Appl. Statist. (2001)
50, *Part* 3, *pp.* 325–335

On the effect of the number of quadrature points in a logistic random-effects model: an example

Emmanuel Lesaffre and Bart Spiessens

Katholieke Universiteit Leuven, Belgium

[Received January 2000. Final revision November 2000]

Summary. Although generalized linear mixed models are recognized to be of major practical importance, it is also known that they can be computationally demanding. The problem is the evaluation of the integral in calculating the marginalized likelihood. The straightforward method is based on the Gauss–Hermite technique, based on Gaussian quadrature points. Another approach is provided by the class of penalized quasi-likelihood methods. It is commonly believed that the Gauss–Hermite method works relatively well in simple situations but fails in more complicated structures. However, we present here a strikingly simple example of a logistic random-intercepts model in the context of a longitudinal clinical trial where the method gives valid results only for a high number of quadrature points (Q). As a consequence, this result warns the practitioner to examine routinely the dependence of the results on Q. The adaptive Gaussian quadrature, as implemented in the new SAS procedure NLMIXED, offered the solution to our problem. However, even the adaptive version of Gaussian quadrature needs careful handling to ensure convergence.

Keywords: Clinical trials; Gaussian quadrature; Generalized linear mixed model; Logistic random-effects model; Longitudinal data

1. A clinical trial in dermatology

A multicentre randomized comparison of two oral treatments for toe-nail infection (dermatophyte onychomycosis) involved 2×189 patients evaluated at seven visits, i.e. on weeks 0, 4, 8, 12, 24, 36 and 48 (De Backer *et al.*, 1998). The primary end point of the study was the absence of toe-nail infection. Here we are interested in the degree of onycholysis which expresses the degree of separation of the nail plate from the nail-bed. This secondary end point was scored in four categories (0, absent; 1, mild; 2, moderate; 3, severe) and was evaluated on 294 patients comprising 1908 measurements. The results are shown in Fig. 1.

In the exploratory phase of the study, various analyses were done. One of these was a longitudinal analysis based on the dichotomized onycholysis end point (no and mild (0) *versus* moderate and severe (1)). We performed two analyses:

(a) a logistic random-effects model including a random intercept and using MIXOR (Hedeker and Gibbons, 1994, 1996);
(b) a type I generalized estimating equation (GEE) method (Liang and Zeger, 1986) using the SAS procedure `GENMOD` (SAS Institute, 1996).

In both analyses the fixed part of the model contains as covariates treatment (0 or 1), time (continuous) and time $*$ treatment. In this model formulation the covariate treatment

Address for correspondence: Emmanuel Lesaffre, Biostatistical Centre, Katholieke Universiteit Leuven, Universitaire Ziekenhuizen Sint-Rafaël, Capucijnenvoer 33, B-3000 Leuven, Belgium.
E-mail: emmanuel.lesaffre@med.kuleuven.ac.be

 0035–9254/01/50325

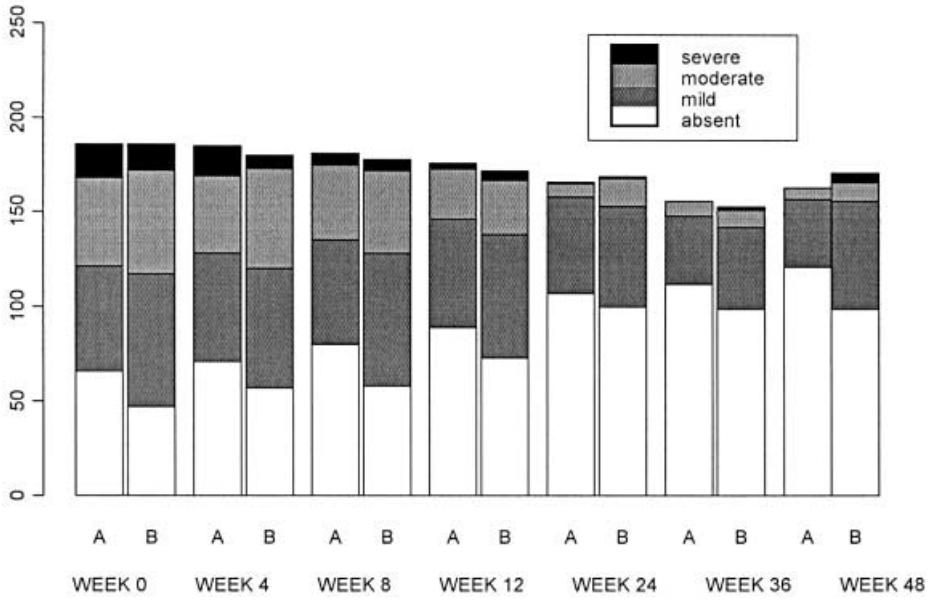

Fig. 1. Number of patients according to the degree of onycholysis during the dermatological study (A *versus* B)

represents the effect of treatment at the base-line, which should be negligible because of the randomization. The estimated treatment effect (and P-value from a Wald test) from the logistic random-effects model was drastically different from the corresponding estimate in the marginal model. With MIXOR, the treatment effect was estimated as -2.23 (standard error SE $= 0.35$; $P < 0.000001$), whereas with GENMOD an estimate of 0.02 (SE $= 0.24$; $P = 0.93$) was obtained with an unstructured working correlation matrix.

The result from the random-effects model implies that the two treatment groups are different at the base-line. This is not confirmed by the GEE analysis, nor by the randomization procedure nor by the data themselves. According to Fig. 1 the proportion of patients with the binary response 1 is practically equal for the two treatment groups at the base-line.

Below, we shall illustrate the problems with Gaussian quadrature by using an even simpler model without the interaction term. With MIXOR, the treatment effect was now estimated as -2.52 (SE $= 0.36$; $P < 0.000001$), whereas with GENMOD an estimate of -0.15 (SE $= 0.22$; $P = 0.49$) was obtained again with the unstructured working correlation matrix.

The logistic random-effects model delivers subject-specific estimates of the parameters, whereas the GEE method provides marginal estimates. It is known that the subject-specific estimate of a parameter is in absolute value larger than the corresponding marginal estimate (Neuhaus *et al.*, 1992). In fact, Diggle *et al.* (1994), page 142, provided an approximate formula for the relationship between the marginal and subject-specific true parameters, namely $\beta^{M} \approx (c^2\sigma^2 + 1)^{-1/2}\beta^{RE}$, where β^{M} and β^{RE} are the marginal and random-effect true regression coefficients respectively, $c = 16\sqrt{3}/15\pi$ and σ^2 is the variance of the random intercept. Using the estimated parameter from the logistic random-effects model (-2.52) and the estimated value for σ (3.57), we would expect for the marginal regression coefficient of treatment a value of -1.08, which is much larger in absolute value than the marginal GEE estimate obtained (taking into account the estimated SE). The discrepancy is so large that we wondered about the validity of our analysis. We shall show in Section 3 that the problem lies

in the evaluation of the marginalized likelihood of the logistic random-effects analysis. But first we review in Section 2 the logistic random-effects model. In Section 4 we compare the performance of several programs on the onychomycosis clinical trial. Finally, in Section 5 we show that replacing the usual Gaussian quadrature by adaptive Gaussian quadrature solved the problem. The data that are analysed in the paper can be obtained from

`http://www.blackwellpublishers.co.uk/rss/`

2. The logistic random-effects model

Suppose that Y_{ij} represents the binary response at the jth visit of the ith subject and that there are n_i measurements on the ith of N subjects. The logistic random-intercepts model for the problem outlined above is given by (with the interaction term omitted at a later stage)

$$\text{logit}\{P(Y_{ij} = 1|b_i, \boldsymbol{\beta})\} = \beta_0 + \beta_1 \,\text{trt}_i + \beta_2\, t_{ij} + \beta_3\, t_{ij} * \text{trt}_i + b_i \quad (i = 1, \ldots, N; j = 1, \ldots, n_i) \tag{1}$$

with $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^{\mathrm{T}}$ the total vector of parameters, t_{ij} the time of the jth visit for the ith subject, $b_i = \sigma z_i$ and $z_i \sim N(0, 1)$. Thus, b_i plays the role of the random intercept and σ is the standard deviation of the random intercept. It is assumed that, conditionally on b_i, the terms of the likelihood involving the ith subject are independent. The evaluation of the (marginalized) likelihood for the ith subject involves integrating out the random intercept and is equal to

$$L_i = L(\mathbf{Y}_i|\boldsymbol{\beta}, \sigma) = \int \prod_{j=1}^{n_i} P(Y_{ij} = y_{ij}|\boldsymbol{\beta}, b_i)\, \phi_\sigma(b_i)\, \mathrm{d}b_i = \int \prod_{j=1}^{n_i} P_\sigma(Y_{ij} = y_{ij}|\boldsymbol{\beta}, z_i)\, \phi(z_i)\, \mathrm{d}z_i, \tag{2}$$

where $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^{\mathrm{T}}$ is the vector of measurements for the ith subject, y_{ij} is the corresponding observed value of the response and $\phi_\sigma(b_i)$ is the normal density with mean 0 and standard deviation σ. In the last integral in equation (2) the scale parameter σ is removed from the normal density and incorporated in the likelihood for the ith subject; $\phi(\cdot)$ is the standard normal density. The total (marginalized) likelihood is the product of the N terms in equation (2) and hence

$$L \equiv L(\mathbf{Y}|\boldsymbol{\beta}, \sigma) = \prod_{i=1}^{N} L(\mathbf{Y}_i|\boldsymbol{\beta}, \sigma) \equiv \prod_{i=1}^{N} L_i, \tag{3}$$

whereby $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \ldots, \mathbf{Y}_N^{\mathrm{T}})^{\mathrm{T}}$ is the total vector of responses.

To determine the maximum likelihood estimates, expression (3) needs to be evaluated and maximized with respect to the parameters and this involves the evaluation of an integral. A classical way to calculate the integral is via Gauss–Hermite polynomials which implies here that

$$\int_{-\infty}^{\infty} P(z)\, \phi(z)\, \mathrm{d}z \approx \sum_{q=1}^{Q} P(z_q)\omega_q, \tag{4}$$

where

$$P(z) = \prod_{j=1}^{n} P(Y_j = y_j|z)$$

and the dependence on the subject and the parameters is omitted from the notation. Expression (4) implies that the integral is approximated by a weighted sum evaluated at Q values z_q, called the quadrature points. The weights ω_q depend only on Q and the normal density. A graphical illustration of expression (4) is given in Fig. 2.

3. The practical implication of the number of quadrature points

3.1. Increasing the number of quadrature points with MIXOR

While running MIXOR, the program indicates that 10 quadrature points are 'often appropriate'. In contrast Longford (1993), page 229, indicated that 'for most purposes 5-point quadrature suffices'. Alternatively, the software program EGRET (Cytel Software Corporation, 1995), which also fits some logistic regression models with random effects, uses 20 quadrature points. Given the simplicity of the model being fitted, together with data comprising 294 subjects and 1908 measurements, at the start of the analysis we did not suspect any problems with 10 quadrature points.

As an illustration of the effect of the number of quadrature points on the calculations, we analysed the logistic random-effects model (1) (omitting the interaction term) with MIXOR for Q ranging from 10 to 50 in steps of 10 (Table 1). For $Q = 20$ the estimated treatment effect has shrunk considerably but is still significant at $P = 0.05$. From $Q = 30$ onwards, the treatment effect was no longer significant, but it was disturbing to see that the estimate of the treatment effect kept changing. Furthermore, increasing Q beyond 100 led to divergence of

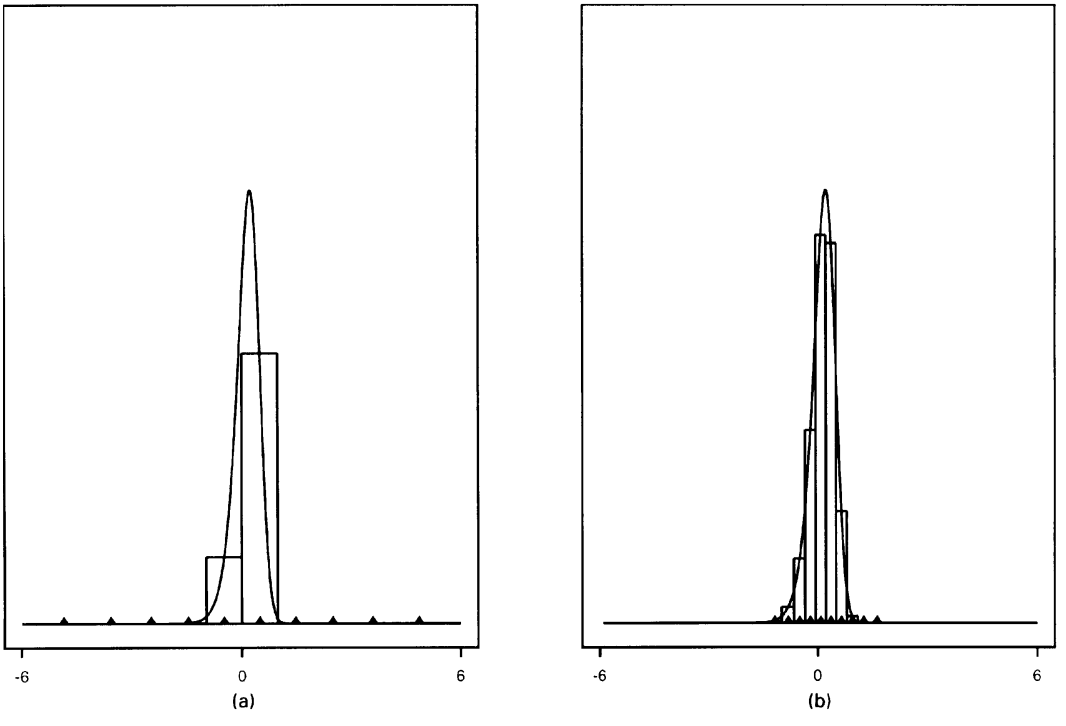

Fig. 2. Comparison of the positions of 10 quadrature points obtained from (a) an ordinary Gaussian quadrature and (b) an adaptive Gaussian quadrature for the same integrand: ▲, position of the quadrature points z_q; □, contribution of each point to the integral, i.e. $f(q)\omega_q$

Table 1. Effect of the number of quadrature points on the output of the logistic random-intercept analysis with MIXOR†

Q	*llik*	*treat*	*SE(treat)*	*P*
10	−635.83	−2.52	0.36	< 0.000001
20	−627.41	−1.02	0.46	0.026
30	−627.38	−0.37	0.53	0.48
40	−627.46	−0.45	0.55	0.42
50	−627.47	−0.51	0.56	0.36

†The number of quadrature points is indicated by *Q*; llik is the log-likelihood; treat is the estimated treatment on convergence; SE(treat) is the estimated standard error calculated from minus the inverse of the (expected) second-derivative matrix at convergence; *P* is the *P*-value corresponding to the Wald statistic at convergence.

MIXOR. When including the interaction term the results were similar, but somewhat less spectacular: at $Q = 10$ the estimated treatment effect is equal to -2.23 (SE $= 0.35$; $P < 0.0001$), whereas at $Q = 20$ it is -0.40 (SE $= 0.46$; $P = 0.38$), and hence non-significant.

With user-defined starting values of 0 for the regression parameters and 1 for σ and with $Q = 10$, the program MIXOR converged to a value of 0.093 for the effect of treatment, which is no longer significant. For these starting values a higher likelihood was obtained at convergence (-630.13). Hence, the previous analyses converged to a local maximum. However, the current problem is not only about local maxima. In another, but related, analysis a global maximum was found with the same problematic behaviour of a highly significant treatment effect, while showing no significant effect for the marginal analysis. Further, the same problems were found when including a random slope and for the ordinal logistic random-effects model involving the four response categories.

Finally, these results were confirmed by our own programs written in GAUSS. Thus, the problem cannot be solely attributed to the program MIXOR.

3.2. The effect of Q on the calculation of the marginalized probabilities

We have calculated the likelihood (4) for model (1) for a grid of parameter values (around the maximum likelihood estimate obtained with $Q = 20$) to indicate the effect of Q when determining the marginalized likelihood for a logistic random-effects model. For this, no iterative procedure is involved and our self-written programs were used. Fig. 3 shows the log-likelihood for a grid keeping the intercept and the regression coefficients of time and time by treatment interaction constant while varying β_1 from -2 to 1.2 in steps of 0.4 and σ from 2.5 to 5.5 in steps of 0.5, both taken around the correct maximum likelihood estimate. The choice of σ corresponds to intraclass correlations ranging from 0.66 to 0.90, which are somewhat on the high side, even for longitudinal data. The conclusion of our exercise was that the value of the likelihood very much depends on Q and that for high values of Q precautions need to be taken against overflow and underflow (illustrated by some lines stopping at $Q = 10$ and $Q = 20$). Other grids were also used and the overall conclusions are as follows (the results are not shown).

(a) Although initial analyses suggested that the problem was only present with a binary covariate, the problem arises with all kinds of covariates. Further, it does not depend on the coding of the binary covariate (0 and 1 *versus* -1 and 1).
(b) The problem disappears with $\sigma = 0$, when no random effect is assumed. Further, on average there is less problem for the smallest values of σ (intraclass correlation around

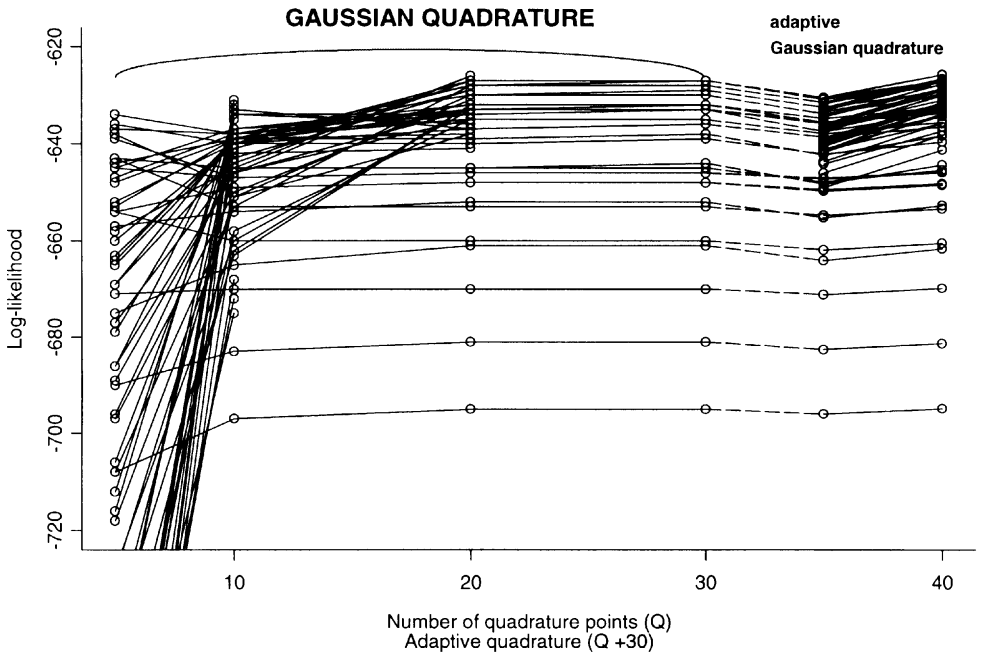

Fig. 3. Log-likelihood as a function of the number of quadrature points, Q, for the Gaussian quadrature method and the adaptive version: the log-likelihoods are evaluated keeping (β_0, β_2, β_3) fixed at (-1.62, -0.38, -0.12) while varying β_1 on the grid $\{-2, -1.6, -1.2, -0.8, -0.4, 0, 0.4, 0.8, 1.2\}$ and σ on the grid {2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5}; the results of the adaptive Gaussian quadrature method with Q quadrature points are plotted with x-co-ordinate $Q + 30$

0.70). However, for each value of σ the numerical problem can be arbitrarily high, depending on the other (regression) parameters.

3.3. *The effect of Q on the maximization procedure*

Finding the maximum likelihood estimate of the logistic random-effects model (1) involves Gaussian quadrature, to calculate

(a) the log-likelihood,
(b) the first derivatives and
(c) the second derivatives.

The expression of the first derivative with respect to the regression parameters is

$$\frac{\partial[\ln\{\mathbf{L}(\mathbf{Y}_i|\boldsymbol{\beta}, \sigma)\}]}{\partial\boldsymbol{\beta}} = \frac{\int \prod_{j=1}^{n_i} P_\sigma(Y_{ij} = y_{ij}|\boldsymbol{\beta}, z_i)\, \mathbf{s}_i(z_i)\, \phi(z_i)\, \mathrm{d}z_i}{\int \prod_{j=1}^{n_i} P_\sigma(Y_{ij} = y_{ij}|\boldsymbol{\beta}, z_i)\, \phi(z_i)\, \mathrm{d}z_i}, \tag{5}$$

with

$$\mathbf{s}_i(z_i) = \sum_{j=1}^{n_i} \{y_{ij} - P_\sigma(Y_{ij} = 1|\boldsymbol{\beta}, z_i)\}\mathbf{x}_{ij}$$

and $\mathbf{x}_{ij}$ is the vector of covariates at the jth measurement for the ith subject.

An expression of the first derivative with respect to σ and the expressions of the second derivative can be found elsewhere (e.g. Longford (1993)). In Fig. 4, we show, for the example in Section 3.2, the 294 log(probabilities) summing up to the log-likelihood and the 294 components of the first derivative with respect to the regression coefficient of treatment, for $Q = 10$ and $Q = 50$. Clearly, although the marginalized log(probabilities) differ for $Q = 10$ and $Q = 50$, a larger discrepancy is seen for the components of the first derivative.

A similar large discrepancy can be seen for the first derivative of β_0, whereas there is less discrepancy for the first derivatives of β_2 and β_3 (the results are not shown). The larger dependence of the first (and second) derivative on Q is easily explained by the fact that

$$\prod_{j=1}^{n_i} P_\sigma(Y_{ij} = y_{ij}|\boldsymbol{\beta}, z_i)\, \mathbf{s}_i(z_i)$$

is a less smooth function of z_i than

$$\prod_{j=1}^{n_i} P_\sigma(Y_{ij} = y_{ij}|\boldsymbol{\beta}, z_i)$$

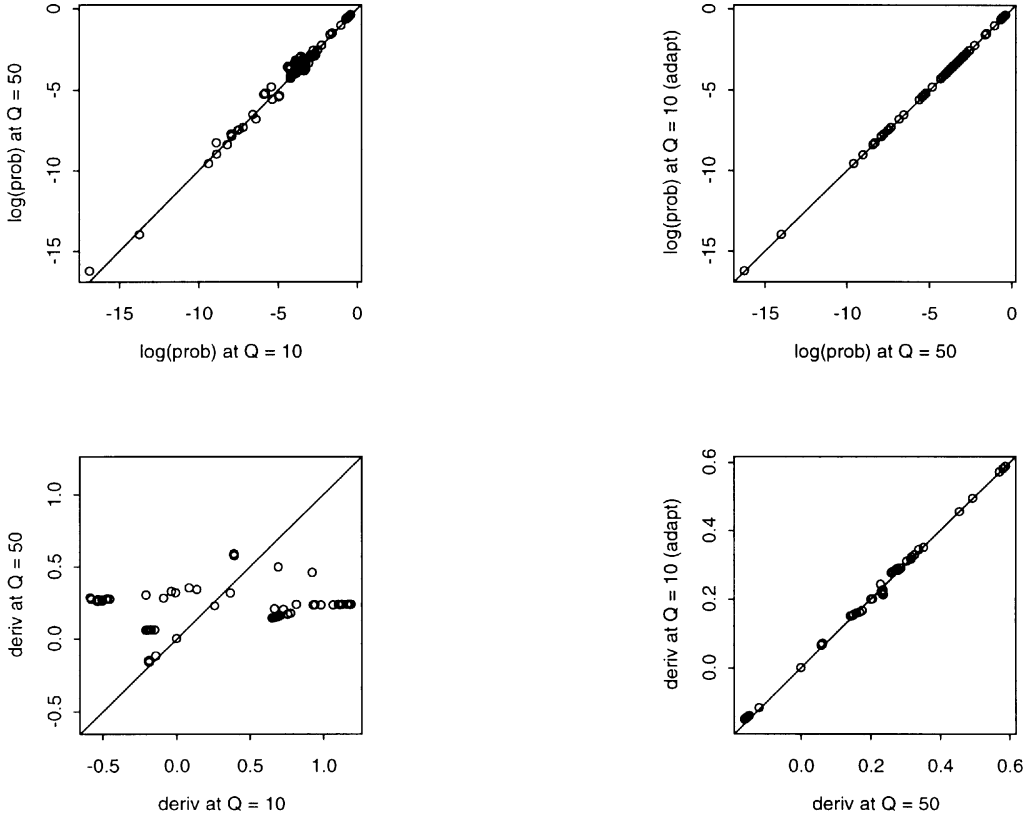

Fig. 4. Log-probabilities of the logistic random-intercept model (1) with $\beta^{\mathrm{T}} = (-1.62, -0.4, -0.39, -0.12)$ and components of the first derivative with respect to β_1 (three situations are considered: for Gaussian quadrature with $Q = 10$ and $Q = 50$, and adaptive Gaussian quadrature with $Q = 10$)

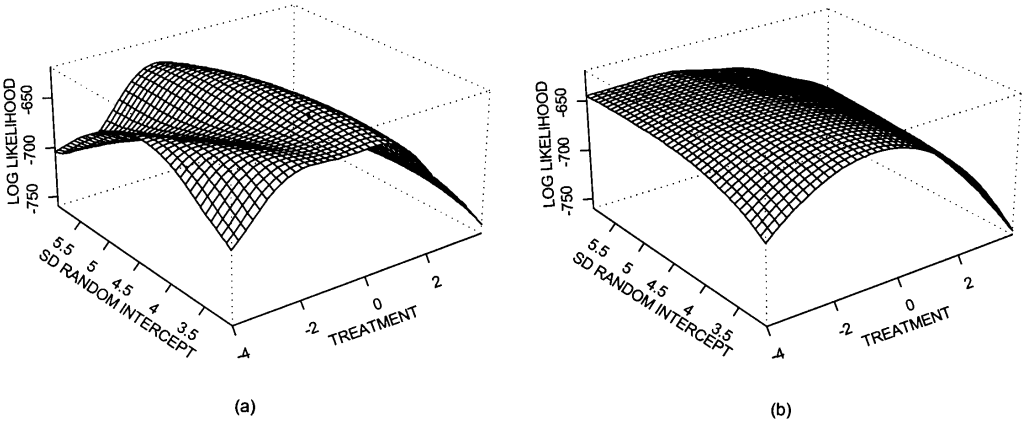

Fig. 5. Log-likelihood surface of the logistic random-intercept model (1), without the interaction term, as a function of the treatment parameter and the standard deviation of the random intercept (for this, the intercept was fixed at -0.70 and the parameter for time was fixed at -0.43): (a) 10 quadrature points; (b) 50 quadrature points

is and many more function evaluations are needed to provide an adequate approximation of the integral. Even with $Q = 50$, it is common for only six function evaluations to be done in the interval where the first derivative function for the ith subject is non-zero.

These findings explain the large difference in the maximum likelihood estimate and associated P-value of the regression coefficient of the treatment effect in model (1) when going from $Q = 10$ to $Q = 50$ (with or without the interaction term). Owing to numerical inaccuracies, the likelihood surfaces for $Q = 10$ and $Q = 50$ differ considerably (Fig. 5). Owing to the multimodality of the likelihood surface, the maximization procedure could converge to a local maximum. Therefore, small differences between the likelihood values for $Q = 10$ and $Q = 50$ will cause different maximum likelihood solutions. The fact that the local curvature is inaccurately calculated for $Q = 10$ will then cause an artificially significant treatment effect.

4. Results from other statistical packages

The computational problems are not unique to MIXOR. We have analysed our model (1) (with and) without the interaction term with MIXOR, EGRET and the new SAS procedure `NLMIXED` (SAS Institute, 1999). For SAS procedure `NLMIXED` we have taken in this section the non-adaptive Gaussian quadrature procedure combined with two maximization procedures: Newton–Raphson and quasi-Newton. The program MIXOR is based on a Fisher scoring algorithm. For technical details about the maximization part, we refer the reader to the manuals of the statistical packages. Finally, for `NLMIXED` and MIXOR user-defined starting values equal to 0 for the regression parameters and 1 for σ were also supplied. The results for the model without the interaction term are summarized in Table 2. Although the SAS macro `GLIMMIX` (SAS Institute, 1996) is not based on Gaussian quadrature, we have applied this program to the onychomycosis data since it was, until recently, the SAS 'solution' for fitting generalized linear mixed models (Wolfinger, 1998).

The outcome of the SAS procedure `NLMIXED` did not depend on the starting values. Note that the default starting values for the SAS procedure `MIXED` are all equal to 1. Table 2 shows clearly that the estimated treatment effects vary considerably, both between programs as well as

Table 2. Estimated treatment effect obtained from various statistical packages performing a logistic random intercept analysis on model (1), without the interaction term†

Program	*Q*	*Starting value*	*Maximization routine*	*llik*	*treat*	*SE(treat)*	*P*	σ^2
MIXOR	10	Program	Fisher scoring	−635.8	−2.52	0.36	< 0.000001	12.7
	20			−627.4	−1.02	0.46	0.026	16.0
	10	User		−630.1	0.093	0.27	0.73	19.8
	20			−627.4	−1.02	0.46	0.026	16.0
	50	Program or user		−627.5	−0.51	0.56	0.36	16.1
NLMIXED	10	Program or user	Newton–Raphson	−909.6	−0.19	0.12	0.10	−0.38
	20			−909.6	−0.19	0.12	0.10	−0.25
	10	Program or user	Quasi-Newton	−635.8	−2.52	0.66	0.0002	12.7
	20			−627.4	−1.02	0.70	0.14	16.0
Adaptive	3	Program	Quasi-Newton	−633.1	−0.58	0.61	0.35	20.2
GLIMMIX	35	Program		−715.0	−0.30	0.33	0.35	5.6

†The first column specifies the program. When the program employs Gaussian quadrature, the number of quadrature points is indicated (Q). The third column indicates the choice of the initial parameter values (program, based on the program's internal procedure; user, 0 for the regression parameters and 1 for σ). The fourth column specifies the maximization procedure. The log-likelihood value at convergence is given in the fifth column (llik). The next three columns report the estimated treatment effect on convergence (treat), the estimated standard error (SE(treat)) and the corresponding *P*-value (from the Wald statistic). In the last column, the estimated variance of the random intercept is given.

within the same program. The same is true for the estimated variance of the random intercept; in some cases the NLMIXED procedure produced even a negative estimate. From the log-likelihood it is clear that the program was trapped in a local maximum. When constraining σ^2 to be non-negative the problem of a negative estimate for σ^2 was solved with the NLMIXED procedure and convergence was obtained to the global maximum, but only for $Q = 20$. Similar highly variable results were found when the interaction term is included. It is also important to observe that, in some cases, the program MIXOR and the SAS procedure NLMIXED produced the same parameter estimates for the treatment effect but with very different estimates for its standard error, which had a considerable effect on the *P*-value. Thus, switching from Fisher scoring to Newton–Raphson algorithms can have a large effect on the interpretation of the results.

No results for EGRET are reported in Table 2 as the program does not allow fitting both the treatment effect and the time effect (and interaction term) with a logistic random-effects model based on a normal distribution for the random effects and in combination with time-dependent covariates. For the model with only the treatment effect, EGRET obtains the same estimates as MIXOR and NLMIXED with the same standard error as NLMIXED. Finally, the results of GLIMMIX differ considerably from the other results, an observation which has also been reported by others (Wolfinger, 1998).

5. The adaptive Gaussian quadrature

For a unimodal, positive-valued function $g(z)$ (e.g. $P(z)$), Liu and Pierce (1994) (see also Pinheiro and Bates (1995)) suggested an adaptive Gauss–Hermite procedure to calculate the integral $\int_{-\infty}^{\infty} g(z)\,\mathrm{d}z$. For a non-adaptive Gauss–Hermite procedure this integral is rewritten as

$$\int_{-\infty}^{\infty} \left\{ \frac{g(z)}{\phi(z)} \right\} \phi(z)\,\mathrm{d}z = \int_{-\infty}^{\infty} h(z)\,\phi(z)\,\mathrm{d}z$$

and is approximated as in equation (4) by $\Sigma_{q=1}^{Q} h(z_q)\,\omega_q$. Thus, the function $h(z)$ is evaluated around 0. However, $g(z)$ is concentrated around its mode $\hat{\mu}$, because of its unimodality. Thus, when $\hat{\mu}$ lies remote from 0, for a small value of Q the above quadrature points z_q will be inappropriate. In that case the integral is better approximated by using quadrature points which are centred at $\hat{\mu}$, with spread depending on the shape of the function. Liu and Pierce (1994) suggested taking $\hat{\mu} + \hat{\tau} z_q \sqrt{2}$ $(q = 1, \ldots, Q)$ as quadrature points, with weights $\omega_q^* = \omega_q \exp(z_q^2)$. As indicated above, the variance parameter $\hat{\tau}^2$ depends on the shape of the function and is equal to $1/\sqrt{\hat{j}}$, with

$$\hat{j} = -\frac{\partial^2}{\partial z^2} \log\{g(z)\}\bigg|_{z=\hat{\mu}}.$$

The integral $\int_{-\infty}^{\infty} g(z)\,\mathrm{d}z$ is now approximated by

$$\hat{\sigma}\sqrt{2} \sum_{q=1}^{Q} \omega_q^*\, g(\hat{\mu} + \hat{\tau} z_q \sqrt{2}).$$

The choice of the quadrature points with the adaptive method is illustrated in Fig. 2. It is clearly seen that now the quadrature points cover the interval of interest (where the density is well above 0) much better.

This is the approach that is adopted by the default option in the new SAS procedure `NLMIXED`. However, the maximization routine also involves the calculation of the first and second derivatives, which imply the calculation of other integrals. These integrands are often not unimodal and often not even positive valued. Hence, the approach of Liu and Pierce (1994) cannot be applied directly here. Therefore the same quadrature points with the same weights could be used to approximate the integral involved in the first and second derivatives. The SAS procedure `NLMIXED` solves this problem by calculating the first and the second derivatives of the approximation instead of approximating the derivatives.

Fig. 3 shows that the adaptive Gaussian quadrature method produces log-likelihood values with $Q = 10$ that are close to those from the non-adaptive Gaussian quadrature with $Q = 50$. Thus, with the adaptive version the integrals are evaluated with lower quadrature points. This does not imply fewer function evaluations, though, as for each subject the mode of the function

$$-\log\left\{\prod_{j=1}^{n_i} P_\sigma(Y_{ij} = y_{ij} | \boldsymbol{\beta}, z_i)\,\phi(z_i)\right\}$$

as a function of z_i needs to be determined. In other words, the empirical Bayes estimate of z_i needs to be calculated for the ith subject (see the manual for `PROC NLMIXED` (SAS Institute, 1999)).

Fig. 4 shows that the individual marginalized probabilities and the components of the first derivative with respect to β_1 very much coincide for the non-adaptive Gaussian quadrature method with $Q = 50$ and the adaptive version with $Q = 10$. Further, Table 2 shows that the ordinary and adaptive Gaussian procedures with $Q = 10$ and $Q = 50$ respectively have a very similar output except for the estimate of σ^2.

For model (1) with the interaction term, the SAS procedure `NLMIXED` did not converge with the default starting values. When starting values for the parameters are equal to the estimates from the ordinary Gaussian quadrature analysis, convergence was easily obtained in eight iterations with the adaptive procedure. However, in this case, the final estimates did not differ much from the starting values.

6. Discussion

At the start of the analysis, we were aware of the dependence of the outcome of a logistic random-effects model on the number of quadrature points. But our experience was similar to the recommendation found in MIXOR, namely that $Q = 10$ is often sufficient and, when differences are found by increasing Q, they are minimal. Hence, we believed that the Gauss–Hermite method is robustly calculating the subject-specific estimates of the parameters and did not need a routine check, as opposed to the methods on which the SAS macro `GLIMMIX` is based (Schall, 1991; Breslow and Clayton, 1993; Wolfinger, 1993). However, the extremeness of the difference in the estimated treatment effect which we obtained under the two methods completely surprised us and others. Further, it was the highly significant treatment effect at the base-line (from the logistic random-effects analysis), combined with the fact that we were dealing with a randomized study, that inspired us to investigate the calculations further. We can imagine that in other circumstances there will be no such incentive. It is therefore important to reiterate that, even for the more robust method based on Gaussian quadrature, the adequacy of the numerical procedure needs to be assessed even in simple random-effects analyses.

We were therefore pleased to see that adaptive Gaussian quadrature can be used in `NLMIXED`, which has been recently released by SAS. However, it is our experience that, even with adaptive Gaussian quadrature and with relatively simple models, convergence to a global maximum can be difficult to obtain. This emphasizes the computational difficulties with random-effects models for categorical outcome data.

Acknowledgements

The authors thank Novartis, Belgium, for permission to use their dermatological data for statistical research, and more specifically Dr Marc De Backer. The authors also thank the Associate Editor and the two referees for helpful comments which improved the paper considerably.

References

Breslow, N. E. and Clayton, D. G. (1993) Approximate inference in generalized linear mixed models. *J. Am. Statist. Ass.*, **88**, 9–25.

Cytel Software Corporation (1995) *EGRET*. Cambridge: Cytel Software Corporation.

De Backer, M., De Vroey, C., Lesaffre, E., Scheys, I. and De Keyser, P. (1998) Twelve weeks of continuous onychomycosis caused by dermatophytes: a double blind comparative trial of terbafine 250 mg/day versus itraconazole 200 mg/day. *J. Am. Acad. Derm.*, **38**, S57–S63.

Diggle, P., Liang, K.-Y. and Zeger, S. L. (1994) *Analysis of Longitudinal Data*. Oxford: Clarendon.

Hedeker, D. and Gibbons, R. D. (1994) A random-effects ordinal regression model for multilevel analysis. *Biometrics*, **50**, 933–944.

———(1996) MIXOR: a computer program for mixed-effects ordinal regression analysis. *Comput. Meth. Programs Biomed.*, **49**, 157–176.

Liang, K.-Y. and Zeger, S. L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Liu, Q. and Pierce, D. A. (1994) A note on Gauss-Hermite quadrature. *Biometrika*, **81**, 624–629.

Longford, N. T. (1993) *Random Coefficient Models*. Oxford: Oxford University Press.

Neuhaus, J. M., Hauck, W. W. and Kalbfleisch, J. D. (1992) The effects of mixture distribution specification when fitting mixed-effects logistic models. *Biometrika*, **79**, 755–762.

Pinheiro, J. C. and Bates, D. M. (1995) Approximations to the log-likelihood function in the non-linear mixed-effects model. *J. Comput. Graph. Statist.*, **4**, 12–35.

SAS Institute (1996) *The SAS System for Windows, Release V6.12*. Cary: SAS Institute.

———(1999) *The SAS System for Windows, Release V8.0*. Cary: SAS Institute.

Schall, R. (1991) Estimation in generalized linear models with random-effects. *Biometrika*, **78**, 719–727.

Wolfinger, R. W. (1993) Laplace's approximation for nonlinear mixed models. *Biometrika*, **80**, 791–795.

———(1998) Towards practical application of generalized linear mixed models. In *Proc. 13th Int. Wrkshp Statistical Modelling, New Orleans* (eds B. Marx and H. Friedl), pp. 388–395.