

On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems

BY YONG CHEN AND KUNG-YEE LIANG

*Department of Biostatistics, Johns Hopkins University, 615 North Wolfe Street,
Baltimore, Maryland 21205, U.S.A.*

yonchen@jhsp.edu kyliang@jhsp.edu

SUMMARY

This paper considers the asymptotic distribution of the likelihood ratio statistic T for testing a subset of parameter of interest θ , $\theta = (\gamma, \eta)$, $H_0 : \gamma = \gamma_0$, based on the pseudolikelihood $L(\theta, \hat{\phi})$, where $\hat{\phi}$ is a consistent estimator of ϕ , the nuisance parameter. We show that the asymptotic distribution of T under H_0 is a weighted sum of independent chi-squared variables. Some sufficient conditions are provided for the limiting distribution to be a chi-squared variable. When the true value of the parameter of interest, θ_0 , or the true value of the nuisance parameter, ϕ_0 , lies on the boundary of parameter space, the problem is shown to be asymptotically equivalent to the problem of testing the restricted mean of a multivariate normal distribution based on one observation from a multivariate normal distribution with misspecified covariance matrix, or from a mixture of multivariate normal distributions. A variety of examples are provided for which the limiting distributions of T may be mixtures of chi-squared variables. We conducted simulation studies to examine the performance of the likelihood ratio test statistics in variance component models and teratological experiments.

Some key words: Asymptotic distribution; Boundary problem; Frailty survival model; Likelihood ratio test; Nuisance parameter; Pseudolikelihood; Teratological experiment; Variance component model.

1. INTRODUCTION

1.1. Motivation

Very often a probability model for data is indexed by two sets of parameters: the parameter of interest, θ , of dimension p , and the nuisance parameter, ϕ , of dimension q . Sometimes the likelihood function $L(\theta, \phi)$, when considered jointly as a function of θ and ϕ , is ill-behaved, yet when considered as a function of θ alone with ϕ held fixed, is well-behaved. When the nuisance parameter cannot be eliminated through conditioning, marginalizing or factorization, one approach is to use the pseudolikelihood proposed by [Gong & Samaniego \(1981\)](#). The key idea of this approach is that inference for θ can be based on $L^*(\theta) = L(\theta, \hat{\phi})$, known as the pseudolikelihood for θ , where $\hat{\phi}$ is a consistent estimator of ϕ , assuming such an estimator exists. Gong & Samaniego derived the asymptotic distribution of the maximum pseudolikelihood estimator $\hat{\theta}$ for θ under regularity conditions. Thus, Wald-based inference for θ can be made based on this asymptotic distribution. However, it has been well documented in the literature that such inference could be ill-behaved, and that this concern can be alleviated by using likelihood-ratio-based inference ([Hauck & Donner, 1977](#)). To this end, [Liang & Self \(1996\)](#) studied the asymptotic behaviour of the likelihood ratio test statistic for testing $H_0 : \theta = \theta_0$ based on $L^*(\theta)$.

The primary purpose of this paper is to extend the work by [Liang & Self \(1996\)](#) to deal with situations where only a subset of θ is tested, and the parameter value of interest or the nuisance parameter may lie on the boundary of its parameter space.

1.2. Example 1: Variance component models

Consider the model

$$y = X\phi + \epsilon, \quad \epsilon \sim N\{0, \Sigma(\theta)\},$$

where y is an $n \times 1$ vector of the response variables, X is the design matrix of $n \times q$ dimension and $\Sigma(\theta)$ is the $n \times n$ covariance matrix indexed by θ , a $p \times 1$ vector. In pedigree analyses, it has been a standard procedure to estimate ϕ by the conventional least-squares estimator $\hat{\phi} = (X^T X)^{-1} X^T y$ and to base inference for θ on $L(\theta, \hat{\phi})$, where $L(\theta, \phi)$ is proportional to the conditional distribution of $Y = y$ given X . Often we are interested in testing the significance of the multiple variance components one at a time and leave others unspecified ([Beatty et al., 1985](#)).

1.3. Example 2: Behrens–Fisher problem

Consider $y_{ij} \sim N(\theta_i, \phi_i)$ ($j = 1, \dots, n_i; i = 1, 2$), when the y_{ij} s are independent of each other. The Behrens–Fisher problem corresponds to the problem of testing the difference between the two means, i.e. $H_0 : \theta_1 = \theta_2$. With ϕ_i replaced by $\hat{\phi}_i = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 / n_i$, the sample variance of the i th group ($i = 1, 2$), the pseudolikelihood for (θ_1, θ_2) has the form

$$L(\theta_1, \theta_2, \hat{\phi}_1, \hat{\phi}_2) \propto \prod_{i=1}^2 \prod_{j=1}^{n_i} (2\pi \hat{\phi}_i)^{1/2} \exp \left\{ -\frac{(y_{ij} - \theta_i)^2}{2\hat{\phi}_i} \right\}.$$

The null hypothesis of interest is $H_0 : \gamma = \theta_1 - \theta_2 = 0$, leaving $\eta = \theta_2$ unspecified.

1.4. Example 3: Frailty survival models

[Bandein-Roche & Liang \(1996\)](#) described a class of models for failure-time data that accounts for multiple levels of clustering. This class of models permits specification of simple distributional forms for all bivariate margins. In addition, the distribution reduces to the univariate frailty model ([Vaupel et al., 1979](#); [Oakes, 1989](#)) when there is only a single level. Consider a cluster with two levels; for example, households and villages. Suppose there are four individual members and the households are clustered as $\{1, 2\}$ and $\{3, 4\}$. When the Clayton copula ([Clayton, 1978](#)) is used, the multivariate survival function has the form

$$S(t_1, \dots, t_4) = \left[\left\{ \sum_{j=1}^2 S_j(t_j)^{1-\theta_1} - 1 \right\}^\alpha + \left\{ \sum_{j=3}^4 S_j(t_j)^{1-\theta_1} - 1 \right\}^\alpha - 1 \right]^{1/(1-\theta_2)}, \quad (1)$$

where $\alpha = (\theta_2 - 1)/(\theta_1 - 1)$, S_j is the marginal survival distribution of the j th member ($j = 1, \dots, 4$), and $\theta_1 \geq 1$ and $\theta_2 \geq 1$ characterize the association within and between households, respectively. To ensure that the above survival function is legitimate, one needs $\theta_2 \leq \theta_1$, i.e. the between-household association not exceeding the within-house association. Let ϕ denote the parameters which specify the S_j s, the pseudolikelihood approach for inference about θ has been proposed ([Bandein-Roche & Liang, 1996](#); [Shih & Louis, 1995](#)) as convenient estimators of ϕ are typically available by, for example, maximizing the likelihood function with respect to ϕ assuming independence, i.e. fixing θ_1 and θ_2 at 1. One hypothesis of interest is $H_0 : \theta_2 = 1$, i.e. no between-household association. Here $\theta_2 = 1$ is on the boundary of the parameter space for θ_2 .

1.5. Example 4: Teratological experiments

In teratological experiments, the response variables are often dichotomous and their sums are commonly modelled by beta-binomial distributions (Weil, 1970; Haseman & Kupper, 1979). The distributions are indexed by π , the probability of experiencing an adverse event, and ϕ , the nuisance parameter characterizing the so-called litter effect, also known as overdispersion, the tendency that the subjects from the same litter to respond more alike than subjects from different litters. Often investigators would like to establish the dose-response relationship between π and x , the dosage level or the treatment status. One may consider the model,

$$g(\pi) = \beta_0 + \beta_1 x, \quad \text{var}(Y) = m\pi(1 - \pi)\{1 + (m - 1)\phi\},$$

where g is known as the prespecified link function. For beta-binomial models, it has been well documented that the nuisance parameter has a profound impact on the parameter of interest (Williams, 1988). One way to alleviate such concern is to base the inference of β on the pseudolikelihood $L(\beta, \hat{\phi})$ where $\hat{\phi}$ is a consistent estimator of ϕ . Such an estimator could be obtained, for example, by using method of moments. One hypothesis of interest is $H_0 : \beta_1 = 0$, i.e. no dose-response relationship. Since the intralitter correlation could be dose-dependent, heterogeneous correlations are often assumed. When the dosage level is zero or the placebo is used, the intralitter correlation is usually close to zero, lying near or on the boundary of its parameter space.

1.6. Example 5: A general setting

Consider the case where the likelihood can be decomposed into two parts,

$$L(\theta, \phi) = L_1(\theta, \phi)L_2(\phi),$$

where both L_1 and L_2 are legitimate likelihood functions. In this case, one could use the maximizer of $L_2(\phi)$, $\hat{\phi}$, to form the pseudolikelihood function $L^*(\theta)$ for θ . Many examples fit into this case. It is easy to see that Example 2 above is a special case with $L_2(\phi)$ proportional to the product of probability density functions of sample variances from both samples. Another example is genetic linkage study where the data consist of genetic marker data, M , and genotyping data, T , from pedigrees. In this situation, the likelihood function is proportional to

$$f(M | T; \theta, \phi)f(T; \phi) = L_1(\theta, \phi)L_2(\phi),$$

where θ_j , ($j = 1, \dots, p$), is the location of the j th disease gene, and ϕ are parameters characterizing an underlying genetic mechanism for the disease. In the context of measurement errors, the true covariate, X , is observed indirectly through surrogate, W , which is measured with errors. With Y as the dependent variable, the likelihood function is proportional to

$$\begin{aligned} f(Y = y | W = w; \theta, \phi)f(W = w; \phi) &= \int_x f(y | x; \theta)f(x | w; \phi) dx \int_x f(w | x; \phi)f(x; \phi) dx \\ &= L_1(\theta, \phi)L_2(\phi), \end{aligned}$$

where θ is the regression coefficient relating X to Y and ϕ is the parameter characterizing the measurement errors for the covariates.

2. MAIN RESULTS

2.1. The parameter of interest and the nuisance parameter are interior points

In this subsection, we consider the cases where only a subset of the parameter of interest is tested while the true values of the parameter of interest and nuisance parameter are interior

points of their parameter spaces. Consider the partition of θ into two parts: the parameters to be tested, γ , of dimension p_1 and the remaining parameters, η , of dimension $p_2 = p - p_1$. The null hypothesis of interest is specified as $H_0 : \gamma = \gamma_0$. Let T be the pseudolikelihood ratio test statistic, i.e. $T = -2 \log [L^* \{ \gamma_0, \hat{\eta}(\gamma_0) \} / L^* \{ \hat{\gamma}, \hat{\eta} \}]$, where $\hat{\eta}(\gamma_0) = \arg \max_{\eta} L^* (\gamma_0, \eta)$ and $(\hat{\gamma}, \hat{\eta}) = \arg \max_{\gamma, \eta} L^* (\gamma, \eta)$. Theorem 1 states the large sample distribution of T under H_0 .

THEOREM 1. *Suppose that under a distribution indexed by (θ_0, ϕ_0) , the consistent estimator of ϕ , $\hat{\phi}$, has the asymptotic distribution $n^{1/2}(\hat{\phi} - \phi_0) \rightarrow N(0, \Sigma_{22})$ as $n \rightarrow \infty$.*

Then, under the regularity conditions (A1)–(A6) in Gong & Samaniego (1981),

(i) *the $p \times p$ matrix*

$$\left\{ I_{11}^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & I_{\eta\eta}^{-1} \end{pmatrix} \right\} I_{11}^*$$

has p_1 positive eigenvalues, $\lambda_1 \geq \dots \geq \lambda_{p_1} > 0$, and $p - p_1$ zero eigenvalues;

(ii) *the asymptotic distribution of the pseudolikelihood ratio statistic T is $U = \sum_{j=1}^{p_1} \lambda_j U_j$, where the U_j s are independent χ^2_1 variables.*

Here

$$I_{11} = \lim_{n \rightarrow \infty} \left[E \left\{ -\frac{1}{n} \frac{\partial^2 \log L(\theta_0, \phi_0)}{\partial \theta^2}; \theta_0, \phi_0 \right\} \right],$$

$$I_{\eta\eta} = \lim_{n \rightarrow \infty} \left[E \left\{ -\frac{1}{n} \frac{\partial^2 \log L(\gamma_0, \eta_0, \phi_0)}{\partial \eta^2}; \gamma_0, \eta_0, \phi_0 \right\} \right],$$

and I_{11}^* is the asymptotic variance of $n^{-1/2} \partial \log L(\theta_0, \hat{\phi}) / \partial \theta$ and can be calculated as $I_{11}^* = I_{11} + I_{12} \Sigma_{22} I_{12}^T$ with

$$I_{12} = \lim_{n \rightarrow \infty} \left[E \left\{ -\frac{1}{n} \frac{\partial^2 \log L(\theta_0, \phi_0)}{\partial \theta \partial \phi}; \theta_0, \phi_0 \right\} \right].$$

The proof is straightforward using Taylor expansions and Theorem 4.4.4 of Graybill (1976) and hence is omitted. Implicitly, we used the fact that $n^{1/2} \{ \hat{\theta}(\phi_0) - \theta_0 \}$ and $n^{1/2} (\hat{\phi} - \phi_0)$ are asymptotically independent because $\hat{\theta}(\phi_0)$ is asymptotically efficient and $n^{1/2} (\hat{\phi} - \phi_0)$ has asymptotic constant mean zero (Pierce, 1982; Parke, 1986). When $p_1 = p$, Theorem 1 reduces to the main result of Liang & Self (1996).

It is easy to show that when $I_{11}^* = I_{11}$, the distribution of U is $\chi^2_{p_1}$, which is the same as the asymptotic distribution of the conventional likelihood ratio statistic. As noted by Liang & Self (1996), this condition is satisfied when $I_{12}(\theta_0, \phi) = 0$ for all ϕ , which means that the score function for θ and ϕ are orthogonal to each other when evaluated at (θ_0, ϕ) . Another situation where we have $I_{11}^* = I_{11}$ is $\Sigma_{22} = 0$. This could occur when ϕ is estimated through external data whose sample size n^* increases at a faster rate than n , i.e. $n^*/n \rightarrow \infty$ as $n \rightarrow \infty$. A less obvious situation, which results in the distribution of U being $\chi^2_{p_1}$, occurs when $I_{\gamma\eta} = I_{\gamma\phi} = 0$. This means that the score function for γ is orthogonal both to the score function for η and to the score function for ϕ .

2.2. The parameter of interest lies on the boundary

In this subsection, we consider cases where the true value of θ lies on the boundary of its parameter space while the true value of the nuisance parameter ϕ is an interior point of its

parameter space. These situations were informally mentioned in [Liang & Self \(1996\)](#) but were not formally treated therein. Denote the parameter space of θ by Ω and rewrite the hypothesis $H_0 : \gamma = \gamma_0$ as $H_0 : \theta \in \Omega_0$, where $\Omega_0 = \{\theta : \theta = (\gamma_0, \eta) \in \Omega\}$. The complement of Ω_0 in Ω is denoted by Ω_1 . In § 2.1 where θ_0 is not on the boundary, Ω , Ω_0 and Ω_1 are simply R^p , R^{p_2} and R^{p_1} , respectively. For any subset of R^p , φ , we define $P_\varphi = \sup_{\theta \in \varphi} L^*(\theta)$. We also define the maximum pseudolikelihood estimator in the parameter space φ , $\hat{\theta}_\varphi$, as that value of θ in the closure of φ which maximizes $L^*(\theta)$.

The pseudolikelihood ratio test statistic can be written as $T = -2 \log(P_{\Omega_0}/P_\Omega)$. [Gong & Samaniego \(1981\)](#) established the $n^{1/2}$ -consistency of the maximum pseudolikelihood estimator for θ when the true value of θ is an interior point of its parameter space. Here we give the consistency result when the true value of θ lies on the boundary.

LEMMA 1. *If the regularity conditions in [Chernoff \(1954\)](#) hold, the intersection of φ and the closure of neighbourhoods centred about the true value of the parameter of interest, θ_0 , constitute closed subsets of R^p , and θ_0 is a limit point of φ , then with probability tending to 1 as $n \rightarrow \infty$, there exists a sequence of points in φ , $\hat{\theta}_\varphi$, at which local maxima of $\log L^*(\theta)$ occur, and that converges to θ_0 in probability. Moreover, $n^{1/2}(\hat{\theta}_\varphi - \theta_0) = O_p(1)$.*

We now establish the equivalence between the asymptotic distribution of the pseudolikelihood ratio and the distribution of the likelihood ratio of testing the restricted mean of a multivariate normal distribution based on one observation from a multivariate normal distribution with misspecified covariance matrix.

THEOREM 2. *If the regularity conditions in [Chernoff \(1954\)](#) are satisfied, θ_0 is a limiting point of both Ω_0 and Ω_1 , and the sets Ω_0 and Ω_1 are approximated by nonnull and disjoint cones C_{Ω_0} and C_{Ω_1} , respectively, then, when $\gamma = \gamma_0$, the asymptotic distribution of the pseudolikelihood ratio statistic, T , is the same as the distribution of the likelihood ratio test of $\theta \in C_{\Omega_0}$ against $\theta \in C_{\Omega_1}$ based on one observation from a population with multivariate normal distribution with mean θ and unknown covariance matrix $I_{11}^{-1} I_{11}^* I_{11}^{-1}$ while the covariance matrix is misspecified as I_{11}^{-1} in the likelihood ratio test.*

A sketch of the proof is given in Appendix A. When there is no nuisance parameter ϕ or $I_{11}^* = I_{11}$, Theorem 2 reduces to Theorem 3 of [Self & Liang \(1987\)](#) in which the likelihood ratio instead of pseudolikelihood ratio is considered.

The result given by Theorem 2 reduces the general problem of computing the limit distribution of a pseudolikelihood ratio to a problem of computing the distribution of

$$T_1(Z) = Q_{C_{\Omega_0}}(Z) - Q_{C_\Omega}(Z), \quad (2)$$

where $Q_\varphi(Z) = \inf_{\theta \in \varphi} (Z - \theta)^T I_{11} (Z - \theta)$, $Z \sim N(0, I_{11}^{-1} I_{11}^* I_{11}^{-1})$ and $C_\Omega = C_{\Omega_0} \cup C_{\Omega_1}$.

The distribution of $T_1(Z)$ in any given case could be rather complicated. We now offer several special cases in which Theorem 2 may be used to compute the asymptotic distribution of the pseudolikelihood ratio statistic.

From here on we assume $\Omega = \Omega_1 \times \dots \times \Omega_p$, where the Ω_j s are either closed, half-open or open intervals in R . To compute the distribution of $T_1(Z)$, we need the following lemma.

LEMMA 2. *Suppose $Z \sim N(0, I_{11}^{-1} I_{11}^* I_{11}^{-1})$ and $\varphi = \{0\}^{p_1} \times R^{p-p_1}$. Let $Q_\varphi(Z) = \inf_{\theta \in \varphi} (Z - \theta)^T I_{11} (Z - \theta)$. Then*

- (i) $Q_\varphi(Z) = Z_{(p_1)}^T e_{(p_1)}^{-1} Z_{(p_1)}$, where $Z_{(p_1)} = (Z_1, \dots, Z_{p_1})^T$ and $e_{(p_1)}$ is the $p_1 \times p_1$ upper left submatrix of I_{11}^{-1} ;

- (ii) the distribution of $Q_\varphi(Z)$ is $U = \sum_{j=1}^{p_1} \lambda_j U_j$, where U_j s are independent χ_1^2 variables and $\lambda_1 \geq \dots \geq \lambda_{p_1} > 0$ are the eigenvalues of matrix $e_{(p_1)}^{-1} e_{(p_1)}^*$, with $e_{(p_1)}^*$ being the $p_1 \times p_1$ upper left submatrix of $I_{11}^{-1} I_{11}^* I_{11}^{-1}$; and
- (iii) when $I_{11}^* = I_{11}$, U is simply $\chi_{p_1}^2$.

Part (i) can be shown by completing the quadratic form, and the rest of the results follow by Theorem 4.4.4 of Graybill (1976).

Now we consider some special cases, adopting the notation in Self & Liang (1987). Partition the parameter vector, $\theta = (\gamma, \eta)$, into four coordinates: $(p_{11}, p_{12}, p_{21}, p - p_{11} - p_{12} - p_{21})$ where the first p_{11} coordinates of θ represent the parameter to be tested with true values on the boundary; the next p_{12} coordinates of θ represent the parameter to be tested with true values not on the boundary; the next p_{21} coordinates of θ represent the first p_{21} component of η with true values on the boundary; and finally the remaining $p - p_{11} - p_{12} - p_{21}$ coordinates of θ represent the last $p - p_{11} - p_{12} - p_{21}$ coordinates of η with the true value not on the boundary. Here $p_1 = p_{11} + p_{12}$.

Case 1. Suppose that the parameter configuration is $(0, p_1, 0, p - p_1)$, i.e. θ_0 is an interior point of Ω . Then $C_{\Omega_0} = \{0\}^{p_1} \times R^{p-p_1}$ and $C_\Omega = R^p$. By Theorem 2, the asymptotic distribution of the pseudolikelihood ratio reduces to the distribution of $T_1(Z) = Q_{C_{\Omega_0}}(Z)$. By Lemma 2, the distribution of $T_1(Z)$ is $U = \sum_{j=1}^{p_1} \lambda_j U_j$, where U_j s are independent χ_1^2 variables and $\lambda_1 \geq \dots \geq \lambda_{p_1} > 0$ are the eigenvalues of $e_{(p_1)}^{-1} e_{(p_1)}^*$. This can be shown to agree with results in Theorem 1. When $I_{11}^* = I_{11}$, the asymptotic distribution of pseudolikelihood ratio statistic is $\chi_{p_1}^2$.

Case 2. Suppose the parameter configuration is $(1, 0, 0, p - 1)$. Then $C_{\Omega_0} = \{0\} \times R^{p-1}$ and $C_\Omega = [0, \infty) \times R^{p-1}$. By Lemma 2, equation (2) can be reduced to

$$T_1(Z) = Z_1^2/e_{(1)} - Z_1^2 I(Z_1 < 0)/e_{(1)} = Z_1^2 I(Z_1 > 0)/e_{(1)},$$

where the first step is due to the fact that $Q_{C_\Omega}(Z) = Q_{C_{\Omega_0}}(Z)$ when $Z_1 \leq 0$. This is true because the quadratic function $(Z - \theta)^T I_{11}(Z - \theta)$ is a continuous function of θ and the infimum in the function $Q_{C_\Omega}(Z)$ is achieved at the boundary of C_Ω , i.e. C_{Ω_0} . Consequently, the asymptotic distribution of T is a 50:50 mixture of χ_0^2 and $e_{(1)}^* \chi_1^2/e_{(1)}$. When $I_{11}^* = I_{11}$, this distribution is a 50:50 mixture of χ_0^2 and χ_1^2 , which agrees with the result in Case 5 of Self & Liang (1987).

Case 3. Suppose the parameter configuration is $(1, p_1 - 1, 0, p - p_1)$. Then $C_{\Omega_0} = \{0\} \times \{0\}^{p_1-1} \times R^{p-p_1}$ and $C_\Omega = [0, \infty) \times R^{p-1}$. By Lemma 2 and a similar argument to that in Case 2, equation (2) can be expressed as $T_1(Z) = Z_{(p_1)}^T e_{(p_1)}^{-1} Z_{(p_1)} - Z_1^2 I(Z_1 < 0)/e_{(1)}$.

Its distribution is a 50:50 mixture of random variables U_a and U_b , where $U_a = \sum_{i=1}^{p_1} \lambda_i U_i$ with the λ_i s being the p_1 positive eigenvalues of the matrix $e_{(p_1)}^{-1} e_{(p_1)}^*$ and $U_b = \sum_{j=1}^{p_1-1} \lambda_j U_j$ with the λ_j s being the $p_1 - 1$ positive eigenvalues of the matrix

$$\left\{ e_{(p_1)}^{-1} - \begin{pmatrix} 1/e_{(1)} & 0_{1 \times (p_1-1)} \\ 0_{(p_1-1) \times 1} & 0_{(p_1-1) \times (p_1-1)} \end{pmatrix} \right\} e_{(p_1)}^*,$$

and the U_i s and U_j s being independent χ_1^2 variables. In variance component models, for example, we are testing a variance component and $p_1 - 1$ regression coefficients simultaneously while leaving the other parameters unspecified. When $I_{11}^* = I_{11}$, the distribution is a 50:50 mixture of $\chi_{p_1-1}^2$ and $\chi_{p_1}^2$. This agrees with the result in Case 3 of Stram & Lee (1994), where one variance component and $p_1 - 1$ covariances are tested simultaneously. When $I_{11}^* = I_{11}$ and $p_1 = 2$, the result reduces to the results in Case 2 of Stram & Lee (1994) and Case 6 of Self & Liang (1987).

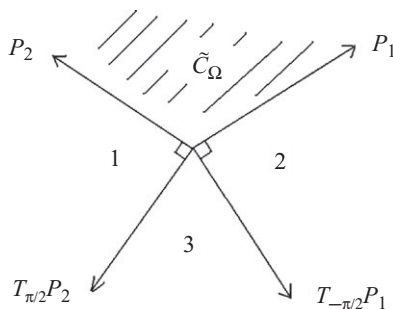


Fig. 1. Diagram of the parameter space for Case 4. The shaded region, \tilde{C}_Ω , represents admissible parameter values under the alternative hypothesis. Under the null hypothesis, the parameter is located at the origin. The asymptotic distribution of T is a mixture of U , $d^* \chi_1^2/d$, $a^* \chi_1^2/a$ and χ_0^2 distributions with mixing probabilities depending on the angles in \tilde{C}_Ω .

In Cases 2 and 3, we take the advantage of the fact that only one component of γ_0 lies on the boundary. In these cases, the parameter spaces C_{Ω_0} and C_Ω , and their boundaries, are straight lines or planes which are preserved under linear transformations. In Cases 4 and 5, we consider the situation where there are two boundary parameters. In Case 4, both boundary parameters are tested while in Case 5, only one boundary parameter is tested. The calculation of the asymptotic distribution of the pseudolikelihood ratio statistic is more complicated than that in Cases 1–3.

Case 4. Suppose the parameter configuration is $(2, 0, 0, 0)$. Then $C_{\Omega_0} = \{0\} \times \{0\}$ and $C_\Omega = [0, \infty) \times [0, \infty)$. Let $I_{11} = P^T P$, where P is a nonsingular matrix and denote $\tilde{C}_\Omega = \{\tilde{\theta} : \tilde{\theta} = P\theta \text{ for any } \theta \in C_\Omega\}$ and $\tilde{Z} = PZ$. Then $T_1(Z)$ can be rewritten as

$$T_1(Z) = \|\tilde{Z}\|^2 - \inf_{\tilde{\theta} \in \tilde{C}_\Omega} \|\tilde{Z} - \tilde{\theta}\|^2, \tag{3}$$

where $\|\cdot\|$ is the usual Euclidean metric. Calculation of the second term in equation (3) depends on the location of \tilde{Z} relative to the boundary of \tilde{C}_Ω . There are four regions that must be considered. The shaded region in Fig. 1 represents \tilde{C}_Ω , and \tilde{C}_{Ω_0} is the origin. The angle in the shaded area is less than 180° . This is due to the fact that the convexity of C_Ω is preserved under linear mapping P . Denote the rotation matrix in R^2 and the information matrices I_{11} and I_{11}^* by

$$T_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad I_{11} = \begin{pmatrix} a & b \\ b & d \end{pmatrix}, \quad I_{11}^* = \begin{pmatrix} a^* & b^* \\ b^* & d^* \end{pmatrix},$$

respectively. Further denote the columns of P by P_1 and P_2 , and the inner product of vectors a and b by $\langle a, b \rangle = a^T b$. Then the distribution of $T_1(Z)$ can be calculated as

$$T_1(Z) = \begin{cases} \|\tilde{Z}\|^2 \sim U = \lambda_1 U_1 + \lambda_2 U_2 & \text{if } \tilde{Z} \text{ is in the shaded region,} \\ \left\{ \frac{\langle \tilde{Z}, P_2 \rangle}{\|P_2\|} \right\}^2 \sim d^* \chi_1^2/d & \text{if } \tilde{Z} \text{ is in region 1,} \\ \left\{ \frac{\langle \tilde{Z}, P_1 \rangle}{\|P_1\|} \right\}^2 \sim a^* \chi_1^2/a & \text{if } \tilde{Z} \text{ is in region 2,} \\ 0 & \text{if } \tilde{Z} \text{ is in region 3,} \end{cases}$$

where λ_1 and λ_2 are eigenvalues of I_{11}^* and I_{11}^{-1} , and U_1 and U_2 are independent χ_1^2 variables.

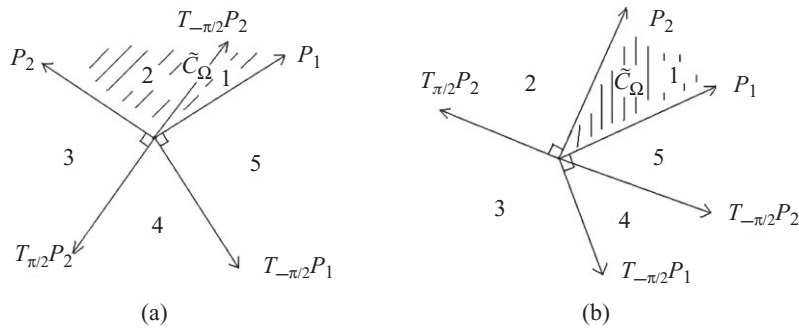


Fig. 2. Diagrams of the parameter space for Case 5 when the angle spanned by P_1 and P_2 is (a) greater than $\pi/2$, or (b) is less than $\pi/2$. The shaded region, \tilde{C}_Ω , represents admissible parameter values under the alternative hypothesis. The half-line in the same direction as the vector P_2 with the endpoint at the origin represents admissible parameter values under the null hypothesis. The asymptotic distribution of T is a mixture of distributions.

The mixing probabilities for the shaded region, region 1 and region 2, are

$$\begin{aligned}
 p_s &= \cos^{-1} \left[\frac{(0, 1)I_{11}\{I_{11}^*\}^{-1}I_{11}\begin{pmatrix} 1 \\ 0 \end{pmatrix}}{\left\{ (0, 1)I_{11}(I_{11}^*)^{-1}I_{11}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}^{1/2} \left\{ (1, 0)I_{11}(I_{11}^*)^{-1}I_{11}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}^{1/2}} \right] / 2\pi, \\
 p_1 &= \cos^{-1} \left[\frac{(0, 1)I_{11}(I_{11}^*)^{-1}P^T T_{\pi/2} P_2}{\left\{ (0, 1)I_{11}(I_{11}^*)^{-1}I_{11}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}^{1/2} \left\{ P_2^T T_{-\pi/2} P(I_{11}^*)^{-1} P^T T_{\pi/2} P_2 \right\}^{1/2}} \right] / 2\pi, \\
 p_2 &= \cos^{-1} \left[\frac{(1, 0)I_{11}(I_{11}^*)^{-1}P^T T_{-\pi/2} P_1}{\left\{ (1, 0)I_{11}(I_{11}^*)^{-1}I_{11}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}^{1/2} \left\{ P_1^T T_{\pi/2} P(I_{11}^*)^{-1} P^T T_{-\pi/2} P_1 \right\}^{1/2}} \right] / 2\pi.
 \end{aligned}$$

Thus the asymptotic distribution of T is a mixture of U , $d^* \chi_1^2/d$, $a^* \chi_1^2/a$ and χ_0^2 with mixing probabilities p_s , p_1 , p_2 and $1 - p_s - p_1 - p_2$, respectively. When $I_{11} = I_{11}^*$, the asymptotic distribution reduces to a p_s , 0.5 and $0.5 - p_s$ mixture of χ_2^2 , χ_1^2 and χ_0^2 , respectively, which agrees with the results in Case 7 of [Self & Liang \(1987\)](#) and Example 3 in [Chernoff \(1954\)](#).

Case 5. Suppose the parameter configuration is $(1, 0, 1, 0)$. Then $C_{\Omega_0} = \{0\} \times [0, +\infty)$ and $C_\Omega = [0, \infty) \times [0, \infty)$. By a similar argument to that in Case 4, we may write

$$T_1(Z) = \inf_{\tilde{\theta} \in \tilde{C}_{\Omega_0}} \|\tilde{Z} - \tilde{\theta}\|^2 - \inf_{\tilde{\theta} \in \tilde{C}_\Omega} \|\tilde{Z} - \tilde{\theta}\|^2. \tag{4}$$

Calculation of the terms in equation (4) depends on the location of \tilde{Z} relative to the boundary of both \tilde{C}_{Ω_0} and \tilde{C}_Ω . In fact, there are two possible partitions of regions that must be considered, depending on whether the angle ψ spanned by P_1 and P_2 is greater than $\pi/2$. It is easy to show $\psi > \pi/2$ if and only if $b < 0$.

If $\psi > \pi/2$, i.e. $b < 0$, there are five different regions to be considered. The shaded region in Fig. 2 represents \tilde{C}_Ω , and \tilde{C}_{Ω_0} is the half-line in the same direction as the vector P_2 with the

endpoint at the origin. The distribution of $T_1(Z)$ can be calculated as

$$T_1(Z) = \begin{cases} \|\tilde{Z}\|^2 \sim \lambda_1 U_1 + \lambda_2 U_2 & \text{if } \tilde{Z} \text{ is in region 1,} \\ \|\tilde{Z}\|^2 - \left\{ \frac{\langle \tilde{Z}, P_2 \rangle}{\|P_2\|} \right\}^2 \sim \lambda_3 \chi_1^2 & \text{if } \tilde{Z} \text{ is in region 2,} \\ 0 & \text{if } \tilde{Z} \text{ is in region 3 or 4,} \\ \left\{ \frac{\langle \tilde{Z}, P_1 \rangle}{\|P_1\|} \right\}^2 \sim a^* \chi_1^2 / a & \text{if } \tilde{Z} \text{ is in region 5,} \end{cases}$$

where λ_1 and λ_2 are eigenvalues of $I_{11}^* I_{11}^{-1}$, U_1 and U_2 are independent χ_1^2 variables, and $\lambda_3 = (a^* d^2 - 2b^* b d + d^* b^2) / \{d(ad - b^2)\}$.

By a similar argument, the distribution of $T_1(Z)$ when $\psi \leq \pi/2$, i.e. $b \geq 0$, is calculated as

$$T_1(Z) = \begin{cases} \|\tilde{Z}\|^2 - \left\{ \frac{\langle \tilde{Z}, P_2 \rangle}{\|P_2\|} \right\}^2 \sim \lambda_1 \chi_1^2 & \text{if } \tilde{Z} \text{ is in region 1,} \\ 0 & \text{if } \tilde{Z} \text{ is in region 2 or 3,} \\ \left\{ \frac{\langle \tilde{Z}, P_1 \rangle}{\|P_1\|} \right\}^2 \sim a^* \chi_1^2 / a & \text{if } \tilde{Z} \text{ is in region 4,} \\ \left\{ \frac{\langle \tilde{Z}, P_1 \rangle}{\|P_1\|} \right\}^2 - \left\{ \frac{\langle \tilde{Z}, P_2 \rangle}{\|P_2\|} \right\}^2 \sim \lambda_2 U_1 + \lambda_3 U_2 & \text{if } \tilde{Z} \text{ is in region 5,} \end{cases}$$

where $\lambda_1 = (a^* d^2 - 2b^* b d + d^* b^2) / \{d(ad - b^2)\}$, U_1 and U_2 are independent χ_1^2 variables, and λ_2 and λ_3 are eigenvalues of

$$I_{11} \begin{pmatrix} 1/a & 0 \\ 0 & -1/d \end{pmatrix} I_{11}^* I_{11}^{-1}.$$

2.3. The nuisance parameter lies on the boundary

In this subsection, we consider cases where the true value of nuisance parameter, ϕ , lies on the boundary of its parameter space. In this situation, the asymptotic distribution of the pseudolikelihood ratio statistic is more complicated since the consistent estimator $\hat{\phi}$ may not be asymptotically normally distributed. Specifically, by a similar argument as in the proof of Theorem 2, the pseudolikelihood ratio statistic can be written as

$$T = \inf_{\theta \in C_{\Omega_0}} (Z_n - \theta)^T I_{11} (Z_n - \theta) - \inf_{\theta \in C_{\Omega}} (Z_n - \theta)^T I_{11} (Z_n - \theta) + o_p(1),$$

where

$$Z_n = n^{-1/2} I_{11}^{-1} \frac{\partial \log L(\theta_0, \phi_0)}{\partial \theta} + I_{11}^{-1} I_{12} n^{1/2} (\hat{\phi} - \phi_0) + o_p(1).$$

The second term in Z_n may not be asymptotically normally distributed and the results by Pierce (1982) and Parke (1986) cannot be applied here to show the asymptotic independence between the two terms in Z_n since the second term in Z_n does not have asymptotic constant mean.

In fact, the asymptotic distribution of Z_n is a mixture of normals with identical mean zero and possibly unequal variances. To show this, denote by $\hat{\phi}^*$ the estimator of ϕ without restriction of parameter space and assume the parameter space for ϕ is $[0, +\infty) \times R^{q-1}$. By applying similar

arguments as in [Self & Liang \(1987\)](#), $\hat{\phi}$ can be represented as

$$\hat{\phi}^* I(\hat{\phi}_1^* > 0) + \begin{pmatrix} 0 \\ \hat{\phi}_2^* - (\Sigma_{22}^{21}/\Sigma_{22}^{11})\hat{\phi}_1^* \\ \vdots \\ \hat{\phi}_q^* - (\Sigma_{22}^{q1}/\Sigma_{22}^{11})\hat{\phi}_1^* \end{pmatrix} I(\hat{\phi}_1^* \leq 0),$$

where Σ_{22}^{ij} are elements of the matrix Σ_{22} , the asymptotic variance of $\hat{\phi}^*$. Therefore, Z_n can be rewritten as

$$\left\{ n^{-1/2} I_{11}^{-1} \frac{\partial \log L(\theta_0, \phi_0)}{\partial \theta} + I_{11}^{-1} I_{12} n^{1/2} (\hat{\phi}^* - \phi_0) \right\} I(\hat{\phi}_1^* > 0) + \left\{ n^{-1/2} I_{11}^{-1} \frac{\partial \log L(\theta_0, \phi_0)}{\partial \theta} + I_{11}^{-1} I_{12} A n^{1/2} (\hat{\phi}^* - \phi_0) \right\} I(\hat{\phi}_1^* \leq 0),$$

where

$$A = \begin{pmatrix} 0 & 0 \dots 0 \\ -\Sigma_{22}^{21}/\Sigma_{22}^{11} & 1 \dots 0 \\ \vdots & \vdots \ddots \vdots \\ -\Sigma_{22}^{q1}/\Sigma_{22}^{11} & 0 \dots 1 \end{pmatrix}.$$

Suppose that as $n \rightarrow \infty$,

$$n^{1/2} \begin{pmatrix} \frac{1}{n} \frac{\partial \log L(\theta_0, \phi_0)}{\partial \theta} \\ \hat{\phi}^* - \phi_0 \end{pmatrix} \rightarrow \begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} I_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix} \right\}, \tag{5}$$

where Σ_{12} is the asymptotic covariance between $n^{-1/2} \partial \log L(\theta_0, \phi_0) / \partial \theta$ and $n^{1/2} (\hat{\phi}^* - \phi_0)$, which is in fact zero by [Pierce \(1982\)](#) and [Parke \(1986\)](#). The asymptotic distribution of Z_n is

$$(I_{11}^{-1} X + I_{11}^{-1} I_{12} Y) I(Y_1 > 0) + (I_{11}^{-1} X + I_{11}^{-1} I_{12} A Y) I(Y_1 \leq 0). \tag{6}$$

Thus the asymptotic distribution of Z_n is a 50:50 mixture of normals with mean zero. For more general cases where the parameter space for ϕ is $[0, +\infty)^{q_0} \times R^{q-q_0}$ with $q_0 \geq 1$, similar arguments can be applied to show that the asymptotic distribution of Z_n is a 2^{q_0} mixture of normals.

The asymptotic distribution of the pseudolikelihood ratio statistic T is

$$\inf_{\theta \in C_{\Omega_0}} (Z - \theta)^T I_{11} (Z - \theta) - \inf_{\theta \in C_{\Omega}} (Z - \theta)^T I_{11} (Z - \theta),$$

where Z is the asymptotic distribution of Z_n . Therefore, the asymptotic distribution of T is that of the likelihood ratio for testing the restricted mean of a multivariate normal distribution with covariance matrix I_{11}^{-1} based on one observation from a population with a mixture of normals specified by equations (5) and (6). When the true value of the nuisance parameter is an interior point, the likelihood ratio reduces to equation (2) since the mixture of normals is in fact a normal distribution.

3. EXAMPLES REVISITED

3.1. Example 1: Variance component models

It can be shown that

$$\frac{\partial \log L(\theta, \phi)}{\partial \phi} = X^T \Sigma^{-1}(\theta)(y - X\phi).$$

Therefore,

$$I_{12}(\theta, \phi) = E \left[-\frac{\partial}{\partial \theta} \{X^T \Sigma^{-1}(\theta)(y - X\phi)\} \right] = \left\{ -\frac{\partial}{\partial \theta} X^T \Sigma^{-1}(\theta) \right\} E(y - X\phi) = 0,$$

for all θ and ϕ , thus $I_{11}^* = I_{11}$. By result in Case 2, the pseudolikelihood ratio testing the j th variance component being zero, i.e. $H_0 : \theta_j = 0$, has an asymptotically 50:50 mixture of χ_0^2 and χ_1^2 .

3.2. Example 2: Behrens–Fisher problem

It can be shown that

$$E \left\{ \frac{\partial^2 \log L(\theta, \phi)}{\partial \theta_{i'} \partial \phi_i} \right\} = -\delta_{i'i} E \left(\sum_{j=1}^{n_i} \frac{y_{ij} - \theta_i}{\phi_i^2} \right) = 0 \quad (i', i = 1, 2),$$

where $\delta_{i'i} = 1$ if $i' = i$ and 0 otherwise. Therefore, $I_{12}(\theta, \phi) = 0$. As indicated in § 2.1, the pseudolikelihood ratio has, asymptotically, a χ_1^2 distribution.

3.3. Example 3: Frailty survival models

Let $y_i = (y_{i1}, \dots, y_{i4})$ be the observations from the i th cluster ($i = 1, \dots, n$), with the joint survival distribution of the form equation (1) where the marginal survival distribution for the j th individual, $S_j = S_j(y_j; \phi)$, is indexed by ϕ . For simplicity, we assume no censoring, so the likelihood function for $(\theta_1, \theta_2, \phi)$ is

$$L(\theta_1, \theta_2, \phi) \propto \frac{\partial^4 S(t_1, t_2, t_3, t_4)}{\partial t_1 \partial t_2 \partial t_3 \partial t_4} = \frac{\partial h_1}{\partial t_1} \frac{\partial h_1}{\partial t_2} \frac{\partial h_2}{\partial t_3} \frac{\partial h_2}{\partial t_4} \frac{\partial^4 S}{\partial h_1^2 \partial h_2^2},$$

where

$$h_1 = \sum_{j=1}^2 S_j(t_j)^{1-\theta_1} - 1, \quad h_2 = \sum_{j=3}^4 S_j(t_j)^{1-\theta_1} - 1,$$

$$S = h^{1/(1-\theta_2)} = (h_1^\alpha + h_2^\alpha - 1)^{1/(1-\theta_2)}, \quad \partial h_1 / \partial t_j = (1 - \theta_1) S_j(t_j)^{1-\theta_1} \lambda_j(t_j) \quad (j = 1, 2),$$

$$\partial h_2 / \partial t_j = (1 - \theta_1) S_j(t_j)^{1-\theta_1} \lambda_j(t_j) \quad (j = 3, 4),$$

$\lambda_j(t_j)$ is the marginal hazard function for the j th individual, and

$$\frac{\partial^4 S}{\partial h_1^2 \partial h_2^2} = \alpha^2 (\alpha - 1)^2 h_1^{\alpha-2} h_2^{\alpha-2} \frac{\partial^2 S}{\partial h^2} + \alpha^3 (\alpha - 1) \left\{ h_1^{2(\alpha-1)} h_2^{\alpha-2} + h_1^{\alpha-2} h_2^{2(\alpha-1)} \right\} \frac{\partial^3 S}{\partial h^3}$$

$$+ \alpha^4 (h_1 h_2)^{2(\alpha-1)} \frac{\partial^4 S}{\partial h^4}$$

with $\partial^l S / \partial h^l = [\prod_{j=1}^l \{1/(1 - \theta_2) + 1 - j\}] h^{1/(1-\theta_2)-l}$ ($l = 2, 3, 4$). A hypothesis of interest is $H_0 : \theta_2 = 1$ which implies the independence of y_{ij} s for two individuals from different households.

The parameters θ_1 and θ_2 must satisfy the inequality $\theta_1 \geq \theta_2 \geq 1$. The cones which approximate Ω_0 and Ω_1 are $[0, \infty) \times \{0\}$ and $\{(t_1, t_2) : t_1 \geq t_2 \geq 0\}$. Since the parameter space cannot be written as the product of intervals, results in Cases 1–5 may not be applied. A simple reparameterization from $\theta = (\theta_1, \theta_2)$ to $\tau = (\theta_2, \delta)$, $\delta = \theta_1 - \theta_2$, yields the cones approximating the parameter space of (θ_2, δ) being $\{0\} \times R$ and $[0, \infty) \times R$. By result in Case 2, the asymptotic distribution of the pseudolikelihood ratio is

$$\frac{1}{2} \chi_0^2 + \frac{1}{2} \frac{e_{(1)}^*}{e_{(1)}} \chi_1^2,$$

where $e_{(1)}$ and $e_{(1)}^*$ are the (1, 1) elements of matrix I_{11}^{-1} and $I_{11}^{-1} I_{11}^* I_{11}^{-1}$, respectively,

$$I_{11} = \lim_{n \rightarrow \infty} \left[E \left\{ -\frac{1}{n} \frac{\partial^2 \log L(\tau_0, \phi_0)}{\partial \tau^2}; \tau_0, \phi_0 \right\} \right], \quad I_{12} = \lim_{n \rightarrow \infty} \left[E \left\{ -\frac{1}{n} \frac{\partial^2 \log L(\tau_0, \phi_0)}{\partial \tau \partial \phi}; \tau_0, \phi_0 \right\} \right],$$

$L(\tau, \phi) = L(\theta_2 + \delta, \theta_2, \phi)$, $I_{11}^* = I_{11} + I_{12} \Sigma_{22} I_{12}^T$, and Σ_{22} is the asymptotic variance of the consistent estimator of ϕ .

3.4. Example 4: Teratological experiments

For simplicity, we consider the special case of a single treatment group and a control group and assume a logistic dose–response model. Denote $x_i = 0$ if the litter i is in the control group and 1 if in the treatment group, and ϕ_c and ϕ_t for correlations within control and treatment groups, respectively. The moment estimates for ϕ_c and ϕ_t can be found by solving the following equations for nonnegative roots,

$$\begin{cases} \sum_{x_i=0} \{(p_i - \pi_c)^2 / [\pi_c(1 - \pi_c)\{1 + (m_i - 1)\phi_c\} / m_i] - (n_0 - 1) / n_0\} = 0, \\ \sum_{x_i=1} \{(p_i - \pi_t)^2 / [\pi_t(1 - \pi_t)\{1 + (m_i - 1)\phi_t\} / m_i] - (n_1 - 1) / n_1\} = 0, \end{cases}$$

where $p_i = y_i / m_i$, $\pi_c = \exp(\beta_0) / \{1 + \exp(\beta_0)\}$ and $\pi_t = \exp(\beta_0 + \beta_1) / \{1 + \exp(\beta_0 + \beta_1)\}$, and n_0 and n_1 are numbers of litters in control and treatment groups, respectively. The parameters π_c and π_t are unknown. One may estimate (ϕ_c, ϕ_t) by substituting (π_c, π_t) by a consistent estimator, such as the maximum likelihood estimator ignoring the correlations, i.e. $\hat{\pi}_c = \sum_{x_i=0} y_i / \sum_{x_i=0} m_i$ and $\hat{\pi}_t = \sum_{x_i=1} y_i / \sum_{x_i=1} m_i$.

To alleviate the impact of the nuisance parameter on the parameter of interest, one could base the inference for β on the pseudolikelihood $L^*(\beta_0, \beta_1) = L(\beta_0, \beta_1, \hat{\phi}_c, \hat{\phi}_t)$. Here we derive the asymptotic distribution of the pseudolikelihood ratio test statistic for $H_0 : \beta_1 = 0$, i.e. no treatment effect, when the intralitter correlation in the control group, ϕ_c , is 0, but the intralitter correlation in the treatment group $\phi_t > 0$. Denote $\theta = (\beta_1, \beta_0)$ and $\phi = (\phi_c, \phi_t)$. The true value of nuisance parameter, ϕ_c , lies on the boundary of its parameter space. According to the results in §2.3, the asymptotic distribution of the pseudolikelihood ratio test statistic is $Z_1^2 / e_{(1)}$, where $e_{(1)}$ is the (1, 1) element of the matrix I_{11}^{-1} , $I_{11} = E[\partial \log L(\theta_0, \phi_0) / \partial \theta \{ \partial \log L(\theta_0, \phi_0) / \partial \theta \}^T]$ and $Z = (Z_1, Z_2)$ is a 50:50 mixture of bivariate normal distributions described by equations (5) and (6). The Fisher information matrix I_{11} and the asymptotic variance of $\hat{\phi}^*$, Σ_{22} , are calculated in Appendix B.

4. SIMULATION STUDY

To explore the theoretical findings empirically, we conducted two simulation studies. In the first, we applied the pseudolikelihood ratio test to test for the statistical significance of a variance component in the model described in Example 1 and compared its performance with the likelihood ratio test based on its asymptotic distribution and the exact finite sample distribution

(Crainiceanu & Ruppert, 2004). The model described in Example 1 belongs to the linear mixed model with one variance component, specifically,

$$y = X\phi + Zb + \epsilon, \quad E \begin{pmatrix} b \\ \epsilon \end{pmatrix} = \begin{pmatrix} 0_K \\ 0_n \end{pmatrix}, \quad \text{cov} \begin{pmatrix} b \\ \epsilon \end{pmatrix} = \begin{pmatrix} \sigma_b^2 I_K & 0 \\ 0 & \sigma_\epsilon^2 I_n \end{pmatrix}, \quad (7)$$

where b is random effects due to specific genetic or nongenetic factors shared among relatives, and ϵ represents random deviations unique to each individual. With nuisance parameter ϕ replaced by the conventional least-squares estimator $\hat{\phi} = (X^T X)^{-1} X^T y$, the pseudolikelihood is $L^*(\sigma_b^2, \sigma_\epsilon^2) = L(\hat{\phi}, \sigma_b^2, \sigma_\epsilon^2)$ and the pseudolikelihood ratio test statistic for testing $H_0 : \sigma_b^2 = 0$ is

$$T = 2 \left\{ \sup_{\{\sigma_b^2 \geq 0, \sigma_\epsilon^2 \geq 0\}} L^*(\sigma_b^2, \sigma_\epsilon^2) - \sup_{\{\sigma_b^2 = 0, \sigma_\epsilon^2 \geq 0\}} L^*(\sigma_b^2, \sigma_\epsilon^2) \right\}.$$

Maximizing $L^*(\sigma_b^2, \sigma_\epsilon^2)$ over σ_b^2 and σ_ϵ^2 simultaneously by Newton–Raphson or Fisher’s scoring method could lead to negative estimates of variances, so we reparameterized σ_b^2 by $\lambda \sigma_\epsilon^2$, and maximized the pseudolikelihood first for fixed λ and then over λ . Testing the null hypothesis $H_0 : \sigma_b^2 = 0$ is equivalent to testing $H_0 : \lambda = 0$ (Crainiceanu & Ruppert, 2004). Twice the log-pseudolikelihood function is

$$2 \log \{L(\hat{\phi}, \lambda, \sigma_\epsilon^2)\} = -n \log(\sigma_\epsilon^2) - \log |V_\lambda| - \frac{(y - X\hat{\phi})^T V_\lambda^{-1} (y - X\hat{\phi})}{\sigma_\epsilon^2},$$

where $V_\lambda = I_n + \lambda Z Z^T$. Under H_0 , $\tilde{\sigma}_\epsilon^2 = \arg \max_{\sigma_\epsilon^2} [\log \{L(\hat{\phi}, 0, \sigma_\epsilon^2)\}] = y^T P_0 y / n$, where $P_0 = I_n - X(X^T X)^{-1} X^T$. For fixed λ , $\hat{\sigma}_\epsilon^2(\lambda) = \arg \max_{\sigma_\epsilon^2} [\log \{L(\hat{\phi}, \lambda, \sigma_\epsilon^2)\}] = y^T P_0^T V_\lambda^{-1} P_0 y / n$. Therefore, the pseudolikelihood ratio test statistic is

$$T = \sup_{\lambda \geq 0} [-\log |V_\lambda| + n \log(\tilde{\sigma}_\epsilon^2) - n \log \{\hat{\sigma}_\epsilon^2(\lambda)\}]. \quad (8)$$

The above supremum can be found by a grid search over possible values of λ . The likelihood ratio test statistic has a form similar to equation (8), except that $\hat{\sigma}_\epsilon^2(\lambda)$ is replaced by $y^T P_\lambda^T V_\lambda^{-1} P_\lambda y / n$, where $P_\lambda = I_n - X(X^T V_\lambda^{-1} X)^{-1} X^T V_\lambda^{-1}$.

To obtain the size of the tests, we drew y from (7) with $\sigma_b^2 = 0$. The family size was set to 3 and the number of families varied from 20 to 40. The true value of ϕ was set to be $(1, 2, 3)^T$ and the matrix X was fixed at randomly drawn numbers from standard normal distribution. The distribution of the pseudolikelihood ratio test statistic does not depend on the magnitude of σ_ϵ^2 because testing the random effect is equivalent to testing the within group correlation $\sigma_b^2 / (\sigma_b^2 + \sigma_\epsilon^2)$, which only depends on $\sigma_b^2 / \sigma_\epsilon^2$. For simplicity, we fixed $\sigma_\epsilon^2 = 1$. For each generated dataset, we computed the pseudolikelihood ratio test statistic and the likelihood ratio test statistic, and rejected the null if the test statistic is larger than the $(1 - \alpha)$ th quantile of 50:50 mixture of χ_0^2 and χ_1^2 , i.e. the $(1 - 2\alpha)$ th quantile of χ_1^2 . We also computed the exact distribution of the likelihood ratio test based on results in Crainiceanu & Ruppert (2004) and obtained the $(1 - \alpha)$ th quantile of its distribution. To obtain the power of the test, a similar procedure was implemented, except that y was drawn from model (7) with $\sigma_b^2 / \sigma_\epsilon^2 = 0.1, 0.2$ and 0.4 , respectively.

Table 1 shows results from 5000 simulations. The nominal level for Type I error, α , was set to 0.05 and 0.01. When the null hypothesis is true, both the pseudolikelihood ratio test and the likelihood ratio test based on its exact distribution have proportions of rejections within 95% confidence intervals for the nominal error rates, i.e. 0.007–0.013 for $\alpha = 0.01$, and 0.044–0.056 for $\alpha = 0.05$, suggesting that the asymptotic approximation for the pseudolikelihood ratio test statistic is adequate for moderate sample sizes. The pseudolikelihood test is slightly less powerful

Table 1. Empirical rejection rates (%) in 5000 simulations for the pseudolikelihood ratio test and the likelihood ratio test for testing for a variance component, based on samples of size n , for different variance ratios

n	$\sigma_b^2/\sigma_\epsilon^2$	PLRT asymptotic		LRT asymptotic		LRT exact	
		$\alpha = 0.05$ Rejection (%)	$\alpha = 0.01$ Rejection (%)	$\alpha = 0.05$ Rejection (%)	$\alpha = 0.01$ Rejection (%)	$\alpha = 0.05$ Rejection (%)	$\alpha = 0.01$ Rejection (%)
20	0	4.4	0.8	5.2	1.2	4.8	0.9
	0.1	15.4	4.3	17.3	5.5	15.8	4.6
	0.2	30.8	12.2	34.4	14.6	32.0	12.4
	0.4	60.8	36.4	63.7	40.6	61.5	36.9
30	0	5.1	0.9	5.9	1.1	5.5	1.1
	0.1	20.0	6.3	21.6	7.5	21.1	6.6
	0.2	41.6	19.5	43.3	21.8	42.8	20.3
	0.4	77.3	54.3	79.0	57.5	77.6	54.9
40	0	4.7	1.1	5.2	1.2	5.0	1.2
	0.1	23.2	8.9	24.8	9.6	24.1	9.0
	0.2	50.8	26.2	52.4	28.3	51.4	26.4
	0.4	86.9	68.5	87.9	70.0	87.4	68.2

n , number of families; PLRT asymptotic, pseudolikelihood ratio test based on its asymptotic distribution; LRT asymptotic, likelihood ratio test based on its asymptotic distribution; LRT exact, likelihood ratio test based on its exact distribution.

than the likelihood ratio test based on its exact distribution. On the other hand, the likelihood ratio test based on its asymptotic distribution is liberal but more powerful compared to the other two tests. In general, the pseudolikelihood ratio test is more conservative and less powerful than the likelihood ratio test when the nuisance parameter is estimated by the maximum likelihood estimator under the null.

In the second simulation study, we applied the pseudolikelihood ratio test to test for the treatment effect in a two-group teratological experiment. We simulated data that mimic the litter size structure of Weil (1970) with $\beta_0 = 1.05$, $\phi_l = 0.05$ and $\phi_c = 0$; see Example 4 of § 1 for detailed descriptions of the statistical models considered. The true value of treatment effect, β_1 , was set to vary from 0 to -1 in order to obtain the size and the power of the test, at nominal levels of 0.05 and 0.01, respectively. The number of litters varied from 16 to 64 per group. In addition to the pseudolikelihood ratio test, we also applied two other methods for comparison; the Wald test, comparing the Wald test statistic to the standard normal distribution; and the naive likelihood ratio test ignoring the intralitter correlations. Table 2 shows results from 5000 simulations. Under all settings considered, the pseudolikelihood ratio test has proportions of rejections within 95% confidence intervals for the nominal error rates. The Wald test was liberal and its performance was not improved as the sample size became larger. This is because the true asymptotic distribution of the Wald test statistic is a mixture of two normals (Self & Liang, 1987). The naive likelihood ratio test has proportions of rejections higher than the nominal level compared to the Wald test. This simulation suggests that neither the Wald test nor the naive likelihood ratio test can be recommended in practice.

5. DISCUSSION

One interesting area of future research is to extend the current results to situations where the nuisance parameters are infinite-dimensional. For instance, in Example 4 of the multilevel frailty

Table 2. Empirical rejection rates (%) in 5000 simulations for the pseudolikelihood ratio test, Wald test and the naive likelihood ratio test for testing the treatment effect in a two-group teratological experiment, based on samples of size n_0/n_1 , for different treatment effects. The true intralitter correlation is 0 for the control group and 0.05 for the treatment group

n_0/n_1	β_1	PLRT asymptotic		Wald test		LRT naive	
		$\alpha = 0.05$ Rejection (%)	$\alpha = 0.01$ Rejection (%)	$\alpha = 0.05$ Rejection (%)	$\alpha = 0.01$ Rejection (%)	$\alpha = 0.05$ Rejection (%)	$\alpha = 0.01$ Rejection (%)
16/16	0	5.2	1.0	5.7	1.3	8.3	2.0
	-0.5	42.0	21.0	41.6	20.4	52.3	30.8
	-1.0	95.7	86.2	95.1	84.9	97.5	92.9
32/32	0	4.7	1.1	5.9	1.3	7.7	2.0
	-0.5	70.9	47.4	69.2	47.2	77.5	58.4
	-1.0	99.9	99.6	99.9	99.5	100.0	99.9
64/64	0	4.7	1.0	6.3	1.6	7.2	2.0
	-0.5	95.0	85.0	93.5	82.9	96.9	90.3
	-1.0	100.0	100.0	100.0	100.0	100.0	100.0

n_0/n_1 , number of litters in control/treatment group; β_1 , log odds ratio between the treatment and the control groups; PLRT asymptotic, pseudolikelihood ratio test based on its asymptotic distribution; LRT naive, likelihood ratio test ignoring the intralitter correlations.

survival models, the marginal survival distribution may be left unspecified. It will be of interest to examine the asymptotic behaviour of the pseudolikelihood ratio test statistic in this situation.

ACKNOWLEDGEMENT

The authors thank the referees and the editor for comments that substantially improved the presentation. This research was supported in part by a grant from the National Institutes of Health, U.S.A.

APPENDIX A

Proof of Theorem 2

Proof. For simplicity of notation, let $\theta_0 = 0$. Under the same regularity conditions as Chernoff (1954), for any θ in a neighbourhood N of $\theta = 0$, we have

$$\begin{aligned} 2\{\log L^*(\theta) - \log L^*(0)\} &= 2\{\log L(\theta, \hat{\phi}) - \log L(0, \hat{\phi})\} \\ &= 2 \left\{ \frac{\partial \log L(0, \hat{\phi})}{\partial \theta} \right\}^\top \theta + \theta^\top \left\{ \frac{\partial^2 \log L(0, \hat{\phi})}{\partial \theta^2} \right\} \theta + O_p(n)\|\theta\|^3 \\ &= 2(n^{1/2}A_1^*)^\top(n^{1/2}\theta) + (n^{1/2}\theta)^\top B_{11}^*(n^{1/2}\theta) + O_p(n)\|\theta\|^3, \end{aligned}$$

where $A_1^* = n^{-1} \sum_{i=1}^n \partial \log L(0, \hat{\phi})/\partial \theta$ and $B_{11}^* = n^{-1} \sum_{i=1}^n \partial^2 \log L(0, \hat{\phi})/\partial \theta^2$.

The asymptotic distribution of $n^{1/2}A_1^*$ is normal with mean zero and variance matrix I_{11}^* , where $I_{11}^* = I_{11} + I_{12}\Sigma_{22}I_{12}^\top$, i.e. $n^{1/2}A_1^* \rightarrow N(0, I_{11}^*)$ and $B_{11}^* \rightarrow -I_{11}$ as $n \rightarrow \infty$.

Let $\theta = I_{11}^{-1}A_1^* + \eta$ with $\eta = O_p(n^{-1/2})$. Then,

$$\begin{aligned} 2\{\log L^*(\theta) - \log L^*(0)\} &= 2(n^{1/2}A_1^*)^\top(n^{1/2}\theta) + (n^{1/2}\theta)^\top I_{11}^*(n^{1/2}\theta) + O_p(1) \\ &= 2(n^{1/2}\eta)^\top I_{11}(n^{1/2}\eta) + (n^{1/2}A_1^*)^\top I_{11}^{-1}(n^{1/2}A_1^*) + O_p(1). \end{aligned}$$

Let $-2 \log \lambda^* = -2 \log\{P_{\Omega_0}/P_{\Omega_1}\}$.

Applying Lemma 1 and the above approximations to Ω_0 and Ω_1 , we get

$$-2 \log \lambda^* = \inf_{\theta \in \Omega_0} (n^{1/2} \eta)^\top I_{11} (n^{1/2} \eta) - \inf_{\theta \in \Omega_1} (n^{1/2} \eta)^\top I_{11} (n^{1/2} \eta) + O_p(1).$$

But for any set φ that can be approximated by C_φ ,

$$\inf_{\theta \in \varphi} (y - \theta)^\top I_{11} (y - \theta) = \inf_{\theta \in C_\varphi} (y - \theta)^\top I_{11} (y - \theta) + o(\|y\|^2),$$

and therefore,

$$\begin{aligned} -2 \log \lambda^* &= \inf_{\theta \in C_{\Omega_0}} (n^{1/2} I_{11}^{-1} A_1^* - n^{1/2} \theta)^\top I_{11} (n^{1/2} I_{11}^{-1} A_1^* - n^{1/2} \theta) \\ &\quad - \inf_{\theta \in C_{\Omega_1}} (n^{1/2} I_{11}^{-1} A_1^* - n^{1/2} \theta)^\top I_{11} (n^{1/2} I_{11}^{-1} A_1^* - n^{1/2} \theta) + O_p(1). \end{aligned}$$

Since C_Ω and C_{Ω_0} are positively homogeneous,

$$-2 \log \lambda^* = \inf_{\theta \in C_{\Omega_0}} (Z_n - \theta)^\top I_{11} (Z_n - \theta) - \inf_{\theta \in C_{\Omega_1}} (Z_n - \theta)^\top I_{11} (Z_n - \theta) + O_p(1),$$

where $Z_n = n^{1/2} I_{11}^{-1} A_1^*$ and $Z_n \rightarrow Z \sim N(0, I_{11}^{-1} I_{11}^* I_{11}^{-1})$ as $n \rightarrow \infty$. The function $\inf_{\theta \in \varphi} (Z - \theta)^\top I_{11} (Z - \theta)$ is a continuous function of Z . Therefore, the asymptotic distribution of $-2 \log \lambda^*$ is the distribution of

$$\inf_{\theta \in C_{\Omega_0}} (Z - \theta)^\top I_{11} (Z - \theta) - \inf_{\theta \in C_{\Omega_1}} (Z - \theta)^\top I_{11} (Z - \theta)$$

under the assumption that $Z \sim N(0, I_{11}^{-1} I_{11}^* I_{11}^{-1})$. By the equation $T = \max\{-2 \log \lambda^*(x), 0\}$, we have that the asymptotic distribution of T is the distribution of

$$\inf_{\theta \in C_{\Omega_0}} (Z - \theta)^\top I_{11} (Z - \theta) - \inf_{\theta \in C_\Omega} (Z - \theta)^\top I_{11} (Z - \theta),$$

where $C_\Omega = C_{\Omega_0} \cup C_{\Omega_1}$. □

APPENDIX B

Calculation of Fisher information matrices and asymptotic variances of moment estimators in Example 4

The loglikelihood function contributed by the i th litter is

$$\begin{aligned} \log L_i(\theta, \phi) &\propto \log Pr(y_i | m_i, x_i; \theta, \phi) \\ &= I(x_i = 0) \left\{ \sum_{k=0}^{y_i-1} \log(\pi_c + k\psi_c) + \sum_{k=0}^{m_i-y_i-1} \log(1 - \pi_c + k\psi_c) - \sum_{k=0}^{m_i-1} \log(1 + k\psi_c) \right\} \\ &\quad + I(x_i = 1) \left\{ \sum_{k=0}^{y_i-1} \log(\pi_t + k\psi_t) + \sum_{k=0}^{m_i-y_i-1} \log(1 - \pi_t + k\psi_t) - \sum_{k=0}^{m_i-1} \log(1 + k\psi_t) \right\}, \end{aligned} \tag{B1}$$

where $\psi_c = \phi_c / (1 - \phi_c)$, $\psi_t = \phi_t / (1 - \phi_t)$, $\pi_c = \exp(\beta_0) / \{1 + \exp(\beta_0)\}$ and $\pi_t = \exp(\beta_0 + \beta_1) / \{1 + \exp(\beta_0 + \beta_1)\}$. Denote the score vector for θ by $S_\theta^i = (S_{\beta_1}^i, S_{\beta_0}^i)^\top$, where $S_{\beta_j}^i = \partial \log L_i(\theta, \phi) / \partial \beta_j$ ($j = 0, 1$); and the score vector for ϕ by $S_\phi^i = (S_{\phi_c}^i, S_{\phi_t}^i)^\top$, where $S_{\phi_j}^i = \partial \log L_i(\theta, \phi) / \partial \phi_j$ ($j = c, t$). Therefore, the Fisher information matrices I_{11} and I_{12} can be estimated by their empirical versions, $n^{-1} \sum_{i=1}^n \{S_\theta^i (S_\theta^i)^\top\}$ and $n^{-1} \sum_{i=1}^n \{S_\phi^i (S_\phi^i)^\top\}$, respectively. The simple representation in equation (B1) avoids having different forms of density functions, binomial if $\phi = 0$ and beta-binomial if $\phi > 0$, and does not require the digamma function when calculating the scores.

To calculate the asymptotic variances of $\hat{\phi}_c^*$ and $\hat{\phi}_i^*$, we first notice that they are asymptotically independent and we only need to calculate the asymptotic variance of $\hat{\phi}_c^*$ since calculation for $\hat{\phi}_i^*$ is similar.

Denote $g(\phi_c, m_i, y_i, \pi_c) = (p_i - \pi_c)^2 / [\pi_c(1 - \pi_c)\{1 + (m_i - 1)\phi_c\} / m_i] - (n_0 - 1) / n_0$. The moment estimator $\hat{\phi}_c^*$ is the solution of $\sum_{x_i=0} g(\hat{\phi}_c^*, m_i, y_i, \hat{\pi}_c) = 0$, where $\hat{\pi}_c = \sum_{x_i=0} y_i / \sum_{x_i=0} m_i$. Apply the Taylor expansion of $g(\hat{\phi}_c^*, m_i, y_i, \hat{\pi}_c)$ around ϕ_c :

$$\sum_{x_i=0} [g(\phi_c, m_i, y_i, \hat{\pi}_c) + \{\partial g(\phi_c, m_i, y_i, \hat{\pi}_c) / \partial \phi_c\}(\hat{\phi}_c^* - \phi_c)] \approx 0.$$

Therefore,

$$n_0^{1/2}(\hat{\phi}_c^* - \phi_c) \approx \left\{ -n_0^{-1} \sum_{x_i=0} \partial g(\phi_c, m_i, y_i, \hat{\pi}_c) / \partial \phi_c \right\}^{-1} n_0^{-1/2} \sum_{x_i=0} g(\phi_c, m_i, y_i, \hat{\pi}_c).$$

Apply again the Taylor expansion of $g(\phi_c, m_i, y_i, \hat{\pi}_c)$, this time around π_c :

$$\begin{aligned} n_0^{1/2}(\hat{\phi}_c^* - \phi_c) &\approx \left\{ -n_0^{-1} \sum_{x_i=0} \frac{\partial g(\phi_c, m_i, y_i, \pi_c)}{\partial \phi_c} \right\}^{-1} \\ &\times \left[\left\{ n_0^{-1/2} \sum_{x_i=0} g(\phi_c, m_i, y_i, \pi_c) \right\} + \left\{ n_0^{-1} \sum_{x_i=0} \frac{\partial g(\phi_c, m_i, y_i, \pi_c)}{\partial \pi_c} \right\} n_0^{1/2}(\hat{\pi}_c - \pi_c) \right] \\ &\approx \left\{ -n_0^{-1} \sum_{x_i=0} \frac{\partial g(\phi_c, m_i, y_i, \pi_c)}{\partial \phi_c} \right\}^{-1} \left[\left\{ n_0^{-1/2} \sum_{x_i=0} g(\phi_c, m_i, y_i, \pi_c) \right\} \right. \\ &\quad \left. - \left\{ n_0^{-1} \sum_{x_i=0} \frac{\partial g(\phi_c, m_i, y_i, \pi_c)}{\partial \pi_c} \right\} \left\{ n_0^{-1} \sum_{x_i=0} m_i \right\}^{-1} n_0^{-1/2} \sum_{x_i=0} (m_i \pi_c - y_i) \right] \\ &= \left\{ -n_0^{-1} \sum_{x_i=0} \frac{\partial g(\phi_c, m_i, y_i, \pi_c)}{\partial \phi_c} \right\}^{-1} \times n_0^{-1/2} \sum_{x_i=0} \\ &\quad \times \left[g(\phi_c, m_i, y_i, \pi_c) - \left\{ n_0^{-1} \sum_{x_i=0} \frac{\partial g(\phi_c, m_i, y_i, \pi_c)}{\partial \pi_c} \right\} \left\{ n_0^{-1} \sum_{x_i=0} m_i \right\}^{-1} (m_i \pi_c - y_i) \right], \end{aligned}$$

where the second to last step is by the fact that $\hat{\pi}_c$ is the solution of the estimating equation $\sum_{x_i=0} (m_i \pi_c - y_i) = 0$.

Finally, the asymptotic variance of $\hat{\phi}_c^*$ is

$$\begin{aligned} &\left\{ -\frac{1}{n_0} \sum_{x_i=0} \frac{\partial g(\phi_c, m_i, y_i, \pi_c)}{\partial \phi_c} \right\}^{-2} \\ &\times \frac{1}{n_0} \sum_{x_i=0} \left[g(\phi_c, m_i, y_i, \pi_c) - \left\{ \frac{1}{n_0} \sum_{x_i=0} \frac{\partial g(\phi_c, m_i, y_i, \pi_c)}{\partial \pi_c} \right\} \left\{ \frac{1}{n_0} \sum_{x_i=0} m_i \right\}^{-1} (m_i \pi_c - y_i) \right]^2. \end{aligned}$$

REFERENCES

BANDEEN-ROCHE, K. J. & LIANG, K.-Y. (1996). Modelling failure-time associations in data with multiple levels of clustering. *Biometrika* **83**, 29–39.
 BEATY, T. H., SELF, S. G., LIANG, K.-Y., CONNOLLY, M. A., CHASE, G. A. & KWITEROVICH, P. O. (1985). Use of robust variance components models to analyse triglyceride data in families. *Ann. Hum. Genet.* **49**, 315–28.

- CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25**, 573–8.
- CLAYTON, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–51.
- CRAINICEANU, C. M. & RUPPERT, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *J. R. Statist. Soc. B* **66**, 165–85.
- GONG, G. & SAMANIEGO, F. J. (1981). Pseudo maximum likelihood estimation: theory and applications. *Ann. Statist.* **9**, 861–9.
- GRAYBILL, F. A. (1976). *Theory and Application of the Linear Model*. North Scituate, MA: Duxbury Press.
- HASEMAN, J. K. & KUPPER, L. L. (1979). Analysis of dichotomous response data from certain toxicological experiments. *Biometrics* **35**, 281–93.
- HAUCK, W. W. & DONNER, A. (1977). Wald's test as applied to hypotheses in logit analysis. *J. Am. Statist. Assoc.* **72**, 851–3.
- LIANG, K.-Y. & SELF, S. G. (1996). On the asymptotic behaviour of the pseudolikelihood ratio test statistic. *J. R. Statist. Soc. B* **58**, 785–96.
- OAKES, D. (1989). Bivariate survival models induced by frailties. *J. Am. Statist. Assoc.* **84**, 487–93.
- PARKE, W. R. (1986). Pseudo maximum likelihood estimation: the asymptotic distribution. *Ann. Statist.* **14**, 355–7.
- PIERCE, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics. *Ann. Statist.* **10**, 475–8.
- SELF, S. G. & LIANG, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Statist. Assoc.* **82**, 605–10.
- SHIH, J. H. & LOUIS, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* **51**, 1384–99.
- STRAM, D. O. & LEE, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–7.
- VAUPEL, J. W., MANTON, K. G. & STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–54.
- WEIL, C. S. (1970). Selection of the valid number of sampling units and a consideration of their combination in toxicological studies involving reproduction, teratogenesis or carcinogenesis. *Food Cosmet. Toxicol.* **8**, 177–82.
- WILLIAMS, D. A. (1988). Estimation bias using the beta-binomial distribution in teratology. *Biometrics* **44**, 305–9.

[Received November 2009. Revised January 2010]