<table>
<thead>
<tr>
<th>Title</th>
<th>Bioassay models with natural mortality and random effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Hinde, John</td>
</tr>
<tr>
<td>Publication Date</td>
<td>2011</td>
</tr>
<tr>
<td>Item record</td>
<td><a href="http://hdl.handle.net/10379/3128">http://hdl.handle.net/10379/3128</a></td>
</tr>
</tbody>
</table>
Bioassays models with natural mortality and random effects

Mariana Ragassi Urbano¹, John Hinde², Clarice Garcia Borges Demétrio¹

¹ Departamento de Ciências Exatas, ESALQ/USP, Piracicaba, São Paulo, Brazil, mrurbano@esalq.usp.br, clarice@esalq.usp.br
² School of Mathematics, Statistics and Applied Mathematics, National University of Ireland, Galway, Ireland, john.hinde@nuigalway.ie

Abstract: In fitting dose-response models to entomological data it is often necessary to take account of natural mortality and/or overdispersion. The standard approach to handle natural mortality is to use Abbott’s formula. Standard overdispersion models include beta-binomial models, logistic-normal, and discrete mixtures. Here we consider combining these two aspects with extensions that allow for the modelling of the natural mortality and overdispersion. Two models are developed: one including a random effect in the linear predictor and other including a random effect in the natural mortality. We consider the application of these models to data from an experiment on the use of a virus (PhopGV) for the biological control of worm larvae (Phthorimaea operculella) in potatoes. Using the models with random effects, we obtained a better fit than that provided by the standard model.

Keywords: Bioassay; Natural mortality; Overdispersion; Random effects.

1 Introduction

Models for binary and binomial response grew out of the needs of a type of experimental investigation known as bioassay. In a typical bioassay, different concentrations of a chemical compound are applied to batches of experimental subjects and the number of subjects in each batch that respond to the chemical is then recorded. These values are regarded as observations on a binomial response variable. Some experiments in entomology exhibit evidence that responses can occur even at zero dose; here the response of interest is death and this phenomenon is referred to as natural mortality. Also the variation of the data may be greater than that predicted by the model, commonly described as overdispersion.

1.1 Natural mortality and overdispersion

Among the available methods for the analysis of data with natural mortality, only a few can also handle overdispersion. According to Collet (2002), this
additional variation can be attributed to relevant explanatory variables that have not been adequately measured or controlled. This situation can be modeled by the inclusion of a random effect and so a mixed model can be used in modelling overdispersion. We modified the usual model, first proposed by Abbott (1925), and considered two other models: one with a random effect in the linear predictor, and another in which the natural mortality was taken to be random.

1.2 Description of the dataset

The application here is to an experiment in which potatoes (*Solanum tuberosum* L.) were each infected with $m_{ij} = 30$ larvae of *Phthorimaea operculella*, and then, different concentrations of a virus (*PhopGV*) $i = 0, ..., D$ were applied to samples of $j = 1, ..., n_i$ potatoes. There was also a control sample (no virus) with $n_i = 9$ potatoes. The experiment was conducted at 18°C, and after 60 days the numbers of dead larvae $Y_{ij}$ were counted.

2 Methodology

In modeling the observed proportions $y_{ij}/m_{ij}$, the $y_{ij}$ can be assumed to have a $B(m_{ij}, \pi_{ij}^*)$ distribution, where $\pi_{ij}^*$ the probability of response depends on the natural mortality and the dose-response relationship. A model for $\pi_{ij}^*$ (Morgan,1992) is therefore

$$\pi_{ij}^* = \omega_{ij} + (1 - \omega_{ij})\pi_{ij}, \quad j = 1, ..., n_i \quad \text{and} \quad i = 0, ..., D \quad (1)$$

where $\pi_{ij}$ is a cumulative distribution function (the normal, logistic or extreme value), it is, the link function for the matrix of explanatory variables and $\omega_{ij}$ is the natural response probability. In general, we can model $\pi_{ij}$ and $\omega_{ij}$ as function of covariates and parameters defining three different models:

(a) Standard model

$$\log \left( \frac{\omega_{ij}}{1 - \omega_{ij}} \right) = G\gamma \quad \text{and} \quad \log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = X\beta,$$

(b) Random effect in the linear predictor

$$\log \left( \frac{\omega_{ij}}{1 - \omega_{ij}} \right) = G\gamma \quad \text{and} \quad \log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = X\beta + \sigma z,$$

(c) Random effect in natural mortality

$$\log \left( \frac{\omega_{ij}}{1 - \omega_{ij}} \right) = G\gamma + \tau v \quad \text{and} \quad \log \left( \frac{\pi_{ij}}{1 - \pi_{ij}} \right) = X\beta,$$
where $z$ and $v$ are random effects with standard normal distribution.
The approach used to estimate the parameters was the EM algorithm
(Dempster et al., 1977), as also used in bioassays with natural mortality
by Hasselblad (1980).
If were possible to label the subjects who responded due to the applied dose
as $y_{ijd}$ and those who responded naturally as $y_{ijc}$ then the total number of
death at dose $d_i$ would be

$$y_{ij} = y_{ijc} + y_{ijd},$$

and in the control group, $y_{0j}$ are the number of larvae that died of a total
of $m_{0j}$ that did not receive the virus.

The log likelihood of Model (a) is given by

$$l(\gamma, \beta; y) \propto \sum_{i=1}^{D} \sum_{j=1}^{m_i} \left\{ (m_{ij} - y_{ij}) \left( \log \left( \frac{1}{1 + e^{x^\beta}} \right) + y_{ijd} \log \left( \frac{e^{x^\beta}}{1 + e^{x^\beta}} \right) \right) + (m_{ij} - y_{ij}) \log \left( \frac{e^{G^\gamma}}{1 + e^{G^\gamma}} \right) + y_{ijd} \log \left( \frac{1}{1 + e^{G^\gamma}} \right) \right\} + \sum_{i=0}^{m_{0j}} \sum_{j=1}^{m_{ij}} y_{0j} \log \left( \frac{e^{G^\gamma}}{1 + e^{G^\gamma}} \right) + (m_{ij} - y_{0j}) \log \left( \frac{1}{1 + e^{G^\gamma}} \right)$$

This log-likelihood is easy to maximize, because $l(\beta; y) + l(\gamma; y)$ can be
maximized separately. With the EM algorithm, the incomplete log-likelihood
(2) is maximized iteratively by alternating between estimating $y_{ijc}$ by its
expectation under the current estimates of $\gamma$ and $\beta$ (E step) and then,
with the $y_{ijc}$s fixed at their expected values from the E step, maximizing
$L(\gamma, \beta; y)$ (M-step), the same procedure used by Todem et al. (2010) and in
other similar mixture models. The $(k + 1)^{th}$ iteration of the EM algorithm
for Model (a) requires three steps:

**E - Step**: Estimate $E(y_{ijc}|y_{ij})$ under the current estimates $\gamma^{(k)}$ and $\beta^{(k)}$

$$E(y_{ijc}|y_{ij})^{(k)} = \begin{cases} 
  y_{0j} & \text{for } d_{0j}; \\
  \frac{e^{G^\gamma y_{0j}}}{e^{G^\gamma} + x^\beta + x^\beta} & \text{for } d_{ij}.
\end{cases}$$

**M - Step for $\beta$**: Find $\beta^{(k+1)}$ by maximizing $l(\beta; y_{ijc}|y_{ij})$, $\beta^{(k+1)}$ can be
found from a weighted binomial regression, with is unit for the control
group and for the experimental group is $E(y_{ijc}|y_{ij})^{(k)}$;

**M - Step for $\gamma$**: Find $\gamma^{(k+1)}$ by maximizing $l(\gamma; y_{ijc}|y_{ij})$, and $\gamma^{(k+1)}$ can be
found from an unweighted binomial logistic regression of the responses
$y_{0j}$ and $E(y_{ijc}|y_{ij})^{(k)}$ with binomial denominators $m_{0j}$ and $m_{ij}$ respectively
on design matrix $G$. 
These three steps must be repeated until the convergence be reached, but in practice 10 iterations are sufficient. Let $\psi = (\gamma, \beta, \sigma)$ be the combined parameter vector. The likelihood of Model (b) is given by

$$L(\psi; y) = \prod_{i=0}^{D} \left\{ \int_{-\infty}^{+\infty} \left[ \prod_{j=1}^{n_i} P(y_{ij}|\psi) \right] \phi(z_j)dz_j \right\}. \quad (3)$$

The integral in the likelihood (3) does not have a closed form except for $Y$ normal, and so for other response models it is approximated by a Gaussian quadrature: the integral is replaced over the normal $Z_j$ by the finite sum over $K$ Gaussian quadrature mass points $z_k$ with masses $\alpha_k$ (Aitkin et al. 2009). The likelihood is then

$$L(\psi; y) = \prod_{i=0}^{D} \left\{ \sum_{k=1}^{K} \left[ \prod_{j=1}^{n_i} P(y_{ij}|\psi) \right] \alpha_k \right\},$$

where $P(y_{ij}|\psi) = \left( \frac{m_{ij}}{y_{ij}} \right)^{\pi_{ij}} \left( 1 - \frac{\pi_{ij}}{y_{ij}} \right)^{m_{ij} - y_{ij}}$. The likelihood is thus (approximately) the likelihood of a finite mixture of exponential families density with known mixture proportions $\alpha_k$ at known mass-points $z_k$, thus $z_k$ becomes another observable variable in the regression, with regression coefficient $\sigma_A$.

The log-likelihood is

$$l(\psi; y) = \sum_{i=0}^{D} \log \left( \sum_{k=1}^{K} \alpha_k \rho_{ik} \right),$$

with $\rho_{ik} = \prod_{j=1}^{n_i} P(y_{ij}|\psi)$. Then

$$\frac{\partial l}{\partial \beta} = \sum_{i=0}^{D} \sum_{k=1}^{K} \frac{\partial \log \rho_{ik}}{\partial \beta} \alpha_k \rho_{ik} = \sum_{i=0}^{D} \sum_{j=1}^{n_i} \sum_{k=1}^{K} w_{ik} s_{ijk}(\beta),$$

where $w_{ik}$ is the posterior probability that observation $y_{ij}$ comes from component $k$, $w_{ik} = \frac{\alpha_k \rho_{ik}}{\sum_{l=1}^{K} \alpha_l \rho_{il}}$ and $s_{ijk}(\beta)$ is the $\beta$-component of the score for observation $(ij)$ in component $k$:

$$s_{ijk}(\beta) = \frac{(y_{ij} - \mu_{ijk})x_{ij}}{\left( \frac{m_i - \mu_i}{m_i} \right) g'_{ijk}}.$$

Equating to zero gives likelihood equations which are simple weighted sums of those for an ordinary GLM with weights $w_{ik}$; alternately solving these equations for given weights $w_{ik}$, and updating these weights from the current parameters in an EM algorithm.

For model (b), the steps of the EM algorithm are the following
E – Step: Estimate E(y_{ijc}) under the current estimates $\gamma^{(k)}$, $\beta^{(k)}$ and $\sigma^{(k)}$

$$E(y_{ijc}|y_j)^{(k)} = \begin{cases} y_{0j} & \text{for } d_{0j}; \\ \frac{e^{\alpha \gamma_{y_{ij}}}}{e^{\alpha \gamma_{y_{ij}}} + x_{ijc} \beta \sigma_{y_{ij}}} & \text{for } d_{ij}. \end{cases}$$

M – Step for $\beta$ and $\sigma$: Find $\beta^{(k+1)}$ and $\sigma^{(k+1)}$ by maximizing $l(\beta, \sigma; y_{ijc}|y_j)$, and $\beta^{(k+1)}$ and $\sigma^{(k+1)}$ can be found from a weighted binomial regression, with weights $w_{ik}$ for the control group and for the experimental group $E(y_{ijc}|y_j)^{(k)} w_{ik}$;

M – Step for $\gamma$: Find $\gamma^{(k+1)}$ by maximizing $l(\gamma; y_{ijc}|y_j)$, and $\gamma^{(k+1)}$ can be found from a unweighted binomial logistic regression of $E(y_{ijc}|y_j)^{(k)}$ with binomial denominator $m_{ij}$ on design matrix $G$.

For model (c), the steps of the EM algorithm are the following:

E – Step: Estimate $E(y_{ijc})$ under the current estimates $\gamma^{(k)}$, $\beta^{(k)}$ and $\sigma^{(k)}$

$$E(y_{ijc}|y_j)^{(k)} = \begin{cases} y_{0j} & \text{for } d_{0j}; \\ \frac{e^{\alpha \gamma_{y_{ij}}}}{e^{\alpha \gamma_{y_{ij}}} + x_{ijc} \beta} & \text{for } d_{ij}. \end{cases}$$

M – Step for $\beta$: Find $\beta^{(k+1)}$ by maximizing $L(\beta; y_{ijc}|y_j)$, and $\beta^{(k+1)}$ can be found from a weighted binomial regression, with is unit for the control group and for the experimental group is $E(y_{ijc}|y_j)^{(k)} w_{ik}$;

M – Step for $\gamma$ and $\tau$: Find $\gamma^{(k+1)}$ and $\tau^{(k+1)}$ by maximizing $l(\gamma, \tau; y_{ijc}|y_j)$, and $\gamma^{(k+1)}$ and $\tau^{(k+1)}$ can be found from a weighted binomial logistic regression of $E(y_{ijc}|y_j)^{(k)}$ with binomial denominator $m_{ij}$ on design matrix $G$ and weights $w_{ik}$.

In both models (b) and (c) were used 10 quadrature points, and the procedures were implemented in the R package.

3 Main Results and Conclusions

We included in the standard model for natural mortality random effects, with the aim to provide a better fit when the dataset exhibits overdispersion. We concluded that data from biological assays that present natural mortality and overdispersion can be more realistically modelled when a random effect is included to account for variability in the larvae and their response to the virus.

Table 1 presents the fit statistics (-2 Log Likelihood, AIC, and BIC) for models (a), (b) and (c).

For these three statistics, the smaller the value the better is the fit.
Bioassays models with natural mortality and random effects

TABLE 1. Fit Statistics: $-2$ Log Likelihood, AIC, and BIC for models (a), (b) and (c)

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th>Model (a)</th>
<th>Model (b)</th>
<th>Model (c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-2$ Log Likelihood</td>
<td>255.83</td>
<td>184.3058</td>
<td>208.41</td>
</tr>
<tr>
<td>AIC</td>
<td>249.83</td>
<td>192.3058</td>
<td>216.41</td>
</tr>
<tr>
<td>BIC</td>
<td>261.00</td>
<td>191.2029</td>
<td>215.31</td>
</tr>
</tbody>
</table>

Can conclude that the model with random effect in the linear predictor in the potato level provides a better fit than the model without the random effect in the linear predictor.

The equation for the fitted model is given by

\[
\hat{\pi}_i = 0.29 + (0.71)\frac{e^{-7.44+1.63\log(d_i)+0.91z_i}}{1+e^{-7.44+1.63\log(d_i)+0.91z_i}}.
\]

Acknowledgments: This work was supported by CAPES - Proc nº 4942/10-8 and Science Foundation Ireland award 07/MI/012.

References


