

REVIEW

Open Access

# On the distribution theory of over-dispersion

Evdokia Xekalaki

Correspondence: exek@aueb.gr  
Department of Statistics, Athens  
University of Economics and  
Business, 76 Patision St., Athens,  
Greece

## Abstract

An overview of the evolution of probability models for over-dispersion is given looking at their origins, motivation, first main contributions, important milestones and applications. A specific class of models called the Waring and generalized Waring models will be a focal point. Their advantages relative to other classes of models and how they can be adapted to handle multivariate data and temporally evolving data will be highlighted.

**Keywords:** Heterogeneity; Contagion; Clustering; Spells model; Accident proneness; Mixtures; Zero-adjusted models; Biased sampling; Generalized Waring distribution; Generalized Waring process

## 1. Introduction

Data analysts have often to deal with data that exhibit a variability that differs from what they expect on the basis of the hypothesized model. The phenomenon is known as *overdispersion* if the observed variability exceeds the expected variability or *underdispersion* if it is lower than expected.

Such differences between observed and nominal variances can be interpreted as brought about by failures of some of the basic assumptions of the model. These can be classified by the mechanism leading to them. As summarized by Xekalaki (2006), in traditional experimental contexts, they may be caused by deviations from the hypothesized structure of the population, due to lack of independence between individual item responses, contagion, clustering, and heterogeneity. In observational study contexts, on the other hand, they are the result of the method of ascertainment, which can lead to partial distortion of the observations. In both contexts, the observed value  $x$  no longer represents an observation on the original variable  $X$ , but constitutes an observation on a random variable  $Y$  whose distribution (*the observed distribution*) is a distorted version of the distribution of  $X$  (*original distribution*).

Such practical situations have been noticed since over a century ago (e.g. Lexis 1879; Student 1919). The *Lexis ratio* appears to be the first statistic suggested for testing for the presence of over- or under-dispersion relative to a binomial hypothesized model in populations structured in clusters. Also, for count data, Fisher (1950) considered using the sample index of dispersion for testing the appropriateness of a Poisson distribution for an observed variable  $Y$ .

The paper is structured as follows. Section 2 introduces the reader to the various approaches to modelling overdispersion in the case of traditional experimental contexts. Section 3 highlights approaches in the case of observational study contexts. Section 4

focuses on the case of heterogeneous populations followed by Sections 5 and 6, which look into a particular type of distribution, the generalized Waring distribution, and its relevance in the context of applications under the various scenaria leading to overdispersion mentioned above. Through the prism of these scenaria, a bivariate version of it is also presented, and its use in applied contexts is discussed in Section 7. A multivariate version of it is also given, and its application potential is outlined in Section 8. Finally, Sections 9 and 10 present a model for temporally evolving data, the multivariate generalized Waring process, and an application illustrating its practical potential.

As the field of accident studies has received much attention, and various theories have been developed for the interpretation of factors underlying an accident situation, most of the models will be presented in accident or actuarial data analysis contexts. Of course the results can be adapted in a great variety of situations with appropriate parameter interpretations so that they can be applied in several other fields ranging from economics, inventory control and insurance through to demometry, biometry, psychometry and web access modeling, as the case is with the application discussed in Section 10.

## 2. Modelling over - or under - dispersion in traditional experimental contexts

One important, but often ignored by data analysts, implication of using single parameter distributions such as the Poisson distribution to analyse data is that the variance can be determined by the mean, a relation that collapses by the presence of overdispersion. If this is ignored in practice, any form of statistical inference may induce low efficiency, although, for modest amounts of overdispersion this may not be the case (Cox 1983). So, insight into the mechanisms that induce over (or under) dispersion is required when dealing with such data. Such insight can be gained by looking at the above-mentioned potential triggering sources as classified by Xekalaki (2006).

### 2.1 Lack of independence between individual responses

In accident study related contexts, where one is interested in the total number of re-

ported accidents  $Y = \sum_{i=1}^n Y_i$  in a total number of accidents,  $n$ , that actually occurred,

when accidents are reported with equal probabilities  $p = P(Y_i = 1) = 1 - P(Y_i = 0)$ , but not independently ( $Cor(Y_i, Y_j) = \rho \neq 0$ ), the mean of  $Y$  will still be  $E(Y) = np$ , but its vari-

ance will be  $V(Y) = V\left(\sum_{i=1}^n Y_i\right) = np(1-p) + 2\binom{n}{2}\rho p(1-p) = np(1-p)(1 + \rho(n-1))$ ,

which exceeds that anticipated under a hypothesized independent trial binomial model if  $\rho > 0$  (over-dispersion) and is exceeded by it if  $\rho < 0$  (under-dispersion).

### 2.2 Contagion

Another common reason for a variance differing from what is anticipated, is that when the assumption that the probability of the occurrence of an event in a very short interval is constant fails. This framework is the classical *contagion* model (Greenwood and Yule 1920; Xekalaki 1983a).

In data modelling problems faced by actuaries, for example, this model postulates that initially all individuals have the same probability of incurring an accident, but later

this probability changes by each accident sustained. It is assumed, specifically, that none of the individuals has had an accident (e.g. new drivers or persons who are just beginning a new type of work), but later the probability with which a person with  $Y = y$  accidents by time  $t$  will have another accident in the time period from  $t$  to  $t + dt$  is of the form  $(k + my)dt$ . This leads to the negative binomial as the distribution of  $Y$  with p.f.  $P(Y=y) = \binom{k/m}{y} e^{-kt} (1 - e^{-mt})^y$  with  $\mu = E(Y) = k(e^{mt} - 1)/m$ , and  $V(Y) = ke^{mt}(e^{mt} - 1)/m = \mu e^{mt}$ .

### 2.3 Clustering

A frequently overlooked clustered structure of the population may also induce over- or under- dispersion.

In an accident context again, an accident is regarded as a *cluster* of injuries:

The number  $Y$  of injuries incurred by persons involved in  $N$  accidents can naturally be thought of as expressed by the sum  $Y = Y_1 + Y_2 + \dots + Y_N$  of the numbers  $Y_i$  of injuries resulting from the  $i$ -th accident, assumed to be i.i.d. independently of the total number of accidents  $N$ , with mean  $\mu$  and variance  $\sigma^2$ . In this case,  $E(Y) = E\left(\sum_{i=1}^N Y_i\right) = \mu E(N)$  and

$$V(Y) = V\left(\sum_{i=1}^N Y_i\right) = \sigma^2 E(N) + \mu^2 V(N).$$

So, when  $N$  is a Poisson variable with mean  $E(N) = \theta = V(N)$ , the last relationship leads to overdispersion or underdispersion according as  $\sigma^2 + \mu^2$  is greater or less than 1.

The first such model was introduced by Cresswell and Froggatt (1963) in a different accident context whereby each person is liable to spells of weak performance during which all of the persons accidents occur. So, if the number  $N$  of spells in a unit time period is Poisson distributed with mean  $\theta$ , and within spells a person can have 0 accidents with probability  $1 - m \log p$ ,  $m > 1/\log p$ ,  $0 < p < 1$  and  $n$  accidents ( $n \geq 1$ ) with probability  $m(1 - p)^n/n$ ,  $m, n > 0$  the observed distribution of accidents is the negative binomial distribution with probability function  $P(Y = y) = \binom{\theta m + y - 1}{y} p^{\theta m} (1 - p)^y$ .

This model, known in the literature as the *spells model*, can also lead to other forms of overdispersed distributions (e.g. Xekalaki 1983a, 1984a).

### 2.4 Heterogeneity

Assuming a homogeneous population when in fact the population is heterogeneous, i.e., when its individuals have constant, but unequal probabilities of sustaining an event can also lead to overdispersion. In this case, each member of the population has its own value of the parameter  $\theta$  and probability density function  $f(\cdot; \theta)$ .

So, with  $\theta$  regarded as the inhomogeneity parameter and varying from individual to individual according to any continuous, discrete, or finite step distribution  $G(\cdot)$  of mean  $\mu$  and variance  $\sigma^2$ , one is led to an observed distribution for  $Y$  with probability density function  $f_Y(y) = E_G(f(y; \theta)) = \int_{\Theta} f(y; \theta) dG(\theta)$ , where  $\Theta$  is the parameter space. Models of this type are known as *mixtures*. (For details on their application in the statistical literature see e.g. Karlis and Xekalaki 2003; McLachlan and Peel 2001; Titterton 1990). Under such models, the variance of  $Y$  consists of two additive components, one representing the variance part due to the variability of  $\theta$  and one due to the inherent variability of  $Y$

if  $\theta$  did not vary, *i.e.*,  $V(Y) = V(E(Y|\theta)) + E(V(Y|\theta))$ . This offers an explanation as to why mixture models are often referred to as *overdispersion models*.

It should be noted that a similar idea forms the basis for *analysis-of-variance* (ANOVA) models, where the total variability can be split into additive components, the *between groups* and the *within groups* components. In the case of the Poisson ( $\theta$ ) distribution, we have in particular that  $V(Y) = E(\theta) + V(\theta)$ . Based on the fact that in this case, the factorial moments of  $Y$  coincide with the moments of  $\theta$  about the origin, Carriere (1993) proposed a test of the hypothesis that a Poisson mixture fits a data set.

Mixed Poisson distributions were first introduced by Greenwood and Woods (1919) in the context of accident studies. Assuming that an individual's accident experience  $Y|\theta$  is Poisson distributed with parameter  $\theta$  that was varying from individual to individual according to a gamma distribution with mean  $\mu$  and index parameter  $\mu/\gamma$ , they obtained a negative binomial distribution for  $Y$  with probability function  $P(Y = y) = \binom{\mu/\gamma + y - 1}{y} \{\gamma/(1 + \gamma)\}^y (1 + \gamma)^{-\mu/\gamma}$  and with mean and variance given respectively by  $E(Y) = \mu$  and  $V(Y) = \mu(1 + \gamma)$ , where  $\gamma$  represents the over-dispersion parameter.

The mixed Poisson process has been popularised in the actuarial literature by Dubourdieu (1938) gamma mixed case was treated by Thyron (1969).

Numerous other mixtures have since then been proposed in the literature for interpreting overdispersion in data, such as binomial mixtures (e.g. Tripathi et al. 1994), negative binomial mixtures (e.g., Xekalaki 1983a, c, 1984a; Irwin 1975), normal mixtures (e.g. Andrews and Mallows 1974) and exponential mixtures (e.g. Jewell 1982). Discrete Poisson mixtures with finite step distributions for the Poisson parameter  $\theta$  have also been proposed, the interest being on creating clusters of data by grouping the observations on  $Y$  according to some criterion (*cluster analysis*). The number of clusters can be decided on the basis of a testing procedure for the number of components in the finite mixture (Karlis and Xekalaki 1999).

#### **2.4.1 Heterogeneity in mixture models treating the parameter $\theta$ as the dependent variable in a regression model**

Heterogeneity in models with explanatory variables can be modelled, by assuming that  $Y$  has a parameter  $\theta$  varying from individual to individual according to some regression model  $\theta = \eta(x; \beta) + \varepsilon$ , where  $x$  is a vector of explanatory variables,  $\beta$  is a vector of regression coefficients,  $\eta$  is a function of a known form and  $\varepsilon$  has some known distribution. Such models are known in the literature as *random effect models* and have been extensively studied within the broad family of Generalized Linear Models. As a simple example in the case of a single covariate, say  $X$ , consider data  $Y_i, i = 1, 2, \dots, n$  coming from a Poisson population with mean  $\theta$  determined by  $\log \theta = \alpha + \beta x + \varepsilon$  for some constants  $\alpha, \beta$  and with  $\varepsilon$  having a distribution with mean 0 and variance say  $\phi$ . In this case, the marginal distribution of  $Y$  is no longer the Poisson distribution. It is a mixed Poisson distribution, with some mixing distribution  $g(\cdot)$  clearly depending on the distribution of  $\varepsilon$ . In particular,  $Y \sim \text{Poisson}(te^{(\alpha + \beta x)}) \hat{g}(t)$  where  $t = e^\varepsilon$ .

Negative Binomial and Poisson Inverse Gaussian regression models have also been proposed as overdispersed alternatives to the Poisson regression model (e.g. Lawless 1987; Dean et al. 1989; Xue and Deddens 1992). The case of a two finite step distribution, the finite Poisson mixture regression model of Wang et al.s (1996) results. The similarity of the

mixture representation and the random effects one is discussed in Hinde and Demetrio (1998).

In meta-analysis contexts, overdispersion (or underdispersion) refers to variance inflation (or deflation) relative to that anticipated by the fixed effects model. Two possible causes of such phenomena are a population structure in clusters or mixing resulting in a compound distribution. Kulinskaya and Olkin (2014) proposed approaching the problem of specification of a random effects model in meta-analysis in terms of a multiplicative model for the distribution of the effect size parameters that allows inflation or deflation. The model considered was motivated by overdispersion induced by intra-class correlation in the model assumed for the distribution of the  $i$ -th effect size estimate. In particular, the variance of the estimator  $\hat{\theta}_i$  of the effect size parameter  $\theta_i$  in the  $i$ -th study is assumed to be of the form  $\sigma_{\hat{\theta}_i}^2 = (1 + \alpha(n_i)\gamma)\sigma_i^2$ , where  $\alpha(n_i)$  are some known functions of the sample sizes  $n_i$ ,  $\sigma_i^2$  is the within the  $i$ -th study variance,  $i = 1, 2, \dots, k$  and  $\gamma$  is interpreted as an intra class correlation parameter.

#### 2.4.2 Estimation and testing for overdispersion under mixture models

The structure of mixture models, including random effect models, entails different forms of variance-to-mean relationships. So, viewing the mean and variance of  $Y$  as represented by  $E(Y) = \mu(\beta)$ , and  $V(Y) = \sigma^2(\mu(\beta), \lambda)$  respectively for some parameters  $\beta$ ,  $\lambda$  a number of estimation approaches have been proposed in the literature based on moment methods (e.g. Breslow 1990; Lawless 1987; Moore 1986) and quasi or pseudo likelihood methods (e.g. Davidian and Carroll 1988; McCullagh and Nelder 1989; Nelder and Pregibon 1987). The above representation for the mean and variance of  $Y$  allows also estimation in the case of multiplicative overdispersion as in McCullagh and Nelder (1989).

Testing for the presence of overdispersion or underdispersion, on the other hand, can be done by means of asymptotic arguments. Let  $f(y; \theta)$  denote the density function of a random variable  $Y$  in the initial model. Cox (1983) showed that, under regularity conditions, the density of  $y$  in the overdispersed model,  $f_Y(y)$ , admits a representation of the form  $f_Y(y) = E_{\Theta}(f(y; \theta)) = f(y; \mu_{\theta}) + \frac{1}{2}\sigma_{\theta}^2 \frac{\partial^2 f(y; \mu_{\theta})}{\partial \mu_{\theta}^2} + O(1/n)$ , with  $\mu_{\theta} = E(\theta)$ ,  $\sigma_{\theta}^2 = V(\theta)$  and  $\Theta$  is the parameter space. This in turn implies that  $f_Y(y)$  can be put in the form  $f(y; \mu_{\theta})(1 + \varepsilon h(y, \phi_{\theta}))$ , where  $h(y, \phi_{\theta}) = \left[ \frac{\partial \log f(y; \mu_{\theta})}{\partial \mu_{\theta}} \right]^2 + \frac{\partial^2 \log f(y; \mu_{\theta})}{\partial \mu_{\theta}^2}$ .

This representation entails overdispersion if  $\varepsilon > 0$ , underdispersion if  $\varepsilon < 0$  and, of course, none of these complications if  $\varepsilon = 0$ . Cox (1983) suggested a testing procedure for the hypothesis  $\varepsilon = 0$ , which can be regarded as a general version of standard dispersion tests.

#### 2.5 Zero adjusted models

It would be interesting to note that another aspect of the population structure that is often responsible for the phenomenon of over-dispersion or under-dispersion is the presence of an excess or a scant number of zeros. Though the models discussed in Sections 2.3 and 2.4 may capture over-dispersion or under-dispersion rather well, they cannot capture excess or scarcity of zeros. In the literature, this question has been addressed by two types of models known as zero-inflated (or zero-deflated) models, and hurdle models. A unified representation of the models is provided by  $f(y; \omega) = \omega I_{\{0\}}(y) + (1 - \omega)f_Y(y)$ , where  $Y$  is the count variable,  $I_{\{0\}}(\cdot)$  is the indicator function and  $\omega$  is a constant, whose values, if in  $(0,1)$

render a hurdle model for  $f_Y(0) = 0$ , a zero-inflated model for  $f_Y(0) \neq 0$ , while negative values of it render a zero-deflated model.

Obviously,  $\omega$  can be interpreted as the proportion of excess zeros in the case of the first two models and the above representation explains why there can be regarded as having a dual nature. They are (finite) mixtures, which account for heterogeneity, while at the same time, they are capturing a population structure in two clusters. However, in the case  $\omega < 0$  (zero-deflation), the model ceases to admit a mixture interpretation.

Zero-inflated and hurdle models have mostly been used for Poisson, generalized Poisson or negative binomial count distributions in various contexts (e.g. Ridout et al. 2001; Gupta et al. 2004; Famoye and Singh 2006). Gupta et al. (1996) proposed a zero-adjusted generalized Poisson distribution and studied the effect of not using an adjusted model for zero-inflation or -deflation when the occurrence of zeroes differs from the anticipated one. Reviews of such models can be found in Ridout et al. (1998), Gschlößl and Czado (2008) and Ngatchou-Wandji and Paris (2011).

### 3. Over- or under-dispersion in observational study contexts - the effect of the method of ascertainment

Often, in connection with data collection based on observation or on recording values as produced by nature, the original distribution may not be reproduced due to various reasons. These may lead to partial destruction or partial enhancement (augmentation) of observations. The models that have been introduced to deal with such situations are respectively known as *damage models* introduced by Rao (1963) and *generating models* introduced by Panaretos (1983). The distortion mechanism is usually assumed to be manifested through the conditional distribution of the resulting random variable  $Y$  given the value of the original random variable  $X$ . Hence, the resulting (observed) distribution is a distorted version of the original distribution that can be represented as a mixture of the distortion mechanism. In particular, in the case of damage,

$$P(Y = r) = \sum_{n=r}^{\infty} P(Y = r|X = n)P(X = n), \quad r = 0, 1, 2, \dots, \text{ while, in the case of enhancement,}$$

$$P(Y = r) = \sum_{n=1}^r P(Y = r|X = n)P(X = n), \quad r = 1, 2, \dots$$

Various forms of distributions have been considered for the distortion mechanism in the above two cases. In the case of damage, the most popular forms have been the binomial distribution Rao (1963), mixtures on  $p$  of the binomial distribution (e.g. Panaretos 1982; Xekalaki and Panaretos 1983) whenever damage can be regarded as additive ( $Y = X - U$ ,  $U$  independent of  $Y$ ) or in terms of the uniform distribution in  $(0, x)$  (e.g. Dimaki and Xekalaki 1990, 1996; Xekalaki 1984b) whenever damage can be regarded as multiplicative ( $Y = [RX]$ ,  $R$  independent of  $X$  and uniformly distributed in  $(0, 1)$ ). The latter case has also been considered in the context of continuous distributions by Krishnaji (1970). The generating model was introduced and studied by Panaretos (1983).

Both, the generating model and the damage model offer a perceptive approach in actuarial contexts where one is interested in modelling the distributions of the numbers of accidents, of the damage claims, and of the claimed amounts. These models become relevant due to the fact that people have in general a tendency to under report their accidents, so that the reported (observed) number  $Y$  is less than or equal to the actual number



$X$  ( $Y \leq X$ ), but tend to over report damages incurred by them, so that the reported damage  $Y$  is greater than or equal to the true damage  $X$  ( $Y \geq X$ ).

Another type of distortion is induced by the adoption of a sampling scheme that assigns to the units in the original distribution unequal probabilities of inclusion in the sample. As a result, the value  $x$  of  $X$  is observed with a frequency that noticeably differs from that anticipated under the original density function  $f_X(x; \theta)$ . It represents an observation on a random variable  $Y$  whose probability distribution is the results of adjusting the probabilities of the anticipated distribution through weighting them with the probability with which the value  $x$  of  $X$  is included in the sample. So, if this probability is proportional to some *weight function*,  $w(x, \beta)$ ,  $\beta \in R$ , the recorded value  $x$  is a value of  $Y$  having density function  $f_Y(x; \theta, \beta) = w(x; \beta)f_X(x; \theta)/E(w(X; \beta))$ .

Distributions of this type are known as *weighted distributions* (see, e.g. Cox 1962; Fisher 1934; Patil and Ord 1976; Rao 1985). For  $w(x; \beta) = x$ , these are known as *size biased distributions*. In actuarial data modelling contexts again, the weight function can represent reporting bias. In the context of reporting accidents or placing damage claims, for example, it can have a value that is directly or inversely analogous to the size  $x$  of  $X$ , the actual number of incurred accidents or the actual size of the incurred damage. The functions  $w(x; \beta) = x$  and  $w(x; \beta) = \beta^x$  ( $\beta > 1$  or  $\beta < 1$ ) are plausible choices. So, for example, in the case of a Poisson ( $\theta$ ) distributed  $X$ , these lead to distributions for  $Y$  that are of Poisson type. In particular, the weight function  $w(x; \beta) = x$  leads to a shifted Poisson distribution with probability function  $P(Y = x) = e^{-\theta} \theta^{x-1} / (x-1)!$ ,  $x = 1, 2, \dots$ , while the choice  $w(x; \beta) = \beta^x$  leads to a Poisson distribution  $P(Y = x) = e^{-\theta\beta} (\theta\beta)^x / x!$ ,  $x = 0, 1, \dots$ . The value of the variance of the observed variable  $Y$  under the first assumption for  $w(x; \beta)$  is  $1 + \theta$  and exceeds that of  $X$  (overdispersion), while under the second assumption it is  $\theta\beta$  implying overdispersion for  $\beta > 1$  or underdispersion for  $\beta < 1$ .

#### 4. Looking closer into the case of heterogeneity

Assuming a specific form for the distribution of the population that generated a data set implies that the mean to variance relation is given for this distribution, e.g. the Poisson distribution with a mean to variance ratio equal to unity. As has become obvious from the above, this relationship ceases to hold in real data sets however. This being rarely the case, flexible families have been sought in the literature by allowing the parameter  $\theta$  of the original distribution to vary according to a distribution with probability density function, say  $g(\cdot)$ .

As mentioned before, a density function  $f_X(\cdot)$  is a mixture on the parameter  $\theta$  of the distribution function  $f(\cdot; \theta)$  with some mixing distribution  $G_\theta(\cdot)$ , which can be continuous, discrete or a finite step distribution, if it can be written in the form  $f_X(x) = E_G(f(x; \theta)) = \int_{\Theta} f(x; \theta) dG(\theta)$ , where  $\Theta$  is the parameter space. An appropriate choice of a mixing distribution allows its parameter to vary and acts as a means of "loosening" the structure of the initial model, thus offering more realistic interpretations of the mechanisms that generated the data.

A large number of Poisson mixtures have been developed. (For an extensive review, see Karlis and Xekalaki 2003, 2005). The derivation of the negative binomial distribution, as a mixture of the Poisson distribution with a gamma distribution as the mixing distribution, originally obtained by Greenwood and Yule (1920) constitutes a typical

example. Mixtures of the negative binomial distribution have also been widely used in connection with applications in a plethora of fields. These include the Yule distribution (Yule 1924; Irwin 1941; Kekalaki 1983c, 1984b) the Waring distribution (Irwin 1963) and the generalized Waring distribution (Irwin 1968, 1975; Kekalaki 1981, 1983a, 1984a), which contains the Yule distribution and the Waring distribution as a special cases.

In what follows, we focus on the generalized Waring distribution and its relevance in accident data modeling contexts.

### 5. The generalized Waring distribution

This was introduced by Irwin (1968) in connection to biological data and later was shown by him to arise as an accident distribution (Irwin 1975). It is the distribution with probability generating function given by

$$G(s) = \frac{\rho_{(k)}}{(a + \rho)_{(k)}} {}_2F_1(a, k; a + k + \rho; s), \quad \alpha, k, \rho > 0$$

with  ${}_2F_1(a, b; c; z)$  denoting the Gauss hypergeometric function  $\sum_{r=0}^x \{a_{(r)}b_{(r)}z^r\} / \{c_{(r)}r!\}$ , where  $h_{(l)} = \Gamma(h + l)/\Gamma(h)$ ,  $h > 0, l \in \mathbb{R}$ .

Irwin's starting point was Waring's expansion (hence the distribution's name) given by  $\frac{1}{x-a} = \sum_{r=0}^{\infty} \frac{a_{(r)}}{x_{(r+1)}}$ , which he then generalized to  $\frac{1}{(x-a)_{(k)}} = \sum_{r=0}^{\infty} \frac{a_{(r)}k_{(r)}}{x_{(k+r)}r!}$ ,  $\alpha, k > 0$ .

Hence, by multiplying both sides by  $\rho_{(k)}$ , where  $\rho = x - a > 0$ , the successive terms of the resulting series could be regarded as defining a probability function, which he termed the generalized Waring distribution with parameters  $\alpha, k, \rho$ . In particular, the probability function of the generalized Waring distribution with parameters  $\alpha, k, \rho$  is given by

$$p_r = \frac{\rho_{(k)}}{(a + \rho)_{(k)}} \frac{a_{(r)}k_{(r)}}{(a + k + \rho)_{(r)}} \frac{1}{r!}, \quad \alpha, k, \rho > 0, r = 0, 1, 2, \dots$$

where  $h_{(l)} = \Gamma(h + l)/\Gamma(h)$ .

Notwithstanding the complexity of its structure, this distribution was shown to offer an insightful tool in the interpretation of accident data as will be seen below. Among its aspects that can be of practical value, is that, as shown by Kekalaki (1983b), it is a discrete self-decomposable distribution in Steutel and van Harns (1979) sense, hence infinitely divisible, implying that its probability generating function can be put in the form  $G(s) = \exp\left\{-\lambda \int_s^1 \frac{1-g(u)}{1-u} du\right\}$ , where  $\lambda = p_1/p_0$  and  $g(\cdot)$  denotes the probability generating function of the distribution with probability function satisfying the recurrence relation

$$q_n = \lambda \left\{ n \frac{ak + \rho(a + k + \rho)}{ak(a + k + \rho + n)} - \sum_{j=0}^{n-1} q_j \binom{n}{j} \binom{a + k + \rho + n - 1}{j} / \left[ \binom{a + n - 1}{j} \binom{k + n - 1}{j} \right] \right\}$$



## 6. The generalized Waring distribution in relation to accident theory

The hypotheses that have formed the basis of investigations into the occurrence of accidents since almost a century ago are

- (i) **Pure chance**, giving rise to the Poisson distribution
- (ii) **True contagion**, i.e. the hypothesis that initially all individuals have the same probability of incurring an accident but that this probability is modified by each accident sustained.
- (iii) **Apparent contagion (heterogeneity)**, i.e. the hypothesis that individuals have constant but unequal probabilities of having an accident - the resultant distribution being a compound Poisson distribution (**“accident proneness” model**).
- (iv) **The “Spells” Model**, i.e. each person is liable to periods of time during which the person’s performance is weak (spells). All of the person’s accidents occur within those spells. The numbers of accidents within different spells are independent and independent of the number of spells.

As already seen, the negative binomial distribution can be given a an accident proneness and a “spells” interpretation in the context of accident theory in terms of a gamma mixed Poisson distribution and a Poisson distribution generalized by a logarithmic distribution (Kemp 1967).

Therefore, a good fit of the negative binomial is no help at all in distinguishing among the proneness”, contagion” and “spells” hypotheses. This is known as the discrimination problem between the compounded, contagion and generalized models for the negative binomial distribution and has been discussed by Arbous and Kerrich (1951); Bates and Neyman (1952); Gurland (1959) and Cane (1974, 1977). For an extensive bibliography on the accident hypotheses mentioned, see Kemp (1970).

### 6.1 Irwin’s “Proneness” model

As evident, in all three of the above models, the data are treated as if the individuals under observation were exposed to equal environmental risk, a fact criticized by Irwin (1968), who suggested a three-parameter distribution, which he called the univariate generalized Waring distribution” (UGWD). He derived this distribution in a framework that allows separately for random factors, differences in the exposure of individuals to external risk of accident, and differences in proneness.

In particular, his model assumes a non homogeneous population with respect to personal and environmental attributes affecting the occurrence of accidents.

Let the distribution of the number,  $X$ , of accidents for individuals of equal proneness  $\nu$ , and of equal exposure to external risk of accident  $\lambda|\nu$ , i.e.  $\lambda$  for given  $\nu$ ), have probability generating function

$$G_{X|\lambda}(s) = \exp\{(\lambda|\nu)(s-1)\}$$

in a unit time interval  $(0, 1)$ . If the distributions of  $\lambda|\nu$  and  $\nu$  in the population at risk can be described by the probability density functions (pdf)

$$\{ \nu^{-k} \exp(-\lambda/\nu) \lambda^{k-1} \} / \Gamma(k), \nu, k > 0$$

and

$$\left\{ \Gamma(a + \rho)v^{a-1}(1 + v)^{-(a+\rho)} \right\} / \left\{ \Gamma(\rho)\Gamma(a) \right\}, a, \rho > 0$$

respectively, the pgf of the resulting distribution of accidents will be  $\{\rho_{(k)2}F_1(a, k; a + k + \rho; s)/(a + \rho)_{(k)}\}$ , i.e. the univariate generalized Waring distribution with parameters  $a$ ,  $k$  and  $\rho$ , which will be denoted by  $UGWD(a, k; \rho)$ . Here,  ${}_2F_1(a, b; c; z)$  denotes the Gauss hypergeometric function  $\sum_{r=0}^x \{a_{(r)}b_{(r)}z^r\} / \{c_{(r)}r!\}$ , where  $h_{(l)} = \Gamma(h + l)/\Gamma(h)$ ,  $h > 0$ ,  $l \in \mathbb{R}$ . For more information about the UGWD the reader is referred to the work of Irwin (1963, 1968, 1975); Xekalaki (1981) and the references therein and Xekalaki (1983a).

### 6.2 The “Contagion” model

Xekalaki (1983a), extended the assumptions of the classical contagion model developed by Greenwood and Yule (1920) by considering a population of individuals exposed to varying accident risk.

In particular, assume that at time  $t = 0$  none of the individuals has had an accident. This would be true if, for example, with a population of new drivers or of individuals just beginning a new type of work. Suppose that during the time period from  $t$  to  $t + dt$  a person with  $x$  accidents by time  $t$  can incur another accident with a probability of  $\{(k + x)/(1 + \lambda t)\}\lambda dt$  (independent of the times of the previous accidents), where  $k$  is a positive constant and  $\lambda$  refers to the individual's risk exposure. At  $t = 0$ , since  $x = 0$ , the probability of an accident is  $k\lambda dt$ . Hence, what the model basically assumes is that, initially, the probability of having an accident is not the same for each individual, but depends on the external conditions; later, the probability is also affected by the number of preceding accidents. Under these assumptions and if differences in the exposure to accident risk can be thought of as governed by a distribution with probability density function given by  $\{\Gamma(a + \rho)v^{a-1}(1 + v)^{-(a+\rho)}\} / \{\Gamma(\rho)\Gamma(a)\}$ , the final distribution of accidents over a unit period of time turns out to be  $UGWD(a, k; \rho)$ .

The above derivation of the generalized Waring distribution closely relates to a modeling approach whereby the distribution of accident occurrences in a time interval  $(0, t)$  is regarded as underpinned by a stochastic process and, in particular, by a pure birth process  $\{X_t, t = 0, 1, 2, \dots\}$  where the probability of a person to incur an accident in  $(t, t + dt)$ , having had  $x$  accidents by time  $t$  is  $P(X_{t+\delta t} = x + 1 | X_t = x) = f_\lambda(x, t)\delta t + o(\delta t)$ .

Irwin (1941), followed later by Arbous and Kerrich (1951), derived the negative binomial distribution on the hypothesis solving the associated Kolmogorov forward differential equations by a method due to McKendrick (1925). Specifically, assuming that individuals can have during the time period from  $t$  to  $dt$ , individuals can have 0 accidents with probability  $1 - f_\lambda(x, t)dt$ , 1 accident with probability  $f_\lambda(x, t)dt$  and  $> 1$  accidents with probability 0, he solved the resulting system of Kolmogorov forward difference-differential equations

$$\begin{aligned} \frac{\partial}{\partial t} P_\lambda(0, t) &= -f_\lambda(0, t)P_\lambda(0, t) \\ \frac{\partial}{\partial t} P_\lambda(x, t) &= -f_\lambda(x, t)P_\lambda(x, t) + f_\lambda(x-1, t)P_\lambda(x-1, t), x \geq 1 \end{aligned}$$

in terms of a single difference-differential equation involving the probability generating function  $G_\lambda(s; t)$  of  $X_t$  given by

$$\frac{\partial}{\partial t} G_\lambda(s; t) = (s-1) \sum_{x=0}^{\infty} s^x f_\lambda(x, t) P_\lambda(x, t)$$

where  $G_\lambda(s; t) = \sum_{x=0}^{\infty} P_\lambda(x, t) s^x$ . (He obtained this equation by multiplying the  $i$ -th equation of the system by  $s^{i-1}$ ,  $i = 1, 2, \dots$  and summing the resulting equations).

Assuming further that  $f_\lambda(x, t) = \lambda(k + mx)$ ,  $k, m > 0$  and subject to the initial conditions  $G_\lambda(1; t) = G_\lambda(s; 0) = 1$ , he obtained for the distribution of accidents

$$G_\lambda(s; t) = [e^{\lambda mt} - s(e^{\lambda mt} - 1)]^{-k/m},$$

i.e. the probability generating function of the negative binomial distribution with parameters  $k/m$  and  $(1 - e^{-\lambda mt})^{-1}$ .

Relaxing Irwin's implicit assumption that all individuals were exposed to the same accident risk, Kekalaki (1981) treated the parameter  $\lambda$  as referring to a variable risk exposure according to an exponential distribution with density  $ae^{-a\lambda}$ ,  $a > 0$  and obtained the generalized Waring distribution as the accident distribution. In particular,

$$\begin{aligned} G_{X_t}(s) &= a \int_0^{\infty} e^{-a\lambda} [e^{\lambda mt} - s(e^{\lambda mt} - 1)]^{-k/m} d\lambda \\ &= \frac{a(1-s)^{-k/m}}{mt} \int_0^{\infty} e^{-\lambda \frac{a+kt}{mt}} \left(1 - \frac{s}{s-1} e^{-\lambda}\right)^{-k/m} d\lambda \\ &= \frac{a(1-s)^{-k/m}}{mt} \frac{\Gamma((a+kt)/(mt))}{\Gamma(1 + (a+kt)/(mt))} {}_2F_1\left(\frac{k}{m}, \frac{a+kt}{mt}; \frac{a+kt}{mt} + 1; \frac{s}{s-1}\right) \\ &= \frac{a}{a+mt} {}_2F_1\left(\frac{k}{m}, 1; \frac{a}{mt} + \frac{k}{m} + 1; s\right) \end{aligned}$$

which is the probability generating function of the  $UGWD\left(\frac{k}{m}, 1; \frac{a}{mt}\right)$ .

This model was considered by Panaretos (1989) for the description of the evolution of surnames. Faddy (1997) provided a unifying approach to under- and over-dispersion relative to the Poisson distribution within a scheme of a similar nature, which generalizes the simple Poisson process that underpins the Poisson distribution. He demonstrated that any count distribution can be obtained by a suitable choice of  $f_\lambda(x, t)$  and provided an expression for the system of Kolmogorov forward differential equations in terms of a matrix-exponential function.

Finally, Winkelmann (1995) looked at under- and over-dispersion using renewal theory by exploring the link between duration dependence and dispersion. He demonstrated that discrepancies between observed and nominal variances are conveyed by a hazard function of the waiting times that is not constant, but instead is a decreasing

function of time inducing over-dispersion or an increasing function of time inducing under-dispersion.

### 6.3 The “Spells” model

Further, Xekalaki (1983a) considered a variant of the “spells” model due to Cresswell and Froggatt (1963) that rejects the presence of proneness and contagion.

Assume that every individual is liable to spells and that the number of spells in a given time period  $(0, t)$  is a Poisson variable with parameter  $\theta t, \theta > 0$ . Suppose that no accidents occur outside spells and that the probability of an accident within a spell depends on the risk exposure of the particular individual. In particular, suppose that within a spell a person can have

$$\begin{aligned} & 0 \text{ accidents with probability } 1 - m \log(1 + \lambda) \\ \text{or} & n \text{ accidents } (n \geq 1) \text{ with probability } m \{ \lambda / (1 + \lambda) \}^n / n \end{aligned}$$

$0 < m < 1 / \log(1 + \lambda), \lambda > 0$ , where  $\lambda$  is the external risk parameter for the given individual. Assume further that the numbers of accidents arising out of different spells are independent and independent of the number of spells. Then, if differences in the risk exposure can be described by a beta distribution of the second kind with probability density function,  $\{ \Gamma(a + \rho) v^{a-1} (1 + v)^{-(a+\rho)} \} / \{ \Gamma(\rho) \Gamma(a) \}$ ,  $a, \rho > 0$ , the resulting accident distribution will have probability generating function given by

$$\left\{ \rho {}_2F_1(a, \theta mt; a + \theta mt + \rho; s) \right\} / (\rho + \theta mt)_{(a)}.$$

Hence, in a unit time period, the number of accidents follows the  $UGWD(a, \theta m; \rho)$ .

It is worth noticing that the form of the distribution of  $\lambda$  in the last two models is more general than that considered by the proneness model. It is however, a reasonable choice as it implies a beta distribution of the first kind (Pearson Type I) for the parameter  $q = \lambda / (1 + \lambda)$  of the negative binomial distribution of  $X | \lambda$ .

### 6.4 Deciding about the underlying model

It is evident from the above, that three completely different sets of hypotheses give rise to exactly the same form of distribution and that while the  $UGWD$  may be a plausible model if accident proneness is accepted as an established fact, a satisfactory fit of it is not to be taken as evidence for the validity of the proneness hypothesis. How can we then discriminate?

Statisticians have always been excited to look for ways of discriminating among different models that give rise to the same distribution. Most attempts seem to have been concentrated on distinguishing between the proneness and contagion models generating the negative binomial distribution. The papers by Bates and Neyman (1952) and Bates (1955) cover part of the work that has been done on the subject, though they primarily focus on distinguishing between different forms

of contagion. Shaw and Sichels (1971) attempt was on proving or disproving proneness by ranking individual accident performance on a scale based on their average interval between successive accidents. However, the first systematic study on how one can discriminate between the proneness and contagion models of the negative binomial distribution appears to be that by Cane (1974).

She demonstrated, however, that one cannot distinguish between the two models, even with knowledge of the time sequence of accidents. She demonstrated, in particular, that the conditional distribution of the times,  $t_i, i = 1, 2, \dots, n$  at which accidents occurred in a time period  $(0, T)$  is the same in both cases, namely that of an ordered sample from a uniform distribution over  $(0, T)$  with probability density function  $n! T^{-n}$ . In fact, this is the case for any compound Poisson accident distribution whose compounding distribution has finite moments (Cane 1977), hence also for the  $UGWD(a, k; \rho)$ .

This implies that the availability of information on the times of the occurrence of accidents is not sufficient to guide one's choice between the proneness and contagion models.

However, as demonstrated by Xekalaki (1983a), there appears to exist a possibility in the framework of the Spells model. Consider, in particular, the problem of finding the joint distribution of times  $t_i, i = 1, 2, \dots, n$  of accidents by individuals with  $n$  accidents in a unit period of time under the spells model. For fixed  $\lambda$ , accidents occur as events in a

generalized Poisson process:  $X(t) = \sum_{i=1}^{N(t)} Y_i, N(t) \sim \text{Poisson}(\theta t)$ , where  $\theta > 0, t \geq 0$  and  $Y_i$

are identically and independently distributed with probability density function given by  $\{\Gamma(a + \rho)v^{a-1}(1+v)^{-(a+\rho)}\}/\{\Gamma(\rho)\Gamma(a)\}, a, \rho > 0$ . Consequently, the required probability

function can be written as  $\int_0^\infty (1+\lambda)^{-\theta m(1-t_n)} \left[ \prod_{i=1}^n \left\{ \lambda m \theta (1+\lambda)^{-\theta m(t_i-t_{i-1})-1} dt_i \right\} \right] dH(\lambda),$

with  $H(\cdot)$  denoting the distribution function of the beta distribution of the second kind defined as above. Hence, the required probability is  $\left\{ \frac{(\theta m)^n \rho_{(a)}^{a(n)}}{(\theta m + \rho)_{(a+n)}} \right\} dt_1 \dots dt_n$ . Therefore, conditional on  $n$  accidents during a time period from 0 to 1, the joint pdf of  $t_i, i = 1, 2, \dots, n$ , is  $n! (\theta m)^n / (\theta m)_{(n)}$ .

The obtained form differs from that arising under the proneness and contagion models. This fact is itself very interesting as far as establishing the presence of spells is concerned, as it implies the following: if an observed accident distribution of the  $UGWD$  type has arisen from the spells model, the time intervals  $(0, t_i), i = 1, 2, \dots, n$ , given a total of  $n$  accidents, will be jointly distributed with the above density function. Any departure from this distribution is, then, evidence against the spells model. Of course, if on the available evidence one has to reject this form in favor of that obtained by Cane, then one is faced again with the question: "proneness or contagion?" This cannot be answered by studying the distribution of  $t_i$ .

### 6.5 What does Irwin's accident model offer beyond a good fit to the data?

The innovation brought by Irwin's accident proneness model does not merely lie in the better fit it provides to accident data, but in the possibility of partitioning the total variance

$(\sigma^2)$  into three additive components due to proneness ( $\sigma_v^2$ ), liability ( $\sigma_\lambda^2$ ) and randomness ( $\sigma_R^2$ ) thus,

$$\sigma^2 = \sigma_\lambda^2 + k^2\sigma_v^2 + \sigma_R^2,$$

Where

$$\begin{aligned} \sigma_\lambda^2 &= ak(a+1)(\rho-1)^{-1}(\rho-2)^{-1} \\ \sigma_v^2 &= a(a+\rho-1)^{-1}(\rho-1)^{-2}(\rho-2)^{-1} \\ \sigma_R^2 &= ak(\rho-1)^{-1} \\ \sigma^2 &= ak(a+\rho-1)(k+\rho-1)(\rho-1)^{-2}(\rho-2)^{-1}. \end{aligned}$$

There is still, however, a problem due to the fact that the  $UGWD(a, k; \rho)$  is symmetrical in  $a$  and  $k$  ( $UGWD(a, k; \rho) \sim UGWD(k, a; \rho)$ ). Hence, although one may consider that  $\sigma_\lambda^2 + k^2\sigma_v^2$  represents the variance component due to all non-random factors, the mathematics alone cannot determine whether  $\sigma_\lambda^2$  represents the liability component and  $k^2\sigma_v^2$  the proneness component or vice versa. As a consequence, distinguishable estimates for the non-random variance components  $\sigma_\lambda^2$  and  $\sigma_v^2$  cannot be obtained unless subjective judgement is made. This problem was addressed by Xekalaki (1984a) with the introduction of her bivariate form of the generalized Waring distribution.

### 7. The bivariate generalized Waring distribution

Generalizing further Irwins (1963) generalization of Waring's expansion, we have for  $k, m, a > 0$ ,

$$\begin{aligned} \frac{1}{(x-a)_{(k+m)}} &= \sum_{\ell=0}^{\infty} \sum_{r=0}^{\ell} \frac{a_{(\ell)}(-1)^\ell}{\ell!} \binom{\ell}{r} \Delta^r \frac{1}{x_{(k)}} \Delta^{\ell-r} \frac{1}{(x+k+r)_{(m)}} \\ &= \sum_{r=0}^{\infty} \sum_{\ell=0}^{\infty} \frac{a_{(r+\ell)}(-1)^{r+\ell}}{r!\ell!} \Delta^r \frac{1}{x_{(k)}} \Delta^\ell \frac{1}{(x+k+r)_{(m)}} \\ &= \sum_{r=0}^{\infty} \sum_{\ell=0}^{\infty} \frac{a_{(r+\ell)}k_{(r)}m_{(\ell)}}{x_{(k+m+r+\ell)}} \frac{1}{r!} \frac{1}{\ell!} \end{aligned}$$

If  $x > a$ , the above series is convergent. Then, by letting  $\rho = x - a > 0$  and multiplying both sides by  $\rho_{(k+m)}$ , leads to a double series of positive terms converging to unity. The general term of the series therefore can be regarded as defining a bivariate discrete probability distribution with probability function

$$p_{r,\ell} = \frac{\rho_{(k+m)}}{(a+\rho)_{(k+m)}} \frac{a_{(r+\ell)}k_{(r)}m_{(\ell)}}{(a+k+m+\rho)_{(r+\ell)}} \frac{1}{r!} \frac{1}{\ell!}, \quad a, k, m, \rho > 0, r, \ell = 0, 1, 2, \dots$$

In the remainder of the paper, we refer to this distribution as the *bivariate generalized Waring distribution* with parameters  $a, k, m$  and  $\rho$  and we denote it by  $BGWD(a; k, m; \rho)$ .



### 7.1 The BGWD in relation to accident theory

Assume that individuals of proneness  $\nu$  and liability  $\lambda_i|\nu$  for a period  $i$  of observation incur, over two non-overlapping time periods, accidents  $X, Y$  according to a double Poisson distribution  $G_{(X,Y)|\lambda_1, \lambda_2, \nu}(s, t) = \exp\{(\lambda_1|\nu)(s-1) + (\lambda_2|\nu)(t-1)\}$ ,  $\lambda_1, \lambda_2 > 0$ . Assume further that the liability parameters  $\lambda_1|\nu, \lambda_2|\nu$  are independently gamma distributed with densities  $(\Gamma(\theta_i)\nu^{\theta_i})^{-1}e^{-\lambda_i|\nu}\lambda_i^{\theta_i-1}$ ,  $\theta_1 \equiv k, \theta_2 \equiv m, \nu > 0$ , whence for individuals with the same proneness  $\nu$ , but varying liabilities, the numbers of occurring accidents over the two periods are jointly distributed as the double negative binomial with probability generating function

$$G_{(X,Y)|\nu}(s, t) = \{1 + \nu(1-s)\}^{-k} \{1 + \nu(1-t)\}^{-m}.$$

Letting now the proneness parameter  $\nu$  be beta distributed with density function  $\{\Gamma(a + \rho)\nu^{a-1}(1 + \nu)^{-(a + \rho)}\}/\{\Gamma(\rho)\Gamma(a)\}$ ,  $a, \rho > 0$ , the probability generating function of the joint distribution of accidents over the two periods takes the form

$$\begin{aligned} G_{(X,Y)}(s, t) &= \frac{\Gamma(\rho + a)}{\Gamma(\rho)\Gamma(a)} \int_0^{+\infty} \nu^{a-1}(1 + \nu)^{-(a+\rho)} \{1 + \nu(1-s)\}^{-k} \{1 + \nu(1-t)\}^{-m} d\nu \\ &= \frac{\rho^{(k+m)}}{(a + \rho)_{(k+m)}} F_1(a; k, m; a + k + m + \rho; s, t) \sim BGWD(a; k, m; \rho), \end{aligned}$$

where  $F_1(a; b, c; d; u, v) = \sum_{r,s=0}^{\infty} \{a_{(r+s)} b_{(r)} c_{(s)} u^r v^s\} / \{d_{(r+s)} r! s!\}$  is Appell's hypergeometric series and  $h(l) = \Gamma(h + l)/\Gamma(h)$ ,  $h > 0, l \in R$ .

Regarding separate estimation of the contribution of proneness, liability and randomness in a given accident situation over a period of observation whenever proneness is accepted as an established fact, Xekalaki (1984a) showed that rearranging the observed distribution in two non-overlapping sub-intervals and fitting the  $BGWD(a; k, m; \rho)$  to the resulting bivariate accident distribution does enable separate estimation of the variance components. This is demonstrated in Table 1.

Further models leading to the  $BGWD$  provided by Xekalaki (1984c), provide the framework within which one can also obtain the  $BGWD$  as an accident distribution under the contagion and the spells accident theories.

**Table 1 Estimators of the components of the variance of the generalized waring distribution**

Component due to	Marginal variance of X	Marginal variance of Y	Variance of X + Y entire period
Random factors	$\frac{\hat{a}\hat{k}}{\hat{\rho}-1}$	$\frac{\hat{a}\hat{m}}{\hat{\rho}-1}$	$\frac{\hat{a}(\hat{k}+\hat{m})}{\hat{\rho}-1}$
Proneness	$\frac{\hat{k}^2\hat{a}(\hat{a}+\hat{\rho}-1)}{(\hat{\rho}-1)^2(\hat{\rho}-2)}$	$\frac{\hat{m}^2\hat{a}(\hat{a}+\hat{\rho}-1)}{(\hat{\rho}-1)^2(\hat{\rho}-2)}$	$\frac{(\hat{k}+\hat{m})^2\hat{a}(\hat{a}+\hat{\rho}-1)}{(\hat{\rho}-1)^2(\hat{\rho}-2)}$
Liability	$\frac{\hat{a}\hat{k}(\hat{a}+1)}{(\hat{\rho}-1)(\hat{\rho}-2)}$	$\frac{\hat{a}\hat{m}(\hat{a}+1)}{(\hat{\rho}-1)(\hat{\rho}-2)}$	$\frac{\hat{a}(\hat{k}+\hat{m})(\hat{a}+1)}{(\hat{\rho}-1)(\hat{\rho}-2)}$
Total	$\frac{\hat{a}\hat{k}(\hat{\rho}+\hat{k}-1)(\hat{\rho}+\hat{a}-1)}{(\hat{\rho}-1)^2(\hat{\rho}-2)}$	$\frac{\hat{a}\hat{m}(\hat{\rho}+\hat{m}-1)(\hat{\rho}+\hat{a}-1)}{(\hat{\rho}-1)^2(\hat{\rho}-2)}$	$\frac{\hat{a}(\hat{k}+\hat{m})(\hat{\rho}+\hat{k}+\hat{m}-1)(\hat{\rho}+\hat{a}-1)}{(\hat{\rho}-1)^2(\hat{\rho}-2)}$

### 8. The multivariate generalized Waring distribution

The n-variate version of the generalized Waring distribution introduced and studied by Xekalaki (1986) is also obtained as an inverse factorial distribution. Its probability generating function is given by

$$G(\underline{t}) = \frac{\rho \binom{\sum k_i}{\sum k_i}}{(a + \rho)^{\binom{\sum k_i}{\sum k_i}}} F_D \left( a; k_1, \dots, k_n; a + \sum_{i=1}^n k_i + \rho; \underline{t} \right)$$

with  $F_D(a; \beta_1, \dots, \beta_n; \gamma; \underline{t})$  denoting Lauricella's hypergeometric function given by

$$F_D(a; \beta_1, \dots, \beta_n; \gamma; \underline{t}) = \sum_{r_1, \dots, r_n} \frac{a \binom{\sum r_i}{\sum r_i}}{\gamma \binom{\sum r_i}{\sum r_i}} \prod_{i=1}^n \frac{(\beta_i)_{(r_i)} t_i^{r_i}}{(r_i)!}$$

The probability function of it is given by

$$P_{\underline{r}} \equiv P(X = \underline{r}) = \frac{\rho \binom{\sum k_i}{\sum k_i} a \binom{\sum r_i}{\sum r_i} (k_1)_{(r_1)} \dots (k_n)_{(r_n)}}{(a + \rho)^{\binom{\sum k_i}{\sum k_i}} (a + \rho + \sum k_i)^{\binom{\sum r_i}{\sum r_i}} (r_1)! \dots (r_n)!},$$

$r_i = 0, 1, 2, \dots; i = 1, \dots, n$

and its probabilities are related by the following first order recurrences, which facilitate their computation

$$\frac{P_{l_1, l_2, \dots, l_{h-1}, l_h+1, l_{h+1}, \dots, l_n}}{P_{l_1, l_2, \dots, l_n}} = \frac{\left[ a + \sum_{i=1}^n l_i \right] (k_n + l_n)}{\left[ a + \sum_{i=1}^n k_i + \sum_{i=1}^n l_i \right] (l_n + 1)}, \quad l = 0, 1, 2, \dots;$$

$i = 1, 2, \dots, n$

An interesting aspect of the bivariate and multivariate versions of the generalized Waring distribution is that their marginal distributions (conditional and unconditional) as well as their convolution are of the same form (*UGWD's*), properties that exhibit a symmetry analogous to that existing in the case of the multivariate normal distribution. Further, the generalized Waring distribution is self-decomposable (Xekalaki 1983b).

### 9. The Generalized Waring Process (gWp)

Looking into how temporally evolving data from the wide spectrum of application contexts that can reasonably be viewed from the perspective of the frameworks discussed in Sections 6, 7 and 8 can be treated, Xekalaki and Zografis (2008) defined and studied the generalized Waring process. In establishing its definition, the structural properties of both the bivariate and the multivariate versions of the generalized Waring distribution played a significant role. This process, analogously to the case of Poisson and Pólya processes, which can be obtained as limiting cases of it, was shown to be a Markov process.

Let  $\{N(t), t \geq 0\}$  be a counting process. This is said to be a *generalized Waring process* with parameters  $a, k, \rho > 0$ , denoted by  $gWp(a, k; \rho)$ , if (i)  $N(0) = 0$ , (ii)  $N(t)$  is a Markov

process, and (iii)  $N(t+h) - N(t)$  has the generalized Waring distribution with parameters  $a, k; \rho$  for  $h > 0, t \geq 0$ . The process starts at 0, it has stationary increments and

$$P(N(t) = n) = \frac{\rho_{(kt)}}{(\rho + a)_{(kt)}} \frac{a_{(n)}(kt)_{(n)}}{(a + \rho + kt)_{(n)}} \frac{1}{n!}$$

i.e.,  $N(t)$  has a generalized Waring distribution with parameters  $a, kt; \rho$ .

The transition probabilities of the generalized Waring process are given by

$$p_{m,n}(s, s+t) = P\{N(s+t) = n | N(s) = m\} = \frac{\Gamma(a+n)}{\Gamma(a+m)} \frac{(kt)_{(n-m)}}{(n-m)!} \frac{(\rho + ks)_{(a+m)}}{(\rho + ks + kt)_{(a+n)}}$$

$$p_{0,n}(0, t) = P\{N(t) = n | N(0) = 0\} = \frac{\rho_{(kt)}}{(\rho + a)_{(kt)}} \frac{a_{(n)}(kt)_{(n)}}{(a + \rho + kt)_{(n)}} \frac{1}{n!} = P(N(t) = n)$$

with the last equality indicating that the generalized Waring process is a non-homogenous Markov process. Its mean and variance are respectively

$$E[N(t)] = \frac{akt}{\rho-1} \text{ and } Var[N(t)] = \frac{akt(\rho + kt-1)(\rho + \alpha-1)}{(\rho-1)^2(\rho-2)}$$

Note that since the generalized Waring process is a stationary process and its mean is of the form  $E[N(t)] = \eta t$ , the above formula implies that its *intensity is*  $\eta = ak/(\rho - 1)$ . Its variance can be split into three additive components, thus

$$Var[N(t)] = \sigma_{\lambda(t)}^2 + (kt)^2 \sigma_v^2 + \sigma_R^2$$

with the liability and random components dependent on time. In particular,

$$\sigma_{\lambda(t)}^2 = akt(a+1)(\rho-1)^{-1}(\rho-2)^{-1}; \sigma_v^2 = a(a+\rho-1)(\rho-1)^{-2}(\rho-2)^{-1};$$

$$\sigma_R^2 = akt(\rho-1)^{-1}.$$

### 9.1 The generalized Waring process in an accident proneness context

We consider a population which is inhomogeneous with respect to personal and environmental attributes affecting the occurrence of accidents. The terms “accident proneness” and “accident liability” are again used to refer respectively to a person’s predisposition to accidents, and to a person’s exposure to external risk of accident with the conditional distribution of the random variable  $\lambda$  given  $\nu$  describing differences in external risk factors among individuals. Liability fluctuations over a time interval  $(t, t+h)$  depend on the length  $h$  of the interval and are described by a distribution for  $\lambda | \nu$  with probability density function  $\lambda^{kh-1} e^{-\lambda/(\nu h)} (\nu h)^{-kh} / \Gamma(kh)$ . Allowing further the parameter  $\nu$  have a beta distribution of the second kind with parameters  $a$  and  $\rho$  and density function  $\phi$  given by

$\phi(\nu) = \Gamma(a+\rho) \nu^{a-1} (1+\nu)^{-(a+\rho)} / [\Gamma(a)\Gamma(\rho)]$ ,  $a, \rho \geq 0$ , we obtain for the distribution of the number of accidents  $N(t)$ :

$$P(N(t+h) - N(t) = n) = \frac{\rho_{(kh)}}{(a + \rho)_{(kh)}} \frac{a_{(n)}(kh)_{(n)}}{(a + \rho + kh)_{(n)}} \frac{1}{n!}$$

and

$$P(N(t) = n) = P_n(t) = \frac{\rho_{(kt)}}{(a + \rho)_{(kt)}} \frac{a_{(n)}(kt)_{(n)}}{(a + \rho + kt)_{(n)}} \frac{1}{n!}, n = 0, 1, \dots$$

So, the process arising in the context of this model, satisfies the defining conditions of the generalized Waring process.

### 9.2 The generalized Waring process in the context of a spells model

Xekalaki and Zografis (2008) showed that the generalized Waring process could also be used in modeling temporally evolving data in the context of a *spells* model. Assume again that each person is liable to spells and that no accidents can occur outside spells. Let  $S(t)$ ,  $t = 0, 1, 2, \dots$ , the number of spells up to a given moment  $t$ , be a homogeneous Poisson process with rate  $k/m$ ,  $k > 0$ , the number  $X_i$  of accidents within a spell  $i$  be a random variable with a logarithmic series distribution with parameters  $m$  and  $\nu$  and probability function given by  $P(X_i = n) = \frac{m}{n} \left( \frac{\nu}{1+\nu} \right)^n$ ,  $n \geq 1$  with  $P(X_i = 0) = 1 - m \log(1 + \nu)$ ,  $\nu > 0$ ,  $0 < m < 1/\log(1 + \nu)$ , and the numbers of accidents arising out of different spells be independent and independent of the number  $S(t)$  of spells. Here  $\nu$  is regarded as the external risk parameter, too, which they assumed varying according to a beta distribution of the second kind with parameters  $a$  and  $\rho$  and probability density function given by  $\Gamma(a + \rho)\nu^{a-1}(1 + \nu)^{-(a+\rho)}/[\Gamma(a)\Gamma(\rho)]$ ,  $a, \rho \geq 0$ . They then showed that the above framework leads to a process conforming with the postulates of the generalized Waring process, thus demonstrating its potential application in the context of the Spells model.

### 10. An application: modeling the counting process $\{N(s), s > 0\}$ associated with the access pattern of a web site

As an illustration of the application potential of the generalized Waring process in other fields by appropriately adjusting the concepts and terminology used in this paper so as to have natural interpretations, we outline an example of a model for temporally evolving data on web access patterns provided by Xekalaki and Zografis (2008).

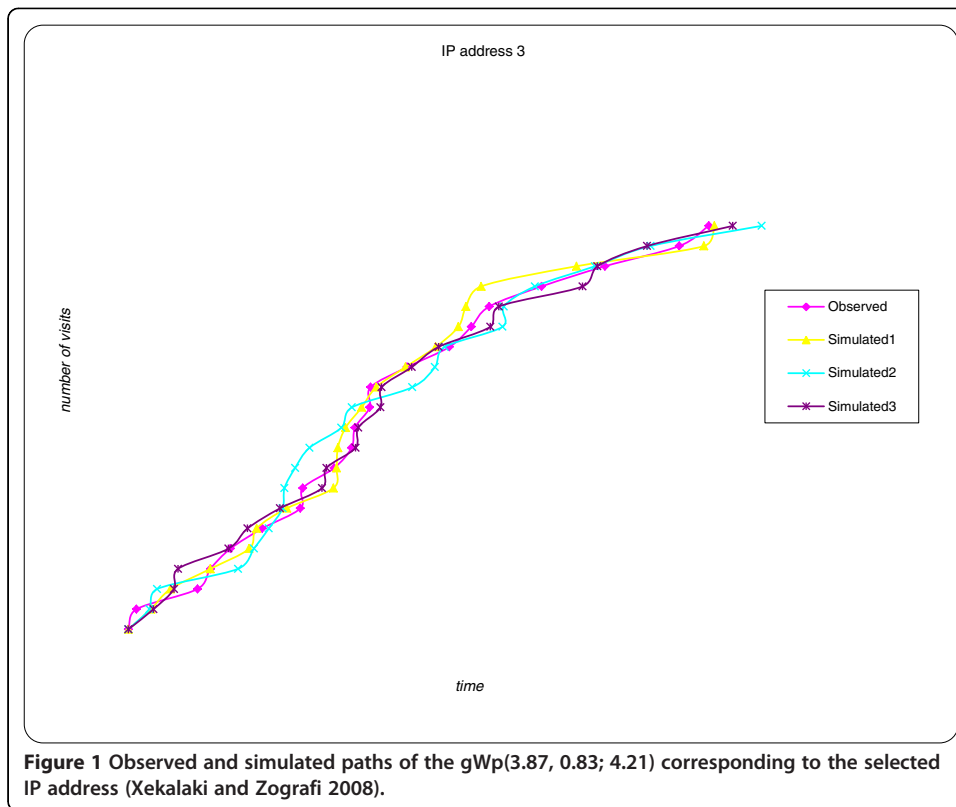
In this context,  $\{N(s), s > 0\}$  is the counting process associated with the access pattern of a web site, where, for any  $t > 0$ ,  $N(t)$  represents the number of visits that the web pages on this particular site get within the interval  $(0, t)$ . Note that the generalized Waring distribution was cited in Ajiferuke et al. (2004) as used by them to fit observed website visitation data for a given period, i.e. to model counts  $N(t_0)$  of web visits on a given fixed time interval  $(0, t_0)$ .

Except for chance, visits to a web site can be regarded as affected by the intrinsic appeal of the particular site to web users (corresponding to proneness) as well as by exogenous factors (corresponding to external factors) such as, links provided by other sites to the particular site, how well the site is advertised etc.

Letting  $\nu$  denote the intrinsic factors and  $\lambda|\nu$  the exogenous factors. Then assuming that  $N(t)|\lambda$  follows a *Poisson*( $\lambda(t)$ ) distribution, where  $\lambda(t) = \lambda t$  with  $\lambda|\nu$  following a gamma distribution with density  $\lambda^{kt-1}e^{-\lambda/(v\nu)}(v\nu)^{-kt}/\Gamma(kt)$ , and with  $\nu$  following a beta distribution of the second kind with density  $\Gamma(a + \rho)\nu^{a-1}(1 + \nu)^{-(a+\rho)}/[\Gamma(a)\Gamma(\rho)]$ ,  $a, \rho \geq 0$ , then the unconditional distribution of  $N(t)$  is the *GWD*( $a, kt; \rho$ ), i.e. the process  $\{N(t), t \geq 0\}$  is a generalized Waring process.

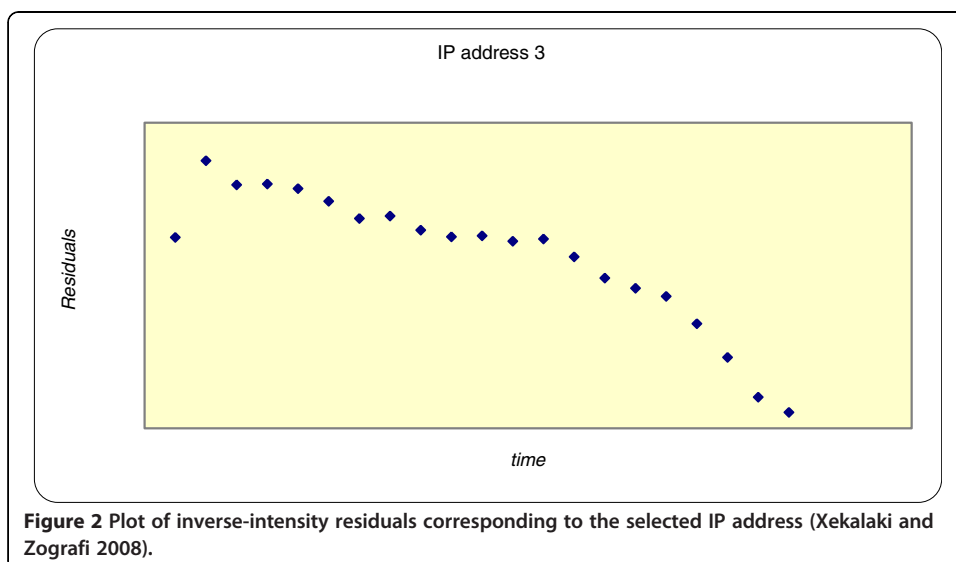
#### 10.1 The data

The log files representing the hits on an e-shop site for the period from March 31, 2006 to April 30, 2006 have been used to fit this model. (A log file typically contains information



on the times of visits per IP address per day). On the basis of such log files, the visits per day made by each of 468 IP addresses to a web site during the above period were enumerated yielding 468 paths of visits  $N_i(t_j)$  made by IP address  $i$  up to and including time  $t_j$  denoted by  $\{N_i(t_j), i = 1, 2, \dots, 468; j = 1, 2, \dots, 31\}$ .

Moment estimates of the parameters of the generalized Waring process were obtained employing an estimation procedure for spatial point process data termed in the literature as the centered reduced moment method. The method introduced and



studied by Ripley (1976, 1977) utilizes the intensity of the process and the mean number of further points within distance  $s$  of an arbitrary point of the process. In particular, the method utilizes the moment estimators  $E(\hat{N}(s)) = \hat{\mu}_1 = \hat{\eta}s = ns/h$ ,

$$E(\hat{N}^2(s)) = \hat{\mu}_2 = X/n_{(2)}, \quad E(\hat{N}^3(s)) = \hat{\mu}_3 = (Z-X)/n_{(3)} \quad \text{with} \quad X = \sum_{i=1}^n \sum_{i \neq j} \phi_s^2(x_i, x_j),$$

$$Z = \sum_{i=1}^n \left( \sum_{j \neq i} \phi_s(x_i, x_j) \right) \left( \sum_{k \neq i} \phi_s(x_i, x_k) \right),$$

where the quantities involved in the above equations represent weights defined, for each value  $x_i$  in the collection of points  $\{x_i : i = 1, 2, \dots, n\}$  of the process within a time interval of length  $h$ , as follows: For each  $x_i$  in  $\{x_i : i = 1, 2, \dots, n\}$  and a given  $s > 0$ , consider the interval of center  $x_i$  and length  $s$  and assign to every point  $x_j$ ,  $j \neq i$  in this interval the weight  $\phi_s(x_i, x_j) = \omega(x_i, x_j)^{-1}$ , where  $\omega(x_i, x_j)$  is the number of other points  $\{x_k, k \neq i, k \neq j\}$  of the process that are included in the interval of length  $|x_i - x_j|$  and center  $x_i$  (see also Diggle and Chetwynd 1991; Chetwynd and Diggle 1998, among others). The standard errors of the thus obtained parameter estimators can in principle be determined by simulation, but the associated computations are formidable. Approximation formulas exist only for the case of homogeneous planar Poisson process, while, for the class of stationary Cox process, there is no obvious way to obtain estimable expressions as noted by Chetwynd and Diggle (1998).

The observed paths were compared to the corresponding time series of simulated realizations of the generalized Waring process over the same time segment. For each IP address, 100 simulated realizations of the  $gWp(a, k; \rho)$ , were obtained and each of the observed time series paths was compared to the corresponding simulated ones. On average, the realizations of the generalized Waring process exhibited similar structural characteristics, notably recognizable, to those of the paths of the observed time series. For illustration purposes, the path of the observed time series associated with one of the IP addresses considered is presented in Figure 1. In the graph, the path is superimposed by a sample of three of the 100 corresponding simulated realizations of the  $gWp(a, k; \rho)$ . Inspection of the graph provides a visual appreciation of the degree of similarity in the structural characteristics of the path of the observed and the realized time series.

Following Lewis (1972), Brillinger (1978) and Andersen et al. (1993), the closeness of the observed and realized time series was also checked using diagnostic plots based on the inverse-intensity residuals computed for each value  $x_j$  in the collection of points

$$\{x_j : j = 1, 2, \dots, n\} \text{ of the process given by } R_{\hat{\theta}}(B_j, \eta^{-1}) = \sum_{x_i \in B_j} \hat{\eta}(x_i) - \int_{B_j} I_{R^+}(\hat{\eta}(x)) dx \text{ where}$$

$B_j = (0, x_j)$ ,  $\hat{\theta} = (\hat{a}, \hat{k}, \hat{\rho})^{-1}$ ,  $\hat{\eta}(x) = \eta(x, \hat{\theta})$  is the fitted intensity and  $I_{R^+}(\cdot)$  is the indicator function. These plots exhibit similar results. The plot corresponding to the data associated to the IP address considered is shown in Figure 2.

#### Competing interests

The authoress declares that she has no competing interests.

#### Acknowledgement

The authoress would like to thank the associate editor and the referees for their constructive comments. This paper is an extended version of the authoress' invited plenary presentation at the International Conference on Statistical Distributions and Applications, Michigan USA, October 10-12, 2013.



Received: 28 February 2014 Accepted: 29 July 2014

Published online: 28 August 2014

## References

- Ajiferuke, I, Wolfram, D, Xie, H: Modelling website visitation and resource usage characteristics by IP address data. In: Julien, H, Thompson, S (eds). *Proceedings of the 32nd Annual Conference of the Canadian Association for Information Science*, Manitoba, Canada (2004). Available at: [www.caais-acsi.ca/proceedings/2004/ajiferuke\\_2004.pdf](http://www.caais-acsi.ca/proceedings/2004/ajiferuke_2004.pdf)
- Andersen, P, Borgan, Ø, Gill, R, Keiding, N: *Statistical Models based on Counting Processes*. Springer, New York (1993)
- Andrews, DF, Mallows, CL: Scale mixtures of normal distributions. *J. R. Stat. Soc. B* **36**, 99–102 (1974)
- Arbous, AG, Kerrich, JE: Accident statistics and the concept of accident proneness. *Biometrics* **7**, 340–432 (1951)
- Bates, GE: Joint distributions of time intervals for the occurrence of successive accidents in a generalized Polya scheme. *Ann. Math. Stat.* **26**, 705–720 (1955)
- Bates, GE, Neyman, J: Contributions to the theory of accident proneness. *U. Calif. Publ. Stat.* **1**, 215–275 (1952)
- Breslow, N: Tests of hypotheses in overdispersed poisson regression and other quasi-likelihood models. *J. Am. Stat. Assoc.* **85**, 565–571 (1990)
- Brillinger, D: Comparative Aspects of the Study of Ordinary Time Series and of Point Processes. In: Krishnaiah, PR (ed.) *Developments in Statistics*, pp. 33–133. Academic Press, New York (1978)
- Cane, VR: The concept of accident proneness. *B. Inst. Math. Bulgarian Acad. Sci.* **15**, 183–188 (1974)
- Cane, VR: A class of non-identifiable stochastic models. *J. Appl. Probab.* **14**, 475–782 (1977)
- Carriere, J: Nonparametric tests for mixed poisson distributions. *Insur. Math. Econ.* **12**, 3–8 (1993)
- Chetwynd, AG, Diggle, PJ: On estimating the reduced second moment measure of a stationary spatial point process. *Aust. N.Z. J. Stat.* **40**(1), 11–15 (1998)
- Cox, DR: *Renewal Theory*. Barnes & Noble, New York (1962)
- Cox, DR: Some remarks on overdispersion. *Biometrika* **70**, 269–274 (1983)
- Cresswell, WL, Froggatt, P: *The Causation of Bus Driver Accidents*. Oxford University Press, London (1963)
- Davidian, M, Carroll, RJ: A note on extended quasi-likelihood. *J. R. Stat. Soc. B* **50**, 74–82 (1988)
- Dean, CB, Lawless, J, Willmot, GE: A mixed poisson-inverse Gaussian regression model. *Can. J. Stat.* **17**, 171–182 (1989)
- Diggle, PJ, Chetwynd, AG: Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics* **47**, 1155–1163 (1991)
- Dimaki, C, Xekalaki, E: Identifiability of income distributions in the context of damage and generating models. *Commun. Stat. A-Theor.* **19**(8), 2757–2766 (1990)
- Dimaki, C, Xekalaki, E: Additive and multiplicative distortion of observations: some characteristic properties. *J. Appl. Stat. Sc.* **5**(2/3), 113–127 (1996)
- Dubourdieu, J: Remarques Relatives sur la Théorie Mathématique de l'Assurance-Accidents. *Bull. Trim. Inst. Actuaries Fran.* **44**, 79–146 (1938)
- Faddy, MJ: Extending poisson process modelling and analysis of count data. *Biometrical J.* **39**(4), 431–440 (1997)
- Famoye, F, Singh, KP: Zero-inflated generalized poisson regression model with an application to domestic violence. *Data. J. Data Sci.* **4**, 117–130 (2006)
- Fisher, RA: The effects of methods of ascertainment upon the estimation of frequencies. *Ann. Eugenics.* **6**, 13–25 (1934)
- Fisher, RA: The significance of deviations from expectation in a poisson series. *Biometrics* **6**, 17–24 (1950)
- Greenwood, M, Woods, HM: On the Incidence of Industrial Accidents upon Individuals with Special Reference to Multiple Accidents. In: Report of the Industrial Fatigue Research Board, **4**, pp. 1–28. His Majesty's Stationary Office, London (1919)
- Greenwood, M, Yule, GU: An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attack of disease or repeated accidents. *J. R. Stat. Soc.* **83**, 255–279 (1920)
- Gschlöbl, S, Czado, C: Modelling count data with overdispersion and spatial effects. *Stat. Pap.* **49**, 531–552 (2008)
- Gupta, PL, Gupta, RC, Tripathi, RC: Analysis of zero adjusted count data. *Comput. Stat. Data An.* **23**, 207–218 (1996)
- Gupta, PL, Gupta, RC, Tripathi, RC: Score test for zero inflated generalized poisson regression model. *Commun. Stat. A-Theor.* **33**, 47–64 (2004)
- Gurland, J: Some applications of the negative binomial and other contagious distributions. *Am. J. Public Health* **49**, 1388–1399 (1959)
- Hinde, J, Demetrio, CGB: Overdispersion: models and estimation. *Comput. Stat. Data An.* **27**, 151–170 (1998)
- Irwin, JO: Discussion on chambers and Yule's paper. *J. R. Stat. Soc. Supplement* **7**, 101–109 (1941)
- Irwin, JO: The place of mathematics in medical and biological statistics. *J. R. Stat. Soc. A* **126**, 1–44 (1963)
- Irwin, JO: The generalized Waring distribution applied to accident theory. *J. R. Stat. Soc. A* **131**, 205–225 (1968)
- Irwin, JO: The Generalized Waring Distribution. *J. R. Stat. Soc. A* **138**, 18–31 (1975). Part I), 204–227 (Part II), 374–384 (Part III)
- Jewell, N: Mixtures of exponential distributions. *Ann. Stat.* **10**, 479–484 (1982)
- Karlis, D, Xekalaki, E: On testing for the number of components in a mixed poisson model. *Ann. I. Stat. Math.* **51**, 149–162 (1999)
- Karlis, D, Xekalaki, E: Mixtures Everywhere. In: Panaretos, J (ed.) *Stochastic Musings: Perspectives from the Pioneers of the Late 20th Century*, pp. 78–95. Laurence Erlbaum, USA (2003)
- Karlis, D, Xekalaki, E: Mixed poisson distributions. *Int. Stat. Rev.* **73**(1), 35–58 (2005)
- Kemp, CD: On a contagious distribution suggested for accident data. *Biometrics* **23**, 241–255 (1967)
- Kemp, CD: "Accident proneness" and discrete distribution theory. In: Patil, GP (ed.) *Random Counts in Scientific Work*, Vol.2, pp. 41–65. State College: Pennsylvania State University Press, University Park, USA (1970)
- Krishnaji, N: Characterization of the Pareto distribution through a model of under-reported incomes. *Econometrica* **38**, 251–255 (1970)
- Kulinskaya, E, Olkin, I: An overdispersion model in meta-analysis. *Stat. Model.* **14**(1), 49–76 (2014)
- Lawless, JF: Negative binomial and mixed poisson regression. *Can. J. Stat.* **15**, 209–225 (1987)

- Lewis, PAW: Recent results in the statistical analysis of univariate point processes. In: Lewis, PAW (ed.) *Stochastic Point Processes*, pp. 1–54. Wiley, New York (1972)
- Lexis, W: Über die Theorie der Statilität Statistischer Reichen. *Jahrb. Nationalökon. u. Statist.* **32**, 60–98 (1879)
- McCullagh, P, Nelder, JA: *Generalized Linear Models*, 2nd edn. Chapman & Hall, London (1989)
- McKendrick, AG: The applications of mathematics to medical problems. *Proc. Edinb. Math. Soc.* **44**, 98–130 (1925). doi:10.1017/S0013091500034428
- McLachlan, JA, Peel, D: *Finite Mixture Models*. Wiley, New York (2001)
- Moore, DF: Asymptotic properties of moment estimators for overdispersed counts and proportions. *Biometrika* **23**, 583–588 (1986)
- Nelder, JA, Pregibon, D: An extended quasi-likelihood function. *Biometrika* **74**, 221–232 (1987)
- Ngatchou-Wandji, J, Paris, C: On the zero-inflated count models with application to modelling annual trends in incidences of some occupational allergic diseases in France. *J. Data Sci.* **9**, 639–659 (2011)
- Panaretos, J: An extension of the damage model. *Metrika* **29**, 189–194 (1982)
- Panaretos, J: A generating model involving Pascal and logarithmic series distributions. *Commun. Stat. A-Theor.* **12**(7), 841–848 (1983)
- Panaretos, J: On the evolution of surnames. *Int. Stat. Rev.* **57**(2), 161–167 (1989)
- Patil, GP, Ord, JK: On size-biased sapling and related form-invariant weighted distributions. *Sankhya* **38**, 48–61 (1976)
- Rao, CR: On discrete distributions arising out of methods of ascertainment. *Sankhya* **A25**, 311–324 (1963)
- Rao, CR: *Weighted Distributions Arising Out of Methods of Ascertainment*. In: Atkinson, AC, Fienberg, SE (eds.) *A Celebration of Statistics*, Chapter 24, pp. 543–569. Springer-Verlag, New York (1985)
- Ridout, M, Demetrio, CGB, Hinde, J: Models for count data with many zeros, pp. 179–190. Invited paper presented at the Nineteenth International Biometric Conference, Cape Town, South Africa (1998)
- Ridout, M, Hinde, J, Demetrio, CGB: A score test for testing zero inflated Poisson regression model against zero inflated negative binomial alternatives. *Biometrics* **57**, 219–223 (2001)
- Ripley, BD: The second-order analysis of stationary point processes. *J. Appl. Probab.* **13**, 255–266 (1976)
- Ripley, BD: Modelling spatial patterns (with Discussion). *J. R. Stat. Soc. B* **39**, 172–212 (1977)
- Shaw, L, Sichel, HS: *Accidents Proneness*. Pergamon Press, Oxford (1971)
- Steutel, FW, van Harn, K: Discrete analogues of self-decomposability and stability. *Ann. Prob.* **7**, 893–899 (1979)
- Student: An explanation of deviations from poisson's law in practice. *Biometrika* **12**, 211–215 (1919)
- Thyriou, P: Extension of the collective risk theory. *Skand. Aktuaritidskrift* **52**(Supplement), 84–98 (1969)
- Titterton, DM: Some recent research in the analysis of mixture distributions. *Statistics* **21**, 619–641 (1990)
- Tripathi, R, Gupta, R, Gurland, J: Estimation of parameters in the beta binomial model. *Ann. I. Stat. Math.* **46**, 317–331 (1994)
- Wang, P, Puterman, M, Cokburn, I, Le, N: Mixed poisson regression models with covariate dependent rates. *Biometrics* **52**, 381–400 (1996)
- Winkelmann, R: Duration dependence and dispersion in count-data models. *J. Bus. Econ. Stat.* **13**(4), 467–474 (1995)
- Xekalaki, E: Chance mechanisms for the univariate generalized Waring distribution and related characterizations. In: Taillie, C, Patil, GP, Baldessari, B (eds.) *Statistical Distributions in Scientific Work*, Vol. 4 (Models, Structures and Characterizations), pp. 157–171. Reidel, Dordrecht (1981)
- Xekalaki, E: The univariate generalized Waring distribution in relation to accident theory: proneness, spells or contagion? *Biometrics* **39**(3), 887–895 (1983a)
- Xekalaki, E: Infinite divisibility, completeness and regression properties of the univariate generalized Waring distribution. *Ann. I. Stat. Math. A* **35**, 279–289 (1983b)
- Xekalaki, E: A property of the Yule distribution and its applications. *Commun. Stat. A-Theor.* **12**(10), 1181–1189 (1983c)
- Xekalaki, E: The bivariate generalized Waring distribution and its application to accident theory. *J. R. Stat. Soc. A* **147**(3), 488–498 (1984a)
- Xekalaki, E: Linear regression and the Yule distribution. *J. Econometrics* **24**(1), 397–403 (1984b)
- Xekalaki, E: Models leading to the bivariate generalized Waring distribution. *Utilitas Math.* **25**, 263–290 (1984c)
- Xekalaki, E: The multivariate generalized Waring distribution. *Commun. Stat. A-Theor.* **15**(3), 1047–1064 (1986)
- Xekalaki, E: Under- and Overdispersion. *Enc. Act. Sci.* **3**, (2006) [<http://onlinelibrary.wiley.com/doi/10.1002/9780470012505.tau003/abstract>]
- Xekalaki, E, Panaretos, J: Identifiability of compound poisson distributions. *Scand. Actuar. J.* **66**, 39–45 (1983)
- Xekalaki, E, Zografis, M: The generalized Waring process and its application. *Commun. Stat. A-Theor.* **37**(12), 1835–1854 (2008)
- Xue, D, Deddens, J: Overdispersed negative binomial models. *Commun. Stat. A-Theor.* **21**, 2215–2226 (1992)
- Yule, GW: A mathematical theory of evolution based on the conclusions of J.C. Willis, F.R.S. *Philos. T. R. Soc. B* **213**, 21–87 (1924)

doi:10.1186/s40488-014-0019-z

Cite this article as: Xekalaki: On the distribution theory of over-dispersion. *Journal of Statistical Distributions and Applications* 2014 1:19.