A simple approach to fit the beta-binomial model

Paulo Guimarães
Medical University of South Carolina

Abstract. In this paper, I show how to estimate the parameters of the beta-binomial distribution and its multivariate generalization, the Dirichlet-multinomial distribution. This approach involves no additional programming, as it relies on an existing Stata command used for overdispersed count panel data. Including covariates to allow for regression models based in these distributions is straightforward.

Keywords: st0089, overdispersion, beta binomial, Dirichlet multinomial, fixed-effects negative binomial

1 Introduction

When fitting parametric models to count data, researchers customarily worry about the undesirable effect of overdispersion. If present and unaccounted for, overdispersion will lead to biased estimates of the variance–covariance matrix that will invalidate the statistical inference performed on the model. One approach to dealing with this problem is to specify parametric models that accommodate overdispersion and that collapse to the simpler models when overdispersion is not present. A well-known example is negative binomial regression. Researchers routinely use this model because it allows for overdispersed count data, yet it collapses to Poisson regression if the data are equidispersed. Less well known are parametric alternatives to deal with overdispersed multinomial-distributed count data. The Dirichlet-multinomial distribution is a natural candidate for this. Applications of this distribution are mostly restricted to binomial distributed data, in which case the Dirichlet-multinomial distribution becomes the beta binomial. Stata does not include routines for estimation of these latter models.

In this paper, I show how you can employ an existing Stata command used for panel (clustered) count data, xtnbreg, to estimate the parameters of the beta-binomial distribution as well as those of the more general Dirichlet-multinomial distribution. This is possible because of a feature that has hitherto remained unnoticed—the conditional-likelihood function used to estimate the parameters of the fixed-effect negative binomial model (FENB) follows the Dirichlet-multinomial distribution. The practical implication of this observation is that the command xtnbreg (with the fe option) can be readily used to estimate the parameters of the Dirichlet-multinomial distribution and, more importantly, of its univariate counterpart, the beta-binomial distribution. I illustrate the use of xtnbreg for this purpose by replicating some of the examples that are published in the literature. I also briefly discuss how to include covariates and how to implement a likelihood-ratio test for overdispersion.
2 The conditional likelihood of the FENB

Using the notation presented in Methods and Formulas in [XT] `xtnbreg`, let \( y_{it} \) be the \( t \)th count observation for the \( i \)th group (cluster or individual). Let \( \lambda_{it} = \exp(x_{it} \beta) \), where the \( x_{it} \) are covariates that change with observation and group and \( \beta \) is the vector of parameters to be estimated. As well described in the Stata manual, the FENB is estimated by conditional maximum likelihood. This is done by constructing for each group the joint probability of the observed counts conditional on the sum of the counts for that group. The contribution of each group to the conditional maximum likelihood function is (see [XT] `xtnbreg`, bottom of page 148)

\[
\frac{\Gamma(\sum_{t=1}^{n_i} \lambda_{it}) \Gamma(\sum_{t=1}^{n_i} y_{it} + 1)}{\Gamma(\sum_{t=1}^{n_i} \lambda_{it} + \sum_{t=1}^{n_i} y_{it})} \prod_{t=1}^{n_i} \frac{\Gamma(\lambda_{it} + y_{it})}{\Gamma(\lambda_{it})\Gamma(y_{it} + 1)}
\]  

The log-likelihood function for the FENB is obtained by taking logs of (1) and adding across groups. The mathematical expression shown in (1) defines the multivariate Dirichlet-multinomial (also known as compound multinomial) distribution. This distribution was introduced by Mosimann (1962) and results from assuming that the cell probabilities \( (p_1, p_2, \ldots, p_k) \) of a multinomial distribution with parameters \( (n; p_1, p_2, \ldots, p_k) \) are distributed according to a multivariate beta distribution (or Dirichlet distribution) with parameters \( (\alpha_1, \alpha_2, \ldots, \alpha_k) \). It follows that

\[
E(p_j) = \frac{\alpha_j}{\alpha_1 + \alpha_2 + \cdots + \alpha_k} \quad \text{with} \quad j = 1, 2, \ldots, k
\]  

The Dirichlet-multinomial distribution is given by Johnson, Kotz, and Balakrishnan (1997, 80, equation 35.152) as

\[
f_{DM}(n_1, n_2, \ldots, n_k) = \frac{n!\Gamma(\alpha_*)}{\Gamma(n + \alpha_*)} \prod_{i=1}^{k} \left\{ \frac{\Gamma(n_i + \alpha_i)}{n_i!\Gamma(\alpha_i)} \right\}
\]  

where \( n = \sum_{i=1}^{k} n_i \) and \( \alpha_* = \sum_{i=1}^{k} \alpha_i \).

Comparison of (1) and (3) makes the equivalence obvious: \( \lambda_{it} = \alpha_i \) and \( y_{it} = n_i \). While applications of the Dirichlet-multinomial distribution are less common, its univariate version, the beta binomial, is more commonly known (see, for example, Agresti [2002]). This distribution is used to model binomial overdispersed data and may be motivated as the composition of a binomial distribution with parameters \( (n, p_i) \), where \( p_i \) follows a beta distribution with parameters \( (a, b) \), with \( a > 0, b > 0 \). It follows that

\[
E(p_i) = \mu = \frac{a}{a + b}, \quad V(p_i) = \mu(1 - \mu)\rho
\]  

where \( \rho = (1 + a + b)^{-1} \) is the intraclass correlation coefficient.

The beta-binomial distribution, presented in a way that highlights the relation with the Dirichlet multinomial, is

\[
f_{BB}(n_1, n_2) = \frac{n!\Gamma(a + b)}{\Gamma(n + a + b)} \cdot \frac{\Gamma(a + n_1)\Gamma(b + n_2)}{n_1!n_2!\Gamma(a)\Gamma(b)}
\]
where \( n = n_1 + n_2 \) represents the total number of trials and \( n_1 \) represents the total number of successes. In practical applications, you observe information for several samples and record the number of trials in the \( i \)th sample, \( n_i \), and the corresponding number of successes, \( n_{1i} \). Interest centers in the estimation of \( E(p_i) \), and consequently of the \( a \) and \( b \) parameters. The equivalence between the likelihood function of the FENB and that of the beta-binomial model suggests that we let \( a = \lambda_{i1} = \exp(\beta_0 + \beta_1) \) and \( b = \lambda_{i2} = \exp(\beta_0) \)—a parametrization that assures the nonnegativity of \( a \) and \( b \). In terms of the FENB, this amounts to introducing an indicator variable that assumes the value 1 when \( t = 1 \) and 0 when \( t = 0 \). It is also possible to introduce covariates that change across groups using a parametrization similar to the one proposed by Heckman and Willis (1977). This is done by linking these covariates to one of the parameters. For example, to introduce a covariate \( x_{1i} \) (where \( i \) is an index for group), you would let \( a = \lambda_{i1} = \exp(\beta_0 + \beta_1 + \beta_2 x_{1i}) \) and thus

\[
E(p_i) = \frac{\exp(\beta_0 + \beta_1 + \beta_2 x_{1i})}{\exp(\beta_0) + \exp(\beta_0 + \beta_1 + \beta_2 x_{1i})} = \frac{\exp(\beta_1 + \beta_2 x_{1i})}{1 + \exp(\beta_1 + \beta_2 x_{1i})}
\]

Now the impact of the covariates on \( E(p_i) \) has the same interpretation as in a conventional logit model (for an additional discussion of beta-binomial regression, see Simonoff [2003]). By the same token, it would be possible to introduce covariates in the Dirichlet-multinomial model following a logic that mimics the one used for the conditional logit model. Note, however, that with the proposed parametrization, the introduction of covariates that change across groups implies that the intraclass correlation coefficient is not constant across groups.

When overdispersion is nonexistent the Dirichlet multinomial (beta binomial) collapses to the multinomial (binomial) distribution. This allows for an easy way to test for overdispersion by means of a likelihood-ratio test comparing the log likelihoods of both distributions.

3 Estimation of the beta binomial

3.1 The simple case

As an example of using `xtmnbreg` for fitting the beta-binomial model, I use the data presented in the numerical example shown in Williams (1975). The data consist of the results of an experiment comparing the proportion of pups alive in two equal-sized sets, each comprising 16 pregnant female rats. Pups from the same litter may be equally affected by unobserved factors in which case the survival outcomes of the different pups will be correlated leading to overdispersed data. The first set (the control set) was fed a control diet during pregnancy and lactation, while the second set was treated with a chemical. The data, reproduced from Williams, show for each litter the number of pups that survived 4 days \( x \) and the number of pups that survived the 21-day lactation period \( n \). The data are shown as they are in Williams (1975) in the format \( x/n \):
Beta-binomial model

Control: 13/13, 12/12, 9/9, 8/8, 8/8, 12/13, 11/12, 9/10, 8/9, 11/13, 4/5, 5/7, 7/10, 7/10.

Treated: 12/12, 11/11, 10/10, 9/9, 10/11, 9/10, 9/10, 8/9, 8/9, 4/5, 7/9, 4/7, 5/10, 3/6, 3/10, 0/7.

You could fit a binomial model to each set and then test for differences between the proportion of pups alive in both sets. However, overdispersion invalidates this statistical comparison. To obviate this, the author proposes estimation of the beta-binomial model and provides parameter estimates for each treatment set and also for the entire sample. Our objective is to show how these results could be replicated using the \texttt{xtnbreg} command. In order to estimate the parameters of the beta-binomial model using \texttt{xtnbreg}, you need to structure the data in a particular way. A (partial) listing of these data is shown here:

\begin{verbatim}
use williams
(Data from Williams (1975) paper)
list

<table>
<thead>
<tr>
<th>class</th>
<th>trt</th>
<th>group</th>
<th>y</th>
<th>class1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>
(\textit{output omitted})
| 59    | 1   | 2     | 30  | 3      |
| 60    | 2   | 2     | 30  | 3      |
| 61    | 1   | 2     | 31  | 1      |
| 62    | 2   | 2     | 31  | 7      |
| 63    | 1   | 2     | 32  | 0      |
| 64    | 2   | 2     | 32  | 7      |
\end{verbatim}

The variable \texttt{class} identifies the two possible outcomes for the pups: 1 is survived, and 2 is died. The variable \texttt{trt} is 1 for control and 2 for the treated set. In addition, \texttt{group} is a group identifier (litter in this case), \texttt{y} is the count for the number of pups for each group in each class (number of successes and number of failures), and \texttt{class1} is an indicator variable that equals 1 for class 1 and 0 otherwise. I now fit the beta-binomial distribution to the treatment group:

\begin{verbatim}(Continued on next page)\end{verbatim}
. keep if trt==2
(32 observations deleted)
. xtnbreg y class1, i(group) fe nolog

Conditional FE negative binomial regression
Number of obs = 32
Group variable (i): group
Number of groups = 16
Obs per group: min = 2
avg = 2.0
max = 2
Wald chi2(1) = 8.36
Prob > chi2 = 0.0038

|         | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|---------|--------|-----------|-------|------|---------------------|
| class1  | 1.046004 | .3618702  | 2.89  | 0.004 | .3367511 - 1.755256 |
| _cons   | -.5815185 | .4784722  | -1.22 | 0.224 | -.1519307  .3562699 |

Williams (1975) reports his results in terms of the expected proportion, $E(p_i) = \mu = a/(a + b)$, and a coefficient $\theta = 1/(a + b)$ that is related to the intraclass correlation coefficient. Estimates for these parameters can be readily obtained by doing the following:

. local a exp(_b[_cons]+_b[class1])
. local b exp(_b[_cons])
. nlcom mu: 'a'/(a + b)

|         | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|---------|--------|-----------|-------|------|---------------------|
| mu      | .7400067 | .0696226  | 10.63 | 0.000 | .6035489  .8764646 |

. nlcom theta: 1/(a + b)

|         | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|---------|--------|-----------|-------|------|---------------------|
| theta   | .4650636 | .2406768  | 1.93  | 0.053 | -.0066542  .9367814 |

The results are identical to those reported in Williams (1975) with the exception of the standard error of $\theta$, which is slightly larger. The value of the log likelihood is different, but this is due to the fact that Williams (1975) did not add the constant terms to the log likelihood.

3.2 Testing for overdispersion

As mentioned earlier, the beta-binomial model is a generalization of the binomial model that allows for overdispersion. As the intraclass correlation coefficient tends to zero, the beta-binomial distribution collapses to the binomial distribution. Thus a test for overdispersion can be easily constructed by means of a likelihood-ratio test comparing
Beta-binomial model

the likelihood of the binomial distribution with that of the beta binomial. To obtain
the log likelihood of the binomial model, we use the `binreg` command. The following
code performs the likelihood-ratio test:

```
. local ll_bb=e(ll) /* log-lik from beta-binomial*/
. by group: egen totaln=sum(y)
. quietly binreg y if class==1, n(totaln) ml nolog
. local ll_bin=e(ll) /* log-lik from binomial*/
. di "Likelihood Ratio --> " 2*('ll_bb'-'ll_bin')
Likelihood Ratio --> 25.559216
```

The value for the likelihood-ratio test is the same as in Williams (1975). The associ-
ated p-value is close to zero, a clear indication that the data for the treatment group are
overdispersed. It should be noted that the null hypothesis for this test is in the boundary
of the parameter space, and thus the limiting distribution of the likelihood-ratio statistic
is a 50:50 mixture of a point mass at zero and a \( \chi^2_1 \). This means that the correct p-
value is one-half that which is obtained from the \( \chi^2_1 \) (see Gutierrez, Carter, and Drukker

3.3 Introduction of covariates

To test for the existence of a difference between the proportion of pups alive in the two
groups, Williams (1975) compares the log likelihood of the model fit with the entire
dataset against the sum of the log likelihoods obtained from fitting the beta binomial
to each individual group. An alternative approach to test for the difference between the
two groups is to fit a beta-binomial model to the full dataset, introducing a covariate
that measures the effect of the treatment. As discussed earlier, we link the covariate to
the \( a \) parameter of the beta-binomial distribution:

```
. use williams, clear
   (Data from Williams (1975) paper)
. gen trt_cov=trt*(class==1)
. xtnbreg y class1 trt_cov, i(group) fe nolog
```

```
Conditional FE negative binomial regression
Group variable (i): group
Number of obs = 64
Number of groups = 32
Obs per group: min = 2
                  avg = 2.0
                  max = 2
Wald chi2(2) = 43.70
Prob > chi2 = 0.0000

Log likelihood = -54.046101

|       | Coef.  | Std. Err. | z     | P>|z|  |  [95% Conf. Interval] |
|-------|--------|-----------|-------|------|----------------------|
| class1| 3.346722| .8159031  | 4.10  | 0.000| 1.746581 - 4.944862  |
| trt  | -1.161821| .4998502  | -2.32 | 0.020| -2.141509 - .1821328 |
| _cons| -.2026163| .4152369  | -0.49 | 0.626| -1.016466 .6112331   |
```
The p-value associated with the treatment covariate provides statistical evidence of
a difference between the treatments. Moreover, as in the conventional logit model, the
coefficients associated with the covariates may be interpreted as odds ratios. For the
treatment variable, we obtain an odds ratio of \( \exp(-1.1618) = 0.3129 \) showing that
treated pups are 3 times as likely to die as untreated ones.

4 Estimating the parameters of the Dirichlet multinomial

Finally, I deal with the estimation of the parameters of the more general Dirichlet-
multinomial distribution. The approach is a natural generalization of the results shown
in the last section. For this example, I use the dataset shown in Mosimann (1962).
The dataset is too large to replicate here. The data in Mosimann (1962) report 73
independent counts (groups) of the frequency of occurrence of four different types of
pollen grains (pine, fir, oak, and alder). All counts have the same total number of
grains equal to 100. The author fits a multinomial distribution to the data and finds
out that the observed variance of the number of counts for each type of pollen grain
has an observed variance that is larger than that implied by the multinomial model. He
considers as an alternative fitting the Dirichlet-multinomial model. A partial listing of
the data is provided below:

```
. use mosimann62, clear
    (Data From Mosimann (1962) paper)
. list in 1/12

<table>
<thead>
<tr>
<th>class</th>
<th>group</th>
<th>y</th>
<th>class1</th>
<th>class2</th>
<th>class3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>94</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>10</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>75</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>81</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
```

The variable `class` identifies the four possible types of grain, `group` is an identifier
variable, `y` shows the actual counts of grains by type, and the remaining variables are
indicator variables for classes. The objective is to show how you can estimate the
parameters of the Dirichlet-multinomial distribution. This can be done by typing
Although Mosimann (1962) does not provide estimates for the parameters of the Dirichlet-multinomial distribution, estimates of the $E(p_j)$ (see [2]) based on these data can be found in a recent paper by Neerchal and Morel (2005). Transforming our estimates to suit the parametrization presented in that paper, we can verify that the values are practically identical with minor differences in the estimates of the standard errors (see table 5 in Neerchal and Morel [2005]):

```
. local alpha1 exp(_b[_cons]+_b[class1])
. local alpha2 exp(_b[_cons]+_b[class2])
. local alpha3 exp(_b[_cons]+_b[class3])
. local alpha4 exp(_b[_cons])
. nlcom Ep1: 'alpha1'/('alpha1'+alpha2'+alpha3'+alpha4')
   Ep1: exp(_b[_cons]+_b[class1])/(exp(_b[_cons]+_b[class1])+exp(_b[_cons]+_b[class2])+exp(_b[_cons]+_b[class3])+exp(_b[_cons]))
    y | Coef. Std. Err.  z  P>|z|   [95% Conf. Interval]
----------|-----------------------------|-----------------|-----------|-----------------------------|
    class1 |   0.8621148 .0065331 131.96 0.000 .84931 .8749195
    class2 |  -.0164256 .0022181  7.41 0.000 .0120781 .020773
    class3 |   0.088799  .0053233 16.68 0.000 .0783656 .0992324
    _cons  |   .6760107  .1855332  3.64 0.000 .3123723 1.039649
```

```
. local alpha1 exp(_b[_cons]+_b[class1])
. local alpha2 exp(_b[_cons]+_b[class2])
. local alpha3 exp(_b[_cons]+_b[class3])
. local alpha4 exp(_b[_cons])
. nlcom Ep2: 'alpha2'/('alpha1'+alpha2'+alpha3'+alpha4')
   Ep2: exp(_b[_cons]+_b[class2])/(exp(_b[_cons]+_b[class1])+exp(_b[_cons]+_b[class2])+exp(_b[_cons]+_b[class3])+exp(_b[_cons]))
    y | Coef. Std. Err.  z  P>|z|   [95% Conf. Interval]
----------|-----------------------------|-----------------|-----------|-----------------------------|
    class1 |  -.0164256 .0022181  7.41 0.000 .0120781 .020773
    class2 |   0.8621148 .0065331 131.96 0.000 .84931 .8749195
    class3 |   0.088799  .0053233 16.68 0.000 .0783656 .0992324
    _cons  |   .6760107  .1855332  3.64 0.000 .3123723 1.039649
```

```
. local alpha1 exp(_b[_cons]+_b[class1])
. local alpha2 exp(_b[_cons]+_b[class2])
. local alpha3 exp(_b[_cons]+_b[class3])
. local alpha4 exp(_b[_cons])
. nlcom Ep3: 'alpha3'/('alpha1'+alpha2'+alpha3'+alpha4')
   Ep3: exp(_b[_cons]+_b[class3])/(exp(_b[_cons]+_b[class1])+exp(_b[_cons]+_b[class2])+exp(_b[_cons]+_b[class3])+exp(_b[_cons]))
    y | Coef. Std. Err.  z  P>|z|   [95% Conf. Interval]
----------|-----------------------------|-----------------|-----------|-----------------------------|
    class1 |   0.088799  .0053233 16.68 0.000 .0783656 .0992324
    class2 |  -.0164256 .0022181  7.41 0.000 .0120781 .020773
    class3 |  -.0164256 .0022181  7.41 0.000 .0120781 .020773
    _cons  |   .6760107  .1855332  3.64 0.000 .3123723 1.039649
```
. nlcrom rho: sqrt(1/(1+'alpha1'+'alpha2'+'alpha3'+'alpha4'))
    rho: sqrt(1/(1+exp(_b[_cons]+_b[class1])+exp(_b[_cons]+_b[class2])+ex
    p(_b[_cons]+_b[class3])+exp(_b[_cons])))

| y | Coef. | Std. Err. | z   | P>|z| | [95% Conf. Interval] |
|---|-------|-----------|-----|-----|----------------------|
| rho | 0.1278323 | 0.0111615 | 11.45 | 0.000 | 0.1059562 .1497084 |

Introducing covariates into the Dirichlet-multinomial model follows a logic similar to that used for the beta-binomial model. On the other hand, implementing the likelihood-ratio test for overdispersion is not as straightforward because to obtain the log likelihood under the null hypothesis, the data will have to be rearranged to allow estimation of the multinomial (or conditional) logit model. However, if the number of groups is not very large, there is a simpler solution that does not require rearranging the data. One can take advantage of the multinomial-Poisson transformation (see Guimarães [2004]) to estimate a Poisson regression and then adjust its log likelihood to obtain the one for the multinomial model.

5 Conclusion

The Stata command `xtnbreg` is intended for estimation of count panel data. In this paper, I show that this command can also be used directly to estimate the parameters of the beta-binomial distribution, as well as those of its multivariate generalization, the Dirichlet-multinomial distribution. I also show how to implement a likelihood-ratio test for overdispersion as well as a way by which covariates may be added to these models.

6 References


**About the Author**

Paulo Guimarães is a research assistant professor at the Department of Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina.