A FAMILY OF MULTIVARIATE
DISCRETE DISTRIBUTIONS

Linda M. Haines

Department of Statistical Sciences, University of Cape Town, South Africa
e-mail: *linda.haines@uct.ac.za*

In this short paper, the multivariate Poisson-Gamma, the multinomial N-mixture and the negative multinomial distributions are shown to have probability mass functions of the same form and thus to share, broadly, the same distributional properties. The three distributions are, however, fundamentally very different in nature, that is, in terms of genesis, interpretation and model building, and these differences are highlighted and discussed.

Key words: Inference, Likelihood, Models, Multinomial N-mixture distribution, Multivariate Poisson-Gamma, Negative multinomial.

1. Introduction

The aim of this paper is to introduce a family of three multivariate discrete distributions, namely the multivariate Poisson-Gamma, the multinomial N-mixture and the negative multinomial, which are related in distribution but not in terms of genesis, interpretation and model building. The paper is organised as follows. A motivation for each of the three distributions is presented in Section 2, together with an outline of the derivation of the probability mass functions and a brief discussion of the common distributional properties. Models which are based on the three distributions of interest, together with details relating to parameter estimation and inference, are introduced in Section 3 and some important differences are highlighted and discussed. Finally, in Section 4, the distributional similarities and the modelling differences between the three distributions are drawn together and some recommendations for further research are made.

2. The family of distributions

2.1 The multivariate Poisson-Gamma distribution

The motivation for the Poisson-Gamma distribution comes from the early papers of Arbous and Kerrich (1951) and Bates and Neyman (1952) on accident proneness. Specifically, consider a vector of counts $y = (y_1, \ldots, y_d)$ representing the numbers of each of d types of accident incurred by an individual. The counts y_j are taken to be Poisson distributed with parameters $\mu_j \nu$, $j = 1, \ldots, d$, where the terms $\mu_j > 0$ are unknown constants relating to the type of accident incurred and $\nu > 0$ is a random term representing the accident proneness of the individual. In order to accommodate variation in accident proneness across individuals, the term ν is taken to follow a gamma distribution with shape parameter α and rate parameter β, that is $\nu \sim \text{Gamma}(\alpha, \beta)$ with mean α/β. In other words the vector

y conditional on v comprises independent Poisson counts. A more meaningful parametrisation of the random term v with $\alpha = \beta$ and thus with $E[v] = 1$ was introduced into the economic literature by Hausman et al. (1984) and is now widely used. The unconditional probability mass function (pmf) of the counts thus follows immediately by integrating out v, where $v \sim \text{Gamma}(r, r)$ with $\alpha = \beta = r$, and is given by

$$\text{Prob}(y_1, \ldots, y_d) = \frac{\Gamma(r + \sum_{j=1}^{d} y_j)}{\prod_{j=1}^{d} y_j! \, \Gamma(r)} \prod_{j=1}^{d} \left(\frac{\mu_j}{r + \sum_{j=1}^{d} \mu_j} \right)^{y_j} \left(\frac{r}{r + \sum_{j=1}^{d} \mu_j} \right)^{r} \tag{1}$$

with overdispersion parameter r (Cameron and Trivedi, 1998, Section 9.4.1; Solis-Trapala and Farewell, 2005).

Applications of the multivariate Poisson-Gamma distribution have been reported somewhat sporadically in the literature. In all cases the associated models are taken to be log-linear, with the vector of counts regressed against an appropriate set of explanatory variables. For example, Hausman et al. (1984) used the distribution to model panel data and Mothafer et al. (2016) to model crash frequency data. In addition, Solis-Trapala and Farewell (2005) provide applications within the context of medical statistics and Schmidt and Schwabe (2020) report the construction of optimal block designs based on the Poisson-Gamma setting.

2.2 The multinomial N-mixture distribution

The motivation for the multinomial N-mixture distribution comes from removal sampling protocols used in ecology (Dorazio et al., 2005). Specifically, suppose that animals are removed from a closed population at d successive times and are counted but not returned to the population. The process stops before all animals have been removed and thus the total number of animals present in the population, say n, remains unknown. The setting can be formulated as a multinomial comprising $d + 1$ classes with the probability of capturing an animal in the jth class given by $\pi_j, j = 1, \ldots, d$, and with $\pi_{d+1} = 1 - \sum_{j=1}^{d} \pi_j$. Thus the counts $y_j, j = 1, \ldots, d$, are recorded at the first d times and the count in the $(d + 1)$th class is then equal to $n - \sum_{j=1}^{d} y_j$, and is unknown. In constructing a distribution for the observed counts $y_j, j = 1, \ldots, d$, it is sensible to adopt a mixing distribution for the unknown total n. In the present study n is taken to follow a negative binomial distribution with pmf of the form

$$\frac{\Gamma(n + r)}{\Gamma(r) \, n!} \left(\frac{\mu}{\mu + r} \right)^{n} \left(\frac{r}{\mu + r} \right)^{r},$$

where μ represents the mean animal abundance in the population and r the overdispersion parameter, that is $n \sim \text{NegBin}(r, \mu)$. The pmf of the observed counts can then be obtained in closed form by a subtle change of indices (Haines, 2020) and is given by

$$\text{Prob}(y_1, \ldots, y_d) = \frac{\Gamma(r + \sum_{j=1}^{d} y_j)}{\prod_{j=1}^{d} y_j! \, \Gamma(r)} \prod_{j=1}^{d} \left(\frac{\mu \pi_j}{r + \mu \sum_{j=1}^{d} \pi_j} \right)^{y_j} \left(\frac{r}{r + \mu \sum_{j=1}^{d} \pi_j} \right)^{r}. \tag{2}$$

There is a range of applications of the multinomial N-mixture distribution within the context of statistical ecology, including removal sampling, multi-observer and capture-recapture protocols, and these are extensively covered in Chapter 7 of the book by Kéry and Royle (2015). In all cases,

the infinite sum embedded in the basic formulation of the pmf of the distribution is approximated by placing an upper bound on that sum and, as a consequence, results are obtained numerically (Dorazio et al., 2005; Kéry, 2018). Note that the pmf of the multinomial N-mixture distribution was included in the early review paper of Sibuya et al. (1964) but with a formulation different to that introduced here. Specifically, the authors considered a sequence of independent Bernoulli trials with a probability of success ρ and took the requisite pmf to be the probability of a success at the rth trial. No applications relating to this parametrisation appear to have been reported in the literature however.

2.3 The negative multinomial distribution

There are a number of ways in which the negative multinomial distribution can be motivated but that based on inverse sampling is, arguably, the most attractive. Specifically, the inverse sampling strategy proceeds by taking independent samples from a multinomial distribution with $d + 1$ classes and attendant probabilities $p_j, j = 1, \ldots, d + 1$, with $\sum_{j=1}^{d+1} p_j = 1$, until the count in the $(d + 1)$th class is equal to r. The pmf of the counts $y_j, j = 1, \ldots, d$, follows from straightforward combinatorial arguments and is given by

$$\text{Prob}(y_1, \ldots, y_d) = \frac{\Gamma(r + \sum_{j=1}^{d} y_j)}{\prod_{j=1}^{d} y_j!\, \Gamma(r)} \prod_{j=1}^{d} p_j^{y_j}\, p_{d+1}^{r}. \tag{3}$$

The motivation is predicated on the assumption that r is an integer but is readily extended to the general case with $r > 0$. Specifically, following Zhou and Lange (2010), consider d independent Poisson processes with intensities $p_j, j = 1, \ldots, d$, run for a time which is gamma distributed with shape parameter r and rate parameter p_{d+1}. Then the joint pmf of the counts associated with the d processes under this strategy is given by (3). More broadly, and in line with the univariate case, the parameter r can be interpreted as an overdispersion parameter.

Applications of the negative multinomial distribution focus on the fact that counts with pmf (3) are positively correlated and involve, at least in general, comparisons with the multinomial and Dirichlet-based multinomial distributions (Johnson et al., 1997, Chapter 36). Such comparisons are, however, delicate and can be inconclusive. This is evidenced by the findings of Zhang et al. (2017) which are based on simulated and real RNA-seq data.

2.4 Commonalities

It is immediately clear that the pmfs for the multivariate Poisson-Gamma, the multinomial N-mixture and the negative multinomial distributions are of the same form. Specifically, consider the negative multinomial with probabilities $p_j, j = 1, \ldots, d, p_{d+1} = 1 - \sum_{j=1}^{d} p_j$, and overdispersion parameter r. Then the distributional properties of the multivariate Poisson-Gamma can be recovered by substituting p_j with $\mu_j/(r + \mu_j)$ and those of the multinomial N-mixture distribution by substituting p_j with $\mu\pi_j/(r + \mu\sum_{j=1}^{d} \pi_j), j = 1, \ldots, d$. In other words, the terms $\mu_j/(r + \mu_j)$, $\mu\pi_j/(r + \mu\sum_{j=1}^{d} \pi_j)$ and p_j in the pmfs (1), (2) and (3), respectively, are interchangeable and are said here to be notationally equivalent. It now follows that the probability generating functions of the three distributions are of the same form and thus that the individual counts $y_j, j = 1, \ldots, d$, and the sum of the counts $\sum_{j=1}^{d} y_j$ are distributed as negative binomials and, crucially, that the counts themselves are positively correlated (Bates and Neyman, 1952; Johnson et al., 1997, Chapter 36).

The multivariate Poisson-Gamma distribution is derived from independent Poisson variates whereas the multinomial N-mixture and the negative multinomial distributions emanate from the multinomial distribution. The three distributions of interest can, however, be linked by invoking the relationships between the Poisson and the multinomial distributions. Specifically, counts of independent Poisson variates given the sum of the counts follow a multinomial distribution and, conversely, counts from a multinomial distribution with a Poisson mixing distribution on the sum of the counts are independent Poisson variates (Dobson and Barnett, 2008, p. 150). The relationships so described are summarised in the flow chart shown in Figure 1.

3.   Models, estimation and inference

The broad-based setting adopted here is that in which n independent samples are taken from a distribution of vectors of d counts, with covariates recorded for each count. The d counts thus define d classes or categories. The counts are denoted here by y_{ij}, the attendant vector of p sample-specific covariates by $x_i = (x_{i1}, \ldots, x_{ip})$ and the vector of t sample- and class-specific covariates, that is cell-specific covariates, by $z_{ij} = (z_{ij,1}, \ldots, z_{ij,t})$ for $i = 1, \ldots, n, j = 1, \ldots, d$. The discussion which follows draws heavily, but not exclusively, on three representative papers: that of Solis-Trapala and Farewell (2005) relating to the multivariate Poisson-Gamma, that of Haines (2020) relating to the multinomial N-mixture distribution, and that of Zhang et al. (2017) relating to the negative multinomial distribution.
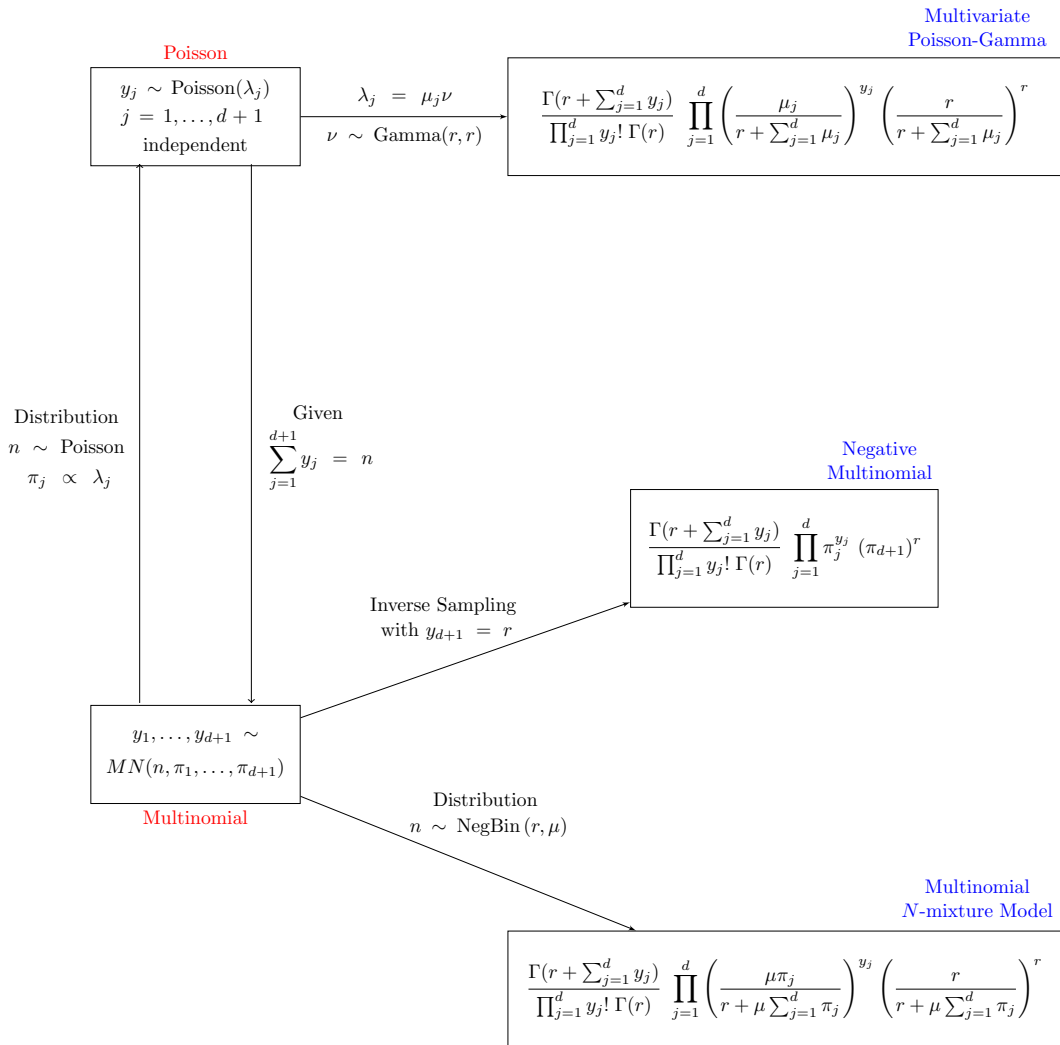
3.1   The multivariate Poisson-Gamma model

Consider first the multivariate Poisson-Gamma setting with $E(y_{ij}) = \mu_{ij}, i = 1, \ldots n, j = 1, \ldots, d$. The associated model for μ_{ij} is, in general, taken to be log-linear, with counts regressed against an appropriate set of explanatory variables, and the overdispersion parameter r is taken to be an unknown constant. The parametrisation of the log-linear model is, however, subtle (Solis-Trapala and Farewell, 2005). Specifically, covariates which are sample-specific and those which are both sample- and class-specific should be distinguished. To fix ideas, suppose that the numbers of animals at n independent sites are recorded over d days. Then covariates can be site-specific, such as distance from water, and site- and day-specific, such as ambient temperature at the time of counting. The correct formulation for μ_{ij} is then given by

$$\ln(\mu_{ij}) = \beta_0 + x_i^\top \beta + (z_{ij} - \bar{z}_{i+})^\top \gamma_W + \bar{z}_{i+}^\top \gamma_B, \tag{4}$$

where x_i is a vector of covariates associated with sites only, z_{ij} is a vector of covariates associated with both sites and days and $\bar{z}_{i+} = \sum_{j=1}^{d} z_{ij}/d$. The term β_0 represents the overall mean, the term β is a vector of parameters conformable with the site covariates x_i, and the terms γ_W and γ_B capture the within- and between-site dependencies emanating from the covariates z_{ij} and are conformable with $z_{ij}, i = 1, \ldots, n, j = 1, \ldots, d$ (Solis-Trapala and Farewell, 2005).

The likelihood for the parameters β, γ and r, where $\gamma = (\gamma_W, \gamma_B)^\top$, follows directly from the pmf for the multivariate Poisson-Gamma distribution, that is pmf (1), and is given by

$$L(\beta, \gamma, r) = \prod_{i=1}^{n} \left[\frac{\Gamma(r + y_{i+})}{\Gamma(r) \prod_{j=1}^{d} y_{ij}!} \prod_{j=1}^{d} \left(\frac{\mu_{ij}}{r + \mu_{i+}} \right)^{y_{ij}} \left(\frac{r}{r + \mu_{i+}} \right)^{r} \right], \tag{5}$$

Poisson

$$y_j \sim \text{Poisson}(\lambda_j)$$
$$j = 1, \ldots, d+1$$
$$\text{independent}$$

$$\lambda_j = \mu_j \nu$$
$$\nu \sim \text{Gamma}(r, r)$$

Multivariate
Poisson-Gamma

$$\frac{\Gamma(r + \sum_{j=1}^{d} y_j)}{\prod_{j=1}^{d} y_j! \, \Gamma(r)} \prod_{j=1}^{d} \left(\frac{\mu_j}{r + \sum_{j=1}^{d} \mu_j} \right)^{y_j} \left(\frac{r}{r + \sum_{j=1}^{d} \mu_j} \right)^{r}$$

Distribution
$$n \sim \text{Poisson}$$
$$\pi_j \propto \lambda_j$$

Given
$$\sum_{j=1}^{d+1} y_j = n$$

Negative
Multinomial

$$\frac{\Gamma(r + \sum_{j=1}^{d} y_j)}{\prod_{j=1}^{d} y_j! \, \Gamma(r)} \prod_{j=1}^{d} \pi_j^{y_j} \, (\pi_{d+1})^{r}$$

Inverse Sampling
with $y_{d+1} = r$

$$y_1, \ldots, y_{d+1} \sim$$
$$MN(n, \pi_1, \ldots, \pi_{d+1})$$

Multinomial

Distribution
$$n \sim \text{NegBin}(r, \mu)$$

Multinomial
N-mixture Model

$$\frac{\Gamma(r + \sum_{j=1}^{d} y_j)}{\prod_{j=1}^{d} y_j! \, \Gamma(r)} \prod_{j=1}^{d} \left(\frac{\mu \pi_j}{r + \mu \sum_{j=1}^{d} \pi_j} \right)^{y_j} \left(\frac{r}{r + \mu \sum_{j=1}^{d} \pi_j} \right)^{r}$$

Figure 1. Flow chart illustrating the relationships between the multivariate discrete distributions of interest.

where $y_{i+} = \sum_{j=1}^{d} y_{ij}$ and $\mu_{i+} = \sum_{j=1}^{d} \mu_{ij}$. The expression for $L(\boldsymbol{\beta}, \boldsymbol{\gamma}, r)$ can in fact be reformulated to some advantage as the product of a conditional and a marginal likelihood (Solis-Trapala and Farewell, 2005). Specifically, the conditional likelihood depends only on the parameter $\boldsymbol{\gamma}_W$ and is given by

$$L_C(\boldsymbol{\gamma}_W) = \prod_{i=1}^{n} \frac{y_{i+}!}{\prod_{j=1}^{d} y_{ij}!} \prod_{j=1}^{d} \left(\frac{\mu_{ij}}{\mu_{i+}} \right)^{y_{ij}},$$

which is the likelihood associated with that of a multinomial distribution. In addition, the marginal likelihood is given by

$$L_M(\boldsymbol{\beta}, \boldsymbol{\gamma}, r) = \prod_{i=1}^{n} \frac{\Gamma(r + y_{i+})}{\Gamma(r) \, y_{i+}!} \left(\frac{\mu_{i+}}{r + \mu_{i+}} \right)^{y_{i+}} \left(\frac{r}{r + \mu_{i+}} \right)^{r}$$

and is that of a negative binomial with mean μ_{i+} and overdispersion parameter r. Both these results are in accord with the distributional properties of the multivariate Poisson-Gamma.

Maximum likelihood estimates (MLEs) of the parameters $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and r, denoted $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}$ and $\hat{r}$, can be obtained by solving the likelihood equations using a Newton or Fisher scoring algorithm (Solis-Trapala and Farewell, 2005). The information matrix for the parameters follows immediately from the log-likelihood and, since $-E\left[\partial^2 \ln L(\boldsymbol{\beta}, \boldsymbol{\gamma}, r)/\partial \boldsymbol{\beta} \partial r \right] = 0$ and $-E\left[\partial^2 \ln L(\boldsymbol{\beta}, \boldsymbol{\gamma}, r)/\partial \boldsymbol{\gamma} \partial r \right] = 0$, is block diagonal (Waller and Zelterman, 1997). Thus the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are information orthogonal to the parameter r or, in other words, the asymptotic covariances of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ with $\hat{r}$ are zero (Pawitan, 2001, p. 287). In addition, Schmidt and Schwabe (2020) have derived an explicit expression for the information matrix for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in terms of that for independently distributed Poisson variates with mean parameters $\mu_{ij}, i = 1, \ldots, n, j = 1, \ldots, d$. Profile likelihoods and confidence sets for the parameters $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and r can be readily constructed, with the latter based on the inverse of the observed information matrix. However, Solis-Trapala and Farewell (2005) noted that the generalised estimating function for the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ coincides with the score function and, as a consequence, recommended that the robust sandwich estimator of the variance matrix of the MLEs $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ be used in preference to that based on the observed information matrix.

Other approaches to estimation and inference for the multivariate Poisson-Gamma model have been introduced in the literature. For example, Dey and Chung (1992) explored Bayesian estimation of the parameters of the log-linear model under normalised square error loss and, more recently, Tsou (2016) examined the applicability of the robust likelihood technique but these approaches have not been developed further.

3.2   The multinomial N-mixture model

Consider now the multinomial N-mixture setting within the context of removal sampling with $E(y_{ij}) = \mu_i \pi_{ij}$ for $i = 1, \ldots, n, j = 1, \ldots, d$ (Haines, 2020). The parameter μ_i represents the mean abundance at site i and the parameter π_{ij} the probability associated with sample i and class j, that is, the cell probability, $i = 1, \ldots, n, j = 1, \ldots, d$. The term μ_i is modelled log-linearly, that is, as $\ln \mu_i = \boldsymbol{x}_i^{\top} \boldsymbol{\beta}$ where $\boldsymbol{x}_i$ is a vector of site-specific covariates and $\boldsymbol{\beta}$ is a vector of unknown parameters conformable with $\boldsymbol{x}_i, i = 1, \ldots, n$. In addition, the overdispersion parameter r is taken to be an unknown constant. Modelling the cell probability π_{ij} is, however, not straightforward. Specifically, π_{ij} is a function of the probability of capture of an animal at site i and removal j and it is the latter

probability which is meaningful in practice. Thus, the logit of p_{ij} is taken to be a linear function of appropriate covariates z_{ij} and the resultant logistic function for p_{ij} is then embedded in the term π_{ij} for $i = 1, \ldots, n, j = 1, \ldots, d$. For example, if the probability of capture is taken to be a constant p over all sites and all removals, then $\pi_{ij} = (1 - p)^{j-1} p$ with logit(p) an appropriate linear function of the cell-specific covariates of the form given in (4).

The development of the likelihood for the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ is given in Haines (2020) and follows closely that of the multivariate Poisson-Gamma but with an additional layer of complexity derived from the cell probabilities. Simplistically, the likelihood corresponds to that of the multivariate Poisson-Gamma given in (5) but with the term μ_{ij} replaced by $\mu_i \pi_{ij}$ for $i = 1, \ldots, n, j = 1, \ldots, d$. MLEs of the parameters can then be obtained by maximising the log-likelihood using a nonlinear optimisation routine and confidence sets derived from the block-diagonal information matrix. In addition, comparisons of results for the multinomial N-mixture model with a Poisson mixing distribution, that is, for the model with independently distributed Poisson variates with mean parameters $\mu_{ij}, i = 1, \ldots, n, j = 1, \ldots, d$, with those for the present setting are of some interest (Haines, 2020). For example, likelihood-based hypothesis tests for the multinomial N-mixture model with a Poisson mixing distribution against that with a negative binomial mixing distribution can be conducted and are based solely on the site totals $y_{i+}, i = 1, \ldots, n$.

3.3 The negative multinomial model

Following Zhang et al. (2017), consider n independent samples of counts taken from the negative multinomial distribution with d classes, denoted $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{id})$, together with p sample-specific covariates, denoted $\boldsymbol{x}_i, i = 1, \ldots, n$. Then $E[y_{ij}] = r p_{ij} / p_{d+1}$ where $p_{i,d+1} = 1 - \sum_{j=1}^{d} p_{ij}$, $i = 1, \ldots, n, j = 1, \ldots, d$. Bishop et al. (1975, Section 13.8) formulated a model for this setting with the probabilities p_{ij} modelled as

$$\frac{\mu_{ij}}{r + \sum_{j=1}^{d} \mu_{ij}} \quad \text{and} \quad p_{d+1} = \frac{r}{r + \sum_{j=1}^{d} \mu_{ij}},$$

where $\mu_{ij} = E[y_{ij}]$. This formulation has been invoked, for example, by Waller and Zelterman (1997) and Zelterman (2004, pp. 49-51), but is somewhat confusing and essentially incorrect. Specifically, the overdispersion parameter is inextricably associated with the probabilities in the model and does not reflect the genesis of the negative multinomial distribution. To compound matters, the resultant likelihood maps directly onto that of the multivariate Poisson-Gamma given in (5), which is in itself misleading. In a recent paper, Zhang et al. (2017) proposed a link function for the probabilities which does not include the overdispersion parameter r and is specified by

$$p_{ij} = \frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j)}{1 + \sum_{j=1}^{d} \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j)}, j = 1, \ldots, d, \quad \text{and} \quad p_{i,d+1} = \frac{1}{1 + \sum_{j=1}^{d} \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j)}$$

for $i = 1, \ldots, n$. Thus $E[y_{ij}] = r \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j)$. The regression parameters for the probabilities in the model can thus be assembled in the $p \times d$ matrix $\boldsymbol{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d)$ and relate directly to those in the standard parametrisation of the multinomial distribution (Dobson and Barnett, 2008, p. 151). In addition, the authors proposed that the overdispersion parameter r be modelled as $r = \ln \boldsymbol{x}_i^\top \boldsymbol{\alpha}$ or as an unknown constant.

The likelihood for the parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d$ and r or $\boldsymbol{\alpha}$ follows directly from the pmf of the negative multinomial distribution given in (3). The number of parameters in the model, that is pd or $p(d+1)$, can, however, be large and Zhang et al. (2017) observed that obtaining MLEs for the parameters is "nontrivial". Specifically, they found Newton and quasi-Newton routines for maximising the log-likelihood to be unstable and the MM-algorithm, while more stable, to be slow. Zhang et al. (2017) therefore introduced a framework for finding the MLEs of the parameters which incorporates a new algorithm, iteratively reweighted Poisson regression (IRPR), coupled optimally with the fast Newton method. The information matrix for the parameters $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_d$ and r or $\boldsymbol{\alpha}$ is derived in the Supplementary Materials of Zhang et al. (2017) and is somewhat complicated. In particular, the only term in the log-likelihood ℓ which involves both $\boldsymbol{\beta}_j$ and r is given by $-r \ln(\sum_{j=1}^{d} \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}_j) + 1)$ and thus $-E[\partial^2 \ell / \partial \boldsymbol{\beta}_j \partial r]$ is non-zero and the information matrix is therefore not block diagonal.

Zhang et al. (2017) compare the multinomial, the negative multinomial, the Dirichlet-multinomial and the generalised Dirichlet-multinomial models in terms of model selection, hypothesis testing and variable selection. Their findings are of interest in the present context, more particularly in relation to the multinomial and negative multinomial models. Specifically, the authors note that information criteria, such as AIC and BIC, and not likelihood-based tests, should be used in choosing a model but that a standard range of tests, as, for example, the likelihood ratio test, the Wald test and the score test, can be invoked to test the significance of the covariates. They also showed by means of extensive simulations that, if the wrong model is chosen, tests of significance for the covariates exhibit highly inflated Type I errors and variable selection by regularisation is poor and, in certain cases, no better than a random selection process.

3.4 Similarities and differences

The multivariate Poisson-Gamma model arises quite naturally from independent Poisson variates. In contrast, the multinomial N-mixture model is built from the multinomial distribution and necessarily incorporates probabilities from that distribution. Nevertheless, the forms of the likelihoods associated with these two distributions are similar and the same broad-based approach to estimation and inference can be adopted. The difference between the likelihoods of the multivariate Poisson-Gamma and the multinomial N-mixture models and that of the negative multinomial is, however, stark. Specifically, the embedded probabilities of the negative multinomial are modelled directly by means of an appropriate link function and, as a consequence, estimation of the parameters is not straightforward and inference must be treated sensitively (Zhang et al., 2017).

4. Discussion

This paper is concerned with three distributions, the multivariate Poisson-Gamma distribution, the multinomial N-mixture distribution and the negative multinomial distribution, all of which can be used to model correlated count data. The distributions have probability mass functions of the same form and therefore share a broad base of common properties, in particular with respect to moments, marginal distributions and positive correlation. At the same time, the distributions are very different in genesis and in terms of interpretation and this is reflected in the attendant model building and analyses.

In drawing the three distributions together, it is interesting to examine how more recent work

might impact on future research and applications. The multivariate Poisson-Gamma distribution has received scant attention in the literature more recently, in part because researchers have found the multivariate Poisson lognormal model to provide a better fit to their data (Egan and Herriges, 2006; Choudhary et al., 2018). There are, however, applications in ecology, such as the example given in Subsection 3.1, for which the multivariate Poisson-Gamma model may well be appropriate. The multivariate N-mixture model was, until very recently, analysed using a numerical approximation to the probability mass function. The paper of Haines (2020) now provides a statistical framework for the model which is based on an explicit expression for the likelihood. Many advantages in terms of computation, analysis and hypothesis testing so accrue and future applications will undoubtedly draw on these results. The negative multinomial distribution has, somewhat surprisingly, been underreported in the statistical literature, with many results embedded in applications and not made explicit. The paper of Zhang et al. (2017) is therefore important in that the authors provide a meaningful link function for the multinomial probabilities and set out a clear strategy for model choice, analysis and hypothesis testing. Practitioners should find the guidelines so provided to be invaluable when choosing and implementing a model from the broad spectrum of multinomial-based models.

Finally it should be noted that the terms multivariate negative binomial distribution, multivariate Poisson-Gamma and negative multinomial have been used somewhat interchangeably in the literature. It is therefore recommended here that the term multivariate negative binomial be used to refer to the family of the three distributions of interest and that the name of the distribution used for a particular setting, that is multivariate Poisson-Gamma, multinomial N-mixture or negative multinomial, should reflect, very strictly, the genesis of the data.

References

Arbous, A. G. and Kerrich, J. E. (1951). Accident statistics and the concept of accident-proneness. *Biometrics*, **7**, 340–432.

Bates, G. E. and Neyman, J. (1952). Contributions to the theory of accident proneness: I. An optimistic model of the correlation between light and severe accidents. *University of California Publications in Statistics*, **1**, 215–253.

Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis. Theory and Practice*. MIT Press, Cambridge.

Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. First edition. Cambridge University Press, Cambridge.

Choudhary, P., Imprialou, M., Velaga, N. R., and Choudhary, A. (2018). Impacts of speed variations on freeway crashes by severity and vehicle type. *Accident Analysis & Prevention*, **121**, 213–222.

Dey, D. K. and Chung, Y. (1992). Compound Poisson distributions: properties and estimation. *Communications in Statistics – Theory and Methods*, **21**, 3097–3121.

Dobson, A. J. and Barnett, A. G. (2008). *An Introduction to Generalized Linear Models*. Third edition. Chapman & Hall/CRC: Boca Raton, FL.

Dorazio, R. M., Jelks, H. L., and Jordan, F. (2005). Improving removal-based estimates of

abundance by sampling a population of spatially distinct subpopulations. *Biometrics*, **61**, 1093–1101.

EGAN, K. AND HERRIGES, J. (2006). Multivariate count data regression models with individual panel data from an on-site sample. *Journal of Environmental Economics and Management*, **52**, 567–581.

HAINES, L. M. (2020). Multinomial *N*-mixture models for removal sampling. *Biometrics*, **76**, 540–548.

HAUSMAN, J. A., HALL, B. H., AND GRILICHES, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, **52**, 909–938.

JOHNSON, N. L., KOTZ, S., AND BALAKRISHNAN, N. (1997). *Discrete Multivariate Distributions*. Wiley, New York.

KÉRY, M. (2018). Identifiability in *N*-mixture models: A large-scale screening test with bird data. *Ecology*, **99**, 281–288.

KÉRY, M. AND ROYLE, J. A. (2015). *Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS: Volume 1: Prelude and Static Models*. Academic Press, London.

MOTHAFER, G. I. M. A., YAMAMOTO, T., AND SHANKAR, V. N. (2016). Evaluating crash type covariances and roadway geometric marginal effects using the multivariate Poisson-gamma mixture model. *Analytic Methods in Accident Research*, **9**, 16–26.

PAWITAN, Y. (2001). *In All Likelihood: Statistical Modelling and Inference using Likelihood*. Oxford University Press, Oxford.

SCHMIDT, M. AND SCHWABE, R. (2020). Optimal designs for Poisson count data with Gamma block effects. *Journal of Statistical Planning and Inference*, **204**, 128–140.

SIBUYA, M., YOSHIMURA, I., AND SHIMIZU, R. (1964). Negative multinomial distribution. *Annals of the Institute of Statistical Mathematics*, **16**, 409–426.

SOLIS-TRAPALA, I. L. AND FAREWELL, V. T. (2005). Regression analysis of overdispersed correlated count data with subject specific covariates. *Statistics in Medicine*, **24**, 2557–2575.

TSOU, T.-S. (2016). Robust likelihood inference for multivariate correlated count data. *Computational Statistics*, **31**, 845–857.

WALLER, L. A. AND ZELTERMAN, D. (1997). Log-linear modeling with the negative multinomial distribution. *Biometrics*, **53**, 971–982.

ZELTERMAN, D. (2004). *Discrete Distributions: Applications in the Health Sciences*. Wiley, New York.

ZHANG, Y., ZHOU, H., ZHOU, J., AND SUN, W. (2017). Regression models for multivariate count data. *Journal of Computational and Graphical Statistics*, **26**, 1–13.

ZHOU, H. AND LANGE, K. (2010). MM algorithms for some discrete multivariate distributions. *Journal of Computational and Graphical Statistics*, **19**, 645–665.