A random effects continuation-ratio model for replicated toxicological data

Martinez, Marie-José
Team MISTIS, INRIA Rhône-Alpes & Laboratoire Jean Kuntzmann
655, Avenue de l'Europe - Montbonnot
38334 Saint-Ismier, FRANCE
E-mail: marie-jose.martinez@iut2.upmf-grenoble.fr

Hinde, John
School of Mathematics, Statistics and Applied Mathematics
National University of Ireland, Galway
University Road
Galway, IRELAND
E-mail: john.hinde@nuigalway.ie

Discrete survival times can be considered as ordered multicategorical data. In the ordinal data modelling context, a variety of multinomial regression models can be used including the baseline-category logit model, the cumulative logit model, the adjacent-category logit model or the continuation-ratio logit model. This last model has been given some attention in the literature (Agresti, 2002). Such a model form is useful when the ordered categories represent a progression through different stages, such as survival through various times. This particular model has the advantage of being a simple decomposition of a multinomial distribution as a succession of hierarchical binomial models. The property of conditional independence enables to fit it by adapting the methods available for binary response data. When one have ordered replicated data, random effects can be incorporated into the linear predictor to account for uncontrolled experimental variation. An increasing number of papers are concerned with random effects models for ordered categorical responses (Ten Have and Uttal, 1994; Tutz and Hennevogl, 1996).

In this paper, we focus on random effects continuation-ratio models. We consider a continuation-ratio model and we include a random intercept into the linear predictor in order to analyse grouped toxicological data. More precisely, the data considered here have been obtained from a biological control essay realized by the Insect Pathology Laboratory of ESALQ-USP, Sao Paulo, Brazil (De Freitas, 2001). In this essay, different isolates of the fungus *Beauveria bassiana* are used as a microbial control for the *Heterotermes tenuis* termite which causes a lot of damage in sugarcane fields in Brazil. In this context, experiments have been carried out to study the pathogenicity and the virulence of the fungus in order to determine effective isolates for the control of this pest population. The obtained data set compares 142 isolates of the fungus. A solution of each isolate is applied to 5 groups of 30 termites and the cumulative mortality in each group is measured daily during an 8-day period after the application of the fungus. A simple graphical representation of the cumulative proportions of dead termites shows different isolate efficacities and different degrees of variability among the replicates within the different isolates. Thus, the aim of this study is to determine effective isolates for use in the field by taking into account the replicated data structure.

Model specification

Suppose the cumulative mortality is measured over D consecutive days. For replicate k of isolate i, $k = 1, \ldots, K$ and $i = 1, \ldots, I$, we denote n_{ik} the initial number of insects. Let Y_{jik} denotes the number of dead insects on day j, $j = 1, \ldots, D$ and $Y_{D+1ik} = n_{ik} - \sum_{j=1}^{D} Y_{jik}$ the number of insects

still alive on day D. The probability of an insect dying on day j for isolate i and replicate k is denoted by π_{jik}. For each replicate, we treat the counts in the $D+1$ categories, $Y_{ik} = (Y_{1ik}, \ldots, Y_{D+1ik})$, as multinomial with probabilities $(\pi_{1ik}, \ldots, \pi_{D+1ik})$ where $\sum_{j=1}^{D+1} \pi_{jik} = 1$.

Consider now w_{jik} the conditional probability that an insect dies on day j given that it has survived up to this day for isolate i and replicate k. This conditional probability is defined by

$$w_{jik} = \frac{\pi_{jik}}{\sum_{j'=j}^{D+1} \pi_{j'ik}}.$$

Let $b(n; y; w)$ denote the binomial probability of obtaining y successes out of n trials with probability w for each trial. The multinomial probability of $p(y_{1ik}, \ldots, y_{D+1ik})$ can be easily expressed in the form

$$b(n_{ik}; y_{1ik}; w_{1ik}) \times b(n_{ik} - y_{1ik}; y_{2ik}; w_{2ik}) \times \ldots \times b(n_{ik} - y_{1ik} - ... - y_{D-1ik}; y_{Dik}; w_{Dik}).$$

Thus, the multinomial model can be expressed as a succession of hierarchical binomial models. The continuation-ratio logits are then defined as

$$\eta_{jik} = \text{logit}(w_{jik}) = \log\Big(\frac{w_{jik}}{1 - w_{jik}}\Big) = \log\left(\frac{\pi_{jik}}{\pi_{j+1ik} + \ldots + \pi_{D+1ik}}\right),$$

and are ordinary logits of the conditional probabilities w_{jik}.

Clearly, a main advantage of the continuation-ratio model is that it can be fitted using methods for binomial logit models merely by a rearrangement of the data. Thus, to fit the different logit models, we require a derived data structure as follows in order to relate the different models.

***Rearrangement of the data for replicate** k **of isolate** i*

Day	No. at risk	No. of deaths	Proba. of death
1	n_{ik}	y_{1ik}	π_{1ik}
2	$n_{ik} - y_{1ik}$	y_{2ik}	$w_{2ik} = \frac{\pi_{2ik}}{1-\pi_{1ik}}$
3	$n_{ik} - y_{1ik} - y_{2ik}$	y_{3ik}	$w_{3ik} = \frac{\pi_{3ik}}{1-\pi_{1ik}-\pi_{2ik}}$
$\vdots$			
D	$n_{ik} - y_{1ik} - \ldots - y_{D-1ik}$	y_{Dik}	$w_{Dik} = \frac{\pi_{Dik}}{1-\sum_{j'=1}^{D-1}\pi_{j'ik}}$

The linear predictor η_{jik} may contains isolate specific factors and covariates in order to model the time dependency. In this work, we consider isolate and time specific linear effects. In addition, the variability observed among the replicates for some isolates leads us to introduce an additive random effect into the linear predictor.
For $j = 1, \ldots, D$, $i = 1, \ldots, I$ and $k = 1, \ldots, K$, we first consider Model I defined by the linear predictor

$$\text{Model I:} \qquad \eta_{jik} = \alpha_i + \beta_j + \sigma\xi_{ik},$$

where α_i is the isolate effect of isolate i, β_j the time effect of day j, $\xi_{ik} \sim \mathcal{N}(0, 1)$ and the ξ_{ik}'s are assumed independent.

Secondly, we consider a second model in which we add a time trend to the previous linear predictor leading to Model II defined by

$$\text{Model II:} \qquad \eta_{jik} = \alpha_i + \beta_j + \gamma_i t_j + \sigma\xi_{ik},$$

where α_i is the baseline effect of isolate i, β_j is the baseline time effect of day j, $t_j = j$ is a quantitative variable for day j and γ_i is the time effect on isolate i.

A third model is also considered. This model has an isolate specific linear time effect defined by the following linear predictor:

$$\text{Model III:} \qquad \eta_{jik} = \alpha_i + \gamma_i t_j + \sigma\xi_{ik}.$$

Model III imposes more structure on the linear predictor than Model II. By omitting the coefficient β_j, it implies a more specific form for the responses over time.

Finally, two additional models will be considered in this paper and compared to the three models defined above. In order to get anywhere near reproducing the general overall pattern, we add a quadratic term in time to the two previous linear predictors. Note that all coefficients α_i, β_j, γ_i and δ_i, $i = 1, \ldots, I$ and $j = 1, \ldots, D$, are assumed to be constant over replicates. These models are random intercept models in which the introduction of an additive random effect allows a random location shift for each replicate of each isolate. In this work, we consider a logit link leading to random effects continuation-ratio logit models. Others link functions can be used. Another common choice is the complementary log-log link yielding the so-called proportional hazards model.

Parameter estimation

In this section, parameter estimation for the random effects continuation-ratio models defined previously is considered based on the EM algorithm. This algorithm is a powerful computational technique for maximizing likelihoods including unobserved variables. However, as with the binary model, the non-conjugate normal distribution for ξ means that the marginal likelihood cannot be worked out analytically. Indeed, assuming that $\varphi(.)$ denotes the normal density function, the likelihood of replicate k of isolate i is given by

$$\begin{aligned}
L_{ik}(\theta, \sigma) &= \int_{-\infty}^{+\infty} \prod_{j=1}^{D} f(y_{jik}|\theta, \sigma, \xi_{ik})\ \varphi(\xi_{ik}; 0, 1)\ d\xi_{ik} \\
&= \int_{-\infty}^{+\infty} \prod_{j=1}^{D} w_{jik}^{y_{jik}}\ (1 - w_{jik})^{n_{ik} - \sum_{j'=1}^{j-1} y_{j'ik}}\ \varphi(\xi_{ik}; 0, 1)\ d\xi_{ik} \\
&= \int_{-\infty}^{+\infty} \prod_{j=1}^{D} \left[\frac{\exp(\eta_{jik})}{1 + \exp(\eta_{jik})}\right]^{y_{jik}} \\
&\qquad \times \left[\frac{1}{1 + \exp(\eta_{jik})}\right]^{n_{ik} - \sum_{j'=1}^{j-1} y_{j'ik}} \varphi(\xi_{ik}; 0, 1)\ d\xi_{ik},
\end{aligned}$$

Clearly, this likelihood function has no closed form and has to be evaluated numerically before being maximized as a function of the fixed effect parameters θ and the random effect parameter σ. In this section, we consider two integration methods for approximating the likelihood which will be then combined with an EM algorithm for the maximization step.

First, we consider classical Gaussian quadrature to evaluate numerically this likelihood integral. The dimension of the integral determining the likelihood function depends on the random effect structure.

When the random effects are assumed normally distributed and the dimension is small, as in the integral defined above, Gaussian-Hermite quadrature methods can approximate the likelihood function. Thus, the likelihood is approximated by

$$L_{ik}(\theta,\sigma) \approx \sum_{r=1}^{R} \pi_r \left\{ \prod_{j=1}^{D} f(y_{jik}|\theta,\sigma,z_r) \right\},$$

with weights π_r and quadrature points z_r that are tabulated. Note that the approximation improves as the number R of quadrature points increases. However, in practice, a large number of quadrature points is often required to approximate correctly the likelihood. Moreover, the approximation can be poor for large random effects variances or can fail for small cluster sizes (Lesaffre and Spiessens, 2001).

To solve these problems associated with ordinary quadrature, we then consider adaptive Gaussian quadrature methods. An adaptive version of the Gauss-Hermite quadrature shifts and scales the quadrature points to place them under the peak of integrand. Note that after normalization with respect to ξ_{ik}, the integrand is the posterior density of ξ_{ik} given the response and can be approximated for large sample sizes by a normal density $\varphi(\xi_{ik};\mu_{ik},\tau_{ik}^2)$ with mean μ_{ik} and variance τ_{ik}^2. In this version, the normal density $\varphi(\xi_{ik};\mu_{ik},\tau_{ik}^2)$ approximating the posterior density is treated as the weight function. The integral is now written as

$$L_{ik}(\theta,\sigma) = \int_{-\infty}^{+\infty} \varphi(\xi_{ik};\mu_{ik},\tau_{ik}^2)\, \frac{\prod_{j=1}^{D} f(y_{jik}|\theta,\sigma,\xi_{ik})\, \varphi(\xi_{ik};0,1)}{\varphi(\xi_{ik};\mu_{ik},\tau_{ik}^2)}\, d\xi_{ik},$$

and applying the standard quadrature rules, the integral is now approximated by

$$L_{ik}(\theta,\sigma) \approx \sum_{r=1}^{R} \pi_{ik\,r} \left\{ \prod_{j=1}^{D} f(y_{jik}|\theta,\sigma,z_{ik\,r}) \right\}.$$

Hence, the adaptive quadrature points are given by $z_{ik\,r} = \tau_{ik} z_r + \mu_{ik}$ with corresponding weights $\pi_{ik\,r} = \sqrt{2\pi}\tau_{ik} \exp(\frac{z_r^2}{2})\varphi(z_{ik\,r})\pi_r$.

Essentially, the posterior density is here approximated by a normal density with the same mean and variance. However, the posterior mean and variance required in this approach are not known and have to be computed. As in Rabe-Hesketh and al. (2005), we obtain these posterior moments using adaptive quadrature leading to an iterative integration. Finally, once the marginal likelihood is evaluated numerically for given parameter values, it has to be maximised with respect to θ and σ. Several methods for maximizing the likelihood can be considered and combined with the two integration methods presented above. Rabe-Hesketh et al. (2005) use for instance a Newton-Raphson algorithm where the Hessian matrix is obtained by numerical differentiation. In this paper, we consider an EM-algorithm which is easy to implement compared to other optimization methods.

Results

The different models presented previously are fitted to the data and compared. More precisely, the models are fitted using ordinary Gaussian quadrature with 3, 5, 10, 20, 40 and 60 quadrature points and adaptive Gaussian quadrature using 3, 5 and 10 quadrature points. For each of these models, we also fit the associated fixed model. For simplicity, we only consider a subset of 30 isolates. Obviously, results for all 142 isolates can be obtained in the same way.

The results obtained by adaptive Gaussian quadrature are the same from using 10 quadrature points. For as few as 3 quadrature points, only very small differences can be observed. On the other hand, the

results, in particular the variance estimates, change considerably using ordinary Gaussian quadrature. Clearly, we need to increase the number of quadrature points to 40 for ordinary quadrature in order to get similar results. Therefore, it is clear that we may be able to achieve good accurary with a smaller number of quadrature points using adaptive Gaussian quadrature instead of ordinary Gaussian quadrature for the different models. Note that simple GLMs were fitted first to the toxicological data. Introducing a random effect into the linear predictors of these models improved these results dramatically. Thus, the large variability in the data has been captured by these random effects models.

For each isolate, we also determine the fitted replicate-specific evolutions and the marginal average evolution implied by the different models. The posterior quantities of interest are the random effects and the corresponding model random linear predictors. One nice feature of using numerical integration via the EM-algorithm is that we can easily calculate these quantities from the estimated posterior distribution of the random effects. For example, using ordinary Gaussian quadrature methods, the posterior distribution of ξ_{ik} is provided by

$$\begin{aligned} f(z_r|y_{ik}) &= \frac{\pi_r \prod_{j=1}^{D} f(y_{jik}|\theta,\sigma,z_r)}{\sum_{l=1}^{R}\pi_l \prod_{j=1}^{D} f(y_{jik}|\theta,\sigma,z_l)}, \qquad r=1,\ldots,R. \\ &= p_{ikr} \end{aligned}$$

These posterior probabilities p_{ikr} that the unobserved ξ_{ik} takes the value z_r correspond to the weights at the final iteration of the EM-algorithm and they provide the posterior distribution of the ξ_{ik} in the empirical Bayes sense by replacing the unknown parameters by their ML estimates. In Model II, for instance, the linear predictors are defined as

$$\log\left(\frac{w_{jik}}{1-w_{jik}}\right)|\xi_{ik} = \eta_{jik} = \alpha_i + \beta_j + \gamma_i t_j + \sigma\xi_{ik},$$

and the corresponding means as $w_{jik} = \dfrac{\exp(\eta_{jik})}{1+\exp(\eta_{jik})}$. In this case, the empirical Bayes predictions are calculated by:

$$\begin{aligned} \hat{p}_{ikr} &= \frac{\pi_r \prod_{j=1}^{D} f(y_{jik}|\hat{\theta},\hat{\sigma},z_r)}{\sum_{l=1}^{R}\pi_l \prod_{j=1}^{D} f(y_{jik}|\hat{\theta},\hat{\sigma},z_l)}, \\ \tilde{\xi}_{ik} &= \sum_{r=1}^{R} \hat{p}_{ikr}\, z_r, \\ \hat{\eta}_{jik} &= \sum_{r=1}^{R} \hat{p}_{ikr}\, \hat{\eta}_{jikr} \quad \text{with} \quad \hat{\eta}_{jikr} = \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_i t_j + \hat{\sigma} z_r, \\ \hat{w}_{jik} &= \frac{\exp(\hat{\eta}_{jik})}{1+\exp(\hat{\eta}_{jik})}. \end{aligned}$$

Note that a similar approach is used when using adaptive Gaussian quadrature. Finally, the fitted probabilities of real interest $\hat{\pi}_{jik}$ are directly obtained from the empirical Bayes predictions $\hat{w}_{jik}$.

Concerning the marginal average evolution, note that it can be derived from averaging the conditional means over the random effects ξ_{ik}. Again, this can be done using numerical integration methods or based on numerical averaging by sampling a large number of random effects from their fitted distribution (Molenberghs and Verbeke, 2005). In this work, we derive the marginal average evolution based on the second method by sampling 1000 random effects ξ_{ik} from their fitted distribution. Note that the approach which consists of plotting the profile for an "average" replicate i.e. a replicate with random intercept $\xi_{ik} = 0$ rather than the marginal average results in different fitted average trends.

Finally, one of the aims of this study is to determine the effective isolates. In this context, one quantity usually used is the lethal time LT_p which is the time required to obtain p % mortality. This quantity can be easily used to summarize and to rank the different isolate effectiveness. More precisely, for each isolate, we plot the marginal median lethal time against the standard deviation of the posterior estimates of the random effect to account for variability among the replicates. Clearly, effective isolates are those with both low lethal time and low replicate variability.

Discussion

In this paper, we have proposed to use random effects continuation-ratio models to model discrete survival times by considering them as ordered multicategorical data. We have seen that this particular model can be easily fitted using the methods available for binary response data by a rearrangement of the data. The use of this specific model also makes possible the generalization of these approaches to replicate measures. In this work, the random effects are assumed to be sampled from a normal distribution. This assumption reflects the prior believe that the random effects are drawn from one homogeneous population. However, the results obtained using ordinary Gaussian quadrature show that the disparity does not decrease monotonically as we increase the number of quadrature points. In other words, bad approximations can give better fits. This behaviour observed for instance in Lesaffre and Spiessens (2001) suggests that the normality assumption is not really convincing in this case. To relax this assumption, we are now considering the use of heterogeneity models as defined by Molenberghs and Verbeke (2005). This extension consists of replacing the normality assumption by a mixture of normal distributions. This model which reflects the prior believe of presence of unobserved heterogeneity among the replicates is also used for classification purposes.

REFERENCES

[1] A. Agresti (2002) *Categorical Data Analysis (2nd ed.)*, John Wiley & Sons, New York.
[2] T. R. Ten Have, D. H. Uttal (1994) Subject-specific and population-averaged continuation ratio logit models for multiple discrete time survival profiles. *Applied Statistics*, 43, 371-384.
[3] G. Tutz, W. Hennevogl (1996) Random effects in ordinal regression models. *Computational Statistics and Data Analysis*, 22, 537-557.
[4] S. De Freitas (2001) *Modelos para proporcoes com superdispersao proveniente de ensaios toxicologicos no tempo*, Ph. Thesis, ESALQ, Universidade de Sao Paulo.
[5] E. Lesaffre, B. Spiessens (2001) On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics*, 50, 325-335.
[6] S. Rabe-Hesketh, A. Skrondal, A. Pickles (2005) Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, 128 301-323.
[7] G. Molenberghs, G. Verbeke (2005) *Models for discrete longitudinal data*, Springer, New York.

ABSTRACT

Discrete survival times can be considered as ordered multicategorical data. In this work, we consider a continuation-ratio model, which is particularly appropriate when the ordered categories represent a progression through different stages, such as survival through various times. In a clustered data context, we incorporate random effects into the linear predictor of the model to account for uncontrolled experimental variation. Assuming a normal distribution for the random effects, we use ordinary and adaptive Gaussian quadrature in an EM-algorithm to estimate the model parameters. This approach is used to analyse grouped toxicological data from a biological control essay where different isolates of a fungus are used as a microbial control for termites.