CHAPTER 7

DEVELOPMENT OF THE BIVARIATE, TRIVARIATE AND MULTIVARIATE MODELS AS A FUNCTION OF TIME

To suplement the new theory developed we investigate what happens if these new models are applied to stochastic processes. In this chapter the dependency of the stationary model as a function of time is explored and attention is paid to the limiting process $t \to \infty$.

For the bivariate, trivariate and multivariate models the development as a function of time is achieved through the distribution fitted to the total of variables. This distribution is dependent on time and has a certain influence on the model.

Since an investigation of the models for different time spans in general is not possible, it is undertaken for specific distributions.

7.1 The Bivariate Model

The model is developed for the case where the distribution on the ascending diagonal arrays is BBD (Beta Binomial distribution) and the distribution on the third marginal is NBD (Negative Binomial distribution).

In this section the following steps are performed:

(a) Calculation of the correlation coefficient if we assume that the distribution for the total of variables is the same as the distribution for the third marginal at $t = 1$.

(b) Calculation of the correlation coefficient if we assume that the distribution of the third marginal is a function of time t.

For $t = 1$ the formulae have been given in Section 2.5.5.

For a general t the distribution of the third marginal S is an NBD and has the following moments around the origin:

$$\mu_1'(S|t) = \frac{t\gamma\theta}{1-\theta} \quad \text{and} \tag{7.1.1}$$

$$\mu_2'(S|t) = \frac{t\gamma\theta}{1-\theta}\left[1+\frac{t(\gamma+1)\theta}{1-\theta}\right] . \tag{7.1.2}$$

By applying the general bivariate model in this case we obtain the following results:

$$\mu_1'(x_1|t) = t\,\frac{\alpha_1}{\alpha_1+\alpha_2}\,\frac{\gamma\theta}{1-\theta} , \tag{7.1.3}$$

$$\mu_1'(x_2|t) = t\,\frac{\alpha_2}{\alpha_1+\alpha_2}\,\frac{\gamma\theta}{1-\theta} , \tag{7.1.4}$$

$$\mu_2(x_1|t) = t\,\frac{\alpha_1}{\alpha_1+\alpha_2}\,\frac{\gamma\theta}{1-\theta}\left[1+\frac{\alpha_1+1}{\alpha_1+\alpha_2+1}\,\frac{t\theta}{1-\theta}+\frac{\alpha_2}{(\alpha_1+\alpha_2)(\alpha_1+\alpha_2+1)}\,\frac{t\gamma\theta}{1-\theta}\right], \tag{7.1.5}$$

$$\mu_2(x_2|t) = t\,\frac{\alpha_2}{\alpha_1+\alpha_2}\,\frac{\gamma\theta}{1-\theta}\left[1+\frac{\alpha_2+1}{\alpha_1+\alpha_2+1}\,\frac{t\theta}{1-\theta}+\frac{\alpha_1}{(\alpha_1+\alpha_2)(\alpha_1+\alpha_2+1)}\,\frac{t\gamma\theta}{1-\theta}\right], \tag{7.1.6}$$

and

$$\mathrm{cov}(x_1,x_2|t) = t^2\,\frac{\gamma\alpha_1\alpha_2}{(\alpha_1+\alpha_2)^2(\alpha_1+\alpha_2+1)}\left[\frac{\theta}{1-\theta}\right]^2(\alpha_1+\alpha_2-\gamma)\ . \tag{7.1.7}$$

By comparing the formulae developed for the bivariate model for $t = 1$ with those developed for a general t we conclude that the following relations hold:

$$\mu_1'(x_1|t) = t\,\mu_1'(x_1|t=1)\ , \tag{7.1.8}$$

$$\mu_1'(x_2|t) = t\,\mu_1'(x_2|t=1) \quad \text{and} \tag{7.1.9}$$

$$\mathrm{cov}(x_1,x_2|t) = t^2\,\mathrm{cov}(x_1,x_2|t=1)\ . \tag{7.1.10}$$

The correlation coefficient for a general t is given by:

$$\rho(x_1,x_2|t) = \frac{t\theta}{\alpha_1+\alpha_2+1}\left[1-\frac{\gamma}{\alpha_1+\alpha_2}\right] \tag{7.1.11}$$

$$\cdot\left[\frac{1-\theta}{\alpha_2}+\frac{t\theta\,(\alpha_1+1)}{\alpha_2(\alpha_1+\alpha_2+1)}+\frac{t\theta\gamma}{(\alpha_1+\alpha_2)(\alpha_1+\alpha_2+1)}\right]^{-1/2}$$

$$\cdot\left[\frac{1-\theta}{\alpha_1}+\frac{t\theta\,(\alpha_2+1)}{\alpha_1(\alpha_1+\alpha_2+1)}+\frac{t\theta\gamma}{(\alpha_1+\alpha_2)(\alpha_1+\alpha_2+1)}\right]^{-1/2}\ .$$

Using the limiting process we obtain:

$$\lim_{t\to\infty}\rho(x_1,x_2|t) = \left[1-\frac{\gamma}{\alpha_1+\alpha_2}\right]\left[\left[\frac{\alpha_1+1}{\alpha_2}+\frac{\gamma}{\alpha_1+\alpha_2}\right]\left[\frac{\alpha_2+1}{\alpha_1}+\frac{\gamma}{\alpha_1+\alpha_2}\right]\right]^{-1/2}\ . \tag{7.1.12}$$

If we denote $w = 1 - \dfrac{\gamma}{\alpha_1+\alpha_2}$ then

$$\lim_{t\to\infty} \rho(x_1,x_2|t) = w\left\{\left[1 - w + \frac{\alpha_1+1}{\alpha_2}\right]\left[1 - w + \frac{\alpha_2+1}{\alpha_1}\right]\right\}^{-1/2} . \quad (7.1.13)$$

For the limiting process $t \to \infty$, the correlation coefficient is a function only of γ, α_1 and α_2. Comparing the correlation coefficient $\rho(x_1,x_2|t=1)$ with $\lim_{t\to\infty} \rho(x_1,x_2|t)$ we conclude that both have the same algebraic form but with different variables. The relation between the variables v and w is:

$v = \theta\, w$.

<u>Note</u>

If we assume that $\alpha_1 \to \infty$ and $\alpha_2 \to \infty$ then $\lim_{\substack{\alpha_1\to\infty \\ \alpha_2\to\infty}} w = 1$. The correlation coefficient $\rho(x_1,x_2|t)$ becomes also 1 in the limiting process. This result was actually expected because if $\alpha_1 \to \infty$ and $\alpha_2 \to \infty$, the ascending diagonal arrays tend to be binomially distributed.

For the third marginal we have the NBD which is a function of t. We know that if we have this situation, then the correlation coefficient tends to 1.

7.2 The Trivariate Model

The trivariate model is developed for the case where the distribution of the ascending diagonal arrays is the BBD (α_1,α_2), the distribution of the third marginal is the BBD $(\alpha_1+\alpha_2,\alpha_0)$ and we assume the NBD (γ,θ) for the total of variables.

As in Section 7.1 the correlation coefficient is calculated at $t = 1$ and for general t.

The trivariate model for the above-mentioned distributions at $t = 1$ has been developed in Section 3.2.

For a general t the distribution of the total of variables n is an NBD and has the following moments around the origin:

$$\mu_1'(n|t) = \frac{t\theta\gamma}{1-\theta} \quad \text{and} \tag{7.2.1}$$

$$\mu_2'(n|t) = \frac{t\theta\gamma}{1-\theta}\left[1 + \frac{t\theta(\gamma+1)}{1-\theta}\right]. \tag{7.2.2}$$

By applying the general trivariate model in this case we obtain the following results:

$$\mu_1'(x_1|t) = \frac{\alpha_1}{\alpha_0+\alpha_1+\alpha_2}\,\frac{t\theta\gamma}{1-\theta}, \tag{7.2.3}$$

$$\mu_1'(x_2|t) = \frac{\alpha_2}{\alpha_0+\alpha_1+\alpha_2}\,\frac{t\theta\gamma}{1-\theta}, \tag{7.2.4}$$

$$\mu_2(x_1|t) = \frac{\alpha_1}{(\alpha_0+\alpha_1+\alpha_2)^2(\alpha_0+\alpha_1+\alpha_2+1)}\,\frac{t\theta\gamma}{(1-\theta)^2} \tag{7.2.5}$$

$$\cdot\left[(\alpha_0+\alpha_1+\alpha_2)(\alpha_0+\alpha_1+\alpha_2+1)(1-\theta) + (\alpha_2+1)(\alpha_0+\alpha_1+\alpha_2)t\theta + (\alpha_0+\alpha_1)t\theta\gamma\right],$$

$$\mu_2(x_2|t) = \frac{\alpha_2}{(\alpha_0+\alpha_1+\alpha_2)^2(\alpha_0+\alpha_1+\alpha_2+1)} \frac{t\theta\gamma}{(1-\theta)^2} \tag{7.2.6}$$

$$\cdot\left[(\alpha_0+\alpha_1+\alpha_2)(\alpha_0+\alpha_1+\alpha_2+1)(1-\theta) + (\alpha_1+1)(\alpha_0+\alpha_1+\alpha_2)t\theta + (\alpha_0+\alpha_2)t\theta\gamma\right]$$

and

$$\operatorname{cov}(x_1,x_2|t) = \frac{\alpha_1\alpha_2}{(\alpha_0+\alpha_1+\alpha_2)^2(\alpha_0+\alpha_1+\alpha_2+1)} t^2 \left[\frac{\theta}{1-\theta}\right]^2 \gamma\ (\alpha_0+\alpha_1+\alpha_2-\gamma) . \tag{7.2.7}$$

By comparing the formulae developed for the trivariate model for $t = 1$ with those developed for general t, we arrive at the following relations:

$$\mu_1'(x_1|t) = t\ \mu_1'(x_1|t=1) , \tag{7.2.8}$$

$$\mu_1'(x_2|t) = t\ \mu_1'(x_2|t=1) \quad \text{and} \tag{7.2.9}$$

$$\operatorname{cov}(x_1,x_2|t) = t^2\ \operatorname{cov}(x_1,x_2|t=1) . \tag{7.2.10}$$

The correlation coefficient for general t is given by:

$$\rho(x_1,x_2|t) = t\ \theta\ (\alpha_0+\alpha_1+\alpha_2-\gamma) \tag{7.2.11}$$

$$\cdot\left[\frac{(\alpha_0+\alpha_1+\alpha_2)(\alpha_0+\alpha_1+\alpha_2+1)(1-\theta)}{\alpha_1} + \frac{(\alpha_2+1)(\alpha_0+\alpha_1+\alpha_2)}{\alpha_1}t\theta + \frac{\alpha_0+\alpha_1}{\alpha_1}t\theta\gamma\right]^{-1/2}$$

$$\cdot\left[\frac{(\alpha_0+\alpha_1+\alpha_2)(\alpha_0+\alpha_1+\alpha_2+1)(1-\theta)}{\alpha_2} + \frac{(\alpha_1+1)(\alpha_0+\alpha_1+\alpha_2)}{\alpha_2}t\theta + \frac{\alpha_0+\alpha_2}{\alpha_2}t\theta\gamma\right]^{-1/2} .$$

By using the limiting process we obtain:

$$\lim_{t\to\infty} \rho(x_1,x_2|t) = \left[1 - \frac{\gamma}{\alpha_0+\alpha_1+\alpha_2}\right] \tag{7.2.12}$$

$$\cdot\left[\left[\frac{\alpha_2+1}{\alpha_1} + \frac{\alpha_0+\alpha_1}{\alpha_1(\alpha_0+\alpha_1+\alpha_2)}\gamma\right]\left[\frac{\alpha_1+1}{\alpha_2} + \frac{\alpha_0+\alpha_2}{\alpha_1(\alpha_0+\alpha_1+\alpha_2)}\gamma\right]\right]^{-1/2} .$$

By substituting $u = 1 - \dfrac{\gamma}{\alpha_0+\alpha_1+\alpha_2}$, the expression (7.2.12) becomes:

$$\lim_{t\to\infty} \rho(x_1,x_2|t) = u\left[\left[\frac{\alpha_2+1}{\alpha_1} + \frac{\alpha_0+\alpha_1}{\alpha_1}(1-u)\right]\left[\frac{\alpha_1+1}{\alpha_2} + \frac{\alpha_0+\alpha_2}{\alpha_2}(1-u)\right]\right]^{-1/2} . \tag{7.2.13}$$

7.3 The Multivariate Model

The multivariate model applied to stochastic processes behaves in a similar way as the trivariate model. The principle to be used to solve this model is to reduce it to several trivariate models.

CHAPTER 8

ESTIMATION

The models discussed in Chapters 2, 3 and 4 are presented in Chapter 9 by means of data set analyses. The fitting of various distributions discussed in Chapter 2 to data sets necessitates the estimation of their parameters. In this chapter we give a number of estimation methods for the BBD, NBD, IGP and GIGP.

8.1 Methods of Estimation for the Beta Binomial Distribution

The probability distribution $f(x|S,\alpha_1,\alpha_2)$ of the BBD is a function of the three parameters S, α_1 and α_2 given by:

$$f(x|S,\alpha_1,\alpha_2) = \begin{bmatrix} S \\ x \end{bmatrix} \frac{B(x+\alpha_1, S-x+\alpha_2)}{B(\alpha_1,\alpha_2)}, \text{ where}$$

$x = 0,1,2,\ldots,S$, $\alpha_1 > 0$, $\alpha_2 > 0$ and $S = 1,2,3,\ldots$.

Skellam (1948) proposed the following method of estimation for the parameters S, α_1 and α_2 based on the ratio:

$$G_k = \frac{\mu'_{(k)}}{\mu'_{(k-1)}} = \frac{(S-k+1)(k+\alpha_1-1)}{(k+\alpha_1+\alpha_2-1)}, \text{ where} \qquad (8.1.1)$$

$\mu'_{(k)}$ is the k-th factorial moment of the Beta Binomial distribution given as

$$\mu'_{(k)} = S(S-1)\dots(S-k+1)\,\frac{B(k+\alpha_1,\alpha_2)}{B(\alpha_1,\alpha_2)}\,. \qquad (8.1.2)$$

In many cases one needs to estimate all three parameters S, α_1 and α_2. This requires three simultaneous equations expressed in terms of the first three factorial moments of the distribution, as follows:

$$G_1 = \frac{S\alpha_1}{\alpha_1+\alpha_2}\,, \qquad (8.1.3)$$

$$G_2 = \frac{(S-1)(\alpha_1+1)}{(\alpha_1+\alpha_2+1)} \quad \text{and} \qquad (8.1.4)$$

$$G_3 = \frac{(S-2)(\alpha_1+2)}{(\alpha_1+\alpha_2+2)}\,. \qquad (8.1.5)$$

The first two equations are equivalent to:

$$\alpha_1 = \frac{G_1G_2 - (S-1)G_1}{(S-1)G_1 - SG_2} \quad \text{and} \qquad (8.1.6)$$

$$\alpha_2 = \alpha_1\left[\frac{S}{G_1} - 1\right]. \qquad (8.1.7)$$

Substitution of the above expressions of α_1 and α_2 in terms of G_1 and G_2 into the third equation gives:

$$S^2(G_1-2G_2+G_3) + S(G_1G_2-3G_1+4G_2-2G_1G_3+G_2G_3-G_3) + 2G_1(G_3-G_2+1) = 0 \qquad (8.1.8)$$

The positive root of the equation (8.1.8), taken to the nearest integer, is an estimate for S, where G_1, G_2 and G_3 are the observed values.

The parameters α_1 and α_2 are determined from equations (8.1.6) and (8.1.7) using the estimate obtained for S. Below, three

methods of finding α_1 and α_2 are presented for the case of S known a priori. They are:

(a) method of moments;

(b) method of equating the first sample proportion and the sample mean to their population values; and

(c) method of equating the first and the last observed sample proportions to their theoretical values.

8.1.1 Method of moments

The method of moments consists of equating the observed first two moments around the origin to their theoretical values. Parameters $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are obtained from the following equations:

$$\hat{\alpha}_1 = \frac{S\,\hat{\mu}'_1 - \hat{\mu}'_2}{S\,(\hat{\mu}'_2/\hat{\mu}'_1 - \hat{\mu}'_1 - 1) + \hat{\mu}'_1} \quad \text{and} \tag{8.1.9}$$

$$\hat{\alpha}_2 = \left[S/\hat{\mu}'_1 - 1\right] \hat{\alpha}_1 \text{ , where} \tag{8.1.10}$$

$\hat{\mu}'_1$ and $\hat{\mu}'_2$ are the first two sample moments around the origin.

8.1.2 Method of equating the first sample proportion and the sample mean

This method consists of equating the first sample proportion and the mean of the sample to their respective theoretical values, i.e.

$$\hat{\mu}_1' = \frac{S\hat{\alpha}_1}{\hat{\alpha}_1+\hat{\alpha}_2} \quad \text{and} \tag{8.1.11}$$

$$\hat{f}(0|S) = \frac{\Gamma(\hat{\alpha}_2+S)\ \Gamma(\hat{\alpha}_1+\hat{\alpha}_2)}{\Gamma(\hat{\alpha}_1+\hat{\alpha}_2+S)\ \Gamma(\hat{\alpha}_2)} . \tag{8.1.12}$$

This technique requires a numerical procedure for the solution of equation (8.1.12). It provides efficient results if $\hat{f}(0) > 0{,}5$.

8.1.3 Method of equating the first and last observed proportions

The method consists of equating the first and the last observed proportions to the theoretical ones. This leads to the following two equations:

$$\hat{f}(0|S) = \frac{\Gamma(\hat{\alpha}_2+S)\ \Gamma(\hat{\alpha}_1+\hat{\alpha}_2)}{\Gamma(\hat{\alpha}_1+\hat{\alpha}_2+S)\ \Gamma(\hat{\alpha}_2)} \quad \text{and} \tag{8.1.13}$$

$$\hat{f}(S|S) = \frac{\Gamma(\hat{\alpha}_1+S)\ \Gamma(\hat{\alpha}_1+\hat{\alpha}_2)}{\Gamma(\hat{\alpha}_1+\hat{\alpha}_2+S)\ \Gamma(\hat{\alpha}_1)} . \tag{8.1.14}$$

This is the most complex procedure, because both equations must be solved by numerical iterations.

In the case of U-shaped distributions the method provides efficient estimates.

If the data in a bivariate table are available from $x_1 = 1$, but x_2 can obtain any value from zero, it may be necessary to use the zero-truncated BBD and the related estimation methods.

8.1.4 Method of maximum likelihood

The maximum likelihood method has been worked out, but it is very complex. It was introduced to the statistical literature by Griffiths (1973) and Smith (1983). These authors proposed the use of the Newton-Raphson method for solving the maximum likelihood equations. For the starting point they took the moment estimates.

8.2 Methods of Estimation for the Generalized Positive Hypergeometric Distribution

Since the Beta-Binomial distribution is the Negative Hypergeometric distribution, the methods of estimation presented in Section 8.1 are also applicable for the Generalized Positive Hypergeometric distribution.

8.3 Method of Estimation for the Parameters of the Ascending Diagonal Arrays Distributions

The distributions that could be applied for the ascending diagonal arrays are the Beta-Binomial and the Generalized Positive Hypergeometric as presented in Section 2.3. These two families of distributions obviously include the Binomial distribution, which is obtained by a limiting process.

The parameters are estimated for each array individually. From this estimation two situations can be encountered:

(a) The estimates for the parameters in each array can be displayed graphically and are dispersed randomly without any relation to the array S. In this case a pooled estimate must be calculated. A practical example for this estimation is presented in Section 9.1.

(b) The estimates of the parameters follow a certain pattern as a function of S. In this case it is necessary to fit a function, so that the parameters vary as a function of S. A practical example for this estimation is presented in Section 9.4.

8.4 Methods of Estimation for the Negative Binomial Distribution

The probability distribution function $f(S|\gamma,\theta)$ of the NBD is a function of its two parameters γ and θ given by:

$$f(S|\gamma,\theta) = \frac{(1-\theta)^{\gamma}}{\Gamma(\gamma)} \frac{\Gamma(S+\gamma)}{\Gamma(S+1)} \theta^{S} , \quad \text{where}$$

$S = 0,1,2,\ldots$ and $0 < \theta < 1$.

Three methods for the estimation of parameters γ and θ are as follows:

(a) method of moments;

(b) method of mean and zero proportion; and

(c) method of maximum likelihood.

8.4.1 Method of moments

The method of moments is based on equating the observed first two moments to the theoretical ones. This leads to the following equations:

$$\hat{\mu}'_1 = \frac{\hat{\gamma}\hat{\theta}}{1-\hat{\theta}} \quad \text{and} \tag{8.4.1}$$

$$\hat{\mu}_2 = \frac{\hat{\gamma}\hat{\theta}}{(1-\hat{\theta})^2} . \tag{8.4.2}$$

The solutions of equations (8.4.1) and (8.4.2) give the following estimates:

$$\hat{\theta} = 1 - \hat{\mu}'_1/\hat{\mu}_2 \quad \text{and} \tag{8.4.3}$$

$$\hat{\gamma} = \frac{(\hat{\mu}'_1)^2}{\hat{\mu}_2 - \hat{\mu}'_1} . \tag{8.4.4}$$

8.4.2 Method of mean and zero proportion

This method consists of equating the sample mean and the sample zero proportion to their respective theoretical values. This leads to the following equations:

$$\hat{\mu}'_1 = \frac{\hat{\gamma}\hat{\theta}}{1-\hat{\theta}} \quad \text{and} \tag{8.4.5}$$

$$\hat{f}(0) = (1-\hat{\theta})^{\gamma} . \tag{8.4.6}$$

The parameter $\hat{\gamma}$ may be found from the equation

$$\ln \hat{f}(0) + \hat{\gamma} \ln(1 + \hat{\mu}'_1/\hat{\gamma}) = 0 \tag{8.4.7}$$

by some numerical method. As noted by Anscombe (1950), this method is efficient only when $\hat{f}(0) > 0.5$.

8.4.1 Method of moments

The method of moments is based on equating the observed first two moments to the theoretical ones. This leads to the following equations:

$$\hat{\mu}_1' = \frac{\hat{\gamma}\hat{\theta}}{1-\hat{\theta}} \quad \text{and} \tag{8.4.1}$$

$$\hat{\mu}_2 = \frac{\hat{\gamma}\hat{\theta}}{(1-\hat{\theta})^2} . \tag{8.4.2}$$

The solutions of equations (8.4.1) and (8.4.2) give the following estimates:

$$\hat{\theta} = 1 - \hat{\mu}_1'/\hat{\mu}_2 \quad \text{and} \tag{8.4.3}$$

$$\hat{\gamma} = \frac{(\hat{\mu}_1')^2}{\hat{\mu}_2 - \hat{\mu}_1'} . \tag{8.4.4}$$

8.4.2 Method of mean and zero proportion

This method consists of equating the sample mean and the sample zero proportion to their respective theoretical values. This leads to the following equations:

$$\hat{\mu}_1' = \frac{\hat{\gamma}\hat{\theta}}{1-\hat{\theta}} \quad \text{and} \tag{8.4.5}$$

$$\hat{f}(0) = (1-\hat{\theta})^{\hat{\gamma}} . \tag{8.4.6}$$

The parameter $\hat{\gamma}$ may be found from the equation

$$\ln \hat{f}(0) + \hat{\gamma} \ln(1 + \hat{\mu}_1'/\hat{\gamma}) = 0 \tag{8.4.7}$$

by some numerical method. As noted by Anscombe (1950), this method is efficient only when $\hat{f}(0) > 0.5$.

8.4.3 Method of maximum likelihood

The solutions of the likelihood equations for the Negative Binomial distribution have been given by Fisher (1941), Haldane (1941) and Sichel (1951). Sichel also generated some useful tables that facilitate the estimation process.

8.5 Zero-Augmented NBD

A large number of observed frequency distributions have what appears to be an excessively large zero cell. Irrespective of the true cause of this effect the ability to model such distributions is required.

In order to model such distributions it is assumed that this cell is made up of two separate groups. The zero-augmented distributions are presented mathematically by the following equations:

$$A(0) = q + (1-q)\, f(0) \quad \text{for } x = 0 \ ; \quad \text{and} \tag{8.5.1}$$

$$A(x) = (1-q)\, f(x) \quad \text{for } x = 1,2,\ldots\ . \tag{8.5.2}$$

The model presented by Morrison (1969) requires for the zero-augmented NBD model the estimate of three parameters γ, θ and q. The probability distribution function of zero-augmented NBD is given by:

$$A(0) = q + (1-q)(1-\theta)^{\gamma} \quad \text{for } x = 0 \quad \text{and}$$

$$A(x) = (1-q)\, \frac{(1-\theta)^{\gamma}}{\Gamma(\gamma)}\, \frac{\Gamma(x+\gamma)}{\Gamma(x+1)}\, \theta^{x} \quad \text{for } x = 1,2,\ldots$$

and satisfies:

$$\sum_{x=0}^{\infty} A(x) = A(0) + \sum_{x=1}^{\infty} A(x)$$

$$= q + (1-q)\, f(0) + \sum_{x=1}^{\infty} (1-q)\, f(x) = 1 \,. \qquad (8.5.3)$$

The first moment around the origin denoted by ${}_0\mu_1'$ is therefore:

$${}_0\mu_1' = 0\, A(0) + \sum_{x=1}^{\infty} x\, A(x) = (1-q) \sum_{x=0}^{\infty} x\, f(x) = (1-q)\, \mu_1' \,, \qquad (8.5.4)$$

where μ_1' is the first moment around the origin of the original distribution.

Similarly, one can calculate the second moment around the origin denoted by ${}_0\mu_2'$ as follows:

$${}_0\mu_2' = (1-q) \sum_{x=0}^{\infty} x^2 f(x) = (1-q)\, \mu_2' \,. \qquad (8.5.5)$$

In order to estimate the parameters γ, θ and q it is necessary to equate the first two observed moments around the origin and the observed frequency of the first cell to the respective theoretical values. This leads to the following equations:

$${}_0\hat{\mu}_1' = (1-\hat{q}) \frac{\hat{\gamma}\hat{\theta}}{1-\hat{\theta}} \,, \qquad (8.5.6)$$

$${}_0\hat{\mu}_2' = (1-\hat{q}) \frac{\hat{\gamma}\hat{\theta}}{(1-\hat{\theta})^2} (1-\hat{\gamma}\hat{\theta}) \quad \text{and} \qquad (8.5.7)$$

$$A(0) = \hat{q} + (1-\hat{q})\, (1-\hat{\theta})^{\hat{\gamma}} \,. \qquad (8.5.8)$$

From the above equations we obtain:

$$\hat{A}(0) = 1 - ({}_0\hat{\mu}_1'/\hat{D})(1 + 1/\hat{\gamma}) \left[1 - \left[1 + \hat{D}/(\hat{\gamma}+1)\right]^{-\hat{\gamma}}\right] \,, \qquad (8.5.9)$$

where $\hat{D} = ({}_0\hat{\mu}_2'/{}_0\hat{\mu}_1') - 1$.

This equation is solved by numerical iterations for $\hat{\gamma}$, giving:

$$\hat{\theta} = \frac{\hat{D}}{\hat{D}+\hat{\gamma}+1} \quad \text{and} \tag{8.5.10}$$

$$\hat{q} = 1 - \frac{{}_0\hat{\mu}'_1(\hat{\gamma}+1)}{\hat{D}\,\hat{\gamma}} . \tag{8.5.11}$$

The expected theoretical frequency of the augmented zero cell is calculated using

$$\tilde{f}_q(0) = N\,\hat{q} , \tag{8.5.12}$$

where N is the original sample size.

The expected theoretical frequency of the zero cell is obtained from

$$\hat{f}(0) - \hat{f}_q(0) , \tag{8.5.13}$$

where $\hat{f}(0)$ is the observed zero cell frequency.

8.6 Inverse Gaussian Poisson Distribution

The probability distribution $f(S|\alpha,\theta)$ of the IGP is a function of two parameters α and θ given by:

$$f(S) = \sqrt{\frac{2\alpha}{\pi}}\, e^{\alpha\sqrt{1-\theta}}\, \frac{(\alpha\theta/2)^S}{S!}\, K_{S-1/2}(\alpha) , \quad \text{where}$$

$S = 0, 1, 2, \ldots$.

The three methods of estimation for these parameters are:

(a) method of moments;

(b) method of zero-proportion and mean in the sample; and

(c) method of maximum likelihood.

8.6.1 Method of moments

Sichel (1982) mentioned this method of estimation as fairly efficient if the observed distribution is unimodal and the upper tail is relatively short, or if the sample coefficient of variation $CV < 0.5$.

The estimates for α and θ were given by Sichel (1974) as:

$$\hat{\theta} = 1 - (2\hat{w} - 1)^{-1} \quad \text{and} \tag{8.6.1}$$

$$\hat{\alpha} = 2\hat{\mu}_1'(1-\hat{\theta})^{1/2} \,/\, \hat{\theta}\,, \quad \text{where} \tag{8.6.2}$$

$\hat{w} = \hat{\mu}_2/\hat{\mu}_1'$ is the sample coefficient of dispersion.

8.6.2 Method of zero-proportion and mean in sample

If the observed zero-proportion in the sample is relatively large it is recommended to estimate the parameters of the distribution by equating the zero proportion and the mean in the sample with the respective theoretical values. A relatively large zero-proportion in the sample is generally considered to be above fifty percent, if we deal with a strongly reverse J-shaped frequency distribution.

In this case Sichel (1973b) proposed the following estimators for θ and α:

$$\hat{\theta} = 1 - \left[\frac{-\ln \hat{f}(0)}{2\hat{\mu}'_1 + \ln \hat{f}(0)} \right]^2 \quad \text{and} \tag{8.6.3}$$

$$\hat{\alpha} = 2\hat{\mu}'_1 \sqrt{1-\hat{\theta}} \; / \; \hat{\theta} \, , \quad \text{where} \tag{8.6.4}$$

$\hat{f}(0)$ is the observed zero-proportion in the sample.

8.6.3 Maximum likelihood method

For the case where the previous two methods are not efficient for the set of data under study, Sichel (1971) developed the maximum likelihood method. This method is however very complex.

8.7 Generalized Inverse Gaussian-Poisson Distribution

The probability distribution function of the GIGP is a function of three parameters γ, α and θ given by:

$$f(S|\gamma,\alpha,\theta) = \frac{(\sqrt{1-\theta})^{\gamma}}{K_{\gamma}(\alpha\sqrt{1-\theta})} \, \frac{(\alpha\theta/2)^{S}}{S!} \, K_{S+\gamma}(\alpha) \, , \quad \text{where}$$

$S = 0,1,2,3,\ldots$, $-\infty < \gamma < \infty$, $0 \leq \theta \leq 1$, $\alpha \geq 0$ and where $K_{\gamma}(z)$ is the modified Bessel function of the second kind of order γ and argument z .

Atkinson and Lam Yeh (1982) proposed an approximate maximum likelihood method for the estimation of the parameters of the GIGP which relies on initial maximum likelihood solutions for three neighbouring half-integer orders of the Bessel function. The final solution is obtained by a parabolic interpolation.

Rubinstein (1985) proposed an algorithm for the maximum likelihood estimation of γ, based on Newton-Raphson iterations. Her method is fully efficient and is a considerable improvement on the Atkinson and Lam Yeh method of estimation.

8.8 Methods of Estimation using Joint Distributions

It is possible to calculate the likelihood function for each one of the families of distributions presented in Section 2.5. By differentiating the likelihood function with respect to the parameters involved, one can develop the maximum likelihood estimators.

In what follows we present the likelihood function, the maximum likelihood equations and their solutions for the two families of distributions:

(a) the NBD - Binomial model and

(b) the NBD - BBD model.

8.8.1 The NBD - Binomial model

The NBD - Binomial model assumes for the third marginal the NBD with the parameters γ and θ given by:

$$f(S) = \frac{(1-\theta)^{\gamma}}{\Gamma(\gamma)} \frac{\Gamma(S+\gamma)}{\Gamma(S+1)} \theta^{S} , \quad \text{where} \tag{8.8.1}$$

$S = 0,1,2,\ldots$ and $0 < \theta < 1$;

and for the ascending diagonal arrays the Binomial distribution with the same parameter p along all the arrays.

The Binomial distribution is given by:

$$\phi(x_1|S) = \binom{S}{x_1} p^{x_1} (1-p)^{S-x_1} , \quad \text{where} \tag{8.8.2}$$

$x_1 = 0,1,2,\ldots,S$.

The joint distribution of x_1 and S is given by:

$$P(x_1,S) = (1-\theta)^{\gamma} \left[\prod_{i=0}^{S-1} (\gamma+i) \right] \frac{\theta^{S} p^{x_1} (1-p)^{S-x_1}}{x_1!\,(S-x_1)!} . \tag{8.8.3}$$

Substitution of $S = x_1 + x_2$ into (8.8.3) leads to the joint distribution of x_1 and x_2:

$$P(x_1,x_2) = (1-\theta)^{\gamma} \left[\prod_{i=0}^{x_1+x_2-1} (\gamma+i) \right] \frac{(\theta p)^{x_1} [\theta(1-p)]^{x_2}}{x_1!\,x_2!} , \tag{8.8.4}$$

where $x_1 = 0,1,2,\ldots$ and $x_2 = 0,1,2,\ldots$.

The likelihood function is given by:

$$L = \prod_{x_1=0}^{\infty} \prod_{x_2=0}^{\infty} \left[P(x_1,x_2)\right]^{f_{x_1x_2}} , \tag{8.8.5}$$

where $f_{x_1x_2}$ are the observed frequencies and

$$\sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_{x_1x_2} = N .$$

The likelihood function for this model is therefore:

$$L = \prod_{x_1=0}^{\infty} \prod_{x_2=0}^{\infty} \left[(1-\theta)^{\gamma} \left[\prod_{i=0}^{x_1+x_2-1} (\gamma+i) \right] \frac{(\theta p)^{x_1} [\theta(1-p)]^{x_2}}{x_1! \, x_2!} \right]^{f_{x_1 x_2}} . \tag{8.8.6}$$

The logarithm of the likelihood function is given by:

$$\ln L = C + \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_{x_1 x_2} \left[\sum_{i=0}^{x_1+x_2-1} \ln(\gamma+i) + \gamma \ln(1-\theta) + (x_1+x_2) \ln\theta + x_1 \ln p + x_2 \ln(1-p) \right] . \tag{8.8.7}$$

By differentiating (8.8.7) with respect to γ, θ and p respectively one obtains the maximum likelihood equations as follows:

$$\frac{\partial \ln L}{\partial \gamma} = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_{x_1 x_2} \Psi(\gamma+x_1+x_2-1) - N \Psi(\gamma-1) + N \ln(1-\theta) , \tag{8.8.8}$$

$$\frac{\partial \ln L}{\partial \theta} = - \frac{\gamma N}{1-\theta} + \frac{N}{\theta} (\overline{x}_1+\overline{x}_2) \quad \text{and} \tag{8.8.9}$$

$$\frac{\partial \ln L}{\partial p} = \frac{N \overline{x}_1}{p} - \frac{N \overline{x}_2}{1-p} , \quad \text{where} \tag{8.8.10}$$

$$\sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_{x_1 x_2} = N ,$$

$$\sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} x_1 f_{x_1 x_2} = N \overline{x}_1 ,$$

$$\sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} x_2 f_{x_1 x_2} = N \overline{x}_2 ,$$

and $\Psi(r)$ is the digamma function given as:

$$\Psi(r) = \left[1 + 1/2 + \ldots + 1/r \right] - \xi , \quad \text{where} \tag{8.8.11}$$

ξ is Euler's constant and its value is $\xi = 0,577215\ldots$.

Equations (8.8.9) and (8.8.10) lead to the maximum likelihood estimates for $\hat{\theta}$ and $\hat{p}$ given by:

$$\hat{\theta} = \left[1 + \frac{\hat{\gamma}}{\bar{x}_1+\bar{x}_2}\right]^{-1} \quad \text{and} \tag{8.8.12}$$

$$\hat{p} = \left[1 + \frac{\bar{x}_2}{\bar{x}_1}\right]^{-1} . \tag{8.8.13}$$

The parameter $\hat{\gamma}$ is estimated by numerical methods from the equation:

$$\sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_{x_1x_2} \Psi(\hat{\gamma}+x_1+x_2-1) - N\,\Psi(\hat{\gamma}-1) - N \ln\left[1 + \frac{\bar{x}_1+\bar{x}_2}{\hat{\gamma}}\right] = 0 . \tag{8.8.14}$$

If γ is known a priori then the maximum likelihood estimates for θ and p are easily obtained from equations (8.8.12) and (8.8.13).

As a particular example we consider the Geometric-Binomial model where $\gamma = 1$. In this case we have only two maximum likelihood equations, which are similar to (8.8.9) and (8.8.10). Since $\gamma = 1$ then:

$$\hat{\theta} = \left[1 + \frac{1}{\bar{x}_1+\bar{x}_2}\right]^{-1} \quad \text{and} \quad \hat{p} = \left[1 + \frac{\bar{x}_2}{\bar{x}_1}\right]^{-1} . \tag{8.8.15}$$

8.8.2 <u>The NBD - BBD model</u>

The NBD - BBD model assumes for the third marginal the NBD with the parameters γ and θ given by (8.8.1) and for the ascending diagonal arrays the BBD with the same parameters α_1 and α_2 along all the arrays.

The BBD is given by:

$$\phi(x_1|S) = \binom{S}{x_1} \frac{B(x_1+\alpha_1,\ S-x_1+\alpha_2)}{B(\alpha_1,\alpha_2)}, \quad \text{where} \tag{8.8.16}$$

$x_1 = 0,1,2,\ldots,S$, $\alpha_1 > 0$ and $\alpha_2 > 0$.

The joint distribution of x_1 and x_2 is given by:

$$P(x_1,x_2) = (1-\theta)^{\gamma} \left[\prod_{i=0}^{x_1+x_2-1} (\gamma+i)\right] \frac{\theta^{x_1+x_2}}{x_1!x_2!}$$

$$\cdot \frac{\left[\prod_{j=0}^{x_1-1} (\nu+j\delta)\right]\left[\prod_{l=0}^{x_2-1} (1-\nu+l\delta)\right]}{\left[\prod_{i=0}^{x_1+x_2-1} (1+i\delta)\right]}, \quad \text{where} \tag{8.8.17}$$

$$\nu = \frac{\alpha_1}{\alpha_1+\alpha_2}, \quad \delta = \frac{1}{\alpha_1+\alpha_2}, \quad x_1 = 0,1,2,\ldots \text{ and } x_2 = 0,1,2,\ldots .$$

The logarithm of the likelihood function is given by:

$$\ln L = C + \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_{x_1x_2} \left[\sum_{i=0}^{x_1+x_2-1} \ln(\gamma+i) + \gamma \ln(1-\theta) + \sum_{j=0}^{x_1-1} \ln(\nu+j\delta) + \sum_{l=0}^{x_2-1} \ln(1-\nu+l\delta) - \sum_{i=0}^{x_1+x_2-1} \ln(1+i\delta)\right]. \tag{8.8.18}$$

The first derivatives of (8.8.18) with respect to γ, θ, ν and δ are:

$$\frac{\partial \ln L}{\partial \gamma} = \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_{x_1x_2}\ \psi(\gamma+x_1+x_2-1) - N\ \psi(\gamma-1) + N \ln(1-\theta) , \tag{8.8.19}$$

$$\frac{\partial \ln L}{\partial \theta} = -\frac{\gamma N}{1-\theta} + \frac{N}{\theta} (\bar{x}_1+\bar{x}_2) , \tag{8.8.20}$$

$$\frac{\partial \ln L}{\partial \nu} = \sum_{x_1=0}^{\infty} f_{x_1 .} \sum_{j=0}^{x_1-1} \frac{1}{\nu+j\delta} - \sum_{x_2=0}^{\infty} f_{.x_2} \sum_{l=0}^{x_2-1} \frac{1}{1-\nu+l\delta} \tag{8.8.21}$$

$$\frac{\partial \ln L}{\partial \delta} = \sum_{x_1=0}^{\infty} f_{x_1 .} \sum_{j=0}^{x_1-1} \frac{j}{\nu+j\delta} - \sum_{x_2=0}^{\infty} f_{.x_2} \sum_{l=0}^{x_2-1} \frac{l}{1-\nu+l\delta}$$

$$- \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_{x_1 x_2} \sum_{i=0}^{x_1+x_2-1} \frac{i}{1+i\delta}, \quad \text{where} \tag{8.8.22}$$

$$\sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_{x_1 x_2} = N ,$$

$$\sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} x_1 f_{x_1 x_2} = N \bar{x}_1 ,$$

$$\sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} x_2 f_{x_1 x_2} = N \bar{x}_2 ,$$

$$\sum_{x_1=0}^{\infty} f_{x_1 x_2} = f_{.x_2} ,$$

$$\sum_{x_2=0}^{\infty} f_{x_1 x_2} = f_{x_1 .} ,$$

and $\psi(r)$ is the digamma function as defined in (8.8.11).

The maximum likelihood estimate for θ is obtained from equation (8.8.20). This estimation method requires a numerical procedure for the solution of (8.8.19), (8.8.21) and (8.8.22).

If the case where $\alpha_2 = c\alpha_1$, $\alpha_1 \to \infty$ and $\alpha_2 \to \infty$ is considered, then the results obtained should be identical to the case where the ascending diagonal arrays are distributed following the Binomial distribution. In this case equations (8.8.10) and

(8.8.21) become identical and equation (8.8.22) has no meaning.

One can consider another particular model of this case, where the distribution for the ascending diagonal arrays is a symmetrical BBD, i.e. $\alpha_1 = \alpha_2 = \alpha$. The joint distribution is therefore a function of only the three parameters γ, θ and α. In this case $\nu = 1/2$ and $\delta = 1/\alpha$.

The logarithm of the likelihood function becomes here:

$$\ln L = C + \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_{x_1x_2} \left[\sum_{i=0}^{x_1+x_2-1} \ln(\gamma+i) + \gamma \ln(1-\theta) \right.$$
$$\left. + (x_1+x_2) \ln\theta + \sum_{j=0}^{x_1-1} \ln(1+j\delta) + \sum_{l=0}^{x_2-1} \ln(1+l\delta) - \sum_{i=0}^{x_1+x_2-1} \ln(2+i\delta) \right]. \tag{8.8.23}$$

The derivatives of ln L with respect to γ, θ and δ can be obtained from equations (8.8.19), (8.8.20) and (8.8.22).

In this section two examples have been brought to demonstrate the procedure of calculating the maximum likelihood estimates for the new joint distributions presented in Chapter 2. In the same way it is possible to obtain the maximum likelihood estimates for any of the joint distributions presented in Section 2.5.

8.9 The Expected Fisher Information Martix

A further step which needs to be considered is the development of the expected Fisher information matrix. This is done for the joint distributions presented in Sections 8.8.1 and 8.8.2.

8.9.1 The NBD - Binomial model

The following expected information matrix for (γ,θ,p) in the NBD - Binomial model was obtained:

$$\begin{bmatrix} i.(\gamma) & i.(\gamma,\theta) & i.(\gamma,p) \\ i.(\theta,\gamma) & i.(\theta) & i.(\theta,p) \\ i.(p,\gamma) & i.(p,\theta) & i.(p) \end{bmatrix}, \quad \text{where} \tag{8.9.1}$$

$$i.(\gamma) = E\left[\sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty} f_{x_1 x_2}\,\Psi(\gamma+x_1+x_2-1)\right]^2 - N^2\left[\Psi(\gamma-1) - \ln(1-\theta)\right]^2, \tag{8.9.2}$$

$$i.(\theta) = \frac{\gamma[\theta(N-1)+1]}{\theta(1-\theta)^2}, \tag{8.9.3}$$

$$i.(p) = \frac{\gamma\theta}{(1-\theta)(1-p)p}, \tag{8.9.4}$$

$$i.(\gamma,\theta) = i.(\theta,\gamma) = \frac{N}{1-\theta}, \tag{8.9.5}$$

$$i.(\gamma,p) = i.(p,\gamma) = 0 \quad \text{and} \tag{8.9.6}$$

$$i.(p,\theta) = i.(\theta,p) = 0\,. \tag{8.9.7}$$

In the Geometric-Binomial model $(\gamma = 1)$ the expected information matrix for (θ,p) is given by:

$$\begin{bmatrix} i.(\theta) & i.(\theta,p) \\ i.(p,\theta) & i.(p) \end{bmatrix}, \quad \text{where} \tag{8.9.8}$$

$$i.(\theta) = \frac{\theta(N-1)+1}{\theta(1-\theta)^2}, \tag{8.9.9}$$

$$i.(p) = \frac{\theta}{(1-\theta)(1-p)p} \quad \text{and} \tag{8.9.10}$$

$$i.(p,\theta) = i.(\theta,p) = 0\ . \tag{8.9.11}$$

8.9.2 <u>The NBD - BBD model</u>

The expected information requires the calculation of the second derivatives of the logarithm of the likelihood function. The second derivatives with respect to γ, θ, ν and δ are:

$$\frac{\partial^2 \ln L}{\partial \gamma^2} = - \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_{x_1 x_2} \sum_{i=0}^{x_1+x_2-1} \frac{1}{(\gamma+i)^2}\ , \tag{8.9.12}$$

$$\frac{\partial^2 \ln L}{\partial \theta^2} = - \frac{\gamma N}{(1-\theta)^2} - \frac{N}{\theta^2}(\overline{x}_1+\overline{x}_2)\ , \tag{8.9.13}$$

$$\frac{\partial^2 \ln L}{\partial \nu^2} = - \sum_{x_1=0}^{\infty} f_{x_1 \cdot} \sum_{j=0}^{x_1-1} \frac{1}{(\nu+j\delta)^2} - \sum_{x_2=0}^{\infty} f_{\cdot x_2} \sum_{l=0}^{x_2-1} \frac{1}{(1-\nu+l\delta)^2}\ , \tag{8.9.14}$$

$$\frac{\partial^2 \ln L}{\partial \delta^2} = - \sum_{x_1=0}^{\infty} f_{x_1 \cdot} \sum_{j=0}^{x_1-1} \frac{j^2}{(\nu+j\delta)^2} - \sum_{x_2=0}^{\infty} f_{\cdot x_2} \sum_{l=0}^{x_2-1} \frac{l^2}{(1-\nu+l\delta)^2}$$

$$+ \sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} f_{x_1 x_2} \sum_{i=0}^{x_1+x_2-1} \frac{i^2}{(1+i\delta)^2}\ , \tag{8.9.15}$$

$$\frac{\partial^2 \ln L}{\partial \gamma\, \partial \theta} = \frac{\partial^2 \ln L}{\partial \theta\, \partial \gamma} = - \frac{N}{1-\theta}\ , \tag{8.9.16}$$

$$\frac{\partial^2 \ln L}{\partial\nu\,\partial\delta} = \frac{\partial^2 \ln L}{\partial\delta\,\partial\nu} = -\sum_{x_1=0}^{\infty} f_{x_1.} \sum_{j=0}^{x_1-1} \frac{j}{(\nu+j\delta)^2}$$

$$+ \sum_{x_2=0}^{\infty} f_{.x_2} \sum_{l=0}^{x_2-1} \frac{l}{(1-\nu+l\delta)^2}\,, \qquad (8.9.17)$$

$$\frac{\partial^2 \ln L}{\partial\gamma\,\partial\nu} = \frac{\partial^2 \ln L}{\partial\nu\,\partial\gamma} = 0\,, \qquad (8.9.18)$$

$$\frac{\partial^2 \ln L}{\partial\gamma\,\partial\delta} = \frac{\partial^2 \ln L}{\partial\delta\,\partial\gamma} = 0\,, \qquad (8.9.19)$$

$$\frac{\partial^2 \ln L}{\partial\theta\,\partial\nu} = \frac{\partial^2 \ln L}{\partial\nu\,\partial\theta} = 0 \quad \text{and} \qquad (8.9.20)$$

$$\frac{\partial^2 \ln L}{\partial\theta\,\partial\delta} = \frac{\partial^2 \ln L}{\partial\delta\,\partial\theta} = 0\,. \qquad (8.9.21)$$

The expected values of the negative derivatives given in equations (8.9.12) to (8.9.21) represent the expected information.

In this model the expected Fisher information matrix for $(\gamma,\theta,\delta,\nu)$ can be written in a similar form to the matrix given in (8.9.1).

C H A P T E R 9

APPLICATIONS OF THE NEW STATISTICAL MODELS

In this chapter several data sets from different application areas are analysed to illuminate the concepts introduced so far and to illustrate the use of the new methodology developed. The data sets are:

(a) data in the accident statistics field, used previously by Bates and Neyman (1952);

(b) data in the accident statistics field, used previously by Mellinger et al. (1965);

(c) data in the library loans field, used previously by Chen (1976); and

(d) data in the library loans field, used previously by Cane and Burrell (1982).

As statistical measures for the fitting we use the χ^2 test, the graphical fitting, the comparison between the theoretical and the empirical correlations and the comparison between the theoretical and the observed regression curves.

9.1 The Data Set from the Accident Field, used by Bates and Neyman (1952)

Following the pioneering work of Greenwood and Yule (1920) and Newbold (1926), Bates and Neyman (1952) developed a multivariate distribution for the number of accidents. Their main purpose was to study whether or not the number of accidents of a specified kind observed in the past has a predictive value for the number of accidents of the same kind that may occur in future. However, for certain purposes this model is not completely relevant and must be modified. Sometimes the question of interest is the influence of the number of mild accidents incurred in the past to severe accidents to which an individual might be exposed in future.

By generalizing the conditions of the problem studied by their predecessors Bates and Neyman developed a new multivariate distribution of the number of accidents. If their general model is reduced to a two-dimensional case it leads to the bivariate negative binomial distribution:

$$P_{x_1,x_2}(x_1,x_2) = \frac{\Gamma(\alpha+x_1+x_2)}{\Gamma(x_1+1)\Gamma(x_2+1)\Gamma(\alpha)}\,\beta^{\alpha}\,A^{x_1}\,(\beta+A+1)^{-(\alpha+x_1+x_2)}$$

where $x_1 = 0,1,2,\ldots$ and $x_2 = 0,1,2,\ldots$. (9.1.1)

The bivariate negative binomial distribution is known in the statistical literature in the following form:

$$P_{x_1,x_2}(x_1,x_2|t,s) = \left[1+\frac{(t+s)m}{k}\right]^{-k} \frac{\Gamma(x_1+x_2+k)}{\Gamma(k)\Gamma(x_1+1)\Gamma(x_2+1)}\, t^{x_1} s^{x_2}$$

$$\cdot \left[\frac{m}{k+(t+s)m}\right]^{x_1+x_2}, \quad \text{where} \qquad (9.1.2)$$

$k > 0$, $m > 0$, $x_1 = 0,1,2,\ldots$ and $x_2 = 0,1,2,\ldots$.

This model is obtained by mixing the product of two univariate Poisson distributions (thus generating a bivariate uncorrelated Poisson distribution) with the Gamma distribution. Because of the term $\Gamma(x_1+x_2+k)$ the distribution cannot be split into two independent distributions of the form $F(x_1)\ G(x_2)$. Hence this term provides a built-in correlation in the model.

The following relations exist between the parameters of (9.1.1) and (9.1.2):

$t = A$, $S = 1$, $m = \alpha/\beta$ and $k = \alpha$.

One of the data sets used by Bates and Neyman (See Table 9.1) is concerned with employees of an industrial establishment. The data lists the number of cases of incapacity suffered during a period of time due to the following two causes:

(a) digestive diseases and

(b) respiratory diseases.

Each case of incapacity caused by any of the two causes was treated as an accident.

Bates and Neyman's fitting of the bivariate negative binomial model with the estimated parameters $\hat{\alpha} = 1,471$, $\hat{\beta} = 1,050$ and

$\hat{A} = 3,798$ gave unsatisfactory results. The χ^2 value obtained for the whole table was 353,1 and the number of degrees of freedom considered was 95. Therefore $P(\chi^2_{(95)} > 353,1) \approx 0$.

Figure 9.6 shows the empirical conditional mean of x_2 given x_1 plotted against x_1. As depicted from the graph, the data show that a nonlinear regression is required.

Mitchell and Paulson (1981) are of the opinion that the above-mentioned data *"strongly suggests that a nonlinear regression function would be the most suitable for describing the empirical relationship between x_1 and x_2. Unfortunately the Bates and Neyman distribution does not admit of nonlinear regressions."* They developed a bivariate negative binomial distribution derived by convoluting an existing bivariate geometric distribution. The probability function for their distribution depends on six parameters and admits positive or negative correlations and linear or nonlinear regressions.

Michell and Paulson tried to fit their model to Bates and Neyman's data. The fit was, however, unsuccesful and they did not publish their results. It is worth quoting the authors: *"Even though the degree of fit as measured by χ^2 increased substantially, the overall fit was still very poor. The conclusion that the Bates and Neyman data possesses characteristics which preclude the possibility of a good representation by a bivariate negative binomial seems inescapable. There is thus no reason to present any of our results concerning this data."*

From the graphical presentation of the data (See Figure 9.6) one can see that a curvilinear regression is more suitable. This was the starting point which led us to investigate the possibility of fitting a new model. The new model, discussed theoretically in Chapter 2, fits the data set well, fulfilling the statistical desiderata discussed above. The model consists of:

- **the NBD for the third marginal and**
- **the BBD for the ascending diagonal arrays.**

In Table 9.1 the empirical and the theoretical frequencies obtained by fitting the developed model are presented. The variables x_1 and x_2 represent the number of cases of digestive and respiratory diseases respectively and $S = x_1 + x_2$.

From the foregoing distributions fitted to the ascending diagonal arrays and to the third marginal, we built a theoretical model for the inside of the table, as described below.

The NBD was fitted to the third marginal using the estimates $\hat{\gamma}$ and $\hat{\theta}$. The parameters γ and θ were estimated using the method of moments. The parameters α_1 and α_2 of the BBD for the ascending diagonal arrays were estimated using also the method of moments. The above estimates are relatively efficient.

Figures 9.1 and 9.2 present a graphical display of the estimates of the parameters α_1 and α_2 for the values tabulated in Table 9.2. The graphical displays show a random dispersion of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ around $\bar{\bar{\alpha}}_1$ and $\bar{\bar{\alpha}}_2$ respectively. The weighted averages

of $\hat{\alpha}_1$ and $\hat{\alpha}_2$ were calculated and on the basis of these values the BBD was fitted with the same $\bar{\hat{\alpha}}_1$ and $\bar{\hat{\alpha}}_2$ along all the ascending diagonal arrays. The weights considered are the observed frequencies for the third marginal from S = 2 to S = 17. From Table 9.2 and Figures 9.1 and 9.2 one notes that the estimates $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are highly positive correlated.

The expected BBD array frequencies were calculated and adjusted proportionally in such a way that they summed up to the <u>expected</u> third marginal frequencies calculated previously.

Expected theoretical frequencies were estimated for the arrays $S \geq 18$ containing too little data for analysis and, finally, expected distributions for the marginals x_1 and x_2 were calculated by summing the table vertically and horizontally. Cells containing expected values of less than 5,0 units were grouped together. The method used for grouping consisted of building groups of cells on each ascending diagonal array starting from the upper right corner such that an expected value of 5,0 or more units was obtained.

For the marginals x_1 and x_2 and the third marginal the same method of grouping was used such that in each cell there were 5,0 units or more.

Table 9.1 The data set of Bates and Neyman (1952) with

$\hat{\gamma} = 1,54$, $\hat{\theta} = 0,8136$, $\hat{\alpha}_1 = 1,73$ and $\hat{\alpha}_2 = 6,79$.

s	f(s)	$x_2 \backslash x_1$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	≥19	$f(x_2)$
0	107 96,8	0	107 96,8	16 24,6	2 7,3	4 2,5	1 0,9	1 0,4	0,2	0,1	0	0	0	0	0	0	0	0	0	0	0	0	131 132,8
1	117 121,2	1	101 96,6	32 36,3	14 13,6	6 5,4	1 2,3	1,0	0,5	2 0,2	0,1	0,1	1 0,1	0	0	0	0	0	0	0	0	0	157 156,2
2	128 125,3	2	94 81,7	38 38,7	13 16,9	3 7,6	3 3,5	2 1,7	0,9	0,4	0,2	0,1	0,1	0,1	0	0	0	0	0,1	0	0	0	153 152,0
3	128 120,3	3	72 65,5	39 36,4	15 17,9	7 8,7	3 4,4	1 2,2	1 1,2	0,6	1 0,4	0,2	0,1	1 0,1	0	0	0	0	0	0	0	0	140 137,7
4	117 111.0	4	58 51,4	18 32,0	14 17,2	1 9,0	6 4,3	4 2,5	1,4	0,8	0,5	0,3	0,2	0,1	0,1	0,1	0	0	0	0	0	0	101 120,5
5	82 100,2	5	44 40,0	30 27,2	16 15,7	9 8,7	8 4,9	1 2,7	1 1,6	0,9	0,5	0,3	0,2	0,1	0,1	0	0	1 0	0	0	0	0	110 102,9
6	79 88,7	6	25 30,9	27 22,6	12 13,8	9 8,1	2 4,7	5 2,7	2 1,5	0,9	1 0,6	0,3	0,2	0,1	0,1	0,1	0,1	0	0	0	0	0	83 86,8
7	70 77,9	7	21 23,9	19 18,5	9 11,8	7 7,2	1 4,3	5 2,6	1 1,6	0,9	1 0,8	3 0,4	1 0,2	0,1	0,1	0,1	0,1	0	0	0	0	0	68 72,4
8	71 67,5	8	22 18,4	14 15,0	12 10,0	8 6,3	3 3,9	1 2,4	1 1,5	0,9	1 0,6	1 0,4	0,2	0,1	0,1	0,1	0	0	0	0	0	0	63 59,9
9	67 58,3	9	22 14,2	7 12,0	6 8,3	7 5,4	3 3,4	2 2,2	1,4	2 0,9	1 0,5	0,3	1 0,2	0,1	1 0,1	0,1	0,1	0,1	1 0	0	0	0	53 49,3
10	36 49,9	10	7 11,0	5 9,6	13 6,9	5 4,6	4 3,0	1,9	1 1,2	0,8	0,5	0,3	1 0,2	0,1	1 0,1	0,1	0,1	0	0	0	0	0	37 40,4
11	38 42,7	11	10 8,5	8 7,7	7 5,6	1 3,8	3 2,6	2 1,7	2 1,1	3 0,7	1 0,5	1 0,3	1 0,2	0,1	0,1	0,1	0	0,1	0	0	0	0	39 33,1
12	49 36,4	12	11 6,6	5 6,1	4 4,6	5 3,2	4 2,2	1 1,4	1 1,0	0,6	0,4	1 0,3	0,2	0,1	0,1	0,1	0	0	0,1	0	1 0	0,1	33 27,1
13	30 30,7	13	8 5,1	4 4,9	5 3,7	8 2,6	2 1,8	1,2	1 0,8	1 0,6	1 0,4	0,3	2 0,2	1 0,1	1 0,1	0,1	0	0	0	0	0	0	34 21,9
14	24 26,1	14	6 4,0	4 3,9	4 3,0	2 2,2	1 1,5	1 1,0	1 0,7	0,5	2 0,3	0,2	0,2	0,1	0,1	0	0	0	0,1	0	0	0	21 17,8
15	20 21,9	15	2 3,1	5 3,1	2 2,4	3 1,8	1 1,3	1 0,9	0,6	0,4	0,3	1 0,2	0,1	0,1	1 0,1	0	0,1	1 0	1 0	0	0	0	18 14,5
16	33 18,5	16	3 2,4	2 2,4	2 2,0	2 1,5	1,0	2 0,7	2 0,5	0,4	0,2	0,2	0,1	0,1	0,1	1 0	0	0	0	0	0	0	14 11,6
17	16 15,4	17	2 1,9	2 1,9	1 1,6	1,2	1 0,9	0,6	0,4	0,3	0,2	0,1	0,1	0,1	0	0,1	0	0	0	0	0	0	6 9,4
18	13 12,9	18	1 1,5	1,6	2 1,3	1 1,0	0,7	1 0,5	0,4	1 0,3	0,2	1 0,1	0,1	0,1	0	0	0	0	0	0	0	0	6 7,7
19	9 10,7	19	1 1,1	1,2	1,0	0,8	1 0,6	0,4	1 0,3	0,2	0,2	0,1	0,1	0,1	0	0	0	0	0	0,1	0,1	0	4 6,3
20	8 9,1	20	0,9	1,0	0,8	0,6	0,5	2 0,3	0,3	1 0,2	0,1	0,1	0,1	1 0	0	0	0,1	0	0	0	0	0	4 5,0
21	9 7,7	21	1 0,7	0,8	3 0,6	0,5	0,4	0,3	0,2	0,1	0,1	0,1	0,1	0	0	0	0	0	0	0	0	0,1	4 4,0
22	5 6,3	22	0,6	0,6	0,5	0,4	0,3	0,2	0,2	0,1	0,1	0,1	0	0	0	0	0	0	0	0	0,1	0	0 3,2
23	7 5,2	23	0,4	0,5	0,4	0,3	0,3	1 0,2	0,1	0,1	0,1	0,1	0	0	0,1	0	0	0	0	0,1	0	0	1 2,7
24	2 4,5	24	0,3	0,4	0,3	[illegible]	0,2	0,2	0,1	0,1	0,1	0	0,1	0	0	0	0	1 0	0	0	0	0	1 2,1
25	6 3,6	25	0,3	0,3	0,3	0,2	0,2	0,1	1 0,1	0,1	0,1	0	0,1	0	0	0	0	0,1	0	0	0	0	1 1,9
26	0 3,0	26	0,2	1 0,2	0,2	0,2	1 0,1	0,1	0,1	0,1	0	0	0	0	0	0	0,1	0	0	0	0	0	2 1,3
27	4 2,5	27	0,2	0,2	0,2	0,1	0,1	0,1	0,1	0	0,1	0,1	0	0,1	0	0	0	0	0	0	0	0	0 1,3
28	1 2,0	28	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0,1	0	1 0	0	0,1	0	0	0	0	0	0	0	0	1 0,9
29	1 1,7	29	0,1	0,1	0,1	0,1	0,1	0,1	1 0	0,1	0	0	0	0,1	0	0	0	0	0	0	0	0	1 0,8
≥30	9 8,0	≥30	0,3	0,5	0,5	0,5	0,1	0,5	0	0,2	0	0,1	0	0	0	0	0	0	0	0	0	0	0 2,5
		$f(x_1)$	618 568,7	276 310,1	156 168,6	68 94,9	49 55,1	33 33,0	17 20,2	10 12,6	10 7,9	8 5,1	7 3,4	3 2,1	4 1,4	1 1,0	0 0,7	3 0,3	2 0,3	0 0,2	1 0,2	0 0,2	1286 1286,0

Table 9.2 Estimates of the parameters α_1 and α_2 for the ascending diagonal arrays

S	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{f}(S)$
2	-4,235	-25,88	128
3	1,04	4,08	128
4	2,364	10,354	117
5	0,79	4,117	82
6	3,593	15,122	79
7	2,422	11,708	70
8	1,106	4,451	71
9	1,325	5,502	67
10	19,603	38,968	36
11	0,975	3,359	38
12	[illegible],524	11,748	49
13	2,696	15,752	30
14	0,85	3,476	24
15	4,535	23,23	20
16	2,061	7,163	33
17	1,663	5,405	16

The estimates used for the parameters γ, θ, α_1 and α_2 were:

$\hat{\gamma} = 1,54$, $\hat{\theta} = 0,8136$, $\bar{\hat{\alpha}}_1 = 1,73$ and $\bar{\hat{\alpha}}_2 = 6,79$.