

BOOTSTRAPPING PSEUDOLIKELIHOOD MODELS FOR CLUSTERED BINARY DATA

MARC AERTS AND GERDA CLAESKENS*

*Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus,
B3590 Diepenbeek, Belgium*

(Received June 3, 1997; revised March 10, 1998)

Abstract. Asymptotic properties of the parametric bootstrap procedure for maximum pseudolikelihood estimators and hypothesis tests are studied in the general framework of associated populations. The technique is applied to the analysis of toxicological experiments which, based on pseudolikelihood inference for clustered binary data, fits into this framework. It is shown that the bootstrap approximation can be used as an interesting alternative to the classical asymptotic distribution of estimators and test statistics. Finite sample simulations for clustered binary data models confirm the asymptotic theory and indicate some substantial improvements.

Key words and phrases: Clustered binary data, developmental toxicity, exponential family, parametric bootstrap, pseudolikelihood.

1. Introduction

In this paper the parametric bootstrap technique is studied for hypothesis tests based on clustered binary data. This type of data arises e.g. from developmental toxicity studies. Different types of models (marginal, conditional, random effects models) are available and estimation methods range from full likelihood to methods based on quasi-likelihood, generalized estimating equations or pseudolikelihood (see e.g. Pengergast *et al.* (1996)). We focus attention on the latter method for conditionally specified models.

Recently, Molenberghs and Ryan (1999) proposed a conditional model for multivariate clustered binary outcomes, based on the multivariate exponential family model as proposed by Cox (1972). The model benefits from the elegance and simplicity of exponential family theory. It allows for flexible response relationships and combines a likelihood basis with numerical stability. A main problem however, particularly with large clusters, is the evaluation of the normalizing constant.

* The research of Gerda Claeskens is supported by the Fund for Scientific Research - Flanders (Belgium) (FWO).

Geys *et al.* (1997, 1999) propose the use of a so-called pseudolikelihood for correlated outcomes. The principal idea is to replace the joint density by a product of conditional densities that do not necessarily multiply to the joint distribution. The advantage of this particular type of misspecification is that the normalizing constant cancels. Pseudolikelihood estimation is widely applicable, see e.g. Besag (1975), Arnold and Strauss (1991), Cressie (1991).

Geys *et al.* (1997, 1999) show that the use of pseudolikelihood estimators results in moderate efficiency loss. They also propose likelihood ratio tests in the pseudolikelihood framework. As a consequence of the likelihood misspecification, the asymptotic distribution of the pseudolikelihood ratio test statistic is a weighted sum of independent χ_1^2 variables. The weights are unknown eigenvalues and have to be estimated. The estimators can be calculated under the null hypothesis but also under the alternative hypothesis. All this complicates the computation of critical points and p -values. In this paper it is shown theoretically that the parametric bootstrap leads to a consistent estimator for the distribution of the pseudolikelihood ratio (PLR) test statistic. The bootstrap approach does not need any additional estimation of unknown eigenvalues. It automatically corrects for the misspecification of the joint distribution. We focus on the parametric bootstrap since it easily allows to generate bootstrap samples reflecting the data mechanism under any null hypothesis (in contrast with a nonparametric bootstrap method).

A small simulation study illustrates the benefit of using the parametric bootstrap for testing purposes in the area of clustered binary data from toxicity studies. There are only a few papers in which bootstrap techniques are used for clustered binary data. They mainly use the nonparametric bootstrap method to estimate the variance of parameter estimators (Lockhart *et al.* (1992), Carr and Portier (1993)). Recently, Frangos and Schucany (1995) proposed a particular parametric bootstrap procedure to construct improved confidence intervals in certain toxicological experiments.

This paper is organized as follows. Section 2 summarizes the basic asymptotic properties of the maximum pseudolikelihood estimators and tests. These results are reconsidered for the bootstrapped pseudolikelihood estimators in Sections 3 (estimation) and 4 (hypothesis tests). Next to the PLR test we also include results for the robust Wald and the robust score test. The theoretical results in these sections are presented in the general framework of multiparameter pseudolikelihood when sampling from a finite number of associated populations. According to Bradley and Gart (1962), associated populations are distinct but related, in the sense that they may have some parameters in common. As indicated in Section 5, the analysis of clustered binary data from toxicity studies fits into this framework. This last section briefly describes the conditional model of Molenberghs and Ryan (1999) and the pseudolikelihood approach of Geys *et al.* (1997, 1999) and presents the results of a finite sample simulation study. It illustrates the estimation problem of the unknown eigenvalues and shows that the χ^2 type PLR test, based on these estimators, has severe problems in achieving the nominal level. The bootstrap PLR test, using bootstrap critical points, seems to nicely correct this towards the nominal size. Similar problems can be remedied by the bootstrap for the Wald

and, less pronounced, for the score test.

2. Pseudolikelihood estimation and inference

The basic asymptotic properties of maximum pseudolikelihood estimators and hypothesis tests, when sampling from a finite number of associated populations, combine classical maximum likelihood theory as presented in e.g. Serfling (1980) and Lehmann (1983), the extension to associated populations (Bradley and Gart (1962)) and pseudolikelihood estimation and testing (Geys *et al.* (1997, 1999)). The proofs of the theorems stated in this section are omitted.

Let $f_i(y, \theta)$ with $y = (y_1, \dots, y_{m_i}) \in R_i$ ($i = 1, \dots, p$) denote p joint density or discrete probability functions. The support R_i does not depend on the unknown parameter vector $\theta = (\theta_1, \dots, \theta_r)^T$. Let Y_{i1}, \dots, Y_{in_i} be independent random variables with common joint density function $f_i(y, \theta)$. Define A_i as the set of all $2^{m_i} - 1$ vectors a of length m_i , consisting solely of zeros and ones, with each vector having at least one non-zero entry. Denote by $Y_{ij}^{(a)}$ the subvector of Y_{ij} corresponding to the non-zero components of a with associated joint density function $f_i^{(a)}(y^{(a)}, \theta)$. The log of the pseudolikelihood is defined as

$$\log \text{PL}_n(\theta) = \sum_{i=1}^p \sum_{a \in A_i} \delta_a \sum_{j=1}^{n_i} \log f_i^{(a)}(y_{ij}^{(a)}, \theta)$$

with $\{\delta_a \mid a \in A_i\}$ a set of $2^{m_i} - 1$ real numbers, not all zero. Classical maximum likelihood corresponds to $\delta_a = 1$ for $a = (1, \dots, 1)$ and zero otherwise. Another typical choice is $\delta_{1_{m_i}} = m_i$ and $\delta_{a_\ell} = -1$ for $\ell = 1, \dots, m_i$ where 1_{m_i} is a vector of ones and a_ℓ consists of ones everywhere, except for the ℓ -th entry. This particular choice is referred to as the “full conditional” log pseudolikelihood function. It has the effect of replacing the joint density function by a product of m_i univariate conditional density functions, thus avoiding the incorporation of a possibly very complicated normalizing constant.

The number p of (possibly) different populations is considered as fixed whereas the numbers n_i of observations from the distinct populations become large as $n = \sum_{i=1}^p n_i$ tends to infinity, according to $n_i/n \rightarrow \lambda_i$ where $\sum_{i=1}^p \lambda_i = 1$ with all $\lambda_i > 0$. Before stating the main asymptotic properties of the maximum pseudolikelihood estimators, the Wald, the score and the pseudolikelihood ratio test, we first list the required regularity conditions on the density functions $f_i^{(a)}(y^{(a)}, \theta)$ for all a in $A_i^\circ = \{a \in A_i \mid \delta_a \neq 0\}$ and each $i = 1, \dots, p$. Let μ denote the Lebesgue measure for continuous $Y_{ij}^{(a)}$ and the counting measure if $Y_{ij}^{(a)}$ is discrete. For θ in an open set Ω containing the true value θ_0 ,

(R1) The densities $f_i^{(a)}(y^{(a)}, \theta)$ are distinct for different values of the parameter θ ; the supports $R_i^{(a)}$ of $f_i^{(a)}(y^{(a)}, \theta)$ do not depend on θ .

(R2) Second order partial derivatives of $f_i^{(a)}(y^{(a)}, \theta)$ w.r.t. θ exist and may be passed under the integral sign in $\int f_i^{(a)}(y^{(a)}, \theta) d\mu(y^{(a)})$.

(R3) For each $k, \ell = 1, \dots, r$, $E_\theta \left| \frac{\partial^2 \log f_i^{(a)}(Y_{i1}^{(a)}, \theta)}{\partial \theta_k \partial \theta_\ell} \right| < \infty$ and the $r \times r$ matrix $J_\delta(\theta_0)$ as defined in Theorem 2.1 is positive definite.

(R4) Third order partial derivatives of $f_i^{(a)}(y^{(a)}, \theta)$ w.r.t. θ exist and there exist functions $K_1(y^{(a)})$ and $K_2(y^{(a)}, y^{(a')})$ such that, for each a, a' in A_i° and each $k, \ell, m = 1, \dots, r$,

$$\left| \frac{\partial^3 \log f_i^{(a)}(y^{(a)}, \theta)}{\partial \theta_k \partial \theta_\ell \partial \theta_m} \right| \leq K_1(y^{(a)}),$$

$$\left| \frac{\partial^2 \log f_i^{(a)}(y^{(a)}, \theta)}{\partial \theta_k \partial \theta_\ell} \frac{\partial \log f_i^{(a')}(y^{(a')}, \theta)}{\partial \theta_m} \right| \leq K_2(y^{(a)}, y^{(a')})$$

and $E_\theta[K_1(Y_{i1}^{(a)})]$ and $E_\theta[K_2(Y_{i1}^{(a)}, Y_{i1}^{(a')})]$ are, as functions of θ , uniformly bounded on Ω .

Theorem 2.1 guarantees the existence of at least one solution to the pseudolikelihood equations

$$(2.1) \quad \frac{\partial}{\partial \theta_\ell} \log \text{PL}_n(\theta) = 0, \quad \ell = 1, \dots, r$$

which is strongly consistent and asymptotically normal.

THEOREM 2.1. *Assume conditions (R1)–(R4). Then there exist solutions $\hat{\theta}_n = (\hat{\theta}_{n1}, \dots, \hat{\theta}_{nr})^T$ to the pseudolikelihood equations (2.1) such that, as $n \rightarrow \infty$ and $n_i/n \rightarrow \lambda_i$ for all $i = 1, \dots, p$,*

- i) $\hat{\theta}_n$ is strongly consistent for θ_0 .
- ii) $n^{1/2}(\hat{\theta}_n - \theta_0)$ is asymptotically normal with mean vector 0 and covariance matrix

$$\Sigma_\delta(\theta_0) = (J_\delta(\theta_0))^{-1} K_\delta(\theta_0) (J_\delta(\theta_0))^{-1}$$

with

$$(J_\delta(\theta))_{k\ell} = \sum_{i=1}^p \lambda_i \sum_{a \in A_i^\circ} \delta_a E_\theta \left[- \frac{\partial^2 \log f_i^{(a)}(Y_{i1}^{(a)}, \theta)}{\partial \theta_k \partial \theta_\ell} \right]$$

and

$$(K_\delta(\theta))_{k\ell} = \sum_{i=1}^p \lambda_i \sum_{a \in A_i^\circ} \sum_{a' \in A_i^\circ} \delta_a \delta_{a'} E_\theta \left[\frac{\partial \log f_i^{(a)}(Y_{i1}^{(a)}, \theta)}{\partial \theta_k} \frac{\partial \log f_i^{(a')}(Y_{i1}^{(a')}, \theta)}{\partial \theta_\ell} \right].$$

Next, consider the following hypothesis

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta \setminus \Theta_0$$

where Θ_0 is a $(r - t)$ dimensional subspace of the parameter space $\Theta \subset \mathbb{R}^r$ such that $\theta = (\theta_1, \dots, \theta_r)^T$ belongs to Θ_0 if and only if $\theta_1 = \dots = \theta_t = 0$, $1 \leq t \leq r$. More general situations, in which H_0 is of the form $H_0 : h_1(\theta) = \dots = h_t(\theta) = 0$ for some smooth real-valued functions h_1, \dots, h_t , can be put into this form by a reparametrization.

In maximum likelihood theory, the Wald, likelihood ratio and score tests are commonly used significance tests. Under appropriate regularity conditions, they are asymptotically equivalent in distribution, both under the null hypothesis and under local alternatives converging sufficiently fast (see e.g. Serfling (1980)). Wald tests however are known not to be invariant to equivalent reparameterizations of nonlinear restrictions (see e.g. Phillips and Park (1988)). Several authors studied and compared the three test procedures in more detail (see e.g. Chandra and Joshi (1983), Chandra and Mukerjee (1984, 1985), Mukerjee (1990a, 1990b) and Cordeiro *et al.* (1994)). In the context of clustered binary data, several papers directly or indirectly compared these and other test procedures (see e.g. Chapter 6 in Morgan (1992)). For pseudolikelihood, Geys *et al.* (1999) propose, similar to the likelihood ratio test, a pseudolikelihood ratio (PLR) test.

The next theorem gives the asymptotic null distributions of the PLR, the robust Wald and the robust score test statistics. Let $C^T = [I_t \ 0_{t,r-t}]$ with I_t the $(t \times t)$ identity matrix and $0_{t,r-t}$ the zero matrix of dimension $t \times (r - t)$. For a $r \times r$ matrix A , define the partitioning

$$A = \begin{pmatrix} A_{LL} & A_{LR} \\ A_{RL} & A_{RR} \end{pmatrix}$$

where $A_{LL} = C^T A C$ and for any $r \times 1$ vector v , let v_L (resp. v_R) denote the subvector of the first t (resp. last $r - t$) elements.

THEOREM 2.2. *Assume conditions (R1)–(R4). Under the null hypothesis H_0 , as $n \rightarrow \infty$ and $n_i/n \rightarrow \lambda_i$ for all $i = 1, \dots, p$, we have*

i) $W_n = n(\hat{\theta}_{nL})^T (\hat{\Sigma}_\delta(\hat{\theta}_n)_{LL})^{-1} \hat{\theta}_{nL}$ converges in distribution to a χ^2_t random variable where

$$\hat{\Sigma}_\delta(\theta) = (\hat{J}_\delta(\theta))^{-1} \hat{K}_\delta(\theta) (\hat{J}_\delta(\theta))^{-1}$$

with

$$(\hat{J}_\delta(\theta))_{k\ell} = -\frac{1}{n} \sum_{i=1}^p \sum_{a \in A_i^\circ} \delta_a \sum_{j=1}^{n_i} \frac{\partial^2 \log f_i^{(a)}(Y_{ij}^{(a)}, \theta)}{\partial \theta_k \partial \theta_\ell}$$

and

$$(\hat{K}_\delta(\theta))_{k\ell} = \frac{1}{n} \sum_{i=1}^p \sum_{a \in A_i^\circ} \sum_{a' \in A_i^\circ} \delta_a \delta_{a'} \sum_{j=1}^{n_i} \frac{\partial \log f_i^{(a)}(Y_{ij}^{(a)}, \theta)}{\partial \theta_k} \frac{\partial \log f_i^{(a')}(Y_{ij}^{(a')}, \theta)}{\partial \theta_\ell}.$$

ii) $S_n = n\mathcal{H}_n(\hat{\theta}_n^\circ)^T (\hat{J}_\delta(\hat{\theta}_n^\circ)^{-1})_{LL} (\hat{\Sigma}_\delta(\hat{\theta}_n^\circ)_{LL})^{-1} (\hat{J}_\delta(\hat{\theta}_n^\circ)^{-1})_{LL} \mathcal{H}_n(\hat{\theta}_n^\circ)_L$ converges in distribution to a χ^2_t random variable where $\hat{\theta}_n^\circ$ is the maximum pseudolikelihood estimator over Θ_0 and

$$\mathcal{H}_n(\theta) = \frac{1}{n} \sum_{i=1}^p \sum_{a \in A_i^\circ} \delta_a \sum_{j=1}^{n_i} \nabla_\theta \log f_i^{(a)}(Y_{ij}^{(a)}, \theta)$$

with $\nabla_\theta \log f_i^{(a)}(y^{(a)}, \theta)$ the $r \times 1$ vector of partial derivatives of $\log f_i^{(a)}(y^{(a)}, \theta)$ w.r.t. θ .

iii) $-2 \log \Lambda_n = 2(\log PL_n(\hat{\theta}_n) - \log PL_n(\hat{\theta}_n^\circ))$ converges in distribution to $\sum_{k=1}^t \alpha_k X_k$ where $\hat{\theta}_n$ is the maximum pseudolikelihood estimator over Θ , X_1, \dots, X_t are independent χ_1^2 random variables and $\alpha_1 \geq \dots \geq \alpha_t$ are the eigenvalues of $(\Sigma_\delta(\theta_0)_{LL})((J_\delta(\theta_0)^{-1})_{LL})^{-1}$.

For the special case that $\delta_a = 1$ for $a = (1, \dots, 1)$ and zero otherwise (classical maximum likelihood theory), the results of Theorems 2.1 and 2.2 simplify to the well-known results. Indeed, for this special case, $J_\delta(\theta_0) = K_\delta(\theta_0)$ such that $\Sigma_\delta(\theta_0) = (J_\delta(\theta_0))^{-1}$. Hence all eigenvalues α_k are equal to 1 and the limiting distribution of $-2 \log \Lambda_n$ reduces to a χ_t^2 distribution. Since for this special case the asymptotic null distribution does not depend on unknown parameters, the bootstrap test is expected to have a smaller asymptotic order of error in level. Beran (1988) shows that the bootstrap likelihood ratio test automatically accomplishes the Bartlett adjustment.

For all other choices, the asymptotic null distribution of the pseudolikelihood ratio test is rather complex (as a consequence of the misspecification of the joint distribution $f_i(y, \theta)$). The bootstrap can play an important role here. As shown in Section 4, the bootstrap estimator is a consistent estimator for the unknown distribution of $-2 \log \Lambda_n$. No eigenvalues α_k have to be estimated.

3. A parametric bootstrap procedure

In this section we discuss the generation of the bootstrap samples and show how they can be used to estimate consistently the distribution of the maximum pseudolikelihood estimators $\hat{\theta}_n$. As before, let Y_{i1}, \dots, Y_{in_i} be a sample from the i -th population. Based on these p samples, the maximum pseudolikelihood estimator $\hat{\theta}_n$ is computed. For each $i = 1, \dots, p$, define the bootstrap sample $Y_{i1}^*, \dots, Y_{in_i}^*$ as n_i independent random variables with common density function $f_i(y, \hat{\theta}_n)$. Denote $\hat{\theta}_n^*$ the maximum pseudolikelihood estimator, which maximizes

$$(3.1) \quad \log PL_n^*(\theta) = \sum_{i=1}^p \sum_{a \in A_i^\circ} \delta_a \sum_{j=1}^{n_i} \log f_i^{(a)}(y_{ij}^{*(a)}, \theta)$$

with $y_{ij}^{*(a)}$ defined as before but now based on the bootstrap sample.

We need the following extra regularity conditions on the density functions $f_i^{(a)}(y^{(a)}, \theta)$, for all a in A_i° and each $i = 1, \dots, p$.

(R5) For each $k, \ell = 1, \dots, r$ and each a, a' in A_i° ,

$$E_\theta \left(\frac{\partial^2 \log f_i^{(a)}(Y_{i1}^{(a)}, \theta)}{\partial \theta_k \partial \theta_\ell} \right) \quad \text{and} \quad E_\theta \left(\frac{\partial \log f_i^{(a)}(Y_{i1}^{(a)}, \theta)}{\partial \theta_k} \frac{\partial \log f_i^{(a')} (Y_{i1}^{(a')}, \theta)}{\partial \theta_\ell} \right)$$

are continuous as functions of θ at $\theta = \theta_0$ and $E_\theta \{ (\frac{\partial^2 \log f_i^{(a)}(Y_{i1}^{(a)}, \theta)}{\partial \theta_k \partial \theta_\ell})^2 \}$ is, as function of θ , uniformly bounded on Ω .

(R6) There exist a function $H^{(a)}(y^{(a)})$ and a $\delta > 0$ such that

$$\left| \left(\frac{\partial}{\partial \theta_k} \log f_i^{(a)}(y^{(a)}, \theta) \right)^{2+\delta} \right| \leq H^{(a)}(y^{(a)})$$

and $E_\theta[H^{(a)}(Y_{i1}^{(a)})]$ is, as function of θ , uniformly bounded on Ω .

In what follows, P^* , E^* , Var^* stand for the bootstrap probability, expectation and variance, conditionally on Y_{i1}, \dots, Y_{in_i} , $i = 1, \dots, p$. The statements in the proofs hold, conditionally on Y_{i1}, \dots, Y_{in_i} , $i = 1, \dots, p$, for almost all sample paths (Y_{i1}, Y_{i2}, \dots) , $i = 1, \dots, p$.

THEOREM 3.1. *Assume conditions (R1)–(R6). Then, for almost all sample paths (Y_{i1}, Y_{i2}, \dots) , $i = 1, \dots, p$, there exist solutions $\hat{\theta}_n^*$ of the pseudolikelihood equations (3.1) such that, as $n \rightarrow \infty$ and $n_i/n \rightarrow \lambda_i$ for all $i = 1, \dots, p$,*

i) $\hat{\theta}_n^*$ converges in bootstrap probability to θ_0

ii) $\sup_{t \in \mathbb{R}^{r'}} |P^*\{g(n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n)) \leq t\} - P\{g(n^{1/2}(\hat{\theta}_n - \theta_0)) \leq t\}| = o(1)$

with $g : \mathbb{R}^r \rightarrow \mathbb{R}^{r'}$ a continuous function such that the distribution function of $g(Z)$ with Z a r' -dimensional normal distributed random variable, is continuous.

PROOF. Consider the expansion

$$(3.2) \quad H_k^*(\theta) = H_k^*(\hat{\theta}_n) + \sum_{\ell=1}^r H_{k\ell}^*(\hat{\theta}_n)(\theta_\ell - \hat{\theta}_{n\ell}) + \frac{1}{2} \sum_{\ell, m=1}^r H_{k\ell m}^*(\tilde{\theta}_n)(\theta_\ell - \hat{\theta}_{n\ell})(\theta_m - \hat{\theta}_{nm})$$

where

$$H_k^*(\theta) = \frac{1}{n} \sum_{i=1}^p \sum_{a \in A_i^c} \delta_a \sum_{j=1}^{n_i} \frac{\partial \log f_i^{(a)}(Y_{ij}^{*(a)}, \theta)}{\partial \theta_k},$$

$$H_{k\ell}^*(\theta) = \frac{1}{n} \sum_{i=1}^p \sum_{a \in A_i^c} \delta_a \sum_{j=1}^{n_i} \frac{\partial^2 \log f_i^{(a)}(Y_{ij}^{*(a)}, \theta)}{\partial \theta_k \partial \theta_\ell}$$

and

$$H_{k\ell m}^*(\theta) = \frac{1}{n} \sum_{i=1}^p \sum_{a \in A_i^c} \delta_a \sum_{j=1}^{n_i} \frac{\partial^3 \log f_i^{(a)}(Y_{ij}^{*(a)}, \theta)}{\partial \theta_k \partial \theta_\ell \partial \theta_m}$$

and with $\tilde{\theta}_n$ an interior point of the line segment joining θ and $\hat{\theta}_n$. The strong consistency of $\hat{\theta}_n$ (Theorem 2.1), (R2) and (R5) imply that as $n \rightarrow \infty$, $E^*(H_k^*(\hat{\theta}_n))$, $\text{Var}^*(H_k^*(\hat{\theta}_n))$, $E^*(H_{k\ell}^*(\hat{\theta}_n)) + (J_\delta(\theta_0))_{k\ell}$ and $\text{Var}^*(H_{k\ell}^*(\hat{\theta}_n))$ all converge to zero from which it immediately follows that both $H_k^*(\hat{\theta}_n)$ and $H_{k\ell}^*(\hat{\theta}_n) + (J_\delta(\theta_0))_{k\ell}$ converge in bootstrap probability to zero as $n \rightarrow \infty$. Condition (R4) is sufficient

to show that $H_{k\ell m}^*(\tilde{\theta}_n)$ is bounded in bootstrap probability. To complete the proof of (i), proceed as in Chanda (1954).

For $u \in \mathbb{R}^r$ an arbitrary vector of unit norm, define for $j = 1, \dots, n_i; i = 1, \dots, p$

$$Z_{nij}^* = n_i^{-1/2} u^T \left(\sum_{a \in A_i^{\circ}} \delta_a \nabla_{\theta} \log f_i^{(a)}(Y_{ij}^{*(a)}, \hat{\theta}_n) - E^* \left(\sum_{a \in A_i^{\circ}} \delta_a \nabla_{\theta} \log f_i^{(a)}(Y_{ij}^{*(a)}, \hat{\theta}_n) \right) \right).$$

$Z_{ni1}^*, \dots, Z_{nin_i}^*$ are (conditionally) independent random variables with mean zero. With $\delta > 0$ from condition (R6), we have that, for n large enough

$$(E^*(|Z_{ni1}^*|^{2+\delta}))^{1/(2+\delta)} \leq 2n_i^{-1/2} \sum_{k=1}^r \sum_{a \in A_i^{\circ}} |\delta_a| \{E^*(H^{(a)}(Y_{i1}^{*(a)}))\}^{1/(2+\delta)}$$

which is $O(n_i^{-1/2})$. Furthermore, using (R5), as $n \rightarrow \infty$,

$$n_i E^*(Z_{ni1}^{*2}) \rightarrow \sum_{a \in A_i^{\circ}} \sum_{a' \in A_i^{\circ}} \delta_a \delta_{a'} E_{\theta_0} \left(\frac{\partial \log f_i^{(a)}(Y_{i1}^{(a)}, \theta_0)}{\partial \theta_k} \frac{\partial \log f_i^{(a')}(Y_{i1}^{(a')}, \theta_0)}{\partial \theta_{\ell}} \right).$$

By an application of the Lyapunov form of the Central Limit Theorem for triangular arrays for each $i = 1, \dots, p$, we get that

$$(3.3) \quad n^{1/2} \mathcal{H}_n^*(\hat{\theta}_n) \text{ converges weakly to } \mathcal{N}(0, K_{\delta}(\theta_0))$$

where $\mathcal{H}_n^*(\theta)$ is the $1 \times r$ vector with elements $H_k^*(\theta)$.

Evaluating (3.2) at $\hat{\theta}_n^*$ gives

$$(3.4) \quad \mathcal{H}_n^*(\hat{\theta}_n) = (J_{\delta}^*(\hat{\theta}_n) + \mathcal{G}^*(\hat{\theta}_n^*, \hat{\theta}_n))(\hat{\theta}_n^* - \hat{\theta}_n)$$

where $J_{\delta}^*(\theta)$ is the $r \times r$ matrix with elements $-H_{k\ell}^*(\theta)$ and $\mathcal{G}^*(\theta^*, \theta)$ the $r \times r$ matrix with elements $\sum_{m=1}^r (\theta_m - \theta_m^*) H_{klm}^*(\tilde{\theta})/2$ ($\tilde{\theta}$ is some interior point of the line segment joining θ and θ^*). From the proof of (i) it follows that,

$$(3.5) \quad J_{\delta}^*(\hat{\theta}_n) \rightarrow J_{\delta}(\theta_0) \text{ in bootstrap probability}$$

and

$$(3.6) \quad \mathcal{G}^*(\hat{\theta}_n^*, \hat{\theta}_n) \rightarrow 0 \text{ in bootstrap probability.}$$

$J_{\delta}^{-1}(\theta_0)$ exists by condition (R3) such that $n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n)$ converges weakly to $\mathcal{N}(0, \Sigma_{\delta}(\theta_0))$. An application of well-known properties of transformed sequences and Pólya's theorem (see e.g. Theorems 1.7 and 1.5.3 in Serfling (1980)) completes the proof. \square

Theorem 3.1 can be used to construct approximate confidence regions for θ_0 . Indeed, choosing e.g. $g(u_1, \dots, u_r) = \|u\| = (\sum_{\ell=1}^r u_\ell^2)^{1/2}$, Theorem 3.1 implies that the confidence region

$$\mathcal{R}_n^\alpha = \{\theta \in \mathbb{R}^r : \|\hat{\theta}_n - \theta\| \leq n^{-1/2} c_{n,\alpha}^*\}$$

where $c_{n,\alpha}^*$ denotes the $(1 - \alpha)$ -th quantile of the (simulated) bootstrap approximation, has a coverage probability converging to $1 - \alpha$. Also “studentized” versions of Theorem 3.1 can be formulated and used to construct confidence regions (percentile t -regions). For the coverage probability of such regions, one would expect a convergence rate at least as high as that of the confidence region based on the normal approximation.

4. Bootstrap pseudolikelihood tests

Recall the testing problem $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta \setminus \Theta_0$ where $\theta = (\theta_1, \dots, \theta_r)$ belongs to Θ_0 if and only if $\theta_1 = \dots = \theta_t = 0$. Critical points and p -values are calculated under the null hypothesis H_0 . Hence, the bootstrap samples should reflect the data generating mechanism under H_0 . Therefore, the sampling mechanism used in the previous section is modified as follows: define $Y_{i1}^*, \dots, Y_{in_i}^*$ to be n_i independent random variables according to the density function $f_i(y, \hat{\theta}_n^\circ)$ where, as in Theorem 2.2, $\hat{\theta}_n^\circ$ is the maximum pseudolikelihood estimator over Θ_0 . Based on this bootstrap sample, $\hat{\theta}_n^*$ denotes the maximum pseudolikelihood estimator, $\hat{\theta}_{nL}^*$ denotes the first t components of this maximum pseudolikelihood estimator and $\hat{\theta}_{n^\circ}^*$ represents the maximum pseudolikelihood estimator over Θ_0 . Further, denote W_n^* , S_n^* and $-2 \log \Lambda_n^*$ the robust Wald, robust score and the pseudolikelihood ratio test statistic based on the bootstrap sample: $W_n^* = n(\hat{\theta}_{nL}^*)^T (\Sigma_\delta^*(\hat{\theta}_n^*)_{LL})^{-1} \hat{\theta}_{nL}^*$, $S_n^* = n \mathcal{H}_n^*(\hat{\theta}_{n^\circ}^*)^T (J_\delta^*(\hat{\theta}_{n^\circ}^*)^{-1})_{LL} (\Sigma_\delta^*(\hat{\theta}_{n^\circ}^*)_{LL})^{-1} \cdot (J_\delta^*(\hat{\theta}_{n^\circ}^*)^{-1})_{LL} \mathcal{H}_n^*(\hat{\theta}_{n^\circ}^*)_L$ and $-2 \log \Lambda_n^* = 2(\log \text{PL}_n^*(\hat{\theta}_n^*) - \log \text{PL}_n^*(\hat{\theta}_{n^\circ}^*))$. Here $\Sigma_\delta^*(\theta)$, $J_\delta^*(\theta)$ and $K_\delta^*(\theta)$ are equal to $\hat{\Sigma}_\delta(\theta)$, $\hat{J}_\delta(\theta)$ and $\hat{K}_\delta(\theta)$ respectively, but with the original sample $\{Y_{ij}\}$ replaced by the bootstrap sample $\{Y_{ij}^*\}$.

The next theorem states that the bootstrap procedure is consistent in estimating the null distribution of W_n , S_n and $-2 \log \Lambda_n$. Hereafter P_{H_0} will denote the probability under the null hypothesis H_0 .

THEOREM 4.1. *Assume conditions (R1)–(R6). Then, for almost all sample paths (Y_{i1}, Y_{i2}, \dots) , $i = 1, \dots, p$, we have, as $n \rightarrow \infty$ and $n_i/n \rightarrow \lambda_i$ for all $i = 1, \dots, p$,*

- i) $\sup_{t \in \mathbb{R}} |P^*\{W_n^* \leq t\} - P_{H_0}\{W_n \leq t\}| = o(1)$,
- ii) $\sup_{t \in \mathbb{R}} |P^*\{S_n^* \leq t\} - P_{H_0}\{S_n \leq t\}| = o(1)$,
- iii) $\sup_{t \in \mathbb{R}} |P^*\{-2 \log \Lambda_n^* \leq t\} - P_{H_0}\{-2 \log \Lambda_n \leq t\}| = o(1)$.

PROOF. From the proof of Theorem 3.1 we have that, under the null hypothesis H_0 , $n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n^\circ)$ converges weakly to $\mathcal{N}(0, \Sigma_\delta(\theta_0))$, where now the first t elements of the true parameter θ_0 equal 0. A one-term expansion shows that

$(J_\delta^*(\hat{\theta}_n^*))_{k\ell} = (J_\delta^*(\hat{\theta}_n^\circ))_{k\ell} + R_{n1}^* \|\hat{\theta}_n^* - \hat{\theta}_n^\circ\|$ and $(K_\delta^*(\hat{\theta}_n^*))_{k\ell} = (K_\delta^*(\hat{\theta}_n^\circ))_{k\ell} + R_{n2}^* \|\hat{\theta}_n^* - \hat{\theta}_n^\circ\|$. Using (R4) it is easy to see that

$$E^*(|R_{n1}^*|) \leq r \sum_{i=1}^p \frac{n_i}{n} \sum_{a \in A_i^\circ} |\delta_a| E^*(K_1(Y_{i1}^{*(a)})),$$

$$E^*(|R_{n2}^*|) \leq 2r \sum_{i=1}^p \frac{n_i}{n} \sum_{a \in A_i^\circ} \sum_{a' \in A_i^\circ} |\delta_a| |\delta_{a'}| E^*(K_2(Y_{i1}^{*(a)}, Y_{i1}^{*(a')})).$$

Similar arguments as in the proof of Theorem 3.1 lead to the asymptotic χ_t^2 distribution of W_n^* .

The proof of (ii) and (iii) is based on the same arguments as the proof of Theorem 3 in Rotnitzky and Jewell (1990). Assume H_0 holds. Then, both $\hat{\theta}_n^*$ and $\hat{\theta}_n^{*\circ}$ are consistent estimators which have asymptotically a normal distribution (by Theorem 3.1). Using the notation introduced in the proof of Theorem 3.1, an expansion of $\log \text{PL}_n^*(\cdot)$ leads to,

$$(4.1) \quad \log \text{PL}_n^*(\hat{\theta}_n^*) - \log \text{PL}_n^*(\hat{\theta}_n^\circ) = n(\hat{\theta}_n^* - \hat{\theta}_n^\circ)^T \mathcal{H}_n^*(\hat{\theta}_n^\circ) - \frac{n}{2}(\hat{\theta}_n^* - \hat{\theta}_n^\circ)^T J_\delta^*(\hat{\theta}_n^\circ)(\hat{\theta}_n^* - \hat{\theta}_n^\circ) + o^*(1).$$

Here and in the remainder of the proof, $o^*(a_n)$ denotes a sequence of random variables of appropriate dimension tending to zero faster than a_n in bootstrap probability.

From (3.4) and (3.6) it follows that

$$(4.2) \quad \mathcal{H}_n^*(\hat{\theta}_n^\circ) = J_\delta^*(\hat{\theta}_n^\circ)(\hat{\theta}_n^* - \hat{\theta}_n^\circ) + o^*(n^{-1/2}).$$

Inserting (4.2) in (4.1) and using (3.5), we get

$$(4.3) \quad 2(\log \text{PL}_n^*(\hat{\theta}_n^*) - \log \text{PL}_n^*(\hat{\theta}_n^\circ)) = n(\hat{\theta}_n^* - \hat{\theta}_n^\circ)^T J_\delta(\theta_0)(\hat{\theta}_n^* - \hat{\theta}_n^\circ) + o^*(1).$$

Identity (3.4) and results (3.5) and (3.6), applied on $\hat{\theta}_n^{*\circ}$ and $\hat{\theta}_n^\circ$, lead to

$$(4.4) \quad \mathcal{H}_{nR}^*(\hat{\theta}_n^\circ) = J_{RR}(\theta_0)(\hat{\theta}_{nR}^{*\circ} - \hat{\theta}_{nR}^\circ) + o^*(n^{-1/2})$$

and (4.2) reformulated for the last $r - t$ elements becomes

$$(4.5) \quad \mathcal{H}_{nR}^*(\hat{\theta}_n^\circ) = J_{RL}(\theta_0)(\hat{\theta}_{nL}^* - \hat{\theta}_{nL}^\circ) + J_{RR}(\theta_0)(\hat{\theta}_{nR}^* - \hat{\theta}_{nR}^\circ) + o^*(n^{-1/2}).$$

The analogue of (4.3) for the estimator $\hat{\theta}_n^{*\circ}$ is

$$(4.6) \quad 2(\log \text{PL}_n^*(\hat{\theta}_n^{*\circ}) - \log \text{PL}_n^*(\hat{\theta}_n^\circ)) = n(\hat{\theta}_{nR}^{*\circ} - \hat{\theta}_{nR}^\circ)^T J_{RR}(\theta_0)(\hat{\theta}_{nR}^{*\circ} - \hat{\theta}_{nR}^\circ) + o^*(1).$$

Subtracting (4.6) from (4.3), isolating $(\hat{\theta}_{nR}^{*\circ} - \hat{\theta}_{nR}^\circ)$ from expressions (4.4) and (4.5) and inserting this, leads to

$$2(\log \text{PL}_n^*(\hat{\theta}_n^*) - \log \text{PL}_n^*(\hat{\theta}_n^{*\circ})) = n(\hat{\theta}_{nL}^* - \hat{\theta}_{nL}^\circ)^T J_{LL}^{-1}(\theta_0)(\hat{\theta}_{nL}^* - \hat{\theta}_{nL}^\circ) + o^*(1).$$

Theorem 3.1 and classical properties of quadratic forms yield the result (iii).

From (4.2) we obtain that

$$(4.7) \quad \hat{\theta}_{nL}^* - \hat{\theta}_{nL}^\circ = (J_\delta^*(\hat{\theta}_n^\circ)^{-1})_{LL} \mathcal{H}_{nL}^*(\hat{\theta}_n^\circ) + (J_\delta^*(\hat{\theta}_n^\circ)^{-1})_{LR} \mathcal{H}_{nR}^*(\hat{\theta}_n^\circ) + o^*(n^{-1/2}).$$

Theorem 3.1 implies that $\sqrt{n}(\Sigma_\delta^*(\hat{\theta}_n^\circ))_{LL}^{-1/2}(\hat{\theta}_{nL}^* - \hat{\theta}_{nL}^\circ)$ converges in distribution to a t -variate standard normal random variable. The result for the robust score statistic now follows by replacing $\hat{\theta}_n^\circ$ in $\mathcal{H}_n^*(\hat{\theta}_n^\circ)$, $J_\delta^*(\hat{\theta}_n^\circ)$ and $\Sigma_\delta^*(\hat{\theta}_n^\circ)$ by its \sqrt{n} consistent bootstrap estimator $\hat{\theta}_n^{*\circ}$. \square

In Section 5 a limited simulation study illustrates the finite sample behaviour of the bootstrap tests for clustered binary data.

5. Simulation study and final remarks

Developmental toxicity experiments are designed to assess the potential adverse effects of drugs or other exposures on developing fetuses of pregnant rodents (dams). A typical study includes a control group and some dosed groups. Exposure usually occurs early in gestation, the dams are sacrificed prior to term and the uterine contents examined for malformations. Denote m_α , $\alpha = 1, \dots, a$ the possible litter sizes and d_β , $\beta = 1, \dots, b$ the possible dose levels. This leads to $p = a \times b$ different ‘‘associated’’ populations. For simplicity, we use a single index $i = 1, \dots, p$ to enumerate these different populations. A litter (cluster) j from population i has a specific size m_i and was given a certain dose d_i . The number of population i litters is n_i and the total number of litters is n . Next to the malformation probability, clustered binary data models all include one or more parameters to describe the association between the outcomes $Y_{ij} = (Y_{ij1}, \dots, Y_{ijm_i})$ of litter j in population i , where Y_{ijk} indicates whether the k -th fetus in litter j is abnormal.

The conditional model of Molenberghs and Ryan (1999) (abbreviated as MR) assumes $Y_{ijk} = 1$ when the k -th fetus in litter j is abnormal and -1 otherwise. This coding provides a parametrisation which more naturally leads to desirable properties when the role of success and failure is reversed and when cluster size is variable (Cox and Wermuth (1994)). Let $Y_{ij}^{(1)}$ denote the total number of malformed fetuses in litter j . The MR probability function of Y_{ij} is

$$f(y_{ij}; \xi_i, \psi_i) = \exp\{\xi_i y_{ij}^{(1)} - \psi_i y_{ij}^{(1)}(m_i - y_{ij}^{(1)}) - A\},$$

with A a normalizing constant and $(\xi_i, \psi_i)^T = X_i \theta$ where X_i is a design matrix based on dose d_i associated with cluster j and θ the coefficient vector. The parameter ξ_i can be interpreted as a main effect and ψ_i as a parameter measuring the intra-litter association.

Instead of the joint probability $f(y_{ij}; \xi_i, \psi_i)$, Geys *et al.* (1997, 1999) (abbreviated as GMR) consider the following product of the m_i conditional probabilities:

$$\prod_{k=1}^{m_i} f(y_{ijk} | y_{ij\ell}, \ell \neq k; \xi_i, \psi_i) = \prod_{k=1}^{m_i} p_{ijs}^{y_{ijk}} p_{ijf}^{1-y_{ijk}}$$

where p_{ijs} is the conditional probability of an additional success, i.e.

$$\begin{aligned} P(y_{ijk} = 1 | y_{ij}^{(1)} - 1 \text{ successes \& } m_i - y_{ij}^{(1)} \text{ failures}) \\ = \frac{\exp[\xi_i - \psi_i(m_i - 2y_{ij}^{(1)} + 1)]}{1 + \exp[\xi_i - \psi_i(m_i - 2y_{ij}^{(1)} + 1)]} \end{aligned}$$

and p_{ijf} the conditional probability of an additional failure, i.e.

$$\begin{aligned} P(y_{ijk} = -1 | y_{ij}^{(1)} \text{ successes \& } m_i - y_{ij}^{(1)} - 1 \text{ failures}) \\ = \frac{\exp[-\xi_i + \psi_i(m_i - 2y_{ij}^{(1)} - 1)]}{1 + \exp[-\xi_i + \psi_i(m_i - 2y_{ij}^{(1)} - 1)]}. \end{aligned}$$

The contribution of the j -th cluster ($j = 1, \dots, n_i$) to the log pseudolikelihood is then given by $y_{ij}^{(1)} \log(p_{ijs}) + (m_i - y_{ij}^{(1)}) \log(p_{ijf})$.

We performed a limited simulation study in order to illustrate the finite sample behaviour of the parametric bootstrap procedure. We used one control group (dose 0) and three active groups (doses 0.25, 0.5 and 1). We experimented with an equal number of $NC = 5, 15$ or 30 clusters assigned to each dose group. The number m_i of fetuses per litter is assumed to follow a local linear smoothed version of the relative frequency distribution given in Table 1 of Kupper *et al.* (1986), which is considered representative of that encountered in actual experimental situations. Realistic parameter values θ were used. We consider two cases. In case 1 data are generated and fitted with

$$X_i = \begin{pmatrix} 1 & d_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \theta^T = (\theta_{10} \quad \theta_{11} \quad \theta_{20})$$

and the hypothesis of interest is $H_0 : \theta_{11} = 0$ (no dose effect on malformation probability). In case 2 we consider

$$X_i = \begin{pmatrix} 1 & d_i & 0 & 0 \\ 0 & 0 & 1 & d_i \end{pmatrix} \quad \text{and} \quad \theta^T = (\theta_{10} \quad \theta_{11} \quad \theta_{20} \quad \theta_{21}),$$

and $H_0 : \theta_{11} = \theta_{21} = 0$ (constant malformation probability and constant intra-litter association).

The pseudolikelihood ratio test can be modified such that it has an approximate χ_t^2 distribution. Similarly to Geys *et al.* (1999), we used the modified test $-2 \log \Lambda_n / \bar{\alpha}$ with $\bar{\alpha}$ the mean of the eigenvalues α_k (see also Rotnitzky

Table 1. GMR model. Case 1 with $\theta_{10} = -2.5$. Simulated type I errors (as %), significance level 0.05.

θ_{20}	Bootstrap tests					χ^2 tests				
	NC	PLR	W(H_1)	W(H_0)	S	PLR(H_1)	W(H_1)	PLR(H_0)	W(H_0)	S
0.1	5	6.0	4.2	5.1	4.8	13.6*	9.6*	5.4	11.1*	3.5*
	30	4.9	5.1	4.7	4.9	6.6*	5.9	5.2	5.5	4.9
0.25	15	5.6	4.0	5.9	4.7	18.5*	14.0*	5.8	12.7*	3.0*
	30	6.5*	4.9	6.3	6.5*	11.5*	9.3*	7.5*	10.1*	6.6*

*denotes the proportion of significant tests (at 5%) which differs significantly from 5%.

Table 2. GMR model. Case 1 with $\theta_{10} = -2.5$, $\theta_{20} = 0.1$. Simulated rejection probabilities (as %), significance level 0.05. Size adjusted values between brackets.

θ_{11}	Bootstrap tests					χ^2 tests				
	NC	PLR	W(H_1)	W(H_0)	S	PLR(H_1)	W(H_1)	PLR(H_0)	W(H_0)	S
0.5	5	8.4	7.8	5.3	7.5	15.6	13.1	7.0	10.2	7.2
		(7.9)	(8.5)	(5.2)	(4.9)	(5.8)	(7.8)	(6.8)	(4.0)	(8.2)
1.0		22.9	19.8	11.1	18.5	16.2	28.7	19.7	20.9	19.6
		(21.1)	(21.9)	(11.1)	(19.0)	(12.5)	(18.9)	(18.9)	(8.8)	(22.4)
0.5	30	26.5	27.6	20.1	25.2	29.8	30.0	25.2	23.3	26.1
		(26.4)	(27.1)	(21.5)	(25.2)	(27.4)	(28.6)	(24.5)	(21.2)	(26.3)
1.0		88.5	88.3	82.8	85.8	89.4	89.3	86.0	85.0	86.5
		(88.6)	(87.7)	(84.4)	(85.9)	(87.6)	(88.5)	(85.6)	(83.6)	(86.7)

and Jewell (1990)). In analogy with the Wald test where the covariance matrix $\Sigma_\delta(\theta_0) = (J_\delta(\theta_0))^{-1}K_\delta(\theta_0)(J_\delta(\theta_0))^{-1}$ can be estimated under H_0 (using $\hat{\Sigma}_\delta(\hat{\theta}_n^o)$) or under H_1 (using $\hat{\Sigma}_\delta(\hat{\theta}_n)$), the eigenvalues α_k can be estimated under H_0 or under H_1 . This leads to five classical χ^2 tests: W(H_0), W(H_1), S, PLR(H_0) and PLR(H_1). There are four bootstrap analogues. All tests have similar first order asymptotic behaviour under H_0 .

For each situation, 1000 datasets were generated and on each dataset the p -values of the different test statistics were computed based on their limiting χ^2 distribution and on the simulated bootstrap distribution, using 1000 bootstrap samples.

Tables 1 and 2 show results for case 1 using the GMR approach. A global inspection of Table 1 clearly shows the superiority of the bootstrap tests in attaining the 5% level. Many χ^2 tests show inflated type I errors which are nicely corrected by their corresponding bootstrap alternatives. Table 2 shows observed and size adjusted power estimates for two alternative values of θ_{11} . Apparently, the boot-

Table 3. GMR model. Case 2 with $\theta_{10} = -2.5$. Simulated type I errors (as %), significance level 0.05.

θ_{20}	Bootstrap tests					χ^2 tests				
	NC	PLR	$W(H_1)$	$W(H_0)$	S	$PLR(H_1)$	$W(H_1)$	$PLR(H_0)$	$W(H_0)$	S
0.1	5	5.3	4.8	5.2	4.9	8.1*	12.7*	2.3*	15.9*	1.3*
	15	3.1*	5.3	5.1	5.9	2.8*	11.7*	0.7*	10.1*	4.9
	30	5.2	3.5*	4.7	5.8	1.3*	9.6*	0.4*	7.5*	5.4
0.25	15	6.1	5.3	6.2	3.7	24.4*	28.6*	8.2*	17.5*	0.5*
	30	5.9	4.9	4.2	6.2	13.1*	19.5*	3.1*	11.5*	2.1*

*denotes the proportion of significant tests (at 5%) which differs significantly from 5%.

Table 4. MR model. Case 1 with $\theta_{10} = -2.5$. Simulated type I errors (as %), significance level 0.05.

θ_{20}	Bootstrap tests					χ^2 tests				
	NC	LR	$W(H_1)$	$W(H_0)$	S	LR	$W(H_1)$	$W(H_0)$	S	
0.1	5	5.5	5.1	4.0	4.5	5.2	10.1*	10.2*	3.9	
	15	5.6	5.8	5.4	5.4	6.1	6.9*	7.3*	5.3	
	30	3.8	4.3	4.2	4.2	3.9	4.6	4.8	4.5	
0.25	15	6.1	6.6*	7.8*	3.3*	4.2	11.0*	11.4*	4.3	
	30	5.6	5.1	4.5	4.8	5.8	9.2*	8.3*	4.6	

*denotes the proportion of significant tests which differs significantly from 5%.

strap tests have, compared with their classical counterparts, comparable power characteristics. As a global conclusion, the bootstrap PLR and $W(H_1)$ and the χ^2 -score tests seem to be the best choices, whereas the χ^2 - $W(H_0)$ and $PLR(H_1)$ tests are the least favourable.

For case 2, Table 3 exhibits the same patterns, but even more pronounced. Also the χ^2 - $PLR(H_0)$ and score tests have size problems while the other χ^2 tests take unacceptably high type I errors, especially in the case where sample information is lowest: a high intra-litter association $\theta_{20} = 0.25$ and only 15 clusters per dose level. We also compared the power estimates and found similar conclusions as in case 1.

Finally, for case 1, Table 4 shows some results for the full likelihood MR model. Recall that, since the joint distribution is correctly specified, no eigenvalues have to be estimated and hence there is only one likelihood ratio test. All test statistics except the robust χ^2 Wald tests, have very comparable simulated type I errors. By using the bootstrap, the poor behaviour of the Wald tests is nicely corrected. We also noticed no substantial differences in power for the corresponding bootstrap

and χ^2 tests.

The main conclusion of this limited simulation study is that the bootstrap tests automatically correct for the likelihood misspecification and they seem to be superior to their classical counterparts. In fact, some improvements in size are quite spectacular. Comparing the χ^2 tests, the robust score test is clearly the preferable one.

Simulations were restricted here to the case of clustered univariate binary data. It would be interesting to see to which extent the χ^2 type PLR test fails in reaching the prescribed significance level and how the parametric bootstrap succeeds in correcting this for the multivariate case of several malformation indicators. Another interesting problem in this general setting is how to define a nonparametric bootstrap method reflecting a specific null hypothesis. One might expect such a method to be more robust to distributional assumptions.

Acknowledgements

The authors would like to thank G. Molenberghs and L. Ryan for interesting discussions and helpful comments and a referee for some valuable suggestions.

REFERENCES

- Arnold, B. C. and Strauss, D. (1991). Pseudolikelihood estimation: Some examples, *Sankhyā Ser. B*, **53**, 233–243.
- Beran, R. (1988). Pivoting test statistics: A bootstrap view of asymptotic refinements, *J. Amer. Statist. Assoc.*, **83**, 687–697.
- Besag, J. (1975). Statistical analysis of non-lattice data, *The Statistician*, **24**, 179–195.
- Bradley, R. A. and Gart, J. J. (1962). The asymptotic properties of ML estimators when sampling from associated populations, *Biometrika*, **49**, 205–214.
- Carr, G. J. and Portier, C. J. (1993). An evaluation of some methods for fitting dose-response models to quantal-response developmental toxicology data, *Biometrics*, **49**, 779–791.
- Chanda, K. C. (1954). A note on the consistency and maxima of the roots of likelihood equations, *Biometrika*, **41**, 56–61.
- Chandra, T. K. and Joshi, S. N. (1983). Comparison of the likelihood ratio, Rao's and Wald's tests and a conjecture of C. R. Rao, *Sankhyā Ser. A*, **45**, 226–246.
- Chandra, T. K. and Mukerjee, R. (1984). On the optimality of Rao's statistic, *Comm. Statist. Theory Methods*, **13**, 1507–1515.
- Chandra, T. K. and Mukerjee, R. (1985). Comparison of the likelihood ratio, Wald's and Rao's tests, *Sankhyā Ser. A*, **47**, 271–284.
- Cordeiro, G. M., Botter, D. A. and Ferrari, S. L. P. (1994). Nonnull asymptotic distributions of three classic criteria in generalized linear models, *Biometrika*, **81**, 709–720.
- Cox, D. R. (1972). The analysis of multivariate binary data, *Applied Statistics*, **21**, 113–120.
- Cox, D. R. and Wermuth, N. (1994). A note on the quadratic exponential binary distribution, *Biometrika*, **81**, 403–408.
- Cressie, N. (1991). *Statistics for Spatial Data*, Wiley, New York.
- Frangos, C. C. and Schucany, W. R. (1995). Improved bootstrap confidence intervals in certain toxicological experiments, *Comm. Statist. Theory Methods*, **24**, 829–844.
- Geys, H., Molenberghs, G. and Ryan, L. (1997). Pseudo-likelihood inference for clustered binary data, *Comm. Statist. Theory Methods*, **26**, 2743–2768.
- Geys, H., Molenberghs, G. and Ryan, L. (1999). Pseudo-likelihood modelling of multivariate outcomes in developmental toxicity, *J. Amer. Statist. Assoc.* (to appear).
- Kupper, L. L., Portier, C., Hogan, M. D. and Yamamoto, E. (1986). The impact of litter effects on dose-response modeling in teratology, *Biometrics*, **42**, 85–98.

- Lehmann, E. L. (1983). *Theory of Point Estimation*, Wiley, New York.
- Lockhart, A.-M. C., Piegorsch, W. W. and Bishop, J. B. (1992). Assessing overdispersion and dose-response in the male dominant lethal assay, *Mutation Research*, **272**, 35-58.
- Molenberghs, G. and Ryan, L. (1999). Likelihood inference for clustered multivariate binary data, *Environmetrics* (to appear).
- Morgan, B. J. T. (1992). *Analysis of Quantal Response Data*, Chapman and Hall, London.
- Mukerjee, R. (1990a). Comparison of tests in the multiparameter case I: Second-order power, *J. Multivariate Anal.*, **33**, 17-30.
- Mukerjee, R. (1990b). Comparison of tests in the multiparameter case II: A third-order optimality property of Rao's test, *J. Multivariate Anal.*, **33**, 31-48.
- Pendergast, J. F., Gange, S. J., Newton, M. A., Linstrom, M. J., Palta, M. and Fisher, M. R. (1996). A survey of methods for analyzing clustered binary response data, *International Statistical Review*, **64**, 89-118.
- Phillips, P. C. B. and Park, J. Y. (1988). On the formulation of Wald tests of nonlinear restrictions, *Econometrica*, **56**, 1065-1083.
- Rotnitzky, A. and Jewell, N. P. (1990). Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data, *Biometrika*, **77**, 485-497.
- Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, Wiley, New York.