & Galbraith, R.F. (1996). A Parametric Model for Heterogeneity in Paired Poisson Counts. *Canadian Journal of Statistics*, **24**, 569–581.
Grandell, J. (1997). *Mixed Poisson Processes*. Chapman and Hall.
Greenwood, M. & Yule, G. (1920). An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents. *Journal of the Royal Statistical Society* A, **83**, 255–279.
Gupta, S. & Huang, W.T. (1981). On Mixtures of Distributions: A Survey and Some New Results on Ranking and Selection. *Sankhyā* B, **43**, 245–290.
Gupta, R.C. & Ong, S.H. (2004). A New Generalization of the Negative Binomial Distribution. *Computational Statistics and Data Analysis*, **45**, 287–300.
Gurland, J. (1957). Some Interrelations among Compound and Generalized Distributions. *Biometrika*, **44**, 263–268.
Gurland, J. (1958). A Generalized Class of Contagious Distributions. *Biometrics*, **14**, 229–249.
Gurmu, S. & Elder, J. (2000). Generalized Bivariate Count Data Regression Models. *Economics Letters*, **68**, 31–36.
Haight, F.A. (1965). On the Effect of Removing Persons with *n* or More Accidents from an Accident Prone Population. *Biometrika*, **52**, 298–300.
Haight, F.A. (1967). *Handbook of Poisson Distributions*. New York: John Wiley and Sons.
Hall, P. (1979). On Measures of the Distance of a Mixture from its Parent Distribution. *Stochastic Processes and Applications*, **8**, 357–365.
Hesselager, O. (1994). A Recursive Procedure for Calculation of Some Compound Distributions. *ASTIN Bulletin*, **24**, 19–32.
Hesselager, O. (1996). A Recursive Procedure for Calculation of Some Mixed Compound Poisson Distributions. *Scandinavian Actuarial Journal*, (1), 54–63.
Hinde, J. & Demetrio, C.G.B. (1998). Overdispersion: Models and Estimation. *Computational Statistics and Data Analysis*, **27**, 151–170.
Holgate, P. (1970). The Modality of Some Compound Poisson Distributions. *Biometrika*, **57**, 666–667.
Holla, M.S. & Bhattacharya, S.K. (1965). On a Discrete Compound Distribution. *Annals of the Institute of Statistical Mathematics*, **15**, 377–384.
Hougaard, P., Lee, M.L.T. & Whitmore, G.A. (1997). Analysis of Overdispersed Count Data by Mixtures of Poisson Variables and Poisson Processes. *Biometrics*, **53**, 1225–1238.
Irwin, J. (1975). The Generalized Waring Distribution Parts I, II, III. *Journal of the Royal Statistical Society* A, **138**, 18–31 (Part I), 204–227 (Part II), 374–384 (Part III).
Johnson, N.L. (1957). Uniqueness of a Result in the Theory of Accident Proneness. *Biometrika*, **44**, 530–531.
Johnson, N.L. (1967). Note on a Uniqueness of a Result in the Theory of Accident Proneness. *Journal of the American Statistical Association*, **62**, 288–289.
Johnson, N.L., Kotz, S. & Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. New York: John Wiley and Sons.
Johnson, N.L., Kotz, S. & Kemp, A.W. (1992). *Univariate Discrete Distributions*. 2nd Edition. New York: John Wiley and Sons.
Karlis, D. (1998). Estimation and Testing Problems in Poisson Mixtures. Ph.D. Thesis, Department of Statistics, Athens University of Economics. ISBN: 960-7929-19-5.
Karlis, D. & Xekalaki, E. (2000). On some Distributions Arising from the Triangular Distribution. Technical Report 111, Department of Statistics, Athens University of Economics and Business, August 2000.
Katti, S. (1966). Interrelations among Generalized Distributions and their Components. *Biometrics*, **22**, 44–52.
Kemp, A.W. (1995). Splitters, Lumpers and Species per Genus. *Mathematical Scientist*, **20**, 107–118.
Kemp, C.D. (1967). Stuttering Poisson Distributions. *Journal of the Statistical and Social Enquiry Society of Ireland*, **21**, 151–157.
Kemperman, J.H.B. (1991). Mixtures with a Limited Number of Modal Intervals. *Annals of Statistics*, **19**, 2120–2144.
Kempton, R.A. (1975). A Generalized Form of Fisher's Logarithmic Series. *Biometrika*, **62**, 29–38.
Khaledi, B.E. & Shaked, M. (2003). Bounds on the Kolmogorov Distance of a Mixture from its Parent Distribution. *Sankhyā*, A, **65**, 317–332.
Kling, B. & Goovaerts, M. (1993). A Note on Compound Generalized Distributions. *Scandinavian Actuarial Journal*, **20**, 60–72.
Kocherlakota, S. (1988). On the Compounded Bivariate Poisson Distribution: A Unified Treatment. *Annals of the Institute of Statistical Mathematics*, **40**, 61–76.
Kocherlakota, S. & Kocherlakota, K. (1992). *Bivariate Discrete Distributions*. New York: Marcel Dekker Inc.
Lawless, J. (1987). Negative Binomial and Mixed Poisson Regression. *Canadian Journal of Statistics*, **15**, 209–225.
Lindsay, B. (1995). *Mixture Models: Theory, Geometry and Applications. Regional Conference Series in Probability and Statistics*, Vol 5. Institute of Mathematical Statistics and American Statistical Association.

Lindsay, B. & Roeder, K. (1993). Uniqueness of Estimation and Identifiability in Mixture Models. *Canadian Journal of Statistics*, **21**, 139–147.
Lynch, J. (1988). Mixtures, Generalized Convexity and Balayages. *Scandinavian Journal of Statistics*, **15**, 203–210.
Maceda, E.C. (1948). On the Compound and Generalized Poisson Distributions. *Annals of Mathematical Statistics*, **19**, 414–416.
Maiti, T. (1998). Hierarchical Bayes Estimation of Mortality Rates for Disease Mapping. *Journal of Statistical Planning and Inference*, **69**, 339–348.
McFadden, J.A. (1965). The Mixed Poisson Process. *Sankhyā*, A, **27**, 83–92.
McLachlan, G. & Basford, K. (1988). *Mixture Models: Inference and Application to Clustering*. New York: Marcel Dekker Inc.
McLachlan, J.A. & Peel, D. (2000). *Finite Mixture Models*. New York: John Wiley and Sons.
McNeney, B. & Petkau, J. (1994). Overdispersed Poisson Regression Models for Studies of Air Pollution and Human Health. *Canadian Journal of Statistics*, **22**, 421–440.
Mellinger, G.D., Sylwester, D.L., Gaffey, W.R. & Manheimer, D.I. (1965). A Mathematical Model with Applications to a Study of Accident Repeatedness among Children. *Journal of the American Statistical Association*, **60**, 1046–1059.
Misra, N., Singh, H. & Harner, E.J (2003). Stochastic Comparisons of Poisson and Binomial Random Variables with their Mixtures. *Statistics and Probability Letters*, **65**, 279–290.
Molenaar, W. & Van Zwet, W. (1966). On Mixtures of Distributions. *Annals of Mathematical Statistics*, **37**, 201–203.
Munkin, M.K. & Trivedi, P.K. (1999). Simulated Maximum Likelihood Estimation of Multivariate Mixed-Poisson Regression Models, with Applications. *Econometrics Journal*, **2**, 29–48.
Nelson J. (1985). Multivariate Gamma-Poisson Models. *Journal of the American Statistical Association*, **80**, 828–834.
Neuts, M.F. & Ramalhoto, M.F. (1984). A Service Model in Which the Server is Required to Search for Customers. *Journal of Applied Probability*, **21**, 157–166.
Nichols, W.G. & Tsokos, C. (1972). Empirical Bayes Point Estimation in a Family of Probability Distributions. *International Statistical Review*, **40**, 147–151.
Ong, S.H. (1995). Computation of Probabilities of a Generalized Log-Series and Related Distributions. *Communication in Statistics-Theory and Methods*, **24**, 253–271.
Ong, S.H. (1996). On a Class of Discrete Distributions Arising from the Birth-Death with Immigration Process. *Metrika*, **43**, 221–235.
Ong, S.H. & Muthaloo, S. (1995). A Class of Discrete Distributions Suited to Fitting Very Long Tailed Data. *Communication in Statistics-Simulation and Computation*, **24**, 929–945.
Ord, K. (1967). Graphical Methods for a Class of Discrete Distributions. *Journal of the Royal Statistical Society* A, **130**, 232–238.
Ord, K. (1972). *Families of Frequency Distributions*. London: Griffin.
Ospina, V. & Gerber, H.U. (1987). A Simple Proof of Feller's Characterization of the Compound Poisson Distribution. *Insurance Mathematics and Economics*, **6**, 63–64.
Panjer, H. (1981). Recursive Evaluation of a Family of Compound Distributions. *ASTIN Bulletin*, **18**, 57–68.
Papageorgiou, H. & Wesolowski, J. (1997). Posterior Mean Identifies the Prior Distribution in NB and Related Models. *Statistics and Probability Letters*, **36**, 127–134.
Patil, G.P. (1964). On Certain Compound Poisson and Compound Binomial Distributions. *Sankhā* A, **27**, 293–294.
Patil, G.P. & Rao, C.R. (1978). Weighted Distributions and Size-Biased Sampling with Applications to Wildlife Populations and Human Families. *Biometrics*, **34**, 179–189.
Patil, G.P., Rao, C.R. & Ratnaparkhi, M.V. (1986). On Discrete Weighted Distributions and Their Use in Model Choice for Observed Data. *Communication in Statistics-Theory and Methods*, **15**, 907–918.
Perline, R. (1998). Mixed Poisson Distributions Tail Equivalent to their Mixing Distributions. *Statistics and Probability Letters*, **38**, 229–233.
Pfeifer, D. (1987). On the Distance between Mixed Poisson and Poisson Distributions. *Statistics and Decision*, **5**, 367–379.
Pielou, E. (1962). Run of One Species with Respect to Another in Transects through Plant Populations. *Biometrics*, **18**, 579–593.
Press, W., Teukolsky, S., Vetterling, W., & Flannery, B. (1992). *Numerical Recipes in FORTRAN: the Art of Scientific Computing*, 2nd Edition. Cambridge University Press.
Quenouille, M.H. (1949). A Relation between the Logarithmic, Poisson and Negative Binomial Series. *Biometrics*, **5**, 162–164.
Rai, G. (1971). A Mathematical Model for Accident Proneness. *Trabajos Estadística*, **22**, 207–212.
Rao, C.R. (1965). On Discrete Distributions Arising out of Methods of Ascertainment. In *Classical and Contagious Discrete Distributions*, Ed. G.P. Patil, pp. 320–332. Pergamon Press and Statistical Publishing Society, Calcutta.
Redner, R. & Walker, H. (1984). Mixture Densities, Maximum Likelihood and the EM Algorithm. *SIAM Review*, **26**, 195–230.
Remillard, B. & Theodorescu, R. (2000). Inference Based on the Empirical Probability Generating Function for Mixtures of Poisson Distributions. *Statistics and Decisions*, **18**, 349–266.
Roos, B. (2003). Improvements in the Poisson Approximations of Mixed Poisson Distributions. *Journal of Statistical Planning and Inference*, **113**, 467–483.
Ruohonen, M. (1988). A Model for the Claim Number Process. *ASTIN Bulletin*, **18**, 57–68.
Sankaran, M. (1969). On Certain Properties of a Class of Compound Poisson Distributions. *Sankhā* B, **32**, 353–362.
Sankaran, M. (1970). The Discrete Poisson–Lindley Distribution. *Biometrics*, **26**, 145–149.
Sapatinas, T. (1995). Identifiability of Mixtures of Power Series Distributions and Related Characterizations. *Annals of the Institute of Statistical Mathematics*, **47**, 447–459.
Schweder, T. (1982). On the Dispersion of Mixtures. *Scandinavian Journal of Statistics*, **9**, 165–169.

Seshadri, V. (1991). Finite Mixtures of Natural Exponential Families. *Canadian Journal of Statistics*, **19**, 437–445.
Shaked, M. (1980). On Mixtures from Exponential Families. *Journal of the Royal Statistical Society* B, **42**, 192–198.
Sibuya, M. (1979). Generalized Hypergeometric, Digamma and Trigamma Distributions. *Annals of the Institute of Statistical Mathematics*, **31**, 373–390.
Sichel, H.S. (1974). On a Distribution Representing Sentence—Length in Written Prose. *Journal of the Royal Statistical Society* A, **137**, 25–34.
Sichel, H.S. (1975). On a Distribution Law for Word Frequencies. *Journal of the American Statistical Association*, **70**, 542–547.
Simon, P. (1955). On a Class of Skew Distributions. *Biometrika*, **42**, 425–440.
Sivaganesan, S. & Berger, J.O. (1993). Robust Bayesian Analysis of the Binomial Empirical Bayes Problem. *The Canadian Journal of Statistics*, **21**, 107–119.
Stein, G.Z., Zucchini, W. & Juritz, J.M. (1987). Parameter Estimation for the Sichel Distribution and its Multivariate Extension. *Journal of the American Statistical Association*, **82**, 938–944.
Stein, G. & Juritz, J.M. (1987). Bivariate Compound Poisson Distributions. *Communications in Statistics-Theory and Methods*, **16**, 3591–3607.
Steyn, H. (1976). On the Multivariate Poisson Normal Distribution. *Journal of the American Statistical Association*, **71**, 233–236.
Subrahmaniam, K. (1966). A Test for Intrinsic Correlation in the Theory of Accident Proneness. *Journal of the Royal Statistical Society* B, **28**, 180–189.
Sundt, B. (1999). *An Introduction to Non-Life Insurance Mathematics*, 4th Edition. Karlsruhe: University of Mannheim Press.
Tallis, G.M. (1969). The Identifiability of Mixtures of Distributions. *Journal of Applied Probability*, **6**, 389–398.
Teicher, H. (1961). Identifiability of Mixtures. *Annals of Mathematical Statistics*, **26**, 244–248.
Teicher, H. (1963). Identifiability of Finite Mixtures. *Annals of Mathematical Statistics*, **28**, 75–88.
Titterington, D.M., Smith, A.F.M. & Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions*. New York: John Wiley and Sons.
Titterington, D.M. (1990). Some Recent Research in the Analysis of Mixture Distributions. *Statistics*, **21**, 619–641.
Wahlin, J.F. & Paris, J. (1999). Using Mixed Poisson Distributions in Connection with Bonus-Malus Systems. *ASTIN Bulletin*, **29**, 81–99.
Wang, S. & Panjer, H. (1993). Critical Starting Points for Stable Evaluation of Mixed Poisson Probabilities. *Insurance Mathematics and Economics*, **13**, 287–297.
Wang, S. & Sobrero, M. (1994). Further Results on Hesselager's Recursive Procedure for Calculation of Some Compound Distributions. *ASTIN Bulletin*, **24**, 160–166.
Wang, P., Puterman, M., Cockburn, I. & Le, N. (1996). Mixed Poisson Regression Models with Covariate Dependent Rates. *Biometrics*, **52**, 381–400.
Willmot, G.E. (1986). Mixed Compound Poisson Distribution. *ASTIN Bulletin Supplement*, **16**, 59–79.
Willmot, G.E. (1990). Asymptotic Tail Behaviour of Poisson Mixtures with Applications. *Advances in Applied Probability*, **22**, 147–159.
Willmot, G.E. (1993). On Recursive Evaluation of Mixed Poisson Probabilities and Related Quantities. *Scandinavian Actuarial Journal*, **18**, 114–133.
Willmot, G.E. & Sundt, B. (1989). On Posterior Probabilities and Moments in Mixed Poisson Processes. *Scandinavian Actuarial Journal*, **14**, 139–146.
Xekalaki, E. (1981). Chance Mechanisms for the Univariate Generalized Waring Distribution and Related Characterizations. In *Statistical Distributions in Scientific Work*, Eds. C. Taillie, G.P. Patil and B. Baldessari, **4**, 157–171. The Netherlands: D. Reidel Publishing Co.
Xekalaki, E. (1983a). The Univariate Generalized Waring Distribution in Relation to Accident Theory: Proneness, Spells Or Contagion? *Biometrics*, **39**, 887–895.
Xekalaki, E. (1983b). Infinite Divisibility, Completeness and Regression Properties of the Univariate Generalized Waring Distribution. *Annals of the Institute of Statistical Mathematics*, **32**, 279–289.
Xekalaki, E. (1984a). The Bivariate Generalized Waring Distribution and its Application to Accident Theory. *Journal of the Royal Statistical Society* A, **147**, 488–498.
Xekalaki, E. (1984b). Models Leading to the Bivariate Generalized Waring Distribution. *Utilitas Mathematica*, **35**, 263–290.
Xekalaki, E. (1986). The Multivariate Generalized Waring Distribution. *Communications in Statistics*, **15**, 1047–1064.
Xekalaki, E. & Panaretos, J. (1983). Identifiability of Compound Poisson Distributions. *Scandinavian Actuarial Journal*, 39–45.
Xue, D. & Deddens, J. (1992). Overdispersed Negative Binomial Models. *Communications in Statistics-Theory and Methods*, **21**, 2215–2226.
Yakowitz, S. & Spragins, J. (1969). On the Identifiability of Finite Mixtures. *Annals of Mathematical Statistics*, **39**, 209–214.

Résumé

Les distributions Poissoniennes mixées ont été utilisées dans plusieurs regions scientifiques pour modeler des populations inhomogènes. Cet article surveille la literature existante sur ces modèles en presentant un grand nombre de propriétés et en donnant d'information tangentiale sur formes des distributions mixées plus généraux. Quelques bien connus modèles Poissoniens mixés sont presentés.

[*Received July 2003, accepted August 2004*]