Biometrika (2003), **90,** 4, *pp.* 765–775

Printed in Great Britain

Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function

By ZENGRI WANG

3M Company, Saint Paul, Minnesota 55144-1000, U.S.A.

zwang@mmm.com

AND THOMAS A. LOUIS

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland 21205-2179, U.S.A.

tlouis@jhsph.edu

Summary

Random effects logistic regression models are often used to model clustered binary response data. Regression parameters in these models have a conditional, subject-specific interpretation in that they quantify regression effects for each cluster. Very often, the logistic functional shape conditional on the random effects does not carry over to the marginal scale. Thus, parameters in these models usually do not have an explicit marginal, population-averaged interpretation. We study a bridge distribution function for the random effect in the random intercept logistic regression model. Under this distributional assumption, the marginal functional shape is still of logistic form, and thus regression parameters have an explicit marginal interpretation. The main advantage of this approach is that likelihood inference can be obtained for either marginal or conditional regression inference within a single model framework. The generality of the results and some properties of the bridge distribution functions are discussed. An example is used for illustration.

Some key words: Bridge distribution function; Clustered data; Gaussian-Hermite quadrature; Marginal model; Random effects model.

1. Introduction

Examples of clustered data include repeated measurements in longitudinal study, paired observations on both eyes in ophthalmology research and family data in genetics study. It is well known that observations within a cluster tend to exhibit intracluster correlation, which has to be appropriately considered for valid inferences.

The random intercept logistic regression model is commonly used to analyse clustered binary data. It is a natural extension of the traditional logistic regression model in which the random intercept is allowed to vary across clusters according to a distribution. The model has the following form:

$$\text{logit}\{\text{pr}(Y_{ij}=1\,|\,b_i, X_{ij})\} = b_i + \alpha_s^{\mathrm{T}} X_{ij}, \tag{1·1}$$

where the binary Y_{ij} is the observation on the jth unit for the ith cluster ($i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n_i$), X_{ij} is a $p \times 1$ covariate vector and α_s is a $p \times 1$ vector of regression parameters. The b_i are cluster-specific random effects that are independent and identically distributed as G given X_i, where X_i contains all covariates for cluster i. Within each cluster, observations Y_{ij} are independent conditional on X_i and b_i. Since the intercept term can be absorbed into $\alpha_s^{\mathrm{T}} X_{ij}$, we assume that $E(b_i \mid X_i) = 0$.

The parameter α_s in model (1·1) measures the change in the conditional logit of the probability of response with the covariate X_{ij} for units in each of the underlying risk groups described by b_i (Neuhaus et al., 1991). Note that, in the marginal scale,

$$\mathrm{pr}(Y_{ij} = 1 \mid X_{ij}) = \int \frac{e^{b_i + \alpha_s^{\mathrm{T}} X_{ij}}}{1 + e^{b_i + \alpha_s^{\mathrm{T}} X_{ij}}} \, dG(b_i \mid X_i), \tag{1·2}$$

which is usually not of logistic form. Thus, the parameter α_s usually does not have an explicit marginal, population-averaged interpretation.

On the other hand, for a marginally specified assessment, the marginal distribution (1·2) may be modelled directly by

$$\mathrm{logit}\{\mathrm{pr}(Y_{ij} = 1 \mid X_{ij})\} = \alpha_p^{\mathrm{T}} X_{ij}, \tag{1·3}$$

where α_p measures a marginal regression effect associated with the covariate X_{ij}. The intracluster correlation is commonly modelled by adopting a 'working' covariance structure and parameter estimation can be carried out by solving generalised estimating equations (Liang & Zeger, 1986).

The relationships of regression parameters between conditional and marginal regression models have been studied by several authors (Neuhaus et al., 1991; Zeger et al., 1988; Chao et al., 1997). It was noted that incorporating a random effect in a conditional model usually induces a very complicated interaction between covariates in the marginal mean. It was observed that parameter estimators in the random intercept logistic regression models are relatively robust with respect to different distributional assumptions (Neuhaus et al., 1992; Heagerty & Kurland, 2001).

In this paper, we study a special distribution function for the random intercept logistic regression model. Under this distributional assumption, the marginal functional shape is still logistic and regression parameters have an explicit marginal interpretation. In § 2 this special distribution is identified and some of its properties are discussed. The generality of the results is investigated in § 3. Parameter estimation and interpretation are given in § 4. In § 5 an example is presented for illustration, followed by a discussion in § 6.

2. Derivation and properties of the bridge function

Consider the random intercept logistic model (1·1) but with a general inverse link function $H(.)$. For ease of discussion, we first assume the same random effects distribution across clusters, that is $G(b_i \mid X_i) := G(b_i)$. We relax this condition to allow b_i to depend on X_i later. It is of interest whether or not there exists a distribution $G(b)$ such that the conditional functional shape is retained in the marginal scale, i.e.

$$\int H(b + \alpha_s^{\mathrm{T}} X) \, dG(b) = H(k + \phi \alpha_s^{\mathrm{T}} X), \tag{2·1}$$

where H is assumed to be monotone, increasing and twice differentiable with finite $H(\pm\infty)$, and k and ϕ are unknown parameters. It is easy to see that $k = 0$ when H is the cumulative distribution function of a symmetric distribution. This is not necessarily true for a general inverse link H. For notational convenience, the subscripts i and j are suppressed.

One can see that equation (2·1) holds for the degenerate distribution $G(b)$. In this scenario, no heterogeneity exists across clusters and the random intercept model is equivalent to a fixed-effects marginal model with independent observations within each cluster.

Let $\eta = \alpha_s^{\mathrm{T}} X$. For any nondegenerate $G(b)$, differentiation of (2·1) with respect to η yields

$$\int h(b+\eta)\, dG(b) = \phi h(k+\phi\eta), \tag{2·2}$$

where $h = H'$. Note that the above equation can be written as

$$h * g_{-b}(\eta) = \phi h(k+\phi\eta),$$

where g_{-b} is the probability density function of $-b$ and the symbol $*$ indicates the convolution operation.

It is easy to see that h and g_{-b} are nonnegative and integrable. Take Fourier transforms of both sides in the above equation, so that

$$\mathscr{F}\{h * g_{-b}(\eta)\} = \mathscr{F}\{\phi h(k+\phi\eta)\},$$

where $\mathscr{F}$ represents the Fourier transform operation, defined as

$$\mathscr{F}\{f(x)\} = \mathscr{F}f(\xi) = \int e^{-i\xi x} f(x)\, dx, \quad f \in L^1(R).$$

Since $\mathscr{F}\{h * g_{-b}(\eta)\} = \mathscr{F}h(\xi)\mathscr{F}g_{-b}(\xi)$ and $\mathscr{F}\{\phi h(k+\phi\eta)\}(\xi) = e^{ik\xi/\phi}\mathscr{F}h(\xi/\phi)$, we obtain

$$\mathscr{F}g_{-b}(\xi) = e^{ik\xi/\phi}\mathscr{F}h(\xi/\phi)/\mathscr{F}h(\xi).$$

Denote the right-hand side of the above equation by $v(\xi)$. For ϕ such that $v(\xi)$ is integrable, by the Fourier Inversion Theorem, the probability density function of b is given by

$$g_b(x) = \frac{1}{2\pi}\int e^{i(k/\phi - x)\xi}\frac{\mathscr{F}h(\xi/\phi)}{\mathscr{F}h(\xi)}\, d\xi. \tag{2·3}$$

We refer to this function as the bridge density function. It is easy to see that $\int g_b(x)\, dx = 1$ since $\mathscr{F}g_b(0) = 1$. Note that h is nonnegative, so $\phi > 0$ from (2·2). For $\phi \geqslant 1$, $\mathscr{F}h(\xi/\phi)/\mathscr{F}h(\xi)$ does not decrease to zero as $|\xi| \to 0$ so it is not integrable. Therefore, in this case, g_b does not exist unless $\phi = 1$, in which case g_b is degenerate. The rescaling parameter ϕ between 0 and 1 is consistent with the effect of attentuation (Zeger et al., 1988; Neuhaus et al., 1991).

The result (2·3) is still valid when b depends on the covariate X since X is held fixed conditionally in (2·1). However, in this situation, the rescaling parameter becomes a function of X, $\phi(X)$ say.

For the logit link, $H(\eta) = e^{\eta}(1+e^{\eta})^{-1}$ and $h(\eta) = e^{\eta}(1+e^{\eta})^{-2}$. The characteristic function of the logistic distribution is $\varphi(t) = 2\pi t/(e^{\pi t} - e^{-\pi t})$, with $\varphi(0)$ defined to be $\lim_{t\to 0}\varphi(t) = 1$. The Fourier transform of the function $f(x) = \sinh(ax)/\sinh(\pi x)$ is given by

$$\mathscr{F}f(\xi) = \sin(a)/\{\cosh(\xi) + \cos(a)\} \quad (-\pi < a < \pi)$$

(Selby, 1974, p. 526). By changing variables, we obtain the bridge density function for the logit link as

$$g_b(x) = \frac{1}{2\pi}\frac{\sin(\phi\pi)}{\cosh(\phi x) + \cos(\phi\pi)} \quad (0 < \phi < 1, -\infty < x < \infty). \tag{2·4}$$

The characteristic function of the above bridge density function is given by

$$\varphi_b(t) = \sinh(\pi t)/\{\phi \sinh(\pi t/\phi)\} \quad (0 < \phi < 1).$$

The moments of the bridge density function are obtainable directly from the Taylor expansion of $\varphi_b(t)$ at $t = 0$, namely

$$\varphi_b(t) = 1 - \frac{\pi^2}{6}\left(\frac{1}{\phi^2} - 1\right) t^2 + O(t^4).$$

Thus the mean is $\mu = 0$ and the variance is $\sigma_b^2 = \pi^2(\phi^{-2} - 1)/3$.

Note that, when $\phi = \frac{1}{2}$, a simple form of the bridge density function is available:

$$g_b(x) = \frac{1}{2\pi \cosh(x/2)} \quad (-\infty < x < \infty).$$

We denote the distribution (2·4) by $B_l(0, \phi)$, where 0 is the mean, ϕ is the rescaling parameter and the subscript 'l' implies 'logistic'. The distribution is symmetric and has a different shape from that of the Gaussian distribution. Figure 1 presents the probability density function for B_l, Gaussian and logistic distributions each with zero mean and unit variance. One can see that B_l has slightly heavier tails than Gaussian and lighter tails than logistic.

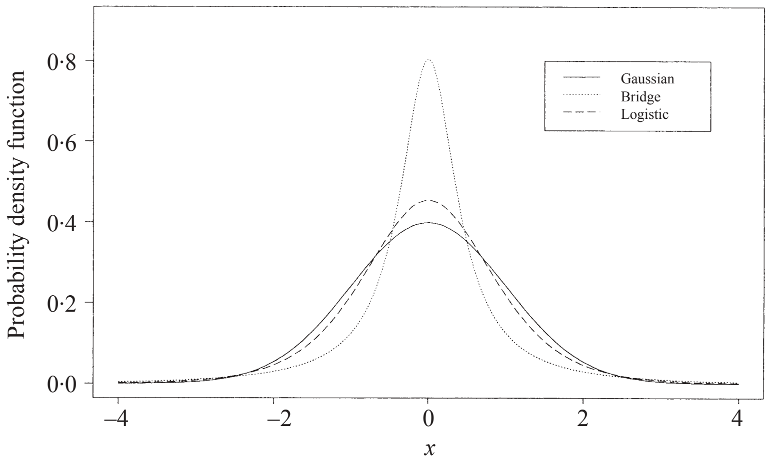

Fig. 1. Probability density functions of the Gaussian, logistic and bridge, for logistic, distributions each with zero mean and unit variance.

PROPOSITION 1. *The bridge distribution for the logit link has the following properties.*

(i) *Function* $\{e^{\phi B_l(0,\phi)} + \cos(\phi\pi)\}/\sin(\phi\pi)$ *has the same distribution as the left-truncated standard Cauchy distribution defined on* $x > \cos(\phi\pi)/\sin(\phi\pi)$ *for* $\phi \in (0, 1)$. *In particular,* $e^{B_l(0,1/2)/2}$ *is distributed as the left-truncated standard Cauchy on* $x > 0$.

(ii) *Distribution* $B_l(0, \phi)$, *for* $\phi \in (0, 1)$, *is a scale mixture of symmetric stable distributions. In particular, it is a scale mixture of Gaussians.*

(iii) *Distribution* $K_{\phi\delta}$ *has the same distribution as* $K_\delta e^{\delta^{-1} B_l(0,\phi)}$ $(\delta > 0, 0 < \phi < 1)$, *where* K_δ *is the Linnik distribution with index parameter* δ.

The proofs can be found in the Appendix. A result similar to (iii) was found by Kotz

& Ostrovskii (1996) by using complex contour integration. Here we offer a different proof and the relationship between the logistic and its bridge distribution is addressed.

Under the bridge distributional assumption, the regression parameters in the marginal and conditional logistic models are related by

$$\alpha_p = \frac{\alpha_s}{\surd(1 + 3\sigma_b^2/\pi^2)}. \tag{2·5}$$

In addition, straightfoward calculations show that

$$\alpha_p = \alpha_s(1 - \rho_Y), \tag{2·6}$$

where $\rho_Y = \text{corr}(Y_{ij}, Y_{ij'} \mid \alpha_s = 0)$ is the intracluster correlation in the binary response scale. This relationship, which was obtained by Neuhaus et al. (1991) using a one-step Taylor series expansion, holds exactly under the bridge distribution.

3. Generality of the results

3·1. *Multivariate distribution*

It does not seem straightforward to generalise the bridge distribution of the random intercept logistic regression model to multivariate random effect distributions. To see this, consider a mixed-effects logistic model

$$\text{logit}\{\text{pr}(Y_{ij} = 1 \mid b_i, X_{ij})\} = b_i^{\mathrm{T}} Z_{ij} + \alpha_s^{\mathrm{T}} X_{ij}, \tag{3·1}$$

where Z_{ij} is $q \times 1$ design vector, which may be a subset of the covariate vector X_{ij}, and b_i is a $q \times 1$ vector of random effects.

In order to retain the logit shape in the marginal mean, $b_i^{\mathrm{T}} Z_{ij}$ must follow the bridge distribution discussed previously. However, as one can easily verify, the bridge distribution B_l is not closed under linear combination. Thus, even if each component of the random vector b_i follows the bridge distribution, $b_i^{\mathrm{T}} Z_{ij}$ does not necessarily follow the same distribution. Therefore, a straightforward generalisation to multivariate random effect distributions does not seem possible.

In practice, there can be covariate group-specific heterogeneity. One simple way of extending of the bridge distribution to accommodate this scenario is to relate the variance of the bridge distribution in (1·1) to covariates via a link function. For example, the log link may be used, so that

$$\log\{\sigma_b^2\} = \beta_s^{\mathrm{T}} Z_{ij}.$$

In this situation, the marginal shape remains logistic and the rescaling parameter becomes dependent on Z_{ij}, that is $\phi(Z_{ij}) = (1 + 3\pi^{-2} e^{\beta_s^{\mathrm{T}} Z_{ij}})^{-\frac{1}{2}}$.

When the random effects are nested, for example of the form $b_i + b_{ij}$, the bridge distribution can still be used to preserve marginalisation. To see this, assume that $b_{ij} \mid (b_i, X_i) \sim B_l(0, \phi_1)$. Then the marginal mean is given by

$$\int H(b_{ij} + b_i + \alpha_s^{\mathrm{T}} X_{ij})\, dG(b_{ij} \mid b_i, X_i)\, dG(b_i \mid X_i) = \int H(\phi_1 b_i + \phi_1 \alpha_s^{\mathrm{T}} X_{ij})\, dG(b_i \mid X_i),$$

where $H(\eta) = e^{\eta}/(1 + e^{\eta})$. If we assume that $b_i \mid X_i \sim \phi_1^{-1} B_l(0, \phi_2)$, not $B_l(0, \phi_2)$, then the marginal mean is given by $H(\phi_1 \phi_2 \alpha_s^{\mathrm{T}} X_{ij})$. However, the marginalisation does not hold if one simply assumes that $b_{ij} \mid (b_i, X_i) \sim B_l(0, \phi_1)$ and $b_i \mid X_i \sim B_l(0, \phi_2)$. The key is that ϕ is not a scale parameter in the bridge distribution and B_l is not closed under linear combination.

3·2. *Other link functions*

The idea of bridge distribution is easily generalised to other link functions. When the link is the probit, it is straightforward to verify that its bridge distribution is Gaussian. Mixed-effects probit regression models with Gaussian random effects have been discussed by many authors, such as Chao et al. (1997) and Heagerty & Zeger (2000). The rescaling parameter in this case is given by

$$\phi = \frac{1}{\sqrt{(1+\sigma_b^2)}},$$

where σ_b^2 is the variance of the bridge distribution $N(0, \sigma_b^2)$.

It is worth noting that, for the probit link, a simple generalisation of the bridge distribution to multivariate bridge distributions is available since the Gaussian family is closed under linear combination. Furthermore, it is straightforward to verify that such a property still holds when the link is the inverse cumulative distribution function of a positive stable distribution.

Another commonly-used link function is the complementary log-log, with inverse function

$$H(\eta) = 1 - e^{-e^{\eta}}.$$

In this case, $h(\eta) = e^{\eta - e^{\eta}}$ and straightforward calculations show that $\mathscr{F}h(\xi) = \Gamma(1 - i\xi)$, where Γ is the gamma function. If we follow previous arguments, the bridge density function is given by

$$g_b(x) = \frac{1}{2\pi} \int e^{i(k/\phi - x)\xi} \frac{\Gamma(1 - i\xi/\phi)}{\Gamma(1 - i\xi)} \, d\xi \quad (0 < \phi < 1).$$

It can be shown that, for $0 < \phi < 1$, $\Gamma(1 - i\xi/\phi)/\Gamma(1 - i\xi)$ is a characteristic function of a random variable and is therefore integrable. Therefore, the above bridge density function $g_b(x)$ exists. In a survival context, survival and hazard functions are closely related to the complementary log-log relationship. Hougaard (1986) showed that the common conditional survival distribution with a random effect, or frailty, is retained marginally in Weibull models when the random effect follows the positive stable distributions. Thus, the bridge distribution for the complementary log-log link is the log-positive stable distribution. It is straightforward to verify that the rescaling parameter in this case is given by

$$\phi = \frac{1}{\sqrt{(1 + 6\pi^{-2}\sigma_b^2)}},$$

where σ_b^2 is the variance of the bridge distribution.

Other link functions exist for which the bridge density functions have explicit analytical forms. For example, when H is the cumulative distribution function of the Cauchy distribution, that is $H(\eta) = \pi^{-1}\{\pi/2 + \arctan(\eta)\}$, $\mathscr{F}h(\xi) = e^{-|\xi|}$ and the corresponding bridge distribution is the Cauchy distribution.

The rescaling parameter ϕ is comparable to the 'attenuation factor' in Neuhaus & Jewell (1993) for the noted links and the 'bias factor' in Chao et al. (1997) for the probit link. For a general link function, direct evaluation of the bridge density function (2·3) seems difficult and explicit analytical forms for the bridge density function and ϕ may not be available.

4. Estimation and interpretation

Parameter estimation for the random intercept logistic model with the bridge distribution can be carried out by evaluating the marginal likelihood function. Assume that we observe the dataset

$$\Omega = \{y_{ij}, x_{ij};\ i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n_i\}.$$

Let $y_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})^{\mathrm{T}}$ and $x_i = (x_{i1}, x_{i2}, \ldots, x_{in_i})^{\mathrm{T}}$. The joint probability function for the ith cluster is given by

$$\begin{aligned}\mathscr{L}_i(\alpha_s, \theta \mid \Omega) &= \int \prod_{j=1}^{n_i} \mathrm{pr}(Y_{ij} = y_{ij} \mid b_i, x_i)\, dG(b_i \mid x_i)\\ &= \int \frac{\exp\{\sum_{j=1}^{n_i} y_{ij}(b_i + \alpha_s^{\mathrm{T}} x_{ij})\}}{\prod_{j=1}^{n_i}\{1 + \exp(b_i + \alpha_s^{\mathrm{T}} x_{ij})\}}\, dG_{B_l \mid X_i},\end{aligned} \tag{4·1}$$

where $G_{B_l \mid X_i}$ is the bridge distribution and θ is a vector containing all the unknown parameters in $G_{B_l \mid X_i}$. The marginal likelihood function becomes

$$\mathscr{L}(\alpha_s, \theta) = \prod_{i=1}^{m} \mathscr{L}_i(\alpha_s, \theta \mid \Omega).$$

The maximum likelihood estimates of the parameter vector (α_s, θ) are obtained by maximising the likelihood function $\mathscr{L}(\alpha_s, \theta)$. The variance-covariance matrix for the parameter estimators is estimated by evaluating the inverse Hessian matrix at the maximum likelihood estimates $(\hat{\alpha}_s, \hat{\theta})$.

The integral over the random effect b_i in (4·1) is commonly evaluated using Gaussian-Hermite quadrature approximation when the random effects are Gaussian (Hedeker & Gibbons, 1994; Pinheiro & Bates, 1995). Computational details are given in the Appendix.

The parameter α_s in (1·1) has a conditional, subject-specific interpretation (Zeger et al., 1988; Neuhaus et al., 1991). As one can see, under the bridge distribution, the parameter α_s also has a marginal, population-averaged interpretation at a special scale $\phi(X)$ that characterises the amount of heterogeneity across clusters. When the magnitude of the heterogeneity is larger, the marginal covariate effects shrink towards zero.

When the rescaling parameter ϕ is independent of X, direct assessment of marginal effects α_p can be made by replacing α_s in (1·1) by $\phi^{-1}\alpha_p$. The model then becomes

$$\mathrm{logit}\{\mathrm{pr}(Y_{ij} = 1 \mid b_i, X_{ij})\} = b_i + \phi^{-1}\alpha_p^{\mathrm{T}} X_{ij}, \tag{4·2}$$

where $b_i \mid X_i \sim B_l(0, \phi)$ and $\mathrm{logit}\{\mathrm{pr}(Y_{ij} = 1 \mid X_{ij})\} = \alpha_p^{\mathrm{T}} X_{ij}$ marginally.

A prediction of $\mathrm{pr}(Y_{ij} = 1 \mid b_i, X_{ij})$ for the jth unit in the ith cluster can be obtained by evaluating its value at $(\hat{\alpha}_s, \hat{b}_i)$, where $\hat{\alpha}_s$ is the maximum likelihood estimate and $\hat{b}_i$ is the empirical Bayes estimate of b_i (Carlin & Louis, 2000, pp. 57–78).

5. An example

A longitudinal dataset from Stokes et al. (2000, p. 461) concerned a clinical trial that was designed to compare two treatments for a respiratory illness. There were two centres and a total of 111 patients. Patients in each of the two centres were randomly assigned to the active treatment or a placebo. The outcome of interest is 'respiratory status', Y, categorised as 0 (poor) or 1 (good). Measurements were taken at four different visits. Potential explanatory variables included a centre indicator X_1, equal to 0 for centre A and 1 for centre B, a sex indicator X_2, 0 for male and 1 for female, age in years X_3, and baseline status X_4, equal to 0 for poor and 1 for good. There were 54 patients in the active group and 57 patients in the placebo group.

We fitted the random intercept logistic model (1·1) to the data assuming the bridge or Gaussian distribution for the random effect. The model is

$$\text{logit}\{\text{pr}(Y_{ij}=1 \mid b_i, X_{ij})\} = \mu_s + b_i + \alpha_{1s}^{\mathrm{T}} X_{ij1} + \alpha_{2s}^{\mathrm{T}} X_{ij2} + \alpha_{3s}^{\mathrm{T}} X_{ij3} + \alpha_{4s}^{\mathrm{T}} X_{ij4},$$

where $i = 1, 2, \ldots, 111$, $j = 1, 2, 3, 4$ and the b_i are distributed as B_l or Gaussian with mean 0 and variance σ_b^2. The resulting parameter estimates are presented in Table 1. The estimates based on the two models are very similar. The estimated conditional log odds ratios measuring the treatment effect are both highly significant, as are the estimated standard deviations of the random effect, indicating considerable heterogeneity across patients. The rescaling parameter ϕ for the model with the bridge distribution is estimated as 0·621, so that in the marginal scale the log odds ratio measuring the marginal treatment effect is approximately $2{\cdot}105 \times 0{\cdot}621 = 1{\cdot}307$.

Table 1: *Respiratory illness data. Estimates of conditional regression parameters and their standard errors,* SE, *for random intercept logistic regression models with the bridge distribution and the Gaussian distribution*

	Gaussian			Bridge		
Covariate	Estimate	SE	p-value	Estimate	SE	p-value
Intercept	−2·380	0·547	<0·001	−2·421	0·560	<0·001
Treatment	2·028	0·545	<0·001	2·105	0·553	<0·001
Age	−0·027	0·020	0·186	−0·029	0·021	0·161
Sex	0·244	0·673	0·717	0·389	0·688	0·573
Centre	0·980	0·546	0·075	0·818	0·553	0·142
Baseline	2·921	0·581	<0·001	2·883	0·578	<0·001
σ_b	1·985	0·326	<0·001	2·290	0·376	<0·001
ϕ				0·621		

We also fitted the marginally specified random intercept logistic model (4·2) to assess the marginal effect directly. For comparison, a marginal generalised estimating equation model with the exchangeable correlation structure was also fitted. The results are presented in Table 2. The parameter estimates and standard errors of the marginal effects shrink towards zero when compared with those from the conditional models. The Wald tests of regression effects appear to be consistent between the marginal and conditional models. The marginal regression effects from the marginally specified model are very comparable to those from the generalised estimating equation model. For the marginally and conditionally specified random intercept models with the bridge distribution, the estimates of the covariate effects approximately satisfy the relationship $\alpha_p = \phi\alpha_s$, which is consistent with our previous discussion: the rescaling factor ϕ reflects the magnitude of the difference between the conditional and marginal regression effects.

Table 2: *Respiratory illness data. Estimates of marginal regression parameters and their standard errors,* SE, *for random intercept logistic regression models with the bridge distribution and the marginal generalised estimating equation,* GEE, *model with the exchangeable correlation structure. Robust standard errors are reported for the generalised estimating equation model*

Covariate	Bridge Estimate	Bridge SE	Bridge p-value	GEE Estimate	GEE SE	GEE p-value
Intercept	$-1{\cdot}507$	0·336	$<0{\cdot}001$	$-1{\cdot}480$	0·343	$<0{\cdot}001$
Treatment	1·309	0·335	$<0{\cdot}001$	1·265	0·347	$<0{\cdot}001$
Age	$-0{\cdot}018$	0·013	0·162	$-0{\cdot}019$	0·013	0·148
Sex	0·242	0·428	0·573	0·137	0·440	0·756
Centre	0·509	0·341	0·139	0·650	0·353	0·066
Baseline	1·793	0·343	$<0{\cdot}001$	1·846	0·346	$<0{\cdot}001$
σ_b	2·299	0·370	$<0{\cdot}001$			
ϕ or ρ	0·621			0·327		

6. Discussion

Other approaches have been proposed for estimating regression effects for clustered binary data for random effects models. One is to use conditional likelihood to estimate the effects of within-cluster covariates; the random effects are eliminated and therefore the distribution of the random effects need not be specified (Breslow & Day, 1980, p. 248). However, the effects of between-cluster covariate effects cannot be estimated by this approach, nor can features of the random effects. Another approach is to model the random effects nonparametrically and to estimate the random effects and covariates effects jointly (Laird, 1978; Follmann & Lambert, 1989). However, marginal effects cannot be assessed in this approach.

The bridge distribution function has a special functional form and this may limit its use in practice. However, it has been observed that regression effects in the random intercept logistic models are quite robust to different distributional assumptions for the random effect (Neuhaus et al., 1992; Heagerty & Kurland, 2001). In practice, it is often unlikely to have enough information to distinguish between different random effects distributions for a given dataset. The bridge distribution offers a useful tool for conditional inference while allowing for meaningful marginal assessments of regression effects.

Acknowledgement

The authors wish to thank the editor and two referees for their constructive comments and suggestions. The authors are grateful to Professors W. Dunsmuir, R. C. Pruitt and W. Pan for reading an early version of this paper.

Appendix

Technical details

Proof of Proposition 1. Part (i) is straightforward. For part (ii), the characteristic function of the symmetric stable distribution S_a is $\varphi_a(t)=e^{-|t|^a/2}$ $(0<a\leqslant 2)$. Note that, when $a=2$, S_a is the standard Gaussian. Consider a class of distributions defined by $L_a=\{X \mid X\sim S_a V, V>0\}$. It can be seen that $X\in L_a$ if and only if its characteristic function $\varphi_X(t)$ is of the form $\varphi_X(t)=$

$\int_0^\infty e^{-|tv|^a}\, dG_v(v)$, where G_v is the distribution of a random variable V. By the result of Keilson & Steutel (1972), it is easy to verify that the class L_a consists of all random variables whose characteristic function $\varphi(t)$ is completely monotone in $|t|^a$, that is $\varphi(|t|^{1/a})$ is completely monotone on $(0, \infty)$. Note that any function that is completely monotone in x is also completely monotone in x^p, for $p \geqslant 1$. Thus $L_{a_1} \subset L_{a_2}$ for $a_1 < a_2$. Also, L_a is closed under mixing and convolution. Since the characteristic function of $B_l(0, \phi)$ is $\varphi_{B_l}(t) = \sinh(\pi t)/\{\phi \sinh(\pi t/\phi)\}$, $\phi \in (0, 1)$, it follows that $\varphi_{B_l}(|t|^{1/a}) = \sinh(\pi|t|^{1/a})/\{\phi \sinh(\pi|t|^{1/a}/\phi)\}$, for $a \in (0, 2]$. Direct verification that $\varphi_{B_l}(|t|^{1/a})$ is completely monotone on $(0, \infty)$ yields part (ii). For part (iii), the Linnik distribution K_δ (Linnik, 1963) has characteristic function $\varphi_\delta(t) = (1 + |t|^\delta)^{-1}$, for $\delta \in (0, 2]$. It is easy to see that $e^{L/(\phi\delta)}$ has the same distribution as $e^{L/\delta + B_l(0,\,\phi)/\delta}$, for $\phi \in (0, 1)$, where L is the logistic distribution and B_l is the bridge distribution. Since the distribution function of $e^{L/(\phi\delta)}$ is $F(x) = 1 - (1 + |x|^{\phi\delta})^{-1}$, for $x > 0$, straightforward calculations yield $(1 + |x|^{\phi\delta})^{-1} = \int (1 + |x\sigma|^\delta)^{-1}\, dG(\sigma)$, where $G(\sigma)$ is the distribution of $e^{B_l/\delta}$. This proves part (iii). □

Evaluation of maximum likelihood estimates. Denote the target integral in (4·1) by $\int r(x; b)\, dG(b|x)$, where b follows the bridge distribution. Let $\tilde{b} = \Phi^{-1}\{G(b|x)\}$, where Φ is the cumulative distribution function of the standard Gaussian distribution. Then $\tilde{b} \sim N(0, 1)$. Let $v(.) = G^{-1}\{\Phi(.)\}$. Then

$$\int r(x; b)\, dG(b|x) = \int r\{x; v(\tilde{b})\}\, dG\{v(\tilde{b}|x)\} = \int r\{x; v(\tilde{b})\}\, d\Phi(\tilde{b}|x). \tag{A·1}$$

Evaluation of the integral $\int r(x; b)\, dG(b|x)$ is therefore equivalent to the evaluation of the integral $\int r\{x; v(\tilde{b})\}\, d\Phi(\tilde{b}|x)$ with $\tilde{b} \sim N(0, 1)$. Thus, the commonly used Gaussian–Hermite quadrature method can be readily applied here.

For the logit link, the cumulative distribution function for the bridge distribution is given by

$$G(x) = 1 - \frac{1}{\pi\phi}\left[\frac{\pi}{2} - \arctan\left\{\frac{e^{\phi x} + \cos(\phi\pi)}{\sin(\phi\pi)}\right\}\right] \quad (0 < \phi < 1,\ -\infty < x < \infty) \tag{A·2}$$

and its inverse is

$$G^{-1}(x) = \frac{1}{\phi}\log\left[\frac{\sin(\phi\pi x)}{\sin\{\phi\pi(1 - x)\}}\right] \quad (0 < \phi,\ x < 1). \tag{A·3}$$

If we combine (A·1), (A·2) and (A·3), it is straightforward to evaluate the maximum of the likelihood in (4·1) using current computing tools. For the example discussed in § 5, a simple implementation of the above algorithm using PROC NLMIXED in SAS (V8.1) is as below.

```
proc nlmixed data=resp;
  parms centre0=1 sex0=0 age0=0 base0=3 trt0=2 mu=-2 s1=2;
  uni = probnorm(b/s1);
  phi = 1.0/sqrt(1 + 3/pi/pi*s1*s1);
  Bl = 1/phi*log(sin(pi*uni*phi)/sin(phi*pi*(1-uni)));
  tmp = mu+centre0*centre+sex0*sex+age0*age+base0*base+trt0*trt;
  expeta = exp(Bl + tmp);
  p = expeta/(1 + expeta);
  model y ~ binary(p);
  random b ~ normal(0,s1*s1) subject=id;
  predict p out=outp;
run;
```

REFERENCES

BRESLOW, N. E. & DAY, N. E. (1980). *Statistical Methods in Cancer Research,* **1**: *The Analysis of Case-Control Studies.* Lyon: International Agency for Research on Cancer.

CARLIN, B. P. & LOUIS, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed. London: Chapman and Hall.

CHAO, W. H., PALTA, M. & YOUNG, T. (1997). Effect of omitted confounders on the analysis of correlated binary data. *Biometrics* **53**, 678–89.

FOLLMANN, D. A. & LAMBERT, D. (1989). Generalized logistic regression by non-parametric mixing. *J. Am. Statist. Assoc.* **84**, 295–300.

HEAGERTY, P. J. & KURLAND, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* **88**, 973–85.

HEAGERTY, P. J. & ZEGER, S. L. (2000). Marginalized multilevel models and likelihood inference (with Discussion). *Statist. Sci.* **15**, 1–26.

HEDEKER, D. & GIBBONS, R. D. (1994). A random effects ordinal regression model for multilevel analysis. *Biometrics* **50**, 933–44.

HOUGAARD, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**, 387–96.

KEILSON, J. & STEUTEL, F. W. (1972). Families of infinitely divisible distributions closed under mixing and convolution. *Ann. Math. Statist.* **43**, 242–50.

KOTZ, S. & OSTROVSKII, I. V. (1996). A mixture representation of the Linnik distribution. *Statist. Prob. Lett.* **26**, 61–4.

LAIRD, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *J. Am. Statist. Assoc.* **73**, 805–11.

LIANG, K.-L. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

LINNIK, J. V. (1963). Linear forms and statistical criteria, I. II. *Selected Translations in Math. Statist. Prob.* **3**, 1–90. (Original paper appeared (1953) in *Ukrainskii Mat. Zhournal* **5**, 207–90.)

NEUHAUS, J. M. & JEWELL, N. P. (1993). A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika* **80**, 807–15.

NEUHAUS, J. M., KALBFLEISCH, J. D. & HAUCK, W. W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int. Statist. Rev.* **59**, 25–35.

NEUHAUS, J. M., HAUCK, W. W. & KALBFLEISCH, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika* **79**, 755–62.

PINHEIRO, J. C. & BATES, D. M. (1995). Approximations to the log-likelihood function in the nonlinear random effects model. *J. Comp. Graph. Statist.* **4**, 12–35.

SELBY, S. M. (1974). *Standard Mathematical Tables*, 22nd ed. Cleveland, OH: Chemical Rubber Co.

STOKES, M. E., DAVIS, C. S. & KOCH, G. G. (2000). *Categorical Data Analysis Using the SAS System*, 2nd ed. Cary, NC: SAS Institute Inc.

ZEGER, S. L., LIANG, K.-Y. & ALBERT, P. A. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44**, 1049–60.

[*Received May* 2002. *Revised May* 2003]