Wiley StatsRef:
Statistics Reference Online

Under- and Overdispersion

By *Evdokia Xekalaki*

Keywords: *heterogeneity, contagion, clustering, spells model, accident proneness, mixtures, biased sampling, zero-adjusted models, Poisson distribution, negative binomial distribution, generalized Waring distribution, generalized Waring process*

Abstract: A synopsis of probability models for over- and underdispersion is provided, looking at their origins, motivation, first main contributions, important milestones, and applications. 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 contexts. Of course, with appropriate parameter interpretations, the results are adaptable in a great variety of other situations in fields ranging from economics, inventory control, and insurance through to demometry, biometry, psychometry, and web access modeling.

1 Introduction

Very often, in connection with applications, one is faced with data that exhibit a variability, which differs from that anticipated under the hypothesized model. This phenomenon is termed *overdispersion*, if the observed variability exceeds the expected variability, or *underdispersion*, if it is lower than expected.

Such discrepancies between the empirical variance and the nominal variance can be interpreted to be induced by a failure of some of the basic assumptions of the model. These failures can be classified by how they arise, that is, by the mechanism leading to them. In traditional experimental contexts, for example, they are caused by deviations from the hypothesized structure of the population, such as lack of independence between individual item responses, contagion, clustering, and *heterogeneity*. In observational study contexts, they are the result of the method of ascertainment, which can lead to partial distortion of the observations. In both cases, the observed value x does not represent an observation on the *random variable* X, but an observation on a random variable Y whose probability distribution (*the observed probability distribution*) is a distorted version of the probability distribution of X (*original distribution*). This can have a variance greater than that anticipated under the original distribution (*overdispersion*) or, much less frequently in practice, lower than expected (*underdispersion*).

Such practical complications have been noticed for over a century[1,2]. The *Lexis ratio*[1] appears to be the first statistic for testing for the presence of over- or underdispersion relative to a *binomial* hypothesized model when, in fact, the population is structured in *clusters*. In such a setup, Y, the number of successes,

Athens University of Economics and Business, Athens, Greece

Update based on original article by Evdokia Xekalaki, Wiley StatsRef: Statistics Reference Online, © 2014, John Wiley & Sons, Ltd

DOI: 10.1002/9781118445112.stat04407.pub2

can be represented as the series of the numbers Y_i of successes in all clusters. Then, assuming N clusters of size n, the Lexis ratio, defined as the ratio of the between-clusters variance to the total variance $Q = n\sum_{i=1}^{N}(p_i - \hat{p})^2/(\hat{p}(1-\hat{p}))$ with $\hat{p}_i = Y_i/n$ and $\hat{p} = Y/(Nn)$, indicates over- or underdispersion if its value exceeds or is exceeded by unity. In addition, for count data, *Fisher*[3] considered using the sample *index of dispersion* $\sum_{i=1}^{n}(Y_i - \overline{Y})/\overline{Y}$, where $\overline{Y} = \sum_{i=1}^{n} Y_i/n$ for testing the appropriateness of a *Poisson distribution* for Y, on the basis of a sample $Y_1, Y_2, \ldots, Y_n$ of observations on Y.

2 Modeling Over- or Underdispersion

Analyzing data using single-parameter distributions implies that the *variance* can be determined by the mean. Over- or underdispersion leads to failure of this relation with an effect that practice has shown not to be ignorable, as using statistics appropriate for the single-parameter family may induce low efficiency, although, for modest amounts of overdispersion, this may not be the case[4]. So, detailed representation of the over- or underdispersion is required. There are many ways in which these two concepts can be represented. The most important are described in the sequel.

2.1 Lack of Independence between Individual Responses

In accident study-related contexts, the mean and variance of $Y = \sum_{i=1}^{x} Y_i$, the total number of reported accidents in a total of x accidents that actually occurred, if individual accidents are reported with probability $p = P(Y_i = 1) = 1 - P(Y_i = 0)$, but not independently ($\mathrm{Cor}(Y_i, Y_j) = \rho \neq 0$), will have mean $E(Y) = np$ and variance $V(Y) = V\left(\sum_{i=1}^{x} Y_i\right) = xp(1-p) + 2\binom{x}{2}\rho p(1-p) = xp(1-p)\,(1+\rho(x-1))$. The variance of Y exceeds that anticipated under a hypothesized independent trial binomial model if $\rho > 0$ (overdispersion) and is exceeded by it if $\rho < 0$ (underdispersion).

2.2 Contagion

Another way in which variance inflation may result is through a failure of the assumption that the probability of the occurrence of an event in a very short interval is constant, that is, not affected by the number of its previous occurrences. This leads to the classical *contagion* model[5,6], which in the framework of data-modeling problems faced by actuaries, for example, postulates that initially, all individuals have the same probability of incurring an accident, but later, this probability changes by each accident sustained. In particular, it assumes that none of the individuals has had an accident (e.g., new drivers or persons who are just beginning a new type of work), and that the probability with which a person with $Y = y$ accidents by time t will have another accident in the time period from t to $t + \mathrm{d}t$ is of the form $(k + my)\mathrm{d}t$, thus yielding the *negative binomial* as the distribution of Y with probability function $P(Y = y) = \binom{k/m}{y} \mathrm{e}^{-kt}(1 - \mathrm{e}^{-mt})^y$ and mean and variance given by $\mu = E(Y) = k(\mathrm{e}^{mt} - 1)/m$, and $V(Y) = k\mathrm{e}^{mt}(\mathrm{e}^{mt} - 1)/m = \mu\mathrm{e}^{mt}$, respectively.

2.3 Clustering

A clustered structure of the population, frequently overlooked by data analysts, may also induce over- or underdispersion. As an example, in an accident context again, the number Y of injuries incurred by persons involved in N accidents can naturally be thought of, as expressed by the sum $Y = Y_1 + Y_2 + \cdots + Y_N$, of independent random variables representing the numbers Y_i of injuries resulting from the ith accident.

DOI: 10.1002/9781118445112.stat04407.pub2

These are assumed to be identically distributed independently of the total number of accidents N. So, an accident is regarded as a *cluster* of injuries. In this case, the mean and variance of Y will be $E(Y) = E\left(\sum_{i=1}^{N} Y_i\right) = \mu E(N)$ and $V(Y) = V\left(\sum_{i=1}^{N} Y_i\right) = \sigma^2 E(N) + \mu^2 V(N)$ with μ and σ^2 denoting the mean and variance of the common distribution of Y_1. So, with N being a Poisson variable with mean $E(N) = \theta$ equal to $V(N)$, the last relationship leads to over- or underdispersion accordingly as $\sigma^2 + \mu^2$ is greater or less than 1.

The first such model was introduced by Cresswell and Froggatt[7] in a different accident context, in which they assumed that each person is liable to spells of weak performance during which all of the person's accidents occur. So, if the number N of spells in a unit time period is a Poisson variate with parameter θ and assuming that within a spell a person can have 0 accidents with probability $1 - m\log p$ and n accidents ($n \geq 1$) with probability $m(1-p)^n/n$; $m, n > 0$, one is led to a negative binomial distribution as the distribution of Y 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 (see, e.g., Refs 6, 8, 9).

2.4 Heterogeneity

Another sort of deviation from the underlying structure of a population that may give rise to over- or underdispersion is assuming a homogeneous population when in fact the population is heterogeneous, that is, when its individuals have constant but unequal probabilities of sustaining an event. In this case, each member of the population has its own value of the parameter θ and probability density function $f(\cdot\,;\theta)$, so that θ can be considered as the inhomogeneity parameter that varies from individual to individual according to a probability distribution $G(\cdot)$ of mean μ and variance σ^2, and the observed distribution of Y has probability density function given by

$$f_Y(y) = E_\theta(f(y;\theta)) = \int_\Theta f(y;\theta)\mathrm{d}G(\theta) \tag{1}$$

where Θ is the parameter space and $G(\cdot)$ can be any continuous, discrete, or finite-step distribution. Models of this type are known as *mixtures* (For details on their application in the statistical literature, see Refs 10–13). Under such models, $V(Y) = V(E(Y \mid \theta)) + E(V(Y|\theta))$, that is, the variance of Y consists of two parts, one representing variance owing to the variability of θ and one reflecting the inherent variability of Y if θ did not vary. Letting θ vary acts as a means of "loosening" the structure of the initial model, thus offering more realistic interpretations of the mechanisms that generated the data. One can recognize that a similar idea is the basis for *ANOVA* models where the total variability is split into the "*between groups*" and the "*within groups*" components. The above variance relationship offers an explanation as to why mixture models are often termed as *overdispersion models*. In the case of the Poisson (θ) distribution, we have, in particular, that $V(Y) = E(\theta) + V(\theta)$. More generally, the factorial moments of Y are the same as the moments about the origin of θ, and Carriere[14] made use of this fact to construct a test of the hypothesis that a Poisson mixture fits a data set.

Historically, the derivation of *mixed Poisson distributions* was first considered by Greenwood and Woods[15] in the context of accidents. Assuming that the accident experience $X|\theta$ of an individual is Poisson distributed with parameter θ varying from individual to individual according to a *gamma distribution* with mean μ and index parameter μ/γ, they obtained a negative binomial distribution as the distribution of Y with probability function $P(Y = y) = \binom{\mu/\gamma + y - 1}{y} \{\gamma/(1+\gamma)\}^y (1+\gamma)^{-\mu/\gamma}$. Under

DOI: 10.1002/9781118445112.stat04407.pub2

such an observed distribution, the mean and variance of Y are $E(Y) = \mu$, $V(Y) = \mu(1 + \gamma)$, where now γ represents the overdispersion parameter. The mixed Poisson process has been popularized in the actuarial literature by Dubourdieu[16] and the gamma mixed case was treated by Thyrion[17].

A large number of other mixtures have been developed for various overdispersed data cases, such as binomial mixtures[18], negative binomial mixtures[6,9,19,20], *normal mixtures*[21], and *exponential mixtures*[22].

Often in practice, a finite-step distribution is assumed for θ in Equation (1), and the interest is on creating clusters of data by grouping the observations on Y (*cluster analysis*). A *finite mixture* model is applied and each observation is allocated to a cluster using the estimated parameters and a decision criterion[12]. The number of clusters can be decided on the basis of a testing procedure for the number of components in the finite mixture[23].

2.5 Treating Heterogeneity in a More General Way

Extending heterogeneity models for models with explanatory variables, one may assume that Y has a parameter θ, which varies from individual to individual according to a *regression model* $\theta = \eta(\mathbf{x};\boldsymbol{\beta}) + \varepsilon$, where $\mathbf{x}$ is a vector of *explanatory variables*, $\boldsymbol{\beta}$ is a vector of regression coefficients, η is a function of a known form, and ε has some known distribution. Such models are known as *random effect* models, and have been extensively studied for the broad family of *generalized linear models*. Consider, for example, the Poisson regression case. For simplicity, we consider only a single *covariate*, say X. A model of this type assumes that the data Y_i, $i = 1, 2, \ldots, n$ follow a Poisson distribution with mean θ such that $\log\theta = \alpha + \beta x + \varepsilon$ for some constants α, β and with ε having a distribution with mean equal to 0 and variance say φ. Now, the marginal distribution of Y is no longer the Poisson distribution, but a mixed Poisson distribution, with mixing distribution clearly depending on the distribution of ε. From the regression equation, one can obtain that $Y \sim$ Poisson $(t \exp(\alpha + \beta x))_t\, g(t)$, where $t = \mathrm{e}^{\varepsilon}$ with a distribution $g(\cdot)$ that depends on the distribution of ε. Negative Binomial and Poisson Inverse Gaussian regression models have been proposed as overdispersed alternatives to the Poisson regression model[24−26]. If the distribution of t is a two finite−step distribution, the finite Poison mixture regression model of Wang *et al.*[27] results. The similarity of the mixture representation and the random effects one is discussed in Ref. 28.

In meta-analysis contexts, overdispersion (or underdispersion) refers to variance inflation (or deflation) relative to that anticipated by the fixed effects model with the population structure in clusters or some form of mixing leading a compound distribution as the two commonly assumed factors. Kulinskaya and Olkin[29] proposed approaching the problem of specification of a random effects model in such contexts with a multiplicative model for the distribution of the effect size parameters. Their model was motivated by overdispersion induced by intra-class correlation in the model assumed for the distribution of the ith effect-size estimate. In particular, the variance of the estimator $\widehat{\theta}_i$ of the effect size parameter θ_i in the ith study is assumed to be of the form $\sigma^2_{\widehat{\theta}_i} = (1 + \alpha(n_i)\gamma)\sigma_i^2$, where $\alpha(n_i)$ are some known functions of the sample sizes n_i, σ_i^2 is the within the ith study variance, $i = 1, 2, \ldots, k$, and γ is interpreted as an intraclass correlation parameter.

2.6 Estimation and Testing for Overdispersion Under Mixture Models

Mixture models including those arising through treating the parameter θ as the dependent variable in a regression model allow for different forms of variance to mean relationships. So, assuming that $E(Y) = \mu(\beta)$, $V(Y) = \sigma^2(\mu(\beta), \lambda)$ for some parameters β, λ, a number of estimation approaches exist in the literature based on moment methods[25,30,31] and *quasi-* or *pseudolikelihood* methods[32−34]. The above formulation also allows for multiplicative overdispersion or more complicated variance parameters as in Ref. 33.

DOI: 10.1002/9781118445112.stat04407.pub2

Testing for the presence of over- or underdispersion in the case of mixtures can be done using asymptotic arguments. Under regularity conditions, the density of Y_i in a sample $Y_1, Y_2, \dots, Y_n$ from the distribution of Y is $f_Y(y) = E_\Theta\ (f(y;\theta)) = f(y;\mu_\theta) + (1/2)\sigma_\theta^2\ (\partial^2 f(y;\mu_\theta)/\partial\mu_\theta^2) + O(1/n)$, where $\mu_\theta = E(\theta)$, $\sigma_\theta^2 = V(\theta)$, that is, of the form $f(y;\mu_\theta)\ (1 + \varepsilon h(y;\phi_\theta))$, where $h(y;\mu_\theta) = [\partial \log f(y;\mu_\theta)/\partial\mu_\theta]^2 + \partial^2 \log f(y;\mu_\theta)/\partial\mu_\theta^2$.

This family represents overdispersion, if $\varepsilon > 0$, underdispersion, if $\varepsilon < 0$, and none of these complications, if $\varepsilon = 0$. The above representation was derived by Cox[4], who suggested a testing procedure for the hypothesis $\varepsilon = 0$, which can be regarded as a general version of standard dispersion tests.

2.7 Zero Adjusted Models

It would be interesting to note that often over- or underdispersion can be thought of as caused by the presence of an excess or a scant number of zeros. The models discussed so far 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 ω 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 (see, e.g., Ref. 35 and the references therein).

Obviously, ω can be interpreted as the proportion of excess zeros in the case of the first two models and as noted by Xekalaki[35], the above representation explains why they 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[36−38]. Gupta *et al.*[39] 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 Refs 40−42.

A shortcoming of all the three of the above accident models is that they treat the data as if the individuals under observation were exposed to the same environmental risk. Being critical of this, Irwin[43] introduced a model that allowed separately for proneness and exposure leading to a Waring distribution. Xekalaki[6] showed that this distribution could also arise in contagion and spells contexts under the assumption of variable exposure to risk.

As noted by Xekalaki[35], the contagion derivation of the negative binomial and generalized Waring distributions closely relates to a modeling approach, whereby the distribution of accident occurrences in a time interval $(0, t)$ is regarded as underpinned by a pure birth process $\{X_t, t = 0, 1, 2, \dots\}$ with the probability of a person to incur an accident in $(t, t + \mathrm{d}t)$, having had x accidents by time t given by $P(X_{t+\delta t} = x + 1 | X_t = x) = f_\lambda(n, t)\delta t + o(\delta t)$.

Irwin[44] derived the negative binomial distribution with parameters k/m and $(1 - \mathrm{e}^{-\lambda m t})^{-1}$ on this hypothesis. Relaxing Irwin's implicit assumption that all individuals were exposed to the same accident risk, Xekalaki[45] treated the parameter λ as referring to a variable risk exposure according to an exponential distribution with density $a\mathrm{e}^{-a\lambda}$, $a > 0$ and obtained the generalized Waring distribution as the accident distribution with parameters k/m, 1 and $a/(mt)$.

Looking more generally into modeling temporally evolving data through the spectra of the proneness and spells models, but in the presence of variable external risk, the generalized Waring process can be employed as demonstrated by Xekalaki and Zografi[46].

Faddy[47] approached under- and overdispersion relative to the Poisson distribution proposing 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 $f_\lambda(x, t)$.

Finally, Winkelmann[48] looked at under- and overdispersion using renewal theory by demonstrating that such discrepancies are conveyed by the monotonicity of an appropriately chosen hazard function of the waiting times.

3 The Effect of the Method of Ascertainment

In collecting a sample of observations produced by nature according to some model, the original distribution may not be reproduced owing to various reasons. These include partial destruction or enhancement (augmentation) of observations. Situations of the former type are known in the literature as *damage models*, while situations of the latter type are known as *generating models*. 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. As a result, the observed distribution is a distorted version of the original distribution obtained as a mixture of the distortion mechanism. In particular, in the case of damage,

$$P(Y = r) = \sum_{n=r}^{\infty} P(Y = r|X = n)P(X = n) \qquad r = 0, 1, \ldots \tag{2}$$

while, in the case of enhancement

$$P(Y = r) = \sum_{n=1}^{r} P(Y = r|X = n)P(X = n) \qquad r = 1, 2, \ldots \tag{3}$$

Various forms of distributions have been considered for the distortion mechanisms in the above two cases. In the case of damage, the most popular forms have been the binomial distribution[49], mixtures on p of the binomial distribution[50,51] whenever damage can be regarded as additive ($Y = X - U$, U independent of Y), or in terms of the *uniform distribution* in $(0, x)$[52–54] 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[55]. The generating model was introduced and studied by Panaretos[56].

In actuarial contexts, where modeling the distributions of the *numbers of accidents*, of the damage claims, and of the *claimed amounts* is important, these models provide a perceptive approach to the interpretation of such data. This is justified by the fact that people have a tendency to underreport 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 overreport damages incurred by them, so that the reported damages Y are greater than or equal to the true damages $X (Y \geq X)$.

Another type of distortion is induced by the adoption of a sampling procedure that gives 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 results by 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 function $w(x;\beta)$, $\beta \in R$(*weight function*), 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., Refs. 57–60). For $w(x;\beta) = x$, these are known as *size-biased distributions* (*see* **Size-Biased Sampling**).

In modeling data that are of interest to actuaries, the weight function can represent reporting *bias* and can help model the over- or underdispersion in the data induced by the reporting bias. In the context of

DOI: 10.1002/9781118445112.stat04407.pub2

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. For a Poisson (θ) distributed X, these lead to distributions for Y of Poisson type. In particular, $w(x;\beta)=x$ leads to

$$P(Y=x)=\frac{e^{-\theta}\theta^{x-1}}{(x-1)!}\quad x=1,2,\ldots \quad \text{(shifted Poisson}(\theta)) \tag{4}$$

while $w(x;\beta)=\beta^x$ leads to

$$P(Y=x)=\frac{e^{-\theta\beta}(\theta\beta)^x}{x!},\quad x=0,1,\ldots \quad (\text{Poisson}(\theta\beta)) \tag{5}$$

The variance of Y under Equation (4) is $1+\theta$ and exceeds that of X (overdispersion), while under Equation (5) it is $\theta\beta$ implying overdispersion for $\beta>1$ or underdispersion for $\beta<1$.

Related Articles

Hidden Markov Models; **Truncated Distributions**; **Zero-Modified Frequency Distributions**.

References

[1] Lexis, W. (1879) Über die Theorie der Statilität Statistischer Reichen. *Jahrbucher ful Nationalökonomie und Statistik*, **32**, 60–98.
[2] Student (1919) An explanation of deviations from Poisson's law in practice. *Biometrika*, **12**, 211–215.
[3] Fisher, R.A. (1950) The significance of deviations from expectation in a Poisson series. *Biometrics*, **6**, 17–24.
[4] Cox, D.R. (1983) Some remarks on overdispersion. *Biometrika*, **70**, 269–274.
[5] Greenwood, M. and Yule, G.U. (1920) 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.
[6] Xekalaki, E. (1983b) The bivariate generalised Waring distribution in relation to accident theory: proneness, spells or contagion? *Biometrics*, **39** (3), 887–895.
[7] Cresswell, W.L. and Froggatt, P. (1963) *The Causation of Bus Driver Accidents*, Oxford University Press, London.
[8] Cane, V.R. (1977) A class of non-identifiable stochastic models. *J. Appl. Probab.*, **14**, 475–482.
[9] Xekalaki, E. (1984a) The bivariate generalized Waring distribution and its application to accident theory. *J. R. Stat. Soc. Ser. A*, **147**, 488–498.
[10] Böhning, D. (1999) *Computer Assisted Analysis of Mixtures (C.A.M.AN)*, Marcel Dekker, New York.
[11] Karlis, D. and Xekalaki, E. (2003) Mixtures everywhere, in *Statistical Musings: Perspectives from the Pioneers of the Late 20th Century* (ed. J. Panaretos), Erlbaum, USA, pp. 78–95.
[12] McLachlan, J.A. and Peel, D. (2001) *Finite Mixture Models*, Wiley, New York.
[13] Titterington, D.M. (1990) Some recent research in the analysis of mixture distributions. *Statistics*, **21**, 619–641.
[14] Carriere, J. (1993) Nonparametric tests for mixed Poisson distributions. *Insur. Math. Econ.*, **12**, 3–8.
[15] Greenwood, M. and Woods, H.M. (1919) On the Incidence of Industrial Accidents upon Individuals with Special Reference to Multiple Accidents. Report of the Industrial Fatigue Research Board, 4, His Majesty's Stationary Office, London, pp. 1–28.
[16] Dubourdieu, J. (1938) Remarques relatives sur la Théorie Mathématique de l' Assurance-Accidents. *Bull Trimest. Inst. Actuaires Francais.*, **44**, 79–146.
[17] Thyrion, P. (1969) Extension of the collective risk theory. *Skandinavisk Aktuarietidskrift*, **52** (Supplement), 84–98.
[18] Tripathi, R., Gupta, R., and Gurland, J. (1994) Estimation of parameters in the beta binomial model. *Ann. Inst. Stat. Math.*, **46**, 317–331.
[19] Irwin, J.O. (1975) The generalized Waring distribution. *J. R. Stat. Soc. Ser. A*, **138**, 18–31 (Part I), 204–227 (Part II), 374–384 (Part III),

DOI: 10.1002/9781118445112.stat04407.pub2

[20] Xekalaki, E. (1983a) A property of the yule distribution and its applications. *Commun. Stat. Theory Methods*, **12**, 1181–1189.
[21] Andrews, D.F. and Mallows, C.L. (1974) Scale mixtures of normal distributions. *J. R. Stat. Soc. Ser. B*, **36**, 99–102.
[22] Jewell, N. (1982) Mixtures of exponential distributions. *Ann. Stat.*, **10**, 479–484.
[23] Karlis, D. and Xekalaki, E. (1999) On testing for the number of components in a mixed Poisson model. *Ann. Inst. Stat. Math.*, **51**, 149–162.
[24] Dean, C.B., Lawless, J., and Willmot, G.E. (1989) A mixed Poisson-inverse Gaussian regression model. *Can. J. Stat.*, **17**, 171–182.
[25] Lawless, J.F. (1987) Negative binomial and mixed Poisson regression. *Can. J. Stat.*, **15**, 209–225.
[26] Xue, D. and Deddens, J. (1992) Overdispersed negative binomial models. *Commun. Stat. Theory Methods*, **21**, 2215–2226.
[27] Wang, P., Puterman, M., Cokburn, I., and Le, N. (1996) Mixed Poisson regression models with covariate dependent rates. *Biometrics*, **52**, 381–400.
[28] Hinde, J. and Demetrio, C.G.B. (1998) Overdispersion: models and estimation. *Comput. Stat. Data Anal.*, **27**, 151–170.
[29] Kulinskaya, E. and Olkin, I. (2014) An overdispersion model in meta-analysis. *Stat. Model.*, **14** (1), 49–76.
[30] Breslow, N. (1990) Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. *J. Am. Stat. Assoc.*, **85**, 565–571.
[31] Moore, D.F. (1986) Asymptotic properties of moment estimators for overdispersed counts and proportions. *Biometrika*, **23**, 583–588.
[32] Davidian, M. and Carroll, R.J. (1988) A note on extended quasi-likelihood. *J. R. Stat. Soc. Ser. B*, **50**, 74–82.
[33] McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, 2nd edn, Chapman & Hall, London.
[34] Nelder, J.A. and Pregibon, D. (1987) An extended quasi-likelihood function. *Biometrika*, **74**, 221–232.
[35] Xekalaki, E. (2014) On the distribution theory of over-dispersion. *J. Stat. Distrib. Appl.*, **1** (19), DOI: 10.1186/s40488-014-0019-z
[36] Ridout, M., Hinde, J., and Demetrio, C.G.B. (2001) A score test for testing zero inflated Poisson regression model against zero inflated negative binomial alternatives. *Biometrics*, **57**, 219–223.
[37] Gupta, P.L., Gupta, R.C., and Tripathi, R.C. (2004) Score test for zero inflated generalized Poisson regression model. *Commun. Stat. Theory Methods*, **33**, 47–64.
[38] Famoye, F. and Singh, K.P. (2006) Zero-inflated generalized Poisson regression model with an application to domestic violence data. *J. Data Sci.*, **4**, 117–130.
[39] Gupta, P.L., Gupta, R.C., and Tripathi, R.C. (1996) Analysis of zero adjusted count data. *Comput. Stat. Data Anal.*, **23**, 207–218.
[40] Ridout, M., Demetrio, C.G.B. and Hinde, J. (1998) Models for Count Data with many Zeros. Invited paper presented at the *Nineteenth International Biometric Conference*, Cape Town, South Africa, pp. 179–190.
[41] Gschlößl, S. and Czado, C. (2008) Modelling count data with overdispersion and spatial effects. *Stat. Paper.*, **49**, 531–552.
[42] Ngatchou-Wandji, J. and Paris, C. (2011) 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.
[43] Irwin, J.O. (1968) The generalized Waring distribution applied to accident theory. *J. R. Stat. Soc. Ser. A*, **131**, 205–225.
[44] Irwin, J.O. (1941) Discussion on Chambers and Yule's paper. *J. R. Stat. Soc. Suppl.*, **7**, 101–109.
[45] Xekalaki, E. (1981) Chance mechanisms for the univariate generalized Waring distribution and related characterizations, in *Statistical Distributions in Scientific Work*, Vol. 4 (Models, Structures and Characterizations) (eds C. Taillie, G.P. Patil, and B. Baldessari), Reidel, Dordrecht, pp. 157–171.
[46] Xekalaki, E. and Zografi, M. (2008) The generalized Waring process and its application. *Commun. Stat. Theory Methods*, **37** (12), 1835–1854.
[47] Faddy, M.J. (1997) Extending Poisson process modelling and analysis of count data. *Biom. J.*, **39** (4), 431–440.
[48] Winkelmann, R. (1995) Duration dependence and dispersion in count-data models. *J. Bus. Econ. Stat.*, **13** (4), 467–474.
[49] Rao, C.R. (1963) On discrete distributions arising out of methods of ascertainment. *Sankhyā, A*, **25**, 311–324.
[50] Panaretos, J. (1982) An extension of the damage model. *Metrika*, **29**, 189–194.
[51] Xekalaki, E. and Panaretos, J. (1983) Identifiability of compound Poisson distributions. *Scand. Actuar. J.*, **66**, 39–45.
[52] Dimaki, C. and Xekalaki, E. (1990) Identifiability of income distributions in the context of damage and generating models. *Commun. Stat. Theory Methods*, **19**, 2757–2766.
[53] Dimaki, C. and Xekalaki, E. (1996) Additive and multiplicative distortion of observations: some characteristic properties. *J. Appl. Stat. Sci.*, **5** (2/3), 113–127.
[54] Xekalaki, E. (1984b) Linear regression and the yule distribution. *J. Econom.*, **24** (1), 397–403.
[55] Krishnaji, N. (1970) Characterization of the Pareto distribution through a model of under-reported incomes. *Econometrica*, **38**, 251–255.
[56] Panaretos, J. (1983) A generating model involving Pascal and logarithmic series distributions. *Commun. Stat. Theory Methods*, **12**, 841–848.
[57] Cox, D.R. (1962) *Renewal Theory*, Barnes & Noble, New York.

DOI: 10.1002/9781118445112.stat04407.pub2

[58] Fisher, R.A. (1934) The effects of methods of ascertainment upon the estimation of frequencies. *Ann. Eugen.*, **6**, 13–25.

[59] Patil, G.P. and Ord, J.K. (1976) On size-biased sampling and related form-invariant weighted distributions. *Sankhyā*, **38**, 48–61.

[60] Rao, C.R. (1985) Weighted distributions arising out of methods of ascertainment, in *A Celebration of Statistics* (eds A.C. Atkinson and S.E. Fienberg), Springer-Verlag Chapter 24, New York, pp. 543–569.

DOI: 10.1002/9781118445112.stat04407.pub2