Dirichlet negative multinomial regression for overdispersed correlated count data

DANIEL M. FAREWELL*

Cochrane Institute of Primary Care and Public Health, Cardiff University, Neuadd Meirionnydd, Heath Park, Cardiff CF14 4YS, UK
farewelld@cf.ac.uk

VERNON T. FAREWELL

MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, UK

SUMMARY

A generic random effects formulation for the Dirichlet negative multinomial distribution is developed together with a convenient regression parameterization. A simulation study indicates that, even when somewhat misspecified, regression models based on the Dirichlet negative multinomial distribution have smaller median absolute error than generalized estimating equations, with a particularly pronounced improvement when correlation between observations in a cluster is high. Estimation of explanatory variable effects and sources of variation is illustrated for a study of clinical trial recruitment.

Keywords: Dirichlet negative multinomial; Longitudinal count data; Regression; Sources of variation.

1. INTRODUCTION

In a recent study of recruitment of patients into clinical trials, multi-disciplinary oncology teams (hereafter termed MDTs) were followed longitudinally. The number of patients approached to enter clinical trials was recorded in successive 6-month periods. Analysis of these figures highlighted the importance of understanding the nature and sources of variation in count data, and underlying approach rates.

A natural approach to modeling longitudinal count data would be through a generalized linear mixed model (GLMM) with Poisson distributions conditional on random effects. However, there are situations when a marginal modeling approach might be preferred. For example, the effect of explanatory variables at the population-averaged level (marginal effects) would be of interest if these variables were defined by genetic marker information, which is time-invariant and for which subject-specific effects would be less attractive conceptually. In our study of trial recruitment, we were interested in intervention effects on the population as a whole, rather than the recruitment performance of individual MDTs. An additional advantage of marginal modeling, especially in contrast to models with subject-specific conditional explanatory variable effects, is that it has been shown to offer some degree of robustness for the estimation of regression parameters when departure from the underlying assumed random effect structure occurs (Heagerty and Kurland, 2001).

*To whom correspondence should be addressed.
One way of employing a marginal model would be to make use of generalized estimating equation (GEE) methods (Solis-Trapala and Farewell, 2005). However, GEE methods do not provide direct information on sources of variation, and there is a cost in efficiency compared with parametric maximum-likelihood estimation. Bearing this in mind, alternatives were investigated, and an analysis based on the Dirichlet negative multinomial distribution appeared to be potentially useful.

The Dirichlet negative multinomial distribution is a discrete multivariate distribution having support on the non-negative integers, and has been well characterized by Mosimann (1963), Leeds and Gelfand (1989), and Johnson and others (1997). However, the typical formulation, in terms of a negative multinomial distribution compounded by a Dirichlet distribution, is unappealing in a regression context; we provide a more natural random effects description. A parameterization for this adaptation is given in the following sections and an illustrative analysis of the motivating data is provided subsequently. We also discuss the results of a small simulation study that explores model robustness and finite-sample behavior.

2. DIRICHLET NEGATIVE MULTINOMIAL DISTRIBUTION

Consider a sequence of independent and identically distributed multinomial trials, each of which has 1 + \( m \) possible outcomes. Call one of the outcomes a “success”, and suppose it has probability \( p_0 \). The other \( m \) outcomes have probabilities \( p_1, \ldots, p_m \) and describe distinct types of “failure”. If the vector \( (y_1, \ldots, y_m) \) counts the \( m \) types of failure until \( y_0 > 0 \) successes are observed, then \( (y_1, \ldots, y_m) \) is said to have a negative multinomial distribution with parameters \( (y_0, p_0, \ldots, p_m) \). Just as in the negative binomial (Pólya) case, \( y_0 \) need not be an integer: the negative multinomial distribution remains well defined when \( y_0 \) takes any positive real value.

Now suppose that, for each set of multinomial trials, the probabilities \( p_0, \ldots, p_m \) are themselves sampled randomly from a Dirichlet distribution, having positive parameters \( \alpha_0, \ldots, \alpha_m \). The resulting distribution of \( (y_1, \ldots, y_m) \) is then called Dirichlet (more generically, compound) negative multinomial. It has \( 2 + m \) parameters, namely \( y_0, \alpha_0, \ldots, \alpha_m \). As shown by Mosimann (1963), the joint probability mass function of \( (y_1, \ldots, y_m) \) so-distributed is

\[
\frac{\Gamma(y_*)}{\Gamma(y_0) \prod_{j=1}^m \Gamma(y_j + 1)} \times \frac{\Gamma(\alpha_*)}{\prod_{j=0}^m \Gamma(\alpha_j)} \times \prod_{j=0}^m \frac{\Gamma(y_j + \alpha_j)}{\Gamma(y_* + \alpha_*)},
\]

where \( y_* = y_0 + \cdots + y_m \) and \( \alpha_* = \alpha_0 + \cdots + \alpha_m \). The Dirichlet negative multinomial distribution has a finite mean when \( \alpha_0 > 1 \), and a finite mean and covariance when \( \alpha_0 > 2 \) (Mosimann, 1963). Heuristically, these constraints can be thought of as placing upper bounds on the variance of the underlying Dirichlet distribution.

An informative alternative to the foregoing, traditional, derivation of the Dirichlet negative multinomial distribution proceeds as follows. Let \( y_1, \ldots, y_m \) be independent Poisson variables with mean values \( \lambda_1, \ldots, \lambda_m \). If \( t \sim \text{Ga}(y_0, y_0) \) (shape = rate = \( y_0 \)) and the \( \lambda_j \) are fixed and known, then \( (y_1, \ldots, y_m) \) has a negative multinomial distribution. Alternatively, if the \( \lambda_j \) are positive random variables of a particular form, then the Dirichlet negative multinomial distribution results.

Interestingly, the required distributional form for the \( \lambda_1, \ldots, \lambda_m \) is not the seemingly natural independent gamma construction, but instead a further hierarchical structure where each \( \lambda_j \) has a gamma distribution with shape \( \alpha_j \) and a common, random rate parameter \( \lambda_0 \sim \text{Ga}(\alpha_0, y_0) \) that induces correlation in the \( \lambda_j \). This three-level formulation results in the same Dirichlet negative multinomial distribution as in (2.1), with parameters \( y_0, \alpha_0, \ldots, \alpha_m \). Marginally, the \( \lambda_j \) have compound-gamma (a special case of the generalized beta-prime) distributions (Dubey, 1970). We see that restrictions on \( \alpha_0 \) correspond directly to restrictions on the skewness of the distribution of \( \lambda_0 \).
3. Regression formulation

Mosimann (1963) reports the mean of the Dirichlet negative multinomial distribution, but omits its derivation. Our alternative, random effects formulation of the Dirichlet negative multinomial provides a direct route: since \( t \) has unit mean, the expectation of \( y_j \) is simply the mean of \( \lambda_j \), which in turn has a compound-gamma distribution whose first moment is \( \alpha_j y_0/(\alpha_0 - 1) \) when \( \alpha_0 > 1 \). Mosimann (1963) also reports variances and covariances of the Dirichlet negative multinomial distribution, and these can similarly be derived from the random effects formulation. Applying the law of total covariance twice, it easily follows that for \( j \neq k \), \( \text{cov}(y_j, y_k) = E(y_j)E(y_k)\text{var}(t) + \text{cov}(\lambda_j, \lambda_k)(1 + \text{var}(t)) \). The required \( \text{cov}(\lambda_j, \lambda_k) \) can either be determined directly from their bivariate compound gamma distribution (Hutchinson, 1981) or by a third application of the same law. Either way, we find (for \( j \neq k \)) that

\[
\text{cov}(y_j, y_k) = \frac{\alpha_j \alpha_k y_0^2}{(\alpha_0 - 1)^2} \times \frac{\alpha_0 - 1 + y_0}{y_0(\alpha_0 - 2)}.
\]

The law of total variance similarly gives \( \text{var}(y_j) = \text{var}(\lambda_j) + E(\lambda_j)(1 + \text{var}(t)) \), which simplifies to

\[
\text{var}(y_j) = \frac{\alpha_j y_0}{\alpha_0 - 1} \times \left( \frac{\alpha_j y_0}{\alpha_0 - 1} + y_0 \right) \times \frac{\alpha_0 - 1 + y_0}{y_0(\alpha_0 - 2)}.
\]

It is apparent from the form of the variance and covariance that the parameterization in terms of the \( \alpha_j \) and \( y_0 \) is far from ideal for regression purposes. At a minimum, we would like to parameterize in terms of the means \( \mu_1, \ldots, \mu_m \) of the vector random variable \( (y_1, \ldots, y_m) \), and have the flexibility to say \( E(y_j) = \mu_j \). This is satisfied if we write \( \alpha_j = \mu_j \times (\alpha_0 - 1)/y_0 \). Additionally, we should like the remaining two parameters to convey something meaningful about the variance structures in the model. We suggest writing \( \varphi_{\text{NM}} = 1/y_0 \) and \( \varphi_{\text{NB}} = y_0/(\alpha_0 - 1) \); the notation is suggestive of the interpretation we ascribe to these parameters in the following section, linked to negative multinomial and negative binomial variation, respectively. A Dirichlet negative multinomial distribution with parameters \( (\varphi_{\text{NM}}^{-1}, \varphi_{\text{NM}}^{-1} \varphi_{\text{NB}}^{-1} + 1, \mu_1/\varphi_{\text{NB}}, \ldots, \mu_m/\varphi_{\text{NB}}) \) is therefore a candidate regression model for correlated count data. The regression specification is completed by setting \( g(\mu_j) = x_j \beta \), where \( g \) is a link function (canonically \( g = \log \)), \( x_j \) is a row vector of explanatory variables, and \( \beta \) is a column vector of regression coefficients to be estimated. The parameter \( \varphi_{\text{NB}} \) provides a convenient way to ensure that \( E(y_j) \) exists: when \( \varphi_{\text{NB}} > 0 \), then \( \alpha_0 > 1 \) and so \( E(y_j) = \mu_j \). Under the regression parameterization, the covariance of two observations \( y_j \) and \( y_k \) within the same unit takes the form

\[
\text{cov}(y_j, y_k) = \mu_j (\varphi_{\text{NM}} + \delta_{jk}) \left( \frac{1 + \varphi_{\text{NB}}}{1 - \varphi_{\text{NM}} \varphi_{\text{NB}}} \right), \tag{3.1}
\]

where \( \delta_{jk} = 1 \) if \( j = k \), and 0 otherwise. This covariance exists provided \( \varphi_{\text{NM}} \varphi_{\text{NB}} < 1 \).

For observations on multiple individuals, the full likelihood function will be defined in the usual way as a product over all Dirichlet negative multinomial observations of terms of the form (2.1). Expressions for likelihood derivatives are provided in supplemental electronic material (available at Biostatistics online), together with \( R \) code to fit the foregoing regression model.

4. Interpretation

The Dirichlet negative multinomial distribution has been used most extensively when modeling the allocation of an unknown number of items (such as product purchases) to a known set of categories (such as individual brands), as for example in Goodhardt and others (1984). In such contexts formulating a model
in terms of a negative binomial total with Dirichlet multinomial category probabilities is sensible and interpretable. However, when used as a model for longitudinal (more generically, repeated measures) count data, the concept of category probabilities has no substantive meaning: the number of "categories" could vary from subject to subject, and the total number of "items" is unlikely to reflect any important aspect of the scientific problem. Neither is it natural to extend the notion of categories to regression on other variables, and Goodhardt and others (1984) provide no such extension. In contrast, the Gamma random effect construction offers a more familiar perspective. Subjects have one shared and as many observation-specific random effects as needed, and these act on the marginal mean in the manner of a generalized linear model. An alternative regression formulation of the Dirichlet negative multinomial distribution to that given in Section 3 was developed by Hausman and others (1984) in the econometric context of studying the relationship between patents and expenditure by research & development. This was based on adopting a negative binomial model for each \( y_j \) with probability function

\[
\frac{\Gamma(\mu_j\delta + y_j)}{\Gamma(\mu_j\delta)\Gamma(y_j + 1)} \times \left( \frac{\delta}{1 + \delta} \right)^{\mu_j\delta} \times \left( \frac{1}{1 + \delta} \right)^{y_j},
\]

where the overdispersion parameter (\( \delta \) in their notation) is the same for all \( y_j \) from the same individual or unit of observation. For reasons of mathematical convenience, Hausman and others (1984) then assumed that the \( \delta \) for an individual or a unit is a random variable where \( \delta/(1 + \delta) \sim \text{Beta}(a, b) \). Conceptually, within-unit variation is reflected in the negative binomial assumption, which corresponds to a Poisson model with gamma random effects while between-unit variation is introduced by random variation in the \( \delta \) values. However, no interpretation was presented for the parameters \( a \) and \( b \), and it was recognized by Hausman and others (1984) that no investigation of the two variance components could be made. Once again, category probabilities, implicit in the underlying Beta distribution, offer little insight into this aspect of the model.

Comparison of the form of the resulting probability function with Equation (2.1) reveals that \( a = \alpha_0 \) and that \( b = y_0 \). This form of the model has been implemented in the Stata package \texttt{xtmnbreg}, where the linear predictor is linked to \( \log(\mu_j/\varphi_{\text{NB}}) \) rather than the \( \mu_j \) and therefore the location term would be the sum of \( \log(\varphi_{\text{NB}}) \) and the location term derived from a fit of the model under our proposed formulation.

Arguably, the random effects formulation of the Dirichlet negative multinomial distribution simply makes it clear that there is no straightforward interpretation of the dispersion parameters. Since \( (\lambda_j | \lambda_0) \sim \text{Ga}(\mu_j/\varphi_{\text{NB}}, \lambda_0) \) and \( \lambda_0 \sim \text{Ga}(1/(\varphi_{\text{NB}}\varphi_{\text{NB}}) + 1, 1/\varphi_{\text{NM}}) \), we see that the parameter \( \varphi_{\text{NM}} \) influences the variance of observation-specific (\( \lambda_j \)) as well as the shared (\( t \)) random effects, while the \( \lambda_j \) themselves induce a small amount of correlation in the \( y_j \).

However, the limiting behavior is informative about their respective roles: consider the limit as \( \varphi_{\text{NM}} \to 0 \). The random effects \( t \) and \( \lambda_0 \) degenerate to 1 and \( 1/\varphi_{\text{NB}} \), respectively, meaning that the only remaining random effects are the (now independent) \( \lambda_j \sim \text{Ga}(\mu_j/\varphi_{\text{NB}}, 1/\varphi_{\text{NB}}) \). We deduce that, as \( \varphi_{\text{NM}} \to 0 \), the distribution of \( (y_1, \ldots, y_m) \) converges to independent negative binomial distributions, having parameters \( \mu_j/\varphi_{\text{NB}} \) specifying the number of successes, \( 1/(1 + \varphi_{\text{NB}}) \) the success probability and \( \varphi_{\text{NB}}/(1 + \varphi_{\text{NB}}) \) the failure probability. As \( \varphi_{\text{NB}} \to 0 \), the \( \lambda_j \) degenerate to point masses on the \( \mu_j \), and hence the distribution of \( (y_1, \ldots, y_m) \) converges to a negative multinomial distribution with parameters \( \mu_0, \mu_0/\mu_0, \ldots, \mu_m/\mu_m \). Here, we have written \( \mu_0 = \mu_0 + \cdots + \mu_m \) and (since it is the correct limit) \( \mu_0 = \varphi_{\text{NM}}^{-1} \).

Taken together, these two limiting results indicate that \( \varphi_{\text{NB}} \) is linked mainly to within-unit variation across time. In contrast, \( \varphi_{\text{NM}} \) is related to variation between units. Aside from the boundary cases just described, the interplay between \( \varphi_{\text{NM}} \) and \( \varphi_{\text{NB}} \) will typically be fairly complex, both capturing some elements of between- and within-cluster variation. We can, however, make use of Mosimann’s observation that the covariance matrix of a Dirichlet negative multinomial distribution is just a scalar multiple of that of the
related negative multinomial distribution. The relevant overdispersion constant is given by the expression

\[
C = \frac{\alpha_0 - 1 + y_0}{\alpha_0 - 2} = \frac{1 + \phi_{NB}}{1 - \phi_{NM}\phi_{NB}}.
\]

In the limiting cases we have discussed, this overdispersion constant tends to \(1 + \phi_{NB}\) (when \(\phi_{NM} \to 0\)) and 1 (when \(\phi_{NB} \to 0\)). For small values of \(\phi_{NM}\phi_{NB}\), then, \(\phi_{NB}\) can be seen crudely as quantifying the amount of extra-negative multinomial variation present in the data.

Another way of understanding the relative contributions of between- and within-unit variation is to compare the variance of \(t\) (which has mean 1) with the covariance matrix of the \(\lambda_j\), holding the mean values of the latter also equal to 1. When \(\mu_j = \mu_k = 1\), the covariance of \(\lambda_j\) and \(\lambda_k\) is \((\delta_{jk} + \phi_{NM})\phi_{NB}/(1 - \phi_{NM}\phi_{NB})\), while the variance of \(t\) is \(\phi_{NM}\) irrespective of the values of the \(\mu_j\). This comparison puts \(t\) and \(\lambda_j\) on an equal footing; it allows us to investigate how the variability in such a set of observations is spread across the between- and within-unit random effects. On this scale, a direct comparison of variances is meaningful. This approach is illustrated in the following section.

Finally, we suggest similar examination of the estimated covariance matrix for two observations having unit or some other convenient mean. From (3.1), we have that when \(\mu_j = \mu_k = 1\),

\[
\text{cov}(y_j, y_k) = (\delta_{jk} + \phi_{NM}) \times \frac{1 + \phi_{NB}}{1 - \phi_{NM}\phi_{NB}}.
\]

Although not at all specific to our random effects formulation, we think that this offers the most direct insight into the variance structures present in the model. For more generality, we could consider the covariance matrix over a range of possible mean values, or plot the variance and covariance as a function of unspecified means \(\mu_j\) and \(\mu_k\).

5. Simulation study

To explore the usefulness of regression based on the Dirichlet negative multinomial distribution, we conducted a small simulation study. We generated data from two, similar models: one was Dirichlet negative multinomial, and the other a related Gamma–Gamma–Poisson model. In both instances, we conducted a small simulation study. We generated data from two, similar models: one was Dirichlet negative multinomial, and the other a related Gamma–Gamma–Poisson model. In both instances, we put \((y_j | \lambda_j, t) \sim \text{Poi}(\lambda_j t)\) for \(j = 1, \ldots, m\) and \(t \sim \text{Ga}(1/\phi_{NM}, 1/\phi_{NM})\), and in both we set \(\log(\mu_j) = \beta_0 + x_1\beta_1 + x_2\beta_2\), where \(x_1\) was an equiprobable binary covariate and \(x_2\) a standard Gaussian covariate. The difference between the models lay in the specification of the \(\lambda_j\); one employed the compound gamma setup that results in our Dirichlet negative multinomial model, the other used independent gamma variables \(\lambda_j \sim \text{Ga}(\mu_j \phi_{NB}, \phi_{NB})\). These competing models allowed us to investigate the performance of Dirichlet negative multinomial regression when it is, and is not, correctly specified.

In all simulations, \(\beta_0 = \beta_1 = \beta_2 = 1\). We used sample sizes \(n = 50, 100\), and unit sizes \(m = 5, 10\), and for each of these four combinations we considered all nine combinations of \(\phi_{NM} = 0.25, 0.5, 0.75\) and \(\phi_{NB} = 0.25, 0.5, 0.75\). We compared the performance of Dirichlet negative multinomial regression with a Poisson GEE using an exchangeable working correlation structure, and focus our reporting on the median absolute error. We use the median absolute error rather than root mean squared error because, in a small percentage of cases, the GEE routine either failed to converge or converged to implausible parameter estimates; since these would be noticed by any competent analyst, median absolute error constituted a fairer comparison of performance, being less sensitive to a few suspect estimates.

Table 1 reports the results for \(n = 100, m = 5\); other results are qualitatively similar, and tables giving details may be found in the online appendix (available at Biostatistics online). We note that, within these simulations, Dirichlet negative multinomial regression is uniformly preferable to the GEE approach.
Table 1. For \( n = 100, m = 5 \), median absolute error of the fixed effects estimates (\( \beta_j \)) over 100 simulations, under different variance structures (\( \psi_{\text{NM}}, \psi_{\text{NB}} \)), data generating models (columns, DNM: Dirichlet negative multinomial, GGP: Gamma–Gamma–Poisson), and regression models (rows, DNM: Dirichlet negative multinomial, GEE: generalized estimating equation)

| \( \psi_{\text{NM}} \) | \( \psi_{\text{NB}} \) | Model  | \( \beta_0 \) GNM | \( \beta_0 \) GGP | \( \beta_1 \) GNM | \( \beta_1 \) GGP | \( \beta_2 \) GNM | \( \beta_2 \) GGP |
|-----------------|-----------------|-------|----------------|----------------|----------------|----------------|----------------|----------------|
| 0.25            | 0.25            | DNM   | 0.051          | 0.047          | 0.031          | 0.031          | 0.018          | 0.019          |
|                 |                 | GEE   | 0.067          | 0.058          | 0.040          | 0.040          | 0.034          | 0.033          |
| 0.5             | 0.25            | DNM   | 0.050          | 0.047          | 0.033          | 0.029          | 0.020          | 0.019          |
|                 |                 | GEE   | 0.074          | 0.059          | 0.052          | 0.041          | 0.035          | 0.028          |
| 0.75            | 0.25            | DNM   | 0.048          | 0.036          | 0.041          | 0.033          | 0.022          | 0.023          |
|                 |                 | GEE   | 0.089          | 0.049          | 0.060          | 0.049          | 0.039          | 0.031          |
| 0.5             | 0.25            | DNM   | 0.064          | 0.059          | 0.035          | 0.032          | 0.015          | 0.016          |
|                 |                 | GEE   | 0.098          | 0.068          | 0.049          | 0.047          | 0.040          | 0.039          |
| 0.5             | 0.25            | DNM   | 0.070          | 0.070          | 0.036          | 0.035          | 0.018          | 0.017          |
|                 |                 | GEE   | 0.114          | 0.093          | 0.065          | 0.044          | 0.066          | 0.050          |
| 0.75            | 0.25            | DNM   | 0.095          | 0.061          | 0.042          | 0.040          | 0.020          | 0.015          |
|                 |                 | GEE   | 0.116          | 0.083          | 0.077          | 0.054          | 0.056          | 0.035          |
| 0.75            | 0.25            | DNM   | 0.081          | 0.078          | 0.034          | 0.029          | 0.019          | 0.020          |
|                 |                 | GEE   | 0.216          | 0.183          | 0.087          | 0.088          | 0.080          | 0.092          |
| 0.5             | 0.25            | DNM   | 0.115          | 0.091          | 0.039          | 0.040          | 0.019          | 0.019          |
|                 |                 | GEE   | 0.153          | 0.106          | 0.107          | 0.093          | 0.074          | 0.053          |
| 0.75            | 0.25            | DNM   | 0.117          | 0.095          | 0.042          | 0.041          | 0.026          | 0.023          |
|                 |                 | GEE   | 0.146          | 0.090          | 0.093          | 0.057          | 0.070          | 0.048          |

in terms of median absolute error, and in almost all cases the best 50% of estimates lie within 10% of the true value. Improvements over GEEs were particularly marked for the coefficient \( \beta_2 \) of the continuous covariate \( x_2 \), in some cases showing a 4-fold improvement in median absolute error.

Interestingly, the biggest gains in efficiency are found when \( \psi_{\text{NM}} = 0.75 \) and \( \psi_{\text{NB}} = 0.25 \). This corresponds to the largest correlation between observations within a unit that we explored in our simulations, and it seems that in such cases modeling the variance (even incorrectly) leads to useful improvements in estimation. Note that the apparently superior performance of both models under the Gamma–Gamma–Poisson setup owes simply to the fact that the variance of the data is smaller here: the compounding gamma distribution is absent, so the only meaningful comparisons in the table are vertical ones.

Table 2 summarizes the estimation of the variance parameters for Dirichlet negative multinomial regression. Recall that \( \psi_{\text{NM}} \) and \( \psi_{\text{NB}} \) have different meanings under the two data-generation setups; nevertheless, the proximity of the mean estimates that arise under the Dirichlet negative multinomial assumption to the true values of \( \psi_{\text{NM}} \) and \( \psi_{\text{NB}} \) even under the Gamma–Gamma–Poisson setup lend empirical support to our interpretation of these two parameters. As expected, the approximation becomes less satisfactory as the product \( \psi_{\text{NM}} \psi_{\text{NB}} \) increases. Even when the model is correctly specified, there is some underestimation of the variance components. This, too, is to be expected; bias of variance component estimates toward the null in small samples is a well-known feature of maximum-likelihood inference in random effect models. When the Dirichlet negative multinomial model was correctly specified, median absolute error in the estimation of \( \psi_{\text{NM}} \) and \( \psi_{\text{NB}} \) (not tabulated) varied between about 0.05 and 0.15, increasing with \( \psi_{\text{NM}} \) and \( \psi_{\text{NB}} \) and decreasing with sample size; this indicates that small, moderate, and large values of the variance components can be fairly reliably distinguished, even in small samples.
Table 2. For \( n = 100, m = 5 \), mean estimates of variance component (columns) over 100 simulations, under different variance structures (rows) and data generating models (columns, DNM: Dirichlet negative multinomial, GGP: Gamma–Gamma–Poisson)

| \( \phi_{NM} \) | \( \phi_{NB} \) | DNM | GGP | DNM | GGP |
|----------------|----------------|-----|-----|-----|-----|
| 0.25           | 0.25           | 0.2440 | 0.1977 | 0.2437 | 0.2426 |
| 0.5            | 0.2466         | 0.1701 | 0.4972 | 0.4799 |
| 0.75           | 0.2439         | 0.1474 | 0.7491 | 0.7424 |
| 0.5            | 0.25           | 0.4680 | 0.3928 | 0.2574 | 0.2406 |
| 0.5            | 0.4721         | 0.3438 | 0.5049 | 0.4789 |
| 0.75           | 0.4782         | 0.3064 | 0.7584 | 0.7313 |
| 0.75           | 0.25           | 0.7265 | 0.6140 | 0.2546 | 0.2325 |
| 0.5            | 0.7207         | 0.5388 | 0.5207 | 0.4990 |
| 0.75           | 0.7342         | 0.4986 | 0.7636 | 0.7419 |

Table 3. For the trial recruitment data, comparative results from a variety of regression models. Standard errors are shown in parentheses for each fixed effect estimate

| Explanatory variables | Hospital type | Nurses > 1 | Number of trials > 5 |
|-----------------------|---------------|------------|----------------------|
| Poisson               | 1.05 (0.087)  | 0.64 (0.093)| 0.64 (0.073)         |
| Negative binomial     | 1.08 (0.189)  | 0.62 (0.212)| 0.61 (0.185)         |
| Negative multinomial  | 1.14 (0.197)  | 0.75 (0.197)| 0.32 (0.115)         |
| Negative multinomial (GEE) | 1.14 (0.273) | 0.75 (0.269)| 0.32 (0.189)         |
| Dirichlet negative multinomial | 0.93 (0.246) | 0.68 (0.276) | 0.36 (0.189) |
| GLMM (Poisson)        | 1.08 (0.259)  | 0.74 (0.292)| 0.44 (0.199)         |
| GLMM (negative binomial) | 1.07 (0.260) | 0.72 (0.295)| 0.45 (0.207)         |

6. Analysis of motivating example

Data available for analysis from the study outlined in the introduction consisted of number of patients approached over 3 consecutive 6-month periods for 12 MDTs and over 4 consecutive 6-month periods for another 10 MDTs. In addition, a variety of factors thought to influence approach rates were recorded, including whether the MDT worked in a specialist center or a district general hospital, whether there was more than one research nurse in the team and the number of trials in the team’s portfolio during the 6-month period (which, for illustration, is used here as a binary indicator variable for more than five trials). The latter variable may vary across periods, while the first two will not.

The first five rows of Table 3 present the results of fitting five different marginal regression models to these data. The regression coefficient for Hospital Type is similar across all models except the Dirichlet negative multinomial, for which the estimate is somewhat smaller. The effect of having multiple nurses increases slightly for the latter three models compared with the first two while the coefficient for number of trials is decreased in the last three models.

The Poisson model, as expected, has the smallest estimated standard errors for the regression coefficients. One extra variance component is present in the negative binomial and negative multinomial models and these models show comparable increases in the estimated standard errors for the fixed explanatory.
variables related to hospital type and the number of research nurses. However, for the variable related to the number of trials recruiting then the standard error is larger for the negative binomial model than for the negative multinomial model. The robust standard errors for the negative multinomial model are larger than the model-based estimates and comparable with those from the Dirichlet negative multinomial model, which has two extra variance components compared with the Poisson model.

The dispersion parameter for the negative binomial model was estimated as 2.35; this corresponds to a variance of 0.43 for a gamma distribution from which random effects are generated for each observed 6-month period of observation. In contrast, the negative multinomial corresponds to random effects associated with each MDT and the variance of the assumed gamma distribution for these effects was estimated as 0.28.

The estimated values of $\varphi_{\text{NM}}$ and $\varphi_{\text{NB}}$ from the Dirichlet negative multinomial were 0.062 and 2.8, respectively. The much smaller value for $\varphi_{\text{NM}}$ suggests that there may be more variation within an MDT’s observations than between teams and the estimated overdispersion factor for extra-negative multinomial variation is $C = 4.6$. Because $\varphi_{\text{NM}}$ and $\varphi_{\text{NB}}$ are not directly comparable, it is also informative to look at the covariance matrix of the underlying $\lambda_j$ assuming unit means, as outlined in the section describing parameter interpretation. The standard deviation of $t$ is 0.25, while the standard deviation of a single $\lambda_j$ when $\mu_j = 1$ is 2.14, nearly an order of magnitude larger. To indicate that this variation is indeed largely within-team, we note that the correlation of distinct $\lambda_j, \lambda_k$ is just 0.059.

As another illustration, consider a period for which the average number of patients approached is 13.8, the observed mean across periods in the data set. With this mean the estimated standard deviation for an observation from the Dirichlet negative multinomial model would be 10.86. In contrast, the estimated standard deviation from the negative binomial model would be 9.75 and from the negative multinomial model would be 8.16. This demonstrates that there is somewhat more variation within teams than between teams, although we emphasize again that either parameter will pick up some of both types of variation if the other is not included in the model. We note that the (model-based) correlation between numbers of patients approached by a team in distinct 6-month periods each with mean 13.8 is estimated to be just 0.13, emphasizing the extent of variability within the observations on a single MDT.

For comparison purposes, the penultimate row of Table 3 gives the results of fitting a generalized linear Poisson mixed model (with both between- and within-MDT random effects) to these data, using the lme4 package in R (Bates and others, 2012; R Core Team, 2012). The estimated coefficients and standard errors can be seen to be similar to the Dirichlet negative multinomial model. The between- and within-team estimated standard deviations of the Gaussian random effects were 0.461 and 0.458. This model can be thought of as $\lambda_j$ having a log-Normal distribution with parameters $\theta_j$ and $\sigma^2_\theta$, and $t$ having a log-Normal distribution with parameters 0 and $\sigma^2_t$. When these both have unit mean, their variances are $\exp(\sigma^2_\theta) - 1 = 0.24$ and $\exp(\sigma^2_t) - 1 = 0.23$, respectively, indicating that the breakdown of variation into these two components is slightly more comparable within this modeling framework than under a marginal model.

The final row of Table 3 shows the results of fitting a GLMM with Gaussian random effects between-MDTs but with negative binomial residuals, using the glmmADMB package in R (Fournier and others, 2012; Skaug and others, 2012); because the standard deviations of the within-team gamma (0.457) and between-team Gaussian (0.467) random effects are comparable with those from the Gaussian GLMM, it is unsurprising that the estimated fixed effects are also rather similar.

A further example of substantive analysis using a Dirichlet negative multinomial regression model can be found in the supplementary material (available at Biostatistics online).

7. Discussion

The purpose of this note is to highlight the availability of a convenient marginal regression model for count data with an explicit introduction of between- and within-unit variation. The choice between a marginal or
conditional random effects model will typically be application-specific, although some possible advantages to the marginal approach have been discussed. If a marginal model is desired, then the Dirichlet negative multinomial model provides a convenient structure for this purpose, and can provide efficiency gains over GEEs. In particular, having separate mean parameters for each component and two variance parameters makes it suitable for use with unbalanced longitudinal count data with a time-invariant covariance structure.

The lack of separability of the variance components is a weakness of the Dirichlet negative multinomial regression model. The resulting difficulty of interpreting these parameters will often be outweighed by the ease of fitting, and comparing, multiple fixed effects structures without the need for numerical approximation of likelihoods. It is also worth reflecting that, outside of linear models, functional links between means and variances almost always result in difficulty in interpreting variance parameters; the Dirichlet negative multinomial is no different in this regard to, say, the Poisson GLMM used for comparison purposes in our motivating example.

SUPPLEMENTARY MATERIAL

Supplementary material is available at http://biostatistics.oxfordjournals.org.

ACKNOWLEDGMENTS

This paper was substantially improved through the perceptive comments of two anonymous referees. Conflict of Interest: None declared.

FUNDING

The authors’ work was supported by the Medical Research Council (VF: U105261167) and Cancer Research UK.

REFERENCES

BATES, D., MAECHLER, M. AND BOLKER, B. (2012). lme4: Linear Mixed-Effects Models using S4 Classes. R package version 0.999999-0.

DUBEY, S. D. (1970). Compound gamma, beta and F distributions. Metrika 16, 27–31.

FOURNIER, D. A., SKAUG, H. J., ANCHETA, J., JANELLI, J., MAGNUSSON, A., MAUNDER, M., NIELSEN, A. AND SIBERT, J. (2012). AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optimization Methods & Software 27, 233–249.

GOODHARDT, G. J., EHRENBERG, A. S. C. AND CHATFIELD, C. (1984). The Dirichlet: a comprehensive model of buying behaviour. Journal of the Royal Statistical Society. Series A (General) 147, 621–655.

HAUSMAN, J., 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.

HEAGERTY, P. J. AND KURLAND, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. Biometrika 88, 973–985.

HUTCHINSON, T. P. (1981). Compound gamma bivariate distributions. Metrika 28, 263–271.

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

LEEDS, S. AND GELFAND, A. E. (1989). Estimation for Dirichlet mixed models. Naval Research Logistics 36, 197–214.
Mosimann, J. E. (1963). On the compound negative multinomial distribution and correlations among inversely sampled pollen counts. *Biometrika* **50**, 47–54.

R Core Team. (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Skaug, H., Fournier, D., Nielsen, A., Magnusson, A. and Bolker, B. (2012). *Generalized Linear Mixed Models using AD Model Builder*. R package version 0.7.2.12.

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.

[Received January 23, 2012; revised October 29, 2012; accepted for publication October 30, 2012]