Use of the Probability Integral Transformation to Fit Nonlinear Mixed-Effects Models With Nonnormal Random Effects

Kerrie P. Nelson, Stuart R. Lipsitz, Garrett M. Fitzmaurice, Joseph Ibrahim, Michael Parzen, and Robert Strawderman

This article describes a simple computational method for obtaining the maximum likelihood estimates (MLE) in nonlinear mixed-effects models when the random effects are assumed to have a nonnormal distribution. Many computer programs for fitting nonlinear mixed-effects models, such as PROC NLMIXED in SAS, require that the random effects have a normal distribution. However, there is often interest in either fitting models with nonnormal random effects or assessing the sensitivity of inferences to departures from the normality assumption for the random effects. When the random effects are assumed to have a nonnormal distribution, we show how the probability integral transform can be used, in conjunction with standard statistical software for fitting nonlinear mixed-effects models (e.g., PROC NLMIXED in SAS), to obtain the MLEs. Specifically, the probability integral transform is used to transform a normal random effect to a nonnormal random effect. The method is illustrated using a gamma frailty model for clustered survival data and a beta-binomial model for clustered binary data. Finally, the results of a simulation study, examining the impact of misspecification of the distribution of the random effects, are presented.

Key Words: Beta-binomial distribution; Frailty model; Newton-Raphson algorithm; Quadrature.

1. INTRODUCTION

Nonlinear mixed-effects (NLME) models are widely used in many areas of application.
We use the term nonlinear mixed-effects models in a broad sense to refer to any statistical model in which the random effects have a nonlinear relationship to the outcome variable. Two common examples of such models are frailty models in survival analysis and generalized linear mixed-effects models for cluster-correlated data. In the latter models, discrete correlated responses (e.g., binary responses or count data) are modeled as a nonlinear function of covariates and random effects. Many computer programs for fitting these models, such as PROC NLMIXED in SAS (SAS Institute Inc. 2000), require that the random effects have a normal distribution. However, there is often interest in either fitting models with nonnormal random effects or assessing the sensitivity of inferences to departures from the normality assumption for the random effects.

To motivate the methodology to be described in later sections, consider the following two examples. The first example is based on a subset of data from an Eastern Cooperative Oncology Group liver cancer clinical trial (Falkson, Cnaan, and Simson 1994). This clinical trial compared two therapies for the treatment of advanced liver cancer with respect to survival time (time from randomization until death). In addition to the treatment comparison, it is also of interest to examine the impact of cardiac status (normal, abnormal) at baseline on survival. A feature of these data that complicates their analyses is that they are clustered, with patients nested within hospitals. In general, observations within a cluster tend to be positively correlated due to various shared characteristics. This correlation can be taken into account in the analysis via the introduction of a random cluster (or frailty) effect. For example, conditional on the frailty, we can assume an exponential model for survival time, where the hazard is an exponential function of the covariates (e.g., treatment and baseline cardiac status) and the frailty. Section 4 considers an exponential model for survival time where the frailty is assumed to have either a normal or gamma distribution; the gamma distribution is a somewhat more natural, and perhaps more realistic, choice of distribution for the frailty. Maximization of the marginal likelihood, obtained by integrating over the distribution of the frailties, yields the MLEs. However, because the model is a nonlinear function of the random effects, the marginal likelihood has no closed form, and numerical approximations are required. Such methods include numerical integration techniques (Liu and Pierce 1994), Monte Carlo Markov chain methods (Kuk and Cheng 1997; McCulloch 1997), approximate maximum likelihood (Breslow and Clayton 1993; McGilchrist 1994; Wolfinger and Lin 1997), and others (Zeger and Karim 1991; Lindstrom and Bates 1990).

The second example is from a study that explored the cardiotoxic effects of doxorubicin chemotherapy for the treatment of acute lymphoblastic leukemia in childhood (Lipshultz et al. 1995). Doxorubicin chemotherapy can be harmful to the heart. Even patients who are cured from their cancer can develop heart problems later in life. At the most recent long-term follow-up visit for each “cured” leukemia patient, six tests of heart function were performed; the results of each test are simply coded as normal/abnormal. In this study it is of interest to model the probability that each test is abnormal as a function of doxorubicin dose (high or low), and the time since the completion of doxorubicin chemotherapy. Investigators believed that a higher dose of the doxorubicin chemotherapy would lead to worse heart function and that, over time, the heart function would get worse. Because the six tests of heart function...
are similar, we can assume the probability that each test is abnormal is the same logistic function of both dose and time, and use the beta-binomial model for these data. In the beta-binomial model, we first define a random probability for each subject which has a beta distribution. Conditional on this random probability, the number of abnormal tests for a subject is assumed to follow a binomial distribution. Maximum likelihood estimates of the logistic regression parameters are obtained from the marginal likelihood, integrating over the distribution of the random effects (or random probabilities). For this particular nonlinear mixed-effects model, the marginal likelihood does have a closed form and so we can compare the MLEs obtained by maximizing the closed form marginal likelihood to the MLEs obtained using numerical integration techniques.

2. NONLINEAR MIXED-EFFECTS MODELS

We suppose there are \( N \) independent clusters, with \( n_i \) members in the \( i \)th cluster. Let \( Y_{ik} \) denote the outcome for the \( k \)th member of the \( i \)th cluster, \( i = 1, \ldots, N; k = 1, \ldots, n_i \). Let \( x_{ik} \) be the covariate vector for the \( k \)th member of the \( i \)th cluster, and \( b_i \) a random effect associated with the \( i \)th cluster. The random effects have density function \( f(b_i | \theta) \), with \( \theta \) the vector of parameters for the distribution of \( b_i \). The distribution of \( Y_{ik} \) given \( x_{ik} \) and \( b_i \) has density \( f(y_{ij} | x_{ij}, b_i, \beta) \), indexed by the parameter vector \( \beta \). We allow both \( f(y_{ij} | x_{ij}, b_i, \beta) \) and \( f(b_i | \theta) \) to have any general parametric density. Given the random effects \( b_i \), we assume the \( Y_{ik} \)'s within a cluster are mutually independent. The MLE of \( (\beta, \theta) \) is obtained by maximizing the marginal likelihood. The \( i \)th cluster’s contribution to the marginal likelihood is

\[
L_i(\beta, \theta) = f(y_{i1}, \ldots, y_{in_i} | x_{i1}, \ldots, x_{in_i}, \beta, \theta) = \int_{b_i} \left[ \prod_{k=1}^{n_i} f(y_{ik} | x_{ik}, b_i, \beta) f(b_i | \theta) \right] db_i ,
\]

and thus, the likelihood to be maximized is

\[
L(\beta, \theta) = \prod_{i=1}^{N} \int_{b_i} \left[ \prod_{k=1}^{n_i} f(y_{ik} | x_{ik}, b_i, \beta) f(b_i | \theta) \right] db_i .
\]}

Clearly, (2.2) involves integration over the distribution of \( b_i \) and, in general, does not have a closed form. A number of methods are available for maximizing this likelihood directly (Geyer and Thompson 1992), or via approximations (Pinheiro and Bates 1995; Lindstrom and Bates 1990; Breslow and Clayton 1993; McGilchrist 1994; Liu and Pierce 1994; Wolfinger and Lin 1997). Many of these methods assume \( f(b_i | \theta) \) is normal. Numerical integration techniques, such as Gaussian quadrature (Davidian and Gallant 1992, 1993; Liu and Pierce 1994), are used increasingly for fitting NLME’s. These methods are available, for example, in PROC NLMIXED in SAS, but only for the case of one or more normally distributed random effects. Both of the examples introduced in Section 1, however, require
a single nonnormal random effect. As we will show, nonnormal random effects can be accommodated within the numerical integration techniques available in PROC NLMIXED, via the use of the probability integral transform.

Suppose that the random effects (assumed continuous) have a nonnormal distribution \( f(b_i|\theta) \), but available software programs are restricted to normal random effects. Let \( a_i \) be a random effect from a standard normal distribution, that is, \( a_i \sim \text{normal}(0,1) \). Then, using the probability integral transform (Hoel, Port, Stone 1971), \( u_i = \Phi(a_i) \) has a uniform(0,1) distribution, where \( \Phi(\cdot) \) is the standard normal cumulative distribution function (CDF). Applying the probability integral transform once more, \( F_\theta(b_i) \) also has a uniform(0,1) distribution, where \( F_\theta(\cdot) \) is the CDF of \( b_i \), with parameter \( \theta \). It then follows that \( b_i = F_\theta^{-1}(u_i) \) has density \( f(b_i|\theta) \), where \( F_\theta^{-1}(\cdot) \) is the inverse CDF of \( b_i \). Thus, \( b_i = F_\theta^{-1}(\Phi(a_i)) \) has the nonnormal distribution of interest. Almost all major statistical software packages have the \( \Phi(\cdot) \) function built-in, as well as most common inverse CDF’s (e.g., beta, gamma, \( t \), and chi-square). Therefore, in terms of the normal random effect \( a_i \), using probability theory for transformations, the \( i \)th cluster’s contribution to the marginal likelihood in (2.1) can be re-written as

\[
L_i(\beta, \theta) = \int_{b_i} \prod_{k=1}^{n_i} f(y_{ik}|x_{ik}, b_i, \beta) f(b_i|\theta) \, db_i.
\]

where \( \phi(\cdot) \) is the standard normal probability density function. The marginal likelihood in (2.3) can be approximated by quadrature and standard maximization methods (e.g., Newton-Raphson, quasi-Newton) can be used to solve for \( (\beta, \theta) \). We discuss quadrature approximations in Section 3.

The idea of using a probability integral transform for modeling a single nonnormally distributed random effect can be extended in a straightforward manner to two or more independent nonnormal random effects \( (b_{i1}, b_{i2}, \ldots) \), with densities \( f_j(b_{ij}) \) and CDFs \( F_j(b_{ij}) \). In the case of two independent nonnormal random effects \( (b_{i1}, b_{i2}) \), the probability integral transform can be applied as follows. Let \( (a_{i1}, a_{i2}) \) be two independent normally distributed random effects, with zero means and each with a variance of one, such that

\[
(a_{i1}, a_{i2}) \sim \text{Normal} \left( 0, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right).
\]

Let \( u_{i1} = \Phi(a_{i1}) \) and \( u_{i2} = \Phi(a_{i2}) \), both with a uniform(0,1) distribution. Then applying the probability integral transform to each variable, \( F_1(b_{i1}) \) and \( F_2(b_{i2}) \) also have independent uniform distributions, so that \( b_{i1} = F_1^{-1}(u_{i1}) \) has density \( f_1(b_{i1}) \) and \( b_{i2} = F_2^{-1}(u_{i2}) \) has density \( f_2(b_{i2}) \). The extension to more than two independent nonnormal random effects follows in a similar manner. When correlation is present between two or more nonnormal random effects, the probability integral transform method becomes more complicated and requires the use of a multivariate probability integral transform approach (Genest and Rivest 2001).
Although we have highlighted that the proposed method can be implemented using PROC NLMIXED in SAS, we note that other software packages that contain flexible approaches for modeling nonlinear mixed-effects models could equally be used. For example, the NLME function in S-Plus, which uses (restricted) maximum likelihood procedures in the fitting process, assumes normal random effects. Although the probability integral transform cannot be readily implemented for use in conjunction with the existing NLME function, it should be relatively straightforward to create a modified NLME function from the original source code.

3. NUMERICAL INTEGRATION METHODS

Numerical integration techniques are a viable approach for approximating the marginal density function in (2.3) and other nonlinear mixed-effects models when no closed-form solution exists. For the method proposed in Section 2, Gaussian quadrature was found to be an appropriate technique for fitting nonlinear mixed-effects models. Gaussian quadrature approximates the marginal density function in (2.3) by a weighted average of the integrand evaluated at a number, \( Q \), of predetermined abscissas (quadrature points) \( d_q \) (\( q = 1, \ldots, Q \)) over the random effects \( a_i \) (Pinheiro and Bates 1995; Davidian and Gallant 1993; Lesaffre and Spiessens 2001; Liu and Pierce 1994). In particular, the integral in (2.3) is approximated as

\[
L_i(\beta, \theta) = \int_{a_i} \left[ \prod_{k=1}^{n_i} f(y_{ik} | x_{ik}, F_{\theta}^{-1}(\Phi(a_i)), \beta) \phi(a_i) \right] \, da_i \\
\approx \sum_{q=1}^{Q} \prod_{k=1}^{n_i} f(y_{ik} | x_{ik}, F_{\theta}^{-1}(\Phi(d_q)), \beta) \phi(d_q) w_q ,
\]

where \( d_q = \sqrt{2} z_q \), \( w_q = \sqrt{2} \eta_q e^{-z_q^2} \), and the standard Gauss-Hermite weights \( \eta_q \) and abscissas \( z_q \) can be obtained from tables (Abramowitz and Stegun 1972, table 25.10) and an algorithm proposed by Golub (1973) and Golub and Welsch (1969). Thus, the likelihood that is maximized is

\[
L(\beta, \theta) = \prod_{i=1}^{N} \sum_{q=1}^{Q} \prod_{k=1}^{n_i} f(y_{ik} | x_{ik}, F_{\theta}^{-1}(\Phi(d_q)), \beta) \phi(d_q) w_q .
\]

4. EXAMPLES

4.1 LIVER CANCER STUDY

The first example considers clustered survival data from a liver cancer study comparing two therapies. In this study, patients are nested within hospitals, with an average of three patients \((n_i)\) per hospital. There is no censoring, so the outcome for the \( k \)th patient in the \( i \)th
hospital is the survival time, here denoted by $T_{ik}$. (This dataset is available for downloading at: http://www.stat.sc.edu/~kerrie/liverdata.html).

Conditional on a random effect (or frailty) $b_i$, we assume that the survival time is exponential,

$$f(t_{ik}|\text{TRT}_{ik}, \text{CARD}_{ik}, b_i, \beta) = \lambda_{ik} e^{-\lambda_{ik} t_{ik}}, \quad (4.1)$$

where

$$\lambda_{ik} = \exp(\beta_0 + \beta_1 \text{TRT}_{ik} + \beta_2 \text{CARD}_{ik} + b_i) \quad (4.2)$$

is the conditional hazard and TRT$_{ik}$ equals 1 if treatment 1, 0 if treatment 0, and CARD$_{ik}$ equals 1 if normal cardiac status, 0 if abnormal. To assess the sensitivity of inferences to assumptions about the distribution of the random effects, we assume: (1) $b_i$ is normal with mean 0 and variance $\theta$ and (2) $b_i = \log(g_i)$, where $g_i > 0$ has the following gamma distribution,

$$f(g_i|\theta_1, \theta_2) = g_i^{1/\theta_1-1} \exp(-g_i/\theta_2)/[\Gamma(1/\theta_1)\theta_2^{1/\theta_1}]. \quad (4.3)$$

For identifiability, we set $\theta_2 = \theta_1$ so that $g_i$ has mean 1, that is,

$$E(g_i) = \theta_2/\theta_1 = 1,$$

and (4.3) reduces to

$$f(g_i|\theta_1) = g_i^{1/\theta_1-1} \exp(-g_i/\theta_1)/[\Gamma(1/\theta_1)\theta_1^{1/\theta_1}]. \quad (4.4)$$

As $\theta_1$ approaches 0, observations within a cluster are independent, while large values of $\theta_1$ induce high within-cluster correlation.

The command syntax for fitting these two models, using Gaussian quadrature, is given in Table 1. Starting values for the parameters were obtained by assuming observations within a cluster were independent. Of note, it was observed that the numerical integration techniques did not perform well if starting values were too far away from the final estimated values; ordinarily, this problem can be circumvented by choosing starting values for the fixed effects from a model fitted without any random effects. To form the exponential likelihood, we use the property that the Poisson and exponential likelihoods are proportional. In particular, the likelihood for an exponential distribution with hazard $\lambda_{ik}$ is equivalent to a Poisson likelihood in which the Poisson outcome is a censoring indicator $\delta_{ik}$, and the Poisson mean parameter is

$$\lambda_{ik}^* = \exp[\log(T_{ik}) + \beta_0 + \beta_1 \text{TRT}_{ik} + \beta_2 \text{CARD}_{ik} + b_i].$$

Because there is no censoring in this dataset, $\delta_{ik} = 1$ for all patients. Furthermore, when assuming gamma random effects, we note that PROC NLMIXED in SAS only allows a special case of the gamma distribution in (4.3), in which $\theta_2 = 1$, that is,

$$f(g_{i2}|\theta_1, \theta_2 = 1) = g_{i2}^{1/\theta_1-1} \exp(-g_{i2})/[\Gamma(1/\theta_1)]. \quad (4.5)$$
Table 1. SAS Command Syntax for Fitting Nonlinear Mixed-Effects Model to the Liver Cancer Dataset

/* Liver Cancer Study dataset */
/* INST = institute, SURVT = survival time (months),
   treat = 1 (yes) or 0 (no) */
/* heart = cardiac status (abnormal = 1, normal = 0) */
data liver;
  input INST SURVT treat heart;
  log_s = log(SURVT);
cens=1;
datalines;
  1 23.286 1 1
  1  6.429 0 0
  1 26.857 1 0
  1 11.143 0 1
  2  3.857 0 1
  2  9.000 1 0
  2  8.714 0 0
  3  1.143 0 1
  3 23.143 1 0
  3  2.571 1 0
  3  2.857 0 1
  4  76.429 1 0
  4 35.857 1 1
  4 25.857 0 0
  4  52.286 0 0
  5 25.143 1 1
  5  25.143 0 0
  6  29.286 1 1
  6  28.857 0 0
  7  1.857 0 1
  7 14.286 1 0
run;

continued
However, if one defines $g_i = \theta_1 g_{i2}$, then $g_i$ has the gamma distribution in (4.4). Letting $F_{\theta_1}^{-1}(\cdot)$ denote the inverse CDF (often referred to as the "quantile function") of the gamma distribution given in (4.5), and $\Phi(\cdot)$ be the standard normal CDF, then, when using PROC NLMIXED in SAS, the gamma random effect $g_i$ is obtained via the following set of transformations:

1. $a_i \sim N(0, 1)$
2. $p_i = \Phi(a_i)$
3. $g_{i2} = F_{\theta_1}^{-1}(p_i)$
4. $g_i = \theta_1 g_{i2}$.

In Table 2 the MLE of $\beta$ for the liver cancer data based on either normal or gamma random effects were found to be very similar. For the subset of data analyzed, both sets of analyses indicate that survival does not depend on treatment or on baseline cardiac status. For this particular example, there is the suggestion that estimation of $\beta$ is not very sensitive
to the choice of the random effect distribution. The sensitivity of inferences to assumptions about the distribution of the random effects is investigated further, via simulation studies, in Section 5. Although the estimates of the fixed effects were not sensitive to assumptions about the random effects, predictions of the hazard in (4.2) for any patient were. For example, the predicted hazard for the eighth patient in Table 1 (the first patient from the third hospital) was 0.216 (SE of prediction = 0.157) when based on the model with normal random effects, while it was 0.136 (SE of prediction = 0.081) when based on the model with gamma random effects. The former prediction is almost 60% larger than the latter. Thus, although the fixed effects estimates are relatively robust to assumptions about the distribution of the random effects, this is not the case for inferences about the random effects.

Finally, for this example the computation time was longer for the model with gamma random effects by a factor of approximately 12. This was most likely the result of the finite difference approximation needed for the derivative of the normal CDF function and the inverse CDF for the gamma distribution, in addition to the numerical integration required after these approximations. In particular, the computing time, on a Pentium IV, 2.4 GHz, 512MB RAM, was 1.2 seconds (real time) for the model with gamma random effects,
and 0.1 seconds for the model with normal random effects when using a relative gradient convergence criterion of $10^{-8}$.

### 4.2 Cardiotoxic Effects of Chemotherapy

The second example considers correlated binary data for a study of the cardiotoxic effects of doxorubicin chemotherapy for the treatment of acute lymphoblastic leukemia in childhood. (This dataset is available for download at: [http://www.stat.sc.edu/~kerrie/cardiodata.html](http://www.stat.sc.edu/~kerrie/cardiodata.html)). In this example, 24 patients who were cured of their leukemia had a long-term follow-up visit to determine how their hearts were functioning. Six tests of heart function were performed for each subject at this visit; the results of each test were coded as normal/abnormal. Thus, we have $N = 24$ clusters, each patient representing a cluster, and $n_i = 5$ (due to a missing test result) or $n_i = 6$ binary observations per cluster. We denote the $k$th heart test on the $i$th patient by $Y_{ik}$ which equals 1 if abnormal and 0 if normal. Because the five or six heart tests are similar, we assume the probability that each test is abnormal is the same logistic function of both dose and time,

$$
\pi_i = \Pr[Y_{ik} = 1 | \text{DOSE}_i, \text{TIME}_i] = \frac{\exp(\beta_0 + \beta_1 \text{DOSE}_i + \beta_2 \text{TIME}_i)}{1 + \exp(\beta_0 + \beta_1 \text{DOSE}_i + \beta_2 \text{TIME}_i)},
$$

(4.6)

where $\text{DOSE}_i$ is 1 if high and 0 if low, and $\text{TIME}_i$ is the time (in years) since the completion of chemotherapy. Further, we let $\rho$ be the correlation between any two heart function measures on the same subject. Using the beta-binomial model, to account for the correlation among observations on the same subject, we first define a random probability $p_i$, which has a beta distribution:

$$
f(p_i | \alpha_{1i}, \alpha_{2i}) = \frac{\Gamma(\alpha_{1i} + \alpha_{2i})p_i^{\alpha_{1i} - 1}(1 - p_i)^{\alpha_{2i} - 1}}{\Gamma(\alpha_{1i})\Gamma(\alpha_{2i})},
$$

(4.7)

where $\alpha_{1i} = \pi_i (1 - \rho)/\rho$ and $\alpha_{2i} = (1 - \pi_i)(1 - \rho)/\rho$, respectively. Next, we let $Y_{i+} = \sum_{k=1}^{n_i} Y_{ik}$ be the number of abnormal tests on the $i$th subject. Then, conditional on $p_i$, it is assumed that $Y_{i+}$ has a binomial distribution with probability $p_i$,

$$
f(Y_{i+} | p_i) = \binom{n_i}{y_{i+}} p_i^{y_{i+}} (1 - p_i)^{n_i - y_{i+}}.
$$

(4.8)

The MLEs are obtained from the marginal likelihood, integrating over the distribution of the random effects. Here, the marginal likelihood for the beta-binomial model has a closed form,

$$
f(y_{i+}) = \int f(Y_{i+} | p_i) f(p_i | \alpha_{1i}, \alpha_{2i}) dp_i
$$

$$
= \binom{n_i}{y_{i+}} \frac{B(\alpha_{1i} + y_{i+}, \alpha_{2i} + n_i - y_{i+})}{B(\alpha_{1i}, \alpha_{2i})}.
$$

(4.9)

Even though there is a closed-form expression for the marginal likelihood, we use PROC NLMIXED in SAS to both directly maximize the marginal likelihood in (4.9), as well as
Table 3. SAS Command Syntax for Fitting Nonlinear Mixed-Effects Model to the Cardiotoxicity Dataset

/* Cardiotoxicity dataset */
/* y = number of abnormal heart tests for a subject */
/* n = number of heart tests, dose = 1 (high) or 0 (low) */
/* time = time since chemotherapy (years) */
data cardtox;
input id y n dose time;
z=1;
datalines;
  1  4  6  1  13.7
  2  0  5  1  15.6
  3  3  5  1  4.6
  4  4  5  1  13.0
  5  0  5  0  6.2
  6  1  6  1  15.4
  7  2  5  0  6.5
  8  0  5  0  4.4
  9  1  5  0  9.6
 10  3  5  1  11.2
 11  3  5  0  8.1
 12  3  5  1  13.1
 13  1  5  0  10.1
 14  4  6  0  8.4
 15  1  5  0  4.2
 16  1  5  1  13.5
 17  1  5  1  17.9
 18  1  5  0  8.6
 19  2  6  0  5.9
 20  3  5  1  13.2
 21  4  5  1  14.5
 22  4  6  0  8.1
 23  0  5  0  8.2
 24  4  6  0  8.1
run;

continued
/* Numerically Integrated Beta-Binomial Marginal Likelihood */
proc nlmixed data=cardtox method=GAUSS NOAD fd qpoints=30;
    parms b0=-0.24 b_dose=1.00 b_time=-0.065 rho=0.1166;

    pi_i = exp(b0 + b_dose*dose + b_time*time)/
           (1 + exp(b0 + b_dose*dose + b_time*time)) ;

    alpha_i1 = pi_i*(1-rho)/rho ;
    alpha_i2 = (1- pi_i)*(1-rho)/rho ;
    prob = CDF('NORMAL',a_i) ;
    p_i = quantile('BETA',prob,alpha_i1,alpha_i2);

    model y ˜ binomial(n,p_i);
    random a_i ˜ normal(0,1) subject=id;
run;

/* Exact Beta-Binomial Marginal Likelihood */
proc nlmixed data=cardtox method=GAUSS NOAD fd qpoints=30;
    parms b0=-0.24 b_dose=1.06 b_time=-0.07 rho=0.11;

    pi_i = exp(b0 + b_dose*dose + b_time*time)/
           (1 + exp(b0 + b_dose*dose + b_time*time)) ;

    alpha_i1 = pi_i*(1-rho)/rho ;
    alpha_i2 = (1- pi_i)*(1-rho)/rho ;

    like1 = gamma(alpha_i1 + y)*gamma(alpha_i2+n-y)/gamma(alpha_i1 + alpha_i2 + n ) ;
    like2 = gamma(alpha_i1) * gamma(alpha_i2)/gamma(alpha_i1 + alpha_i2) ;

    llik = log(like1) - log(like2);

    model z ˜ general(llik);
    /* Note: When using general(), the outcome ‘z’ */
    /* is not actually used, and we have */
    /* arbitrarily set z=1 */
run;
Table 4. Selected Output from PROC NLMIXED in SAS for the Commands and Data Given in Table 3

/* Numerically Integrated Beta-Binomial Marginal Likelihood */

| Standard | Parameter | Estimate | Error   | DF  | t Value | Pr > |t| |
|----------|-----------|----------|---------|-----|---------|------|---|
|          | b0        | -0.3007  | 0.7107  | 23  | -0.42   | 0.6761 |
|          | b_dose    | 0.9602   | 0.6785  | 23  | 1.42    | 0.1704 |
|          | b_time    | -0.05856 | 0.08659 | 23  | -0.68   | 0.5056 |
|          | rho       | 0.1140   | 0.07903 | 23  | 1.44    | 0.1625 |

/* Exact Beta-Binomial Marginal Likelihood */

| Standard | Parameter | Estimate | Error   | DF  | t Value | Pr > |t| |
|----------|-----------|----------|---------|-----|---------|------|---|
|          | b0        | -0.3007  | 0.7107  | 24  | -0.42   | 0.6760 |
|          | b_dose    | 0.9602   | 0.6785  | 24  | 1.42    | 0.1699 |
|          | b_time    | -0.05856 | 0.08659 | 24  | -0.68   | 0.5054 |
|          | rho       | 0.1140   | 0.07903 | 24  | 1.44    | 0.1620 |

Note: The difference in reported degrees of freedom is due to the fact that PROC NLMIXED in SAS subtracts a degree of freedom for the number of random effects in the model.

to maximize the approximation to the marginal likelihood using Gaussian quadrature. The command syntax is given in Table 3; selected output is displayed in Table 4. To obtain the approximate marginal likelihood, the beta random probability $p_i$ is created via the following set of transformations:

1. $a_i \sim N(0, 1)$
2. $e_i = \Phi(a_i)$
3. $p_i = F_{\beta,\rho}^{-1}(e_i)$,

where $F_{\beta,\rho}^{-1}(\cdot)$ is the inverse CDF of the beta distribution given in (4.7).

In Table 4 the MLEs of $\beta$ and the variance estimates using either the exact marginal likelihood or its numerical approximation are identical, when rounded to the fifth decimal point, confirming that the method we have proposed yields the MLE in this example. Both sets of estimates indicate that the log odds of an abnormal test is unrelated to dose or time since completion of chemotherapy. Naturally, the computing time was longer for the approximate method because of the finite difference approximation used for the derivative of the normal CDF function and the inverse CDF for the beta distribution. We note that on
In this section we conduct simulation studies based on the two examples to investigate the impact of nonnormally distributed random effects on the resulting parameter estimates and standard errors. Comparisons are also made with normally distributed random effects, and when gamma random effects are incorrectly modeled as normal random effects. The effects of cluster size and numbers of clusters are examined.

To determine if the proposed method yields unbiased estimates in the liver cancer data when using the gamma random effects model, we performed a simulation study. The clustered data were generated as follows: We let each subject in a cluster have two covariates, that is, \( x_{ik} = (x_{ik1}, x_{ik2}) \) represents covariate values for the \( k \)th subject in the \( i \)th cluster. The covariate \( x_{ik1} \) was distributed as Bernoulli(0.5), and the covariate \( x_{ik2} \) was distributed as normal(0,1), and was generated independently of \( x_{ik1} \). The covariates generated were designed to specify group membership and some other characteristic measured at baseline such as a quantitative measure of heart function. To induce dependence within a cluster, we generated a random cluster effect, \( \epsilon_i \), and let \( \lambda_{ik} = \exp(\beta_1 x_{ik1} + \beta_2 x_{ik2} + \epsilon_i) \).

Then, conditional on \( \epsilon_i \), \( n_i \) survival times \( (T_{i1}, \ldots, T_{in_i}) \) in a cluster were generated independently from an exponential distribution with conditional hazard \( \lambda_{ik} \). Also, the censoring distribution was uniform(0, 25), which resulted in approximately 20% censored observations. One thousand datasets were generated for each combination of the following numbers of clusters \( N = 10, 50, 100, \) or 250, and the number of subjects within each cluster, \( n_i = 5 \) or 10. The random effects \( \epsilon_i \) were generated using either a normal(0, \( \theta_1 \)) distribution or the log of a gamma(\( (1/\theta_1), \theta_1 \)) distribution. In both cases \( \theta_1 = 1.0 \). Finally, we let \( \beta_1 = \beta_2 = -1 \). In all simulations, Gaussian quadrature was used. The random effects were modeled in two ways: (1) assuming a normal distribution, and (2) assuming a gamma distribution, with the use of the probability integral transformation technique.

Table 5 presents the means and standard deviations of the estimates for the parameters \( (\theta_1, \beta_1, \beta_2) \) for the models that correctly assume gamma and normally distributed random effects, and also for the case where gamma random effects are incorrectly modeled as normal. The use of Gaussian quadrature led to an almost 100% convergence rate for all sets of simulations. The estimates of the parameters \( (\beta_1, \beta_2) \) when the random effects were correctly assumed to be gamma or normally distributed were almost unbiased for all values of \( N \) and \( n_i \). These results suggest that fitting nonlinear mixed-effects models with nonnormal random effects via numerical integration provides almost unbiased estimates of both the regression coefficients and the parameters for the random effects distribution.
Table 5. Results of Simulation Study (1,000 replications) Based on the Liver Cancer Model, with Parameters $\theta_1$, $\beta_1$, $\beta_2$: (1) Gamma distributed random effects correctly assumed to be gamma (left column); (2) Normally distributed random effects correctly assumed to be normal (middle column); (3) Gamma distributed random effects incorrectly assumed to be normal (right column). $N =$ number of clusters, $n_i =$ number of units within each cluster. Results are presented in terms of the mean estimate (standard deviation).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$n_i$</th>
<th>$N$</th>
<th>Gamma $\left(\frac{1}{\theta_1}, \theta_1\right)$</th>
<th>Normal($0, \theta_1$)</th>
<th>Gamma* $\left(\frac{1}{\theta_1}, \theta_1\right)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_1 = 1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_i = 5$</td>
<td>10</td>
<td>0.981 (0.548)</td>
<td>1.042 (0.730)</td>
<td>1.763 (1.397)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.995 (0.231)</td>
<td>1.018 (0.298)</td>
<td>1.644 (0.522)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>1.000 (0.163)</td>
<td>1.008 (0.203)</td>
<td>1.646 (0.365)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1.000 (0.098)</td>
<td>1.001 (0.114)</td>
<td>1.635 (0.219)</td>
<td></td>
</tr>
<tr>
<td>$n_i = 10$</td>
<td>10</td>
<td>0.967 (0.479)</td>
<td>0.994 (0.536)</td>
<td>1.888 (1.434)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>1.000 (0.201)</td>
<td>0.986 (0.228)</td>
<td>1.747 (0.522)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.993 (0.136)</td>
<td>0.990 (0.157)</td>
<td>1.721 (0.352)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>1.010 (0.098)</td>
<td>1.000 (0.101)</td>
<td>1.729 (0.226)</td>
<td></td>
</tr>
<tr>
<td>$\beta_1 = -1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_i = 5$</td>
<td>10</td>
<td>-1.027 (0.413)</td>
<td>-1.003 (0.388)</td>
<td>-1.216 (0.426)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-1.004 (0.168)</td>
<td>-1.005 (0.162)</td>
<td>-1.198 (0.173)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-1.004 (0.125)</td>
<td>-1.002 (0.118)</td>
<td>-1.197 (0.126)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>-1.000 (0.076)</td>
<td>-0.997 (0.068)</td>
<td>-1.196 (0.077)</td>
<td></td>
</tr>
<tr>
<td>$n_i = 10$</td>
<td>10</td>
<td>-1.012 (0.251)</td>
<td>-0.999 (0.240)</td>
<td>-1.096 (0.259)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-1.007 (0.112)</td>
<td>-0.997 (0.108)</td>
<td>-1.093 (0.114)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-1.002 (0.081)</td>
<td>-1.000 (0.075)</td>
<td>-1.086 (0.084)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>-1.002 (0.006)</td>
<td>-0.997 (0.049)</td>
<td>-1.085 (0.055)</td>
<td></td>
</tr>
<tr>
<td>$\beta_2 = -1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n_i = 5$</td>
<td>10</td>
<td>-1.008 (0.238)</td>
<td>-1.014 (0.230)</td>
<td>-0.998 (0.238)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-1.001 (0.106)</td>
<td>-1.002 (0.096)</td>
<td>-0.988 (0.105)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-1.001 (0.075)</td>
<td>-1.002 (0.067)</td>
<td>-0.987 (0.077)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>-1.001 (0.046)</td>
<td>-1.002 (0.038)</td>
<td>-0.987 (0.045)</td>
<td></td>
</tr>
<tr>
<td>$n_i = 10$</td>
<td>10</td>
<td>-1.007 (0.141)</td>
<td>-0.998 (0.142)</td>
<td>-0.998 (0.145)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-1.001 (0.060)</td>
<td>-1.001 (0.059)</td>
<td>-0.994 (0.063)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-1.002 (0.045)</td>
<td>-1.002 (0.042)</td>
<td>-0.996 (0.045)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>-1.001 (0.046)</td>
<td>-1.001 (0.026)</td>
<td>-0.994 (0.032)</td>
<td></td>
</tr>
</tbody>
</table>
Table 6. Results of Simulation Study (1,000 replications) Based on the Cardiotoxicity Model. \( N \) = number of subjects, \( n_i \) = number of tests per subject, \( \rho = 0.5 \). Results are presented in terms of the mean estimate (standard deviation).

<table>
<thead>
<tr>
<th>Parameter ( \beta )</th>
<th>( n_i )</th>
<th>( N )</th>
<th>Gaussian quadrature</th>
<th>Exact solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 = -1 )</td>
<td>10</td>
<td>0.429 (0.262)</td>
<td>0.353 (0.227)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.500 (0.113)</td>
<td>0.474 (0.087)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.492 (0.070)</td>
<td>0.483 (0.060)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.494 (0.039)</td>
<td>0.494 (0.037)</td>
<td></td>
</tr>
<tr>
<td>( \beta_1 = 2 )</td>
<td>10</td>
<td>-1.300 (1.837)</td>
<td>-2.224 (5.147)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-1.060 (0.563)</td>
<td>-1.055 (0.568)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-1.022 (0.398)</td>
<td>-1.019 (0.384)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>-1.010 (0.254)</td>
<td>-1.008 (0.243)</td>
<td></td>
</tr>
<tr>
<td>( \beta_2 = -0.1 )</td>
<td>10</td>
<td>2.433 (1.825)</td>
<td>3.880 (5.718)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>2.059 (0.393)</td>
<td>2.067 (0.554)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>2.031 (0.382)</td>
<td>2.034 (0.336)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>2.015 (0.242)</td>
<td>2.009 (0.235)</td>
<td></td>
</tr>
<tr>
<td>( \beta_3 = 0 )</td>
<td>10</td>
<td>-0.111 (0.062)</td>
<td>-0.159 (0.446)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-0.095 (0.088)</td>
<td>-0.100 (0.088)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.102 (0.061)</td>
<td>-0.102 (0.059)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>-0.101 (0.039)</td>
<td>-0.100 (0.037)</td>
<td></td>
</tr>
<tr>
<td>( \beta_4 = -0.5 )</td>
<td>10</td>
<td>-0.112 (0.248)</td>
<td>-0.137 (0.337)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-0.101 (0.083)</td>
<td>-0.100 (0.084)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.102 (0.062)</td>
<td>-0.101 (0.055)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>-0.101 (0.037)</td>
<td>-0.101 (0.035)</td>
<td></td>
</tr>
</tbody>
</table>
When the gamma random effects were incorrectly modeled as normal and the cluster size was small \((n_i = 5)\), regardless of the number of clusters \(N\), moderate bias \((\sim 20\%)\) was found in the estimate of \(\beta_1\); for larger cluster size \((n_i = 10)\), this bias was less than 10%. Variability of the parameter estimates also improved with increasing number of clusters and as the number of subjects within a cluster increased.

Simulation studies based on the cardiotoxicity data were also carried out to investigate the effect of sample size and \(\rho\), the within-cluster correlation, on parameter estimation for the beta-binomial model. Each cluster was assumed to have \(n_i\) Bernoulli outcomes, \(Y_{ij}\), and two covariates: the covariate \(x_{i1}\) was distributed as Bernoulli(0.5), and the covariate \(x_{i2}\) was distributed as normal(5,3), independent of \(x_{i1}\). The covariates generated were designed to specify whether the subject (or cluster) received treatment or not, and the time since completion of chemotherapy. The probability that \(Y_{ij}\) equals 1, given \((x_{i1}, x_{i2})\), is specified as

\[
\pi_i = \Pr[Y_{ik} = 1|x_{i1}, x_{i2}] = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})},
\]

(5.1)

where \((\beta_0, \beta_1, \beta_2)\) is set at \((-1, 2, -0.10)\). For each cluster, the random probability \(p_i\) was generated from the beta distribution described in Equation (4.7), with \(\pi_i\) as defined in Equation (5.1), and \(\rho = 0, 0.2, 0.5\), or 0.8. The responses \(Y_{i+}\) are then generated from the binomial distribution described in (4.8). One thousand datasets were generated for each combination of the following numbers of clusters \(N = 10, 50, 100\), or 250, and the number of subjects within each cluster, \(n_i = 5\) or 10. We used PROC NLMIXED in SAS to both directly maximize the closed form marginal likelihood, as well as to maximize the approximation to the marginal likelihood using Gaussian quadrature.

Results from these simulations are presented in Table 6. The use of Gaussian quadrature led to estimates of the regression parameters \((\beta_0, \beta_1, \beta_2)\) that were almost unbiased in moderate and larger sized samples. When the number of clusters was small \((N = 10)\), the estimates from the exact likelihood were actually more biased than those from Gaussian quadrature. When \(N > 50\), the estimates from the exact and approximate likelihoods were unbiased and very similar. As in the modeling of the liver cancer data, these results suggest that numerical integration provides unbiased estimation of the regression coefficients for a beta-binomial model.

6. CONCLUSION

Nonnormal random effects models are increasingly used in many areas of application, for example, models for survival data with gamma frailties. The technique we have proposed extends the use of readily available nonlinear mixed-effects software (e.g., PROC NLMIXED in SAS) to a wider range of distributions for a single random effect. Recently, Piepho and McCulloch (2000) described a method for transforming random effects to have nonnormal distributions; however, their method is restricted to lognormal and exponential distributions so that quadrature techniques can be used.
The method proposed in this article is based on a fundamental concept that is taught in every first course in probability: the probability integral transformation. This method provides a computationally feasible approach; for the two examples considered, the average time for fitting the models was at most a few seconds. We have found that with nonnormal random effects \( b_i = F^{-1}(\Phi(a_i)) \), Gaussian quadrature is the preferred numerical integration method for approximating the likelihood. Also, the results of the simulations indicate that the proposed method performs well with respect to bias and variability of the regression estimates for moderate sized samples. Section 2 described how the proposed method can be easily extended to two or more independent random effects. The proposed method can also be applied to multilevel models with more than a single level of nesting; for example, although PROC NLMIXED in SAS restricts realizations of random effects to a single level, this does not preclude the fitting of multilevel nonlinear mixed effects models provided the dimension of the lower-level units is relatively small. Finally, from the examples presented in Section 4, we see that the command syntax for implementing the technique within PROC NLMIXED in SAS is straightforward. The proposed method alleviates the need to write special-purpose software for nonlinear mixed-effects models with nonnormal random effects.

ACKNOWLEDGMENTS

The authors are grateful for the support provided by the following grants from the U.S. National Institutes of Health: GM 29745, HL 69800, AI 60373, CA 74015, CA 70101, and CA 68484. They also thank the referees for their helpful comments and suggestions.

[Received August 2001. Revised June 2005.]

REFERENCES


