Under- and Overdispersion

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 *over dispersion* if the observed variability exceeds the expected variability or *under dispersion* 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 (*over dispersion*) or, much less frequently in practice, lower than expected (*under dispersion*).

Such practical complications have been noticed for over a century (see, e.g. [25, 35]). The *Lexis ratio* [25] appears to be the first statistic for testing for the presence of over or under dispersion relative to a binomial hypothesized model when, in fact, the population is structured in clusters. In such a setup, Y, the number of successes, 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 under dispersion if its value exceeds or is exceeded by unity. Also, for **count data**, Fisher [15] 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.

Modeling Over or Under dispersion

Analyzing data using single-parameter distributions implies that the variance can be determined by the mean. Over or under dispersion leads to failure of this relation with an effect that practice has shown not to be ignorable, since using statistics appropriate for the single-parameter family may induce low efficiency, although, for modest amounts of over dispersion, this may not be the case [7]. So, detailed representation of the over or under dispersion is required. There are many ways in which these two concepts can be represented. The most important are described in the sequel.

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 $\mathrm{E}(Y) = np$ and variance

$$V(Y) = V(\textstyle\sum_{i=1}^{x} Y_i) = xp(1 - p) + 2\binom{x}{2} \rho p(1 - p) = xp(1 - p)(1 + \rho(x - 1)).$$ The variance of Y exceeds that anticipated under a hypothesized independent trial binomial model if $\rho > 0$ (over dispersion) and is exceeded by it if $\rho < 0$ (under dispersion).

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 [17, 41], 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 $t + dt$ is of the form $(k + my)dt$, thus yielding the negative binomial as the distribution of Y with probability function $P(Y = y) = \binom{k/m}{y} e^{-kt}(1 - e^{-mt})^y$ and mean and variance given by $\mu = \mathrm{E}(Y) = k(e^{mt} - 1)/m$, and $V(Y) = ke^{mt}(e^{mt} - 1)/m = \mu e^{mt}$, respectively.

Clustering

A clustered structure of the population, frequently overlooked by data analysts, may also induce over or under dispersion. 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, which 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 $\mathrm{E}(Y) = \mathrm{E}(\sum_{i=1}^{N} Y_i) = \mu\mathrm{E}(N)$ and $V(Y) = V(\sum_{i=1}^{N} Y_i) = \sigma^2\mathrm{E}(N) + \mu^2 V(N)$ with μ, σ^2 denoting the mean and variance of the common distribution of Y_1. So, with N being a Poisson variable with mean $\mathrm{E}(N) = \theta$ equal to $V(N)$, the last relationship leads to over dispersion or under dispersion according as $\sigma^2 + \mu^2$ is greater or less than 1.

The first such model was introduced by Cresswell and Froggatt [8] 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 over dispersed distributions (see, e.g. [4, 41, 42]).

Heterogeneity

Another sort of deviation from the underlying structure of a population that may give rise to over or under dispersion 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

$$f_Y(y) = \mathrm{E}_G(f(y; \theta)) = \int_{\Theta} f(y; \theta) dG(\theta), \quad (1)$$

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 [2, 22, 27, 36]). Under such models, $V(Y) = V(E(Y|\theta)) + E(V(Y|\theta))$, that is, the variance of Y consists of two parts, one representing variance due to the variability of θ and one reflecting the inherent variability of Y if θ did not vary. 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 *over dispersion 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 [5] 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 [16] 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 $P(Y = y) = \begin{pmatrix} \mu/\gamma + y - 1 \\ y \end{pmatrix} \{\gamma/(1 + \gamma)\}^y (1 + \gamma)^{-\mu/\gamma}$. Under such an observed distribution, the mean and variance of Y are $E(Y) = \mu$, $V(Y) = \mu(1 + \gamma)$, where now γ represents the over dispersion parameter. The mixed Poisson process has been popularized in the actuarial literature by Dubourdieu [13] and the gamma mixed case was treated by Thyrion [37].

A large number of other mixtures have been developed for various over dispersed data cases, such as binomial mixtures (e.g. [38]), negative binomial mixtures (e.g. [19, 40–42]), normal mixtures (e.g. [1]), and **mixtures of exponential** (e.g. [20]).

Often in practice, a finite-step distribution is assumed for θ in (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** (e.g. [27]). The number of clusters can be decided on the basis of a testing procedure for the number of components in the finite mixture [21].

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** $\theta = \eta(x; \beta) + \varepsilon$, where x is a vector of explanatory variables, β 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 $Y \sim Poisson(t \exp(\alpha + \beta x))_i g(t)$, where $t = e^\varepsilon$ with a distribution $g(\cdot)$ that depends on the distribution of ε. Negative Binomial and Poisson Inverse Gaussian regression models have been proposed as over dispersed alternatives to the Poisson regression model [10, 24, 45]. If the distribution of t is a two finite–step distribution, the finite Poison mixture regression model of Wang et al. [39] results. The similarity of the mixture representation and the random effects one is discussed in [18].

Estimation and Testing for Over dispersion 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 (e.g. [3, 24, 28]) and quasi- or pseudolikelihood methods (e.g. [9, 26, 29]). The above formulation also allows for multiplicative over dispersion or more complicated variance parameters as in [26].

Testing for the presence of over dispersion or under dispersion in the case of mixtures can be done

using asymptotic arguments. Under regularity conditions, the density of Y_i in a sample $Y_1, Y_2, \ldots, Y_n$ from the distribution of Y is $f_Y(y) = E_\Theta (f(y;\theta)) = f(y;\mu_\theta) + (1/2)\sigma_\theta^2 \quad (\partial^2 f(y;\mu_\theta)/\partial\mu^2) + O(1/n)$, where $\mu_\theta = E(\theta)$, $\sigma_\theta^2 = E(\theta)$, that is, of the form $f(y;\mu_\theta) (1 + \varepsilon h(y;\varphi_\theta))$, where $h(y;\mu_\theta) = [\partial \log f(y;\mu_\theta)/\partial\mu_\theta]^2 + \partial^2 \log f(y;\mu_\theta)/\partial\mu^2$.

This family represents over dispersion if $\varepsilon > 0$, under dispersion if $\varepsilon < 0$, and none of these complications if $\varepsilon = 0$. The above representation was derived by Cox [7], who suggested a testing procedure for the hypothesis $\varepsilon = 0$, which can be regarded as a general version of standard dispersion tests.

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,

$$P(Y = r) = \sum_{n=r}^{\infty} P(Y = r|X = n)P(X = n),$$

$$r = 0, 1, \ldots, \tag{2}$$

while, in the case of enhancement

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

$$r = 1, 2, \ldots \tag{3}$$

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 [33], mixtures on p of the binomial distribution (e.g., [30,

44]) 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. [11, 12, 43]) 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 [23]. The generating model was introduced and studied by Panaretos [31].

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. [6, 14, 32, 34]). For $w(x;\beta) = x$, these are known as *size-biased distributions*.

In modeling data that are of interest to actuaries, the weight function can represent reporting bias and can help model the over or under dispersion in the data induced by the 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. For a Poisson (θ) distributed X, these lead to distributions for Y of Poisson type. In particular, $w(x;\beta) = x$

leads to

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

$$(\text{shifted Poisson}(\theta)), \tag{4}$$

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

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

$$(\text{Poisson}(\theta\beta)). \tag{5}$$

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

References

[1] Andrews, D.F. & Mallows, C.L. (1974). Scale mixtures of normal distributions, *Journal of the Royal Statistical Society, Series B* **36**, 99–102.

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

[3] Breslow, N. (1990). Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models, *Journal of the American Statistical Association* **85**, 565–571.

[4] Cane, V.R. (1977). A class of non-identifiable stochastic models, *Journal of Applied Probability* **14**, 475–482.

[5] Carriere, J. (1993). Nonparametric tests for mixed Poisson distributions, *Insurance: Mathematics and Economics* **12**, 3–8.

[6] Cox, D.R. (1962). *Renewal Theory*, Barnes & Noble, New York.

[7] Cox, D.R. (1983). Some remarks on overdispersion, *Biometrika* **70**, 269–274.

[8] Cresswell, W.L. & Froggatt, P. (1963). *The Causation of Bus Driver Accidents*, Oxford University Press, London.

[9] Davidian, M. & Carroll, R.J. (1988). A note on extended quasi-likelihood, *Journal of the Royal Statistical Society, Series B* **50**, 74–82.

[10] Dean, C.B., Lawless, J. & Willmot, G.E. (1989). A mixed Poisson-inverse Gaussian regression model, *Canadian Journal of Statistics* **17**, 171–182.

[11] Dimaki, C. & Xekalaki, E. (1990). Identifiability of income distributions in the context of damage and generating models, *Communications in Statistics, Part A (Theory and Methods)* **19**, 2757–2766.

[12] Dimaki, C. & Xekalaki, E. (1996). Additive and multiplicative distortion of observations: some characteristic properties, *Journal of Applied Statistical Science* **5**(2/3), 113–127.

[13] Dubourdieu, J. (1938). Remarques relatives sur la Théorie Mathématique de l' Assurance-Accidents, *Bulletin Trimestriel de l'Institit des Actuaires Francais* **44**, 79–146.

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

[15] Fisher, R.A. (1950). The significance of deviations from expectation in a Poisson series, *Biometrics* **6**, 17–24.

[16] Greenwood, M. & 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.

[17] Greenwood, M. & 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, *Journal of the Royal Statistical Society* **83**, 255–279.

[18] Hinde, J. & Demetrio, C.G.B. (1998). Overdispersion: models and estimation, *Computational Statistics and Data Analysis* **27**, 151–170.

[19] Irwin, J.O. (1975). The generalized Waring distribution, *Journal of the Royal Statistical Society, Series A* **138**, 18–31 (Part I), 204–227 (Part II), 374–384 (Part III).

[20] Jewell, N. (1982). Mixtures of exponential distributions, *Annals of Statistics* **10**, 479–484.

[21] Karlis, D. & Xekalaki, E. (1999). On testing for the number of components in a mixed Poisson model, *Annals of the Institute of Statistical Mathematics* **51**, 149–162.

[22] Karlis, D. & Xekalaki, E. (2003). Mixtures everywhere, in *Statistical Musings: Perspectives from the Pioneers of the Late 20th Century*, J. Panaretos, ed., Erlbaum, USA, pp. 78–95.

[23] Krishnaji, N. (1970). Characterization of the Pareto distribution through a model of under-reported incomes, *Econometrica* **38**, 251–255.

[24] Lawless, J.F. (1987). Negative binomial and mixed Poisson regression, *Canadian Journal of Statistics* **15**, 209–225.

[25] Lexis, W. (1879). Über die Theorie der Statilität Statischer Reichen, *Jahrbucher ful Nationalökonomie und Statistik* **32**, 60–98.

[26] McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*, 2nd Edition, Chapman & Hall, London.

[27] McLachlan, J.A. & Peel, D. (2001). *Finite Mixture Models*, Wiley, New York.

[28] Moore, D.F. (1986). Asymptotic properties of moment estimators for overdispersed counts and proportions, *Biometrika* **23**, 583–588.

[29] Nelder, J.A. & Pregibon, D. (1987). An extended quasi-likelihood function, *Biometrika* **74**, 221–232.

[30] Panaretos, J. (1982). An extension of the damage model, *Metrika* **29**, 189–194.

[31] Panaretos, J. (1983). A generating model involving Pascal and logarithmic series distributions, *Communications in Statistics, Part A (Theory and Methods)* **12**, 841–848.

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

[33] Rao, C.R. (1963). On discrete distributions arising out of methods of ascertainment, *Sankhyā, A* **25**, 311–324.

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

[35] Student (1919). An explanation of deviations from Poisson's law in practice, *Biometrika* **12**, 211–215.

[36] Titterington, D.M. (1990). Some recent research in the analysis of mixture distributions, *Statistics* **21**, 619–641.

[37] Thyrion, P. (1969). Extension of the collective risk theory, *Skandinavisk Aktuarietidskrift* **52** Supplement, 84–98.

[38] Tripathi, R., Gupta, R. & Gurland, J. (1994). Estimation of parameters in the beta binomial model, *Annals of the Institute of Statistical Mathematics* **46**, 317–331.

[39] Wang, P., Puterman, M., Cokburn, I. & Le, N. (1996). Mixed Poisson regression models with covariate dependent rates, *Biometrics* **52**, 381–400.

[40] Xekalaki, E. (1983a). A property of the yule distribution and its applications, *Communications in Statistics, Part A (Theory and Methods)* **12**, 1181–1189.

[41] Xekalaki, E. (1983b). The bivariate generalised Waring distribution in relation to accident theory: proneness, spells or contagion? *Biometrics* **39**, 887–895.

[42] Xekalaki, E. (1984a). The bivariate generalized Waring distribution and its application to accident theory, *Journal of the Royal Statistical Society, Series* **A147**, 488–498.

[43] Xekalaki, E. (1984b). Linear regression and the yule distribution, *Journal of Econometrics* **24**(1), 397–403.

[44] Xekalaki, E. & Panaretos, J. (1983). Identifiability of compound Poisson distributions, *Scandinavian Actuarial Journal* **66**, 39–45.

[45] Xue, D. & Deddens, J. (1992). Overdispersed negative binomial models, *Communications in Statistics, Part A (Theory and Methods)* **21**, 2215–2226.

(*See also* **Discrete Multivariate Distributions; Hidden Markov Models; Truncated Distributions; Zero-modified Frequency Distributions**)

EVDOKIA XEKALAKI