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. Pendergast 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 $\chi^2_1$ 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 $\chi^2$ 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 $y = (y_1, \ldots, y_{m_i}) \in R_i$ $(i = 1, \ldots, p)$ denote $p$ joint density or discrete probability functions. The support $R_i$ does not depend on the unknown parameter vector $\theta = (\theta_1, \ldots, \theta_r)^T$. Let $Y_{i1}, \ldots, Y_{in_i}$ 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

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

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 $a = (1, \ldots, 1)$ and zero otherwise. Another typical choice is $\delta_{m_i} = m_i$ and $\delta_{a_\ell} = -1$ for $\ell = 1, \ldots, m_i$ where $1_{m_i}$ is a vector of ones and $a_\ell$ consists of ones everywhere, except for the $\ell$-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 $n_i/n \to \lambda_i$ 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 $A_i = \{a \in A_i \mid \delta_a \neq 0\}$ and each $i = 1, \ldots, p$. Let $\mu$ denote the Lebesgue measure for continuous $Y_{ij}^{(a)}$ and the counting measure if $Y_{ij}^{(a)}$ is discrete. For $\theta$ in an open set $\Omega$ containing the true value $\theta_0$,

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

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

(R3) For each $k, \ell = 1, \ldots, r$, $E_\theta |\partial^2 \log f_i^{(a)}(Y_{ij}^{(a)}, \theta)| < \infty$ and the $r \times r$ matrix $J_\theta(\theta_0)$ as defined in Theorem 2.1 is positive definite.
(R4) Third order partial derivatives of $f_{i}^{(a)}(y^{(a)}, \theta)$ w.r.t. $\theta$ exist and there exist functions $K_{1}(y^{(a)})$ and $K_{2}(y^{(a)}, y^{(a)'})$ such that, for each $a, a'$ in $A_{i}^{\ell}$ and each $k, \ell, m = 1, \ldots, r$,

$$\left| \frac{\partial^{3} \log f_{i}^{(a)}(y^{(a)}, \theta)}{\partial \theta_{k} \partial \theta_{\ell} \partial \theta_{m}} \right| \leq K_{1}(y^{(a)})$$

$$\left| \frac{\partial^{2} \log f_{i}^{(a)}(y^{(a)}, \theta)}{\partial \theta_{k} \partial \theta_{\ell}} \log f_{i}^{(a')}(y^{(a')}, \theta) \right| \leq K_{2}(y^{(a)}, y^{(a')})$$

and $E_{\theta}[K_{1}(Y_{i1}^{(a)})]$ and $E_{\theta}[K_{2}(Y_{i1}^{(a)}, Y_{i1}^{(a')})]$ are, as functions of $\theta$, uniformly bounded on $\Omega$.

Theorem 2.1 guarantees the existence of at least one solution to the pseudo-likelihood equations

$$(2.1) \quad \frac{\partial}{\partial \theta_{\ell}} \log PL_{n}(\theta) = 0, \quad \ell = 1, \ldots, r$$

which is strongly consistent and asymptotically normal.

**Theorem 2.1.** Assume conditions (R1)–(R4). Then there exist solutions $\hat{\theta}_{n} = (\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nr})^{T}$ to the pseudo-likelihood equations (2.1) such that, as $n \to \infty$ and $n_{i}/n \to \lambda_{i}$ for all $i = 1, \ldots, p$,

i) $\hat{\theta}_{n}$ is strongly consistent for $\theta_{0}$.

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

$$(J_{\delta}(\theta))_{kl} = \sum_{i=1}^{p} \lambda_{i} \sum_{a \in A_{i}^{\ell}} \delta_{a} E_{\theta} \left[ -\frac{\partial^{2} \log f_{i}^{(a)}(Y_{i1}^{(a)}, \theta)}{\partial \theta_{k} \partial \theta_{\ell}} \right]$$

and

$$(K_{\delta}(\theta))_{kl} = \sum_{i=1}^{p} \lambda_{i} \sum_{a \in A_{i}^{\ell}} \sum_{a' \in A_{i}^{\ell}} \delta_{a} \delta_{a'} E_{\theta} \left[ \frac{\partial \log f_{i}^{(a)}(Y_{i1}^{(a)}, \theta)}{\partial \theta_{k}} \frac{\partial \log f_{i}^{(a')}(Y_{i1}^{(a')}, \theta)}{\partial \theta_{\ell}} \right].$$

Next, consider the following hypothesis

$$H_{0} : \theta \in \Theta_{0} \quad \text{versus} \quad H_{1} : \theta \in \Theta \setminus \Theta_{0}$$

where $\Theta_{0}$ is a $(r - t)$ dimensional subspace of the parameter space $\Theta \subset \mathbb{R}^{r}$ such that $\theta = (\theta_{1}, \ldots, \theta_{r})^{T}$ belongs to $\Theta_{0}$ if and only if $\theta_{1} = \cdots = \theta_{t} = 0$, $1 \leq t \leq r$. More general situations, in which $H_{0}$ is of the form $H_{0} : h_{1}(\theta) = \cdots = h_{t}(\theta) = 0$ for some smooth real-valued functions $h_{1}, \ldots, h_{t}$, 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 $n \to \infty$ and $n_i/n \to \lambda_i$ for all $i = 1, \ldots, p$, 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 $\chi^2_t$ random variable where

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

with

$$(\hat{J}_\delta(\theta))_{kt} = -\frac{1}{n} \sum_{i=1}^{p} \sum_{a \in A_\theta^i} \delta_k \sum_{a' \in A_\theta^i} \frac{\partial^2 \log f_i^{(a)}(y_{ij}^{(a)}, \theta)}{\partial \theta_k \partial \theta_{t'}}$$

and

$$(\hat{K}_\delta(\theta))_{kt} = \frac{1}{n} \sum_{i=1}^{p} \sum_{a \in A_\theta^i} \sum_{a' \in A_\theta^i} \delta_k \delta_{a'} \sum_{a' \in A_\theta^i} \frac{\partial \log f_i^{(a)}(y_{ij}^{(a)}, \theta) \partial \log f_i^{(a')}(y_{ij}^{(a')}, \theta)}{\partial \theta_k \partial \theta_{t'}}.$$ 

ii) $S_n = n \mathcal{H}_n(\hat{\theta}_n^o)^T (\hat{J}_\delta(\hat{\theta}_n^o)^{-1})_{LL} (\hat{\Sigma}_\delta(\hat{\theta}_n^o)_{LL}^{-1} (\hat{J}_\delta(\hat{\theta}_n^o)^{-1})_{LL} \mathcal{H}_n(\hat{\theta}_n^o)_{LL}$ converges in distribution to a $\chi^2_t$ random variable where $\hat{\theta}_n^o$ is the maximum pseudolikelihood estimator over $\Theta_0$ and

$$\mathcal{H}_n(\theta) = \frac{1}{n} \sum_{i=1}^{p} \sum_{a \in A_\theta^i} \delta_k \sum_{a' \in A_\theta^i} \nabla_\theta \log f_i^{(a)}(y_{ij}^{(a)}, \theta)$$

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. $\theta$. 


iii) \(-2 \log \Lambda_n = 2(\log PL_n(\hat{\theta}_n) - \log PL_n(\hat{\theta}_n^0))\) converges in distribution to 
\(\sum_{k=1}^t \alpha_k X_k\) where \(\hat{\theta}_n\) is the maximum pseudolikelihood estimator over \(\Theta\), \(X_1, \ldots, X_t\) are independent \(\chi_k^2\) random variables and \(\alpha_1 \geq \cdots \geq \alpha_t\) 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 \(a = (1, \ldots, 1)\) 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 \(\Sigma_{\delta}(\theta_0) = (J_{\delta}(\theta_0)^{-1})^{-1}\). Hence all eigenvalues \(\alpha_k\) are equal to 1 and the limiting distribution of \(-2 \log \Lambda_n\) reduces to a \(\chi_t^2\) 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 \(\alpha_k\) 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 \(Y_{i_1}, \ldots, Y_{i_n}\) be a sample from the \(i\)-th population. Based on these \(p\) samples, the maximum pseudolikelihood estimator \(\hat{\theta}_n\) is computed. For each \(i = 1, \ldots, p\), define the bootstrap sample \(Y_{i_1}^*, \ldots, Y_{i_n}^*\) 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

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

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

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

(R5) For each \(k, \ell = 1, \ldots, r\) and each \(a, a' \in A_i^*\),

\[
E_\theta \left( \frac{\partial^2 \log f_i^{(a)}(Y_{i_1}^{(a)}, \theta)}{\partial \theta_k \partial \theta_{\ell}} \right) \quad \text{and} \quad E_\theta \left( \frac{\partial \log f_i^{(a)}(Y_{i_1}^{(a)}, \theta)}{\partial \theta_k} \frac{\partial \log f_i^{(a')} (Y_{i_1}^{(a')}, \theta)}{\partial \theta_{\ell}} \right)
\]

are continuous as functions of \(\theta\) at \(\theta = \theta_0\) and \(E_\theta \{ ( \frac{\partial^2 \log f_i^{(a)}(Y_{i_1}^{(a)}, \theta)}{\partial \theta_k \partial \theta_{\ell}} )^2 \} \) is, as function of \(\theta\), uniformly bounded on \(\Omega\).
(R6) There exist a function $H^{(a)}(y^{(a)})$ and a $\delta > 0$ such that

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

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

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

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

i) $\hat{\theta}_n^*$ converges in bootstrap probability to $\theta_0$.

ii) $\sup_{t \in \mathbb{R}^r} |P^*\{g(n_1^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n)) \leq t\} - P\{g(n_1^{1/2}(\hat{\theta}_n - \theta_0)) \leq t\}| = o(1)$

with $g : \mathbb{R}^r \to \mathbb{R}^{r'}$ 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

$$H_k^*(\theta) = H_k^*(\hat{\theta}_n) + \sum_{\ell=1}^r H_k^{*\ell}(\hat{\theta}_n)(\theta_\ell - \hat{\theta}_n)$$

$$+ \frac{1}{2} \sum_{\ell, m=1}^r H_{k\ell m}(\hat{\theta}_n)(\theta_\ell - \hat{\theta}_n)(\theta_m - \hat{\theta}_n)$$

where

$$H_k^*(\theta) = \frac{1}{n} \sum_{i=1}^p \sum_{a \in A_i^0} \delta_a \sum_{j=1}^{n_i} \frac{\partial \log f_i^{(a)}(Y_{ij}^{(a)}, \theta)}{\partial \theta_k},$$

$$H_k^{*\ell}(\theta) = \frac{1}{n} \sum_{i=1}^p \sum_{a \in A_i^0} \delta_a \sum_{j=1}^{n_i} \frac{\partial^2 \log f_i^{(a)}(Y_{ij}^{(a)}, \theta)}{\partial \theta_k \partial \theta_\ell},$$

and

$$H_{k\ell m}(\theta) = \frac{1}{n} \sum_{i=1}^p \sum_{a \in A_i^0} \delta_a \sum_{j=1}^{n_i} \frac{\partial^3 \log f_i^{(a)}(Y_{ij}^{(a)}, \theta)}{\partial \theta_k \partial \theta_\ell \partial \theta_m},$$

and with $\hat{\theta}_n$ an interior point of the line segment joining $\theta$ and $\hat{\theta}_n$. The strong consistency of $\hat{\theta}_n$ (Theorem 2.1), (R2) and (R5) imply that as $n \to \infty$, $E^*(H_k^*(\hat{\theta}_n))$, $\text{Var}^*(H_k^*(\hat{\theta}_n))$, $E^*(H_k^{*\ell}(\hat{\theta}_n)) + (J_\delta(\theta_0))_{k\ell}$ 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 $n \to \infty$. Condition (R4) is sufficient.
to show that $H^*_{k\ell m}(\hat{\theta}_n)$ 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 $j = 1, \ldots, n_i; \ i = 1, \ldots, p$

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

$$- E^* \left( \sum_{a \in A_i^j} \delta_a \nabla_\theta \log f_i^{(a)}(Y_{ij}^{(a)}, \hat{\theta}_n) \right).$$

$Z_{n11}^*, \ldots, Z_{n1n_i}^*$ are (conditionally) independent random variables with mean zero. With $\delta > 0$ from condition (R6), we have that, for $n$ large enough

$$(E^*(|Z_{n11}^*|^{2+\delta}))/2n_i^{-1/2} \leq 2n_i^{-1/2} \sum_{k=1}^r \sum_{a \in A_i^k} |\delta_a| \{E^*(H'(a)(Y_{i1}^{(a)}))\}^{1/(2+\delta)}$$

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

$$n_i E^*(Z_{n11}^*2) \to \sum_{a \in A_i^k} \sum_{a' \in A_i^k} \delta_a \delta_{a'} E_{\theta_0} \left( \frac{\partial \log f_i^{(a)}(Y_{i1}^{(a)}, \theta_0)}{\partial \theta_k} \frac{\partial \log f_i^{(a')}(Y_{i1}^{(a')}, \theta_0)}{\partial \theta_{a'}} \right).$$

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

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

where $H^*_i(\theta)$ is the $1 \times r$ vector with elements $H^*_{ki}(\theta)$.

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

$$(3.4) \quad H^*_n(\hat{\theta}_n) = \left( J^*_\delta(\hat{\theta}_n) + G^*(\hat{\theta}_n, \hat{\theta}_n) \right)(\hat{\theta}_n - \hat{\theta}_n)$$

where $J^*_\delta(\theta)$ is the $r \times r$ matrix with elements $-H^*_{\delta \ell m}(\theta)$ and $G^*(\theta^*, \theta)$ the $r \times r$ matrix with elements $\sum_{m=1}^r (\theta - \theta^m) H^*_{km}(\hat{\theta})/2$ ($\hat{\theta}$ is some interior point of the line segment joining $\theta$ and $\theta^*$). From the proof of (i) it follows that,

$$(3.5) \quad J^*_\delta(\hat{\theta}_n) \to J^*_\delta(\theta_0) \text{ in bootstrap probability}$$

and

$$(3.6) \quad G^*(\hat{\theta}_n, \hat{\theta}_n) \to 0 \text{ in bootstrap probability}.$$
Theorem 3.1 can be used to construct approximate confidence regions for \( \theta_0 \). Indeed, choosing e.g. \( g(u_1, \ldots, u_r) = \|u\| = (\sum_{t=1}^{r} u_t^2)^{1/2} \), Theorem 3.1 implies that the confidence region

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

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 \( \theta = (\theta_1, \ldots, \theta_r) \) belongs to \( \Theta_0 \) if and only if \( \theta_1 = \cdots = \theta_t = 0 \). 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 \( Y^*_i, \ldots, Y^*_n \) to be \( n_i \) independent random variables according to the density function \( f_i(y, \hat{\theta}_n^0) \) where, as in Theorem 2.2, \( \hat{\theta}_n^0 \) is the maximum pseudolikelihood estimator over \( \Theta_0 \). Based on this bootstrap sample, \( \hat{\theta}_n^* \) denotes the maximum pseudolikelihood estimator, \( \hat{\theta}_n^L \) denotes the first \( t \) components of this maximum pseudolikelihood estimator and \( \hat{\theta}_n^o \) represents the maximum pseudolikelihood estimator over \( \Theta_0 \). 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:

\[
W_n^* = n(\hat{\theta}_n^L)^T (\Sigma_\delta^*(\hat{\theta}_n^L))^{-1} \hat{\theta}_n^L, \quad S_n^* = n \mathcal{H}_n^*(\hat{\theta}_n^o)^T (J_\delta^*(\hat{\theta}_n^o))^{-1} (J_\delta^*(\hat{\theta}_n^o))^{-1/2} \mathcal{H}_n^*(\hat{\theta}_n^o)^{-1/2},
\]

\( (J_\delta^*(\hat{\theta}_n^o))\mathcal{H}_n^*(\hat{\theta}_n^o) \) and \( -2 \log \Lambda_n^* = \log \mathcal{P}_n^*(\hat{\theta}_n^o) - \log \mathcal{P}_n^*(\hat{\theta}_n^0) \).

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 \( (Y_{i1}, Y_{i2}, \ldots) \), \( i = 1, \ldots, p \), we have, as \( n \to \infty \) and \( n_i/n \to \lambda_i \) for all \( i = 1, \ldots, p \),

\[
i) \sup_{t \in \mathbb{R}} |P^*\{W_n^* \leq t\} - P_{H_0}\{W_n \leq t\}| = o(1),
\]

\[
ii) \sup_{t \in \mathbb{R}} |P^*\{S_n^* \leq t\} - P_{H_0}\{S_n \leq t\}| = o(1),
\]

\[
iii) \sup_{t \in \mathbb{R}} |P^*\{-2 \log \Lambda_n^* \leq t\} - P_{H_0}\{-2 \log \Lambda_n \leq t\}| = o(1).
\]

**Proof.** From the proof of Theorem 3.1 we have that, under the null hypothesis \( H_0 \), \( n^{1/2}(\hat{\theta}_n^* - \theta_0) \) converges weakly to \( \mathcal{N}(0, \Sigma_\delta(\theta_0)) \), where now the first \( t \) elements of the true parameter \( \theta_0 \) equal 0. A one-term expansion shows that
(Jₜₙ(δₙ)₀)ₜₙ = (Jₜₙ(δₙ)₀)ₜₙ + Rₙ₁∥δₙₙ₀ - δₙₙ₀∥ and (Kₜₙ(δₙ)₀)ₜₙ = (Kₜₙ(δₙ)₀)ₜₙ + Rₙ₂∥δₙₙ₀ - δₙₙ₀∥. Using (R4) it is easy to see that

\[ E^*(|Rₙ₁|) \leq r \sum_{i=1}^{p} \frac{n_i}{n} \sum_{a \in A^*_i} |\delta_a| E^*(K_1(Y_{i₁}^{*(a)})) , \]
\[ E^*(|Rₙ₂|) \leq 2r \sum_{i=1}^{p} \frac{n_i}{n} \sum_{a \in A^*_i} \sum_{a' \in A^*_i} |\delta_a||\delta_{a'}| E^*(K_2(Y_{i₁}^{*(a)}, Y_{i₁}^{*(a')})) . \]

Similar arguments as in the proof of Theorem 3.1 lead to the asymptotic χ² distribution of Wₙ.

The proof of (ii) and (iii) is based on the same arguments as the proof of Theorem 3 in Rotnitzky and Jewell (1990). Assume H₀ holds. Then, both \( \hat{\theta}_n^* \) and \( \hat{\theta}_n^{0*} \) 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 PLₙ(·) leads to,

\[ \log PLₙ(\hat{\theta}_n^*) - \log PLₙ(\hat{\theta}_n^{0*}) = n(\hat{\theta}_n^* - \hat{\theta}_n^{0*})^T \mathcal{H}_n^*(\hat{\theta}_n^{0*}) - \frac{n}{2}(\hat{\theta}_n^* - \hat{\theta}_n^{0*})^T Jₚₙ(\hat{\theta}_n^{0*})(\hat{\theta}_n^* - \hat{\theta}_n^{0*}) + o^*(1). \]  

(4.1)

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

\[ \mathcal{H}_n^*(\hat{\theta}_n^{0*}) = Jₚₙ(\hat{\theta}_n^{0*})(\hat{\theta}_n^* - \hat{\theta}_n^{0*}) + o^*(n^{-1/2}). \]  

(4.2)

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

\[ 2(\log PLₙ(\hat{\theta}_n^*) - \log PLₙ(\hat{\theta}_n^{0*})) = n(\hat{\theta}_n^* - \hat{\theta}_n^{0*})^T Jₚₙ(\theta_0)(\hat{\theta}_n^* - \hat{\theta}_n^{0*}) + o^*(1). \]  

Identity (3.4) and results (3.5) and (3.6), applied on \( \hat{\theta}_n^{0*} \) and \( \hat{\theta}_n^* \), lead to

\[ \mathcal{H}_{n_{RR}}(\hat{\theta}_n^{0*}) = J_{RR}(\theta_0)(\hat{\theta}_{n_{RR}}^{0*} - \hat{\theta}_{n_{RR}}^{0*}) + o^*(n^{-1/2}) \]  

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

\[ \mathcal{H}_{n_{RR}}(\hat{\theta}_n^{0*}) = J_{RL}(\theta_0)(\hat{\theta}_{n_{RL}}^{0*} - \hat{\theta}_{n_{RL}}^{0*}) + J_{RR}(\theta_0)(\hat{\theta}_{n_{RR}}^{0*} - \hat{\theta}_{n_{RR}}^{0*}) + o^*(n^{-1/2}). \]  

(4.5)

The analogue of (4.3) for the estimator \( \hat{\theta}_n^{0*} \) is

\[ 2(\log PLₙ(\hat{\theta}_n^{0*}) - \log PLₙ(\hat{\theta}_n^{0*})) = n(\hat{\theta}_{n_{RR}}^{0*} - \hat{\theta}_{n_{RR}}^{0*})^T J_{RR}(\theta_0)(\hat{\theta}_{n_{RR}}^{0*} - \hat{\theta}_{n_{RR}}^{0*}) + o^*(1). \]  

(4.6)
Subtracting (4.6) from (4.3), isolating \((\hat{\theta}^{\ast}_n - \hat{\theta}_n)\) from expressions (4.4) and (4.5) and inserting this, leads to

\[
2(\log \mathrm{PL}_n^* (\hat{\theta}_n) - \log \mathrm{PL}_n^* (\hat{\theta}_n^{\ast})) = n(\hat{\theta}_n^{\ast} - \hat{\theta}_n^{\ast})^T J_{\delta}^{-1}(\theta_0)(\hat{\theta}_n^{\ast} - \hat{\theta}_n^{\ast}) + o^*(1).
\]

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

From (4.2) we obtain that

\[
\hat{\theta}_n^{\ast} - \hat{\theta}_n = (J_{\delta}^{\ast}(\hat{\theta}_n^{\ast})^{-1})_{\delta_{\delta}} \mathcal{H}_n^{\ast}(\hat{\theta}_n^{\ast}) + (J_{\delta}^{\ast}(\hat{\theta}_n^{\ast})^{-1})_{LR} \mathcal{H}_n^{\ast}(\hat{\theta}_n^{\ast}) + o^*(n^{-1/2}).
\]

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

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 \(m_\alpha, \alpha = 1, \ldots, a\) the possible litter sizes and \(d_\beta, \beta = 1, \ldots, b\) the possible dose levels. This leads to \(p = a \times b\) different "associated" populations. For simplicity, we use a single index \(i = 1, \ldots, p\) 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 \(Y_{ij} = (Y_{ij1}, \ldots, Y_{ijm_i})\) 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 \(\theta\) the coefficient vector. The parameter \(\xi_i\) can be interpreted as a main effect and \(\psi_i\) as a parameter measuring the intra-litter association.
Instead of the joint probability \( f(y_{ijk} \mid y_{ij\ell}, \ell \neq k; \xi_i, \psi_i) \), Geys et al. (1997, 1999) (abbreviated as GMR) consider the following product of the \( m_i \) conditional probabilities:

\[
\prod_{k=1}^{m_i} f(y_{ijk} \mid y_{ij\ell}, \ell \neq k; \xi_i, \psi_i) = \prod_{k=1}^{m_i} p_{ij\ell}^{y_{ijk}} (1 - p_{ij\ell})^{y_{ijk} - y_{ijk}}
\]

where \( p_{ij\ell} \) is the conditional probability of an additional success, i.e.

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

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

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

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

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 \( \theta \) were used. We consider two cases. In case 1 data are generated and fitted with

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

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

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

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 \( \chi^2 \) 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 \( \alpha_k \) (see also Rotnitzky
Table 1. GMR model. Case 1 with $\theta_{10} = -2.5$. Simulated type I errors (as %), significance level 0.05.

<table>
<thead>
<tr>
<th>$\theta_{20}$</th>
<th>NC</th>
<th>PLR</th>
<th>$W(H_1)$</th>
<th>$W(H_0)$</th>
<th>S</th>
<th>$\chi^2$ tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PLR($H_1$)</td>
</tr>
<tr>
<td>0.1</td>
<td>5</td>
<td>6.0</td>
<td>4.2</td>
<td>5.1</td>
<td>4.8</td>
<td>13.6*</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>4.9</td>
<td>5.1</td>
<td>4.7</td>
<td>4.9</td>
<td>6.6*</td>
</tr>
<tr>
<td>0.25</td>
<td>15</td>
<td>5.6</td>
<td>4.0</td>
<td>5.9</td>
<td>4.7</td>
<td>18.5*</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>6.5*</td>
<td>4.9</td>
<td>6.3</td>
<td>6.5*</td>
<td>11.5*</td>
</tr>
</tbody>
</table>

*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.

<table>
<thead>
<tr>
<th>$\theta_{11}$</th>
<th>NC</th>
<th>PLR</th>
<th>$W(H_1)$</th>
<th>$W(H_0)$</th>
<th>S</th>
<th>$\chi^2$ tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PLR($H_1$)</td>
</tr>
<tr>
<td>0.5</td>
<td>5</td>
<td>8.4</td>
<td>7.8</td>
<td>5.3</td>
<td>7.5</td>
<td>15.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(7.9)</td>
<td>(8.5)</td>
<td>(5.2)</td>
<td>(4.9)</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>22.9</td>
<td>19.8</td>
<td>11.1</td>
<td>18.5</td>
<td></td>
<td>16.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(21.1)</td>
<td>(21.9)</td>
<td>(11.1)</td>
<td>(19.0)</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>30</td>
<td>26.5</td>
<td>27.6</td>
<td>20.1</td>
<td>25.2</td>
<td>29.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(26.4)</td>
<td>(27.1)</td>
<td>(21.5)</td>
<td>(25.2)</td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>88.5</td>
<td>88.3</td>
<td>82.8</td>
<td>85.8</td>
<td></td>
<td>89.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(88.6)</td>
<td>(87.7)</td>
<td>(84.4)</td>
<td>(85.9)</td>
<td></td>
</tr>
</tbody>
</table>

and Jewell (1990)). In analogy with the Wald test where the covariance matrix $\Sigma_\theta = (J_\theta(\theta_0))^{-1}K_\theta(\theta_0)(J_\theta(\theta_0))^{-1}$ can be estimated under $H_0$ (using $\hat{\Sigma}_\theta(\hat{\theta}_n)$) or under $H_1$ (using $\tilde{\Sigma}_\theta(\tilde{\theta}_n)$), the eigenvalues $\alpha_k$ can be estimated under $H_0$ or under $H_1$. This leads to five classical $\chi^2$ 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 $\chi^2$ 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 $\chi^2$ 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 $\theta_{11}$. Apparently, the boot-
Table 3. GMR model. Case 2 with $\theta_{10} = -2.5$. Simulated type I errors (as %), significance level 0.05.

<table>
<thead>
<tr>
<th>$\theta_{20}$</th>
<th>NC</th>
<th>PLR</th>
<th>$W(H_1)$</th>
<th>$W(H_0)$</th>
<th>S</th>
<th>$\chi^2$</th>
<th>$W(H_1)$</th>
<th>PLR($H_0$)</th>
<th>$W(H_0)$</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>5</td>
<td>5.3</td>
<td>4.8</td>
<td>5.2</td>
<td>4.9</td>
<td>8.1*</td>
<td>12.7*</td>
<td>2.3*</td>
<td>15.9*</td>
<td>1.3*</td>
</tr>
<tr>
<td>15</td>
<td>3.1*</td>
<td>5.3</td>
<td>5.1</td>
<td>5.9</td>
<td>2.8*</td>
<td>11.7*</td>
<td>0.7*</td>
<td>10.1*</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>5.2</td>
<td>3.5*</td>
<td>4.7</td>
<td>5.8</td>
<td>1.3*</td>
<td>9.6*</td>
<td>0.4*</td>
<td>7.5*</td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>15</td>
<td>6.1</td>
<td>5.3</td>
<td>6.2</td>
<td>3.7</td>
<td>24.4*</td>
<td>28.6*</td>
<td>8.2*</td>
<td>17.5*</td>
<td>0.5*</td>
</tr>
<tr>
<td>30</td>
<td>5.9</td>
<td>4.9</td>
<td>4.2</td>
<td>6.2</td>
<td>13.1*</td>
<td>19.5*</td>
<td>3.1*</td>
<td>11.5*</td>
<td>2.1*</td>
<td></td>
</tr>
</tbody>
</table>

*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.

<table>
<thead>
<tr>
<th>$\theta_{20}$</th>
<th>NC</th>
<th>LR</th>
<th>$W(H_1)$</th>
<th>$W(H_0)$</th>
<th>S</th>
<th>$\chi^2$</th>
<th>$W(H_1)$</th>
<th>LR</th>
<th>$W(H_0)$</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>5</td>
<td>5.5</td>
<td>5.1</td>
<td>4.0</td>
<td>4.5</td>
<td>5.2</td>
<td>10.1*</td>
<td>10.2*</td>
<td>3.9</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>5.6</td>
<td>5.8</td>
<td>5.4</td>
<td>5.4</td>
<td>6.1</td>
<td>6.9*</td>
<td>7.3*</td>
<td>5.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>3.8</td>
<td>4.3</td>
<td>4.2</td>
<td>4.2</td>
<td>3.9</td>
<td>4.6</td>
<td>4.8</td>
<td>4.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>15</td>
<td>6.1</td>
<td>6.6*</td>
<td>7.8*</td>
<td>3.3*</td>
<td>4.2</td>
<td>11.0*</td>
<td>11.4*</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>5.6</td>
<td>5.1</td>
<td>4.5</td>
<td>4.8</td>
<td>5.8</td>
<td>9.2*</td>
<td>8.3*</td>
<td>4.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

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

For case 2, Table 3 exhibits the same patterns, but even more pronounced. Also the $\chi^2$-PLR($H_0$) and score tests have size problems while the other $\chi^2$ 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 $\chi^2$ 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 $\chi^2$ 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 $\chi^2$ 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 $\chi^2$ 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


Lockhart, A.-M. C., Piegorsch, W. W. and Bishop, J. B. (1992). Assessing overdispersion and
Mukerjee, R. (1990b). Comparison of tests in the multiparameter case II: A third-order optimality
Statistical Review*, **64**, 89–118.
Rotnitzky, A. and Jewell, N. P. (1990). Hypothesis testing of regression parameters in semipara-