VARIANCE ESTIMATES IN LOGISTIC REGRESSION USING THE BOOTSTRAP

Herwig Friedl and Norbert Tilg

Institute of Statistics, Technical University of Graz
A-8010 Graz, Austria

Key Words and Phrases: Binary Regression; Overdispersion; Maximum-Likelihood Estimator; Bootstrap Estimator.

ABSTRACT

Large-sample confidence intervals for the parameter β under the binomial and extra-binomial variance model are presented. Alternative estimates of $\mathrm{Var}(\hat{\beta})$ are discussed, which all have a nice bootstrap interpretation in the context of resampling from residuals or score components. The latter approach yields both a model-based and a robust estimate. Some properties of these estimates and their corresponding confidence intervals are also discussed. In an extensive simulation study we compare the coverage probabilities of the intervals supposing binomial variation as well as overdispersion.

1. INTRODUCTION

In the standard setting of a logistic regression model the data consists of pairs (y_i, x_i), $i = 1, \ldots, n$, where the responses y_i are independent proportions and the p-dimensional regressors $x_i = (x_{i1}, \ldots, x_{ip})'$ are considered to be fixed. If the total numbers m_i of trials are assumed to be known, then it is often appropriate to use binomial variation for the number of events, that means $m_i y_i \overset{i}{\sim} \mathrm{B}(m_i, \mu_i)$, implying $\mathrm{E}(y_i) = \mu_i$ and

$$\mathrm{Var}(y_i) = \tfrac{1}{m_i}\mu_i(1 - \mu_i). \tag{1}$$

The functional relationship between each mean μ_i and its corresponding regressor x_i is defined as the inverse of the logit link $\mu_i(\beta) = \exp(x_i'\beta)/(1+\exp(x_i'\beta))$, where $\beta = (\beta_1, \ldots, \beta_p)'$ denotes the vector of unknown parameters.

Under the binomial assumption the maximum likelihood estimate (MLE) $\hat{\beta}$ is given by the root of the p-dimensional score function

$$s(\beta) = \sum_{i=1}^{n} s_i(\beta) = \sum_{i=1}^{n} x_i m_i (y_i - \mu_i(\beta)).$$

In practical applications of binary regression models the variability in the data often exceeds the binomial variation. Some well known stochastic mechanisms implying extra-binomial variability are discussed in section 2. They are all conceived in a nature such that a global dispersion factor σ^2 is incorporated and the quasi-likelihood approach is appropriate. The asymptotic properties of the estimates are outlined in the third section. With these characteristics the construction of estimates for the variance of $\hat{\beta}$ is straightforward.

If we want to handle overdispersed data without considering an extra parameter σ^2, we can use bootstrap techniques. Moulton and Zeger's (1991) residual resampling approach is motivated in section 4. Further, it is shown that this one-step procedure uses weighted least squares estimation according to a linear model to approximate $\hat{\beta}$. Assuming binomial variability only, their procedure automatically includes a dispersion estimate in the bootstrap variance. In section 5 Lele's (1992) bootstrap technique is applied to logistic regression models. The so called score resampling approach yields variance estimates, which in general do not depend on a specific variance structure. We will show that this parametric bootstrap procedure is a residual resampling technique based on artificially generated errors instead of observed residuals.

The primary purpose of this paper is to analyse the variance estimates in the case of binomial and extra-binomial variation. A Monte Carlo study in section 6 examines the theoretical results.

2. EXTRA-BINOMIAL VARIATION

A simple way to model extra-binomial variation is to introduce an additional parameter σ^2 that reflects common overdispersion. The corresponding variance model can be written as

$$\mathrm{Var}(y_i) = \sigma^2 \tfrac{1}{m_i} \mu_i (1 - \mu_i). \tag{2}$$

Note that the formal restriction $\max_i(0, (1 - m_i\mu_i)/(1 - \mu_i)) \leq \sigma^2 \leq \min_i(m_i)$ must hold. For a further discussion we refer to McCullagh and Nelder (1989).

Williams (1982) gives some interpretations of the term σ^2. He uses unobservable continuous variables p_i independently distributed on $(0, 1)$. Assuming $\mathrm{E}(p_i) = \mu_i$, $\mathrm{Var}(p_i) = \phi_i \mu_i (1 - \mu_i)$ and $m_i y_i | p_i \sim \mathrm{B}(m_i, \mu_i)$, the unconditional moments of the responses are given by $\mathrm{E}(y_i) = \mu_i$ and

$$\mathrm{Var}(y_i) = (1 + (m_i - 1)\phi_i) \tfrac{1}{m_i} \mu_i (1 - \mu_i). \tag{3}$$

The same variance model can be obtained by a sum of m_i correlated homogenious Bernoulli variates, e.g. $y_i = \sum y_{ij}$. In this case the parameters ϕ_i in (3) are determined by the coefficients of correlation $\rho_i = \text{Corr}(y_{ij}, y_{ik})$, for all $j \neq k$. For this reason, some authors call this a 'correlated-binomial model'.

A common choice is to use the beta(a_i, b_i) distribution for the p_i's. The parameterization $\text{E}(p_i) = a_i/(a_i + b_i) = \mu_i$ yields variance (3) with $\phi_i = (1 + a_i + b_i)^{-1}$. Furthermore, the unconditional distribution of $m_i y_i$ is beta-binomial. Recently, Liang and McCullagh (1993) worked out some case studies of such variance models.

The correlated-binomial as well as the beta-binomial model can be described by (2) only in the case of $m_i = m$ and $\rho_i = \rho$ or $\phi_i = \phi$. The latter means that in our model we additionally have to assume equal correlations or constant dispersion, both yielding one global dispersion parameter σ^2. In all these situations the binomial distribution is no longer adequate. Therefore, we replace any exact distributional model by assumptions concerning the first two moments of y_i only, which are $\text{E}(y_i) = \mu_i$ and $\text{Var}(y_i) = \sigma^2 \frac{1}{m_i} \mu_i (1 - \mu_i)$.

The quasi-likelihood technique, introduced by Wedderburn (1974), provides a method to construct a maximum quasi-likelihood estimate (MQE) $\hat{\beta}$ considering only the mean-variance relationship of the responses. Now the MQE $\hat{\beta}$ under model (2) is based on the quasi-score function

$$\sum_{i=1}^{n} \frac{\partial \mu_i(\beta)}{\partial \beta} \frac{y_i - \mu_i(\beta)}{\text{Var}(y_i)} = \frac{1}{\sigma^2} \sum_{i=1}^{n} x_i m_i (y_i - \mu_i(\beta)) = \frac{1}{\sigma^2} s(\beta).$$

It is obvious that MLE and MQE are defined in the same way.

3. ESTIMATING THE PARAMETERS

A Taylor expansion of the score function in the true β yields the linear approximation $s(\hat{\beta}) = 0 \approx s(\beta) + \frac{\partial s(\beta)}{\partial \beta}(\hat{\beta} - \beta)$ or rather

$$\frac{1}{\sqrt{n}} s(\beta) \approx \frac{1}{n} F(\beta) \sqrt{n} (\hat{\beta} - \beta) \tag{4}$$

with

$$F(\beta) = -\frac{\partial s(\beta)}{\partial \beta} = \sum_{i=1}^{n} x_i m_i \mu_i (1 - \mu_i) x_i'.$$

Note that F is not random. Since $\text{E}(s(\beta)) = 0$ and $\text{Var}(s(\beta)) = F(\beta)$ under model (1) and by applying the central limit theorem it follows that

$$\frac{1}{\sqrt{n}} s(\beta) \xrightarrow{d} \text{N}_p \left(0, \frac{1}{n} F(\beta) \right).$$

This asymptotic property of the score together with (4) yields

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} \mathrm{N}_p\left(0, nF^{-1}(\beta)\right).$$

In the case of variance model (2) we get $\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} \mathrm{N}_p\left(0, n\sigma^2 F^{-1}(\beta)\right)$. A rigorous discussion concerning those asymptotics in the context of generalized linear models (GLM's) is given in Fahrmeir and Kaufmann (1985).

Corresponding to the asymptotics above, binomial responses induce 95% confidence intervals for every β_j, $j = 1, \ldots, p$, of the form

$$\hat{\beta}_j \pm 1.96\sqrt{F_{jj}^{-1}(\hat{\beta})}. \tag{5}$$

In the presence of extra-binomial variation, $\mathrm{Var}(\hat{\beta})$ is proportional to σ^2. So all standard errors are multiplied by $\sqrt{\sigma^2}$ or by an estimate thereof. One suggested estimate for σ^2 is based on the mean Pearson statistic

$$\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(y_i-\hat{\mu}_i)^2}{\frac{1}{m_i}\hat{\mu}_i(1-\hat{\mu}_i)}.$$

Like in the binomial case (5), confidence intervals may be defined by

$$\hat{\beta}_j \pm 1.96\sqrt{\hat{\sigma}^2 F_{jj}^{-1}(\hat{\beta})}. \tag{6}$$

The MLE $\hat{\beta}$ can be computed using the approximation $\hat{\beta} = \beta + F^{-1}(\beta)s(\beta)$ which yields $\hat{\beta} = \beta + (G'G)^{-1}G'e$ with weighted design $G = V^{1/2}X$, variance matrix $V = \mathrm{diag}(m_i\mu_i(1-\mu_i))$ and Pearson residuals $e_i = m_i(y_i-\mu_i)/\sqrt{V_i}$. This can be rewritten in weighted least squares (WLS) notation as

$$\hat{\beta} = (G'G)^{-1}G'z \tag{7}$$

with some standardized adjusted dependent variates

$$z = G\beta + e. \tag{8}$$

If $\{\mu_i\}$ are known, then $\mathrm{E}_y(z) = G\beta$ and we have a linear model with known design matrix G. In this case the parameter β can be estimated by WLS. However, in practice the μ_i's are unknown and we use the iterative WLS procedure $\hat{\beta}^{(t+1)} = (\hat{G}'\hat{G})^{-1}\hat{G}'\hat{z}$, where $G = G^{(t)}$ and $z = z^{(t)}$.

4. THE TECHNIQUE OF RESIDUAL RESAMPLING

In ordinary linear models (OLM's), $y = X\beta + \epsilon$, the density of the response can be written as $f_y(y) = f_\epsilon(y - X\beta)$, because the key characteristic is the additive error. Let f_{ϵ_i} correspond to an unknown distribution function

F_{ϵ_i} with zero mean and variance σ_i^2. If we further assume independence and homoscedasticity of the y_i's, that is $\sigma_i^2 = \sigma^2$, which implies exchangeability of the error terms, residual resampling is known to be a well suited bootstrap procedure for estimating the distributional properties of the least squares estimate $\hat{\beta} = (X'X)^{-1}X'y$. This fact is thoroughly documented in Efron (1979), Freedman (1981) or in Wu (1986) among others.

Actually, in the logistic regression case such additivity is not assumed. Hence, the exchangeability of the error terms is typically no longer valid under this model. Nevertheless, Moulton and Zeger (1991) have introduced a one-step bootstrap procedure for the whole class of GLM's based on residual resampling.

The standard approach of residual resampling in homoscedastic OLM's is to approximate the distribution of $\hat{\beta}$ by the distribution of the bootstrap estimate $\hat{\beta}^*$ with respect to the empirical distribution function $\hat{F}_e$, which gives probability mass $1/n$ on every observed error $e_i = y_i - x_i'\hat{\beta}$. Under these assumptions the resampled bootstrap vector $\hat{\beta}^*$ corresponding to $\hat{\beta}$ is given by $\hat{\beta}^* = (X'X)^{-1}X'y^*$ with $y^* = X\hat{\beta} + e^*$. Here and in the following, e_i^* are sampled with replacement from the set $\{e_i\}$. The resulting bootstrap variance $\text{Var}_{\hat{F}_e}(\hat{\beta}^*) = \hat{\sigma}_e^2(X'X)^{-1}$ includes the variance estimate $\hat{\sigma}_e^2 = 1/n\sum e_i^2$, which is biased for σ^2. To get the moment based estimate, we have to resample from the standardized residuals $(y_i - x_i'\hat{\beta})/\sqrt{1-p/n}$.

In heteroscedastic OLM's Wu (1986) used resampled values $y_i^* = x_i'\hat{\beta} + \hat{\sigma}_i t_i^*$, where $t_i = (e_i - \bar{e})/(1/n\sum(e_j - \bar{e})^2)$ with some variance estimates $\hat{\sigma}_i^2$ like $e_i^2/(1-p/n)$ or $e_i^2/(1-h_{ii})$. h_{ii} denotes the i-th diagonal element of the hat matrix $H = X(X'X)^{-1}X'$.

Studying bootstrap techniques in GLM's, Moulton and Zeger (1991) proposed to resample from the observed standardized (leverage adjusted) Pearson residuals, which can be defined in the logistic regression model as

$$e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\frac{1}{m_i}\hat{\mu}_i(1-\hat{\mu}_i)(1-\hat{h}_{ii})}},$$

where $H = G(G'G)^{-1}G'$ is now a generalized version of the hat matrix of (8).

Hinde (1992) adapted Wu's idea to GLM's, which under the binomial assumption results in resampling from

$$y_i^* = x_i'\hat{\beta} + \sqrt{\frac{1}{m_i}\hat{\mu}_i(1-\hat{\mu}_i)(1-\hat{h}_{ii})}\; e_i^*. \tag{9}$$

However, in the case of discrete distributions we run into problems. Regarding binomial or Poisson distributed responses the bootstrap value y_i^* cannot represent an admissible replication under the model (logistic or loglinear), because (9) allows for negative and non-integer replications.

In modelling Poisson and negative binomial variates, Firth et al. (1991) introduced a modification to guarantee positive integer responses. This includes rounding to the nearest integer and replacing negative values by zero. The paper is concerned with goodness of fit statistics in order to check the models. Davison and Snell (1991) used the same technique to generate envelopes for Cook statistics. As we are interested in obtaining estimates for the variance of $\hat{\beta}$ we take a closer look at the so called one-step procedure of Moulton and Zeger (1991), which does not use the replications y_i^* in an explicit way.

If we resample from the observed errors of the estimated model, which is the bootstrap model with $\hat{\beta}$ as true parameter, then the WLS estimate (7) with the OLM (8) is convenient for getting a linear approximation of the bootstrap estimate $\hat{\beta}^*$. Note that this is a 'one-step' procedure and may not be iterated until convergence. Therefore, we should regard the behaviour of

$$\hat{\beta}^* = (\hat{G}'\hat{G})^{-1}\hat{G}'\hat{z}^* = \hat{\beta} + (\hat{G}'\hat{G})^{-1}\hat{G}'e^*, \tag{10}$$

where $\hat{G} = G(\hat{\beta})$ denotes the likelihood based estimated weighted design matrix. The assumed exchangeability of the error terms can be motivated by considering the variability of the OLM (8). Under model (1) the variance of z_i is determined by the variance of the independent Pearson residuals, which is $\mathrm{Var}(y_i)/(\mu_i(1-\mu_i)/m_i) = 1$ for all i. In particular, we have to resample under a homoscedastic OLM. Obvious candidates for a bias-reducing choice of the residuals are the Pearson residuals scaled by $\sqrt{1-h_{ii}}$ or by $\sqrt{1-p/n}$, where p/n represents the mean value of the h_{ii}. Our simulation study below shows that the differences in the behaviour of both proposals are neglectable, but a complete renunciation of the scaling factor yields poorer results.

The variance of (10) with respect to $\hat{F}_e$ is given by

$$\mathrm{Var}_{\hat{F}_e}(\hat{\beta}^*) = \hat{\sigma}_e^2(\hat{G}'\hat{G})^{-1} = \hat{\sigma}_e^2 F^{-1}(\hat{\beta}). \tag{11}$$

The dispersion estimate

$$\hat{\sigma}_e^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} e_i\right)^2$$

is incorporated automatically into the variance term by the bootstrap, although we had assumed model (1) with no overdispersion. 95% confidence intervals based on (11) are given by

$$\hat{\beta}_j \pm 1.96\sqrt{\hat{\sigma}_e^2 F_{jj}^{-1}(\hat{\beta})}, \quad j = 1, \ldots, p.$$

The normal limit can be avoided by generating a large number B of replications $\hat{\beta}^{*b}$, $b = 1, \ldots, B$ and utilizing the quantiles of the empirical distribution functions $\hat{R}_j$ of $\hat{\beta}_j^*$ defined in (10) to approximate the 95% confidence interval by the percentile interval

$$\left(\hat{R}_{j,\alpha/2}^{-1}, \hat{R}_{j,1-\alpha/2}^{-1}\right). \tag{12}$$

5. THE TECHNIQUE OF SCORE RESAMPLING

As in the residual resampling approach we make use of the linear approximation (4). Now the distributional properties of $s(\beta)$ are employed to describe the behaviour of the MLE $\hat{\beta}$. For instance, improving the estimation of $\mathrm{Var}(s(\beta))$ should result in a better estimate of

$$\mathrm{Var}(\hat{\beta}) \approx F^{-1}(\beta)\mathrm{Var}(s(\beta))F^{-1}(\beta). \tag{13}$$

The following bootstrap method is mentioned in Lele (1992) in the context of estimating equations. In the sequel we will call it score resampling technique.

To estimate $\mathrm{Var}(s(\beta))$, we consider the total score function in terms of an arithmetic mean of independent but not identically distributed variates with equal zero means but different variances, that is

$$\bar{s}(\beta) = \frac{1}{n}s(\beta) = \frac{1}{n}\sum_{i=1}^{n} s_i(\beta).$$

Liu (1988) applied Wu's bootstrap procedure in order to estimate a common mean from different distributions. Instead of resampling from observed errors under an OLM, she generated independent pseudo random numbers t_i with zero mean and unit variance, e.g. $t_i \sim F_t(0,1)$. Therefore, this strategy is a parametric bootstrap procedure. Details concerning the generation of the t_i's are discussed in her paper. Strictly speaking, she used bootstrap scores

$$s_i^* = \bar{s}(\hat{\beta}) + (s_i(\hat{\beta}) - \bar{s}(\hat{\beta}))t_i^* = s_i(\hat{\beta})t_i^*,$$

since $\bar{s}(\hat{\beta}) = 0$. This yields $\mathrm{E}_{F_t}(s_i^*) = 0$ but $\mathrm{Var}_{F_t}(s_i^*) = s_i(\hat{\beta})s_i'(\hat{\beta})$, which in it's own right is an estimate of the variance of $s_i(\hat{\beta})$. In our case we use

$$\mathrm{Var}_{F_t}(s^*) = \sum_{i=1}^{n} x_i m_i^2 (y_i - \hat{\mu}_i)^2 x_i' \tag{14}$$

to approximate $\mathrm{Var}(s(\beta))$, instead of the standard estimate

$$F(\hat{\beta}) = \sum_{i=1}^{n} x_i m_i \hat{\mu}_i (1 - \hat{\mu}_i) x_i'. \tag{15}$$

Comparing (14) with (15) we can see that both estimate $\sum \mathrm{Var}(x_i m_i y_i)$, but the bootstrap variance estimator is independent of the assumed variance model and obviously more robust against misspecifications in the variance. The limiting normal together with (13) and (14) results in a 95% confidence interval

$$\hat{\beta}_j \pm 1.96\sqrt{\left(F^{-1}(\hat{\beta})\mathrm{Var}_{F_t}(s^*)F^{-1}(\hat{\beta})\right)_{jj}}. \tag{16}$$

In a much more general context Liang and Zeger (1986) used the same estimate that remains consistent even if the wrong variance structure is chosen.

On the other hand, since $\mathrm{Var}(s(\beta)) = F(\beta)$ under model (1) we get $\mathrm{Var}(\hat{\beta}) \approx F^{-1}(\beta)$. Therefore, bootstrapping $\mathrm{Var}(s(\beta))$ can be interpreted as bootstrapping $F(\beta)$ and the corresponding confidence intervals would be

$$\hat{\beta}_j \pm 1.96\sqrt{\mathrm{Var}^{-1}_{jj,F_t}(s^*)}. \tag{17}$$

This model-based estimation is only valid if the variance holds. In case of overdispersion the supposed model differs from the assumed one by the factor σ^2. Provided that the $\hat{\mu}_i$'s are the true means, the expected value of (14) under (2) is given by $\sigma^2 F(\hat{\beta})$. This implies that the estimate of $\mathrm{Var}(\hat{\beta})$ used in (17) is of the magnitude $1/\sigma^2 F^{-1}(\hat{\beta})$ but not $\sigma^2 F^{-1}(\hat{\beta})$ as stated in (6).

Regarding the arguments at the beginning of this section, the bootstrap vector is given by $\hat{\beta}^* = \hat{\beta} + F^{-1}(\hat{\beta})s^*$. Let $W^{1/2} = \mathrm{diag}(m_i(y_i - \mu_i))$, then $s^* = X'\hat{W}^{1/2}t^*$ with $t^* = (t_1^*, \ldots, t_n^*)$. Hence, $\hat{\beta}^*$ is defined in a similar way as in the one-step residual resampling procedure, namely

$$\hat{\beta}^* = \hat{\beta} + F^{-1}(\hat{\beta})X'\hat{W}^{1/2}t^*. \tag{18}$$

Comparing this with the bootstrap coefficient (10), which was

$$\hat{\beta}^* = \hat{\beta} + F^{-1}(\hat{\beta})X'\hat{V}^{1/2}e^*,$$

it is easy to recognize some slight differences between both approaches. The score resampling technique uses the robust variance estimate $\hat{W}$ instead of the model-based $\hat{V}$. Moreover, we resample from 'pseudo-residuals' t_i instead of observed Pearson residuals e_i. Therefore, the empirical distribution function $\hat{F}_e$ is replaced by the theoretical F_t, both reflecting zero mean and unit variance.

Additionally, Lele (1992) proposed to improve the robust confidence intervals (16) by using the bootstrap to 'estimate' the normal approximation. In a Monte Carlo simulation the sequence s^{*b}, $b = 1, \ldots, B$, is generated, which permits to calculate the empirical distribution function $\hat{S}_j$ of the j-th component of $\hat{\beta}^{*b}$ given in (18). As in (12) we approximate the confidence interval by the corresponding percentile interval

$$\left(\hat{S}^{-1}_{j,\alpha/2}, \hat{S}^{-1}_{j,1-\alpha/2}\right). \tag{19}$$

6. SIMULATION RESULTS

We carried out an extensive simulation study to obtain knowledge on the coverage probabilities of the confidence intervals. Our primary aim was to investigate their behaviour under the assumption of either binomial (1) or extra-binomial variability (EBV) (2). A simulation design close to that of Moulton and Zeger (1991) was used to enable comparisons with their results.

In a further Monte Carlo approach, we calculated the MLE $\hat{\beta}$ and generated bootstrap replications $m_i y_i^*$ from $\mathrm{B}(m_i, \mu_i(\hat{\beta}))$ to get the bootstrap MLE $\hat{\beta}^*$. This binomial resampling procedure is a fully parametric bootstrap technique. Again we used the histogram of the replications $\hat{\beta}^{*b}$, $b = 1, \ldots, B$, to determine percentile intervals denoted by MLE(p).

6.1 DATA GENERATION

Only the two parameter model $\log(\mu_i/(1-\mu_i)) = \beta_0 + \beta_1 x_i = \eta_i$ with $x_i = (2i-1)/20$, $i = 1, \ldots, 20$, is regarded. To compare the behaviour of small and large sample sizes we choose $n = 20$ and copy the set $x_1, \ldots, x_{20}$ three times yielding $n = 80$, respectively. Our attention is confined to the special example of equal binomial denominators $m_i = m = 10$.

Binomial variance is reflected by $m y_i \sim \mathrm{B}(m, \mu_i)$ with probability of success $\mu_i = \exp(\eta_i)/(1 + \exp(\eta_i))$.

Extra-binomial variability in the data is obtained from beta-binomial variates $m y_i \sim \mathrm{betaB}(a_i, b_i)$. If $p_i \sim \mathrm{beta}(a_i, b_i)$ and $m y_i | p_i{=}\mu_i \sim \mathrm{B}(m_i, \mu_i)$, then $\mathrm{E}(p_i) = a_i/(a_i + b_i)$ and assuming the logit link model results in $\mathrm{E}(y_i | p_i{=}\mu_i) = \mu_i = \exp(\eta_i)/(\exp(\eta_i) + 1)$.

Moulton and Zeger (1991) set simply $a_i = \exp(\eta_i) = \mu_i/(1-\mu_i)$ and $b_i = 1$. Because of the special choice of the parameters, the generation of the p_i's can easily be done via the inversion method: If $u_i \sim \mathrm{U}(0,1)$, then $p_i = u_i^{\exp(-\eta_i)} \sim \mathrm{beta}(\exp(\eta_i), 1)$, since p^a represents the $\mathrm{beta}(a, 1)$ distribution function. Then, the beta-binomial variate $m y_i$ is generated as a binomial variate with probability of success p_i.

But this implies $\mathrm{E}(m y_i) = m\mu_i$ and $\mathrm{Var}(m y_i) = \sigma_i^2 m \mu_i (1-\mu_i)$ with different dispersion factors $\sigma_i^2 = 1 + (m-1)\rho_i$, where $\rho_i = (\exp(\eta_i) + 2)^{-1}$ in the context of a correlated-binomial model. Note that the degree of correlation is determined by μ_i. For that reason, varying the parameters (β_0, β_1) or considering different x_i's both yield distinct factors of overdispersion as it is outlined in Table I. Consequently, the quasi-likelihood approach under the variance model (2) is not the appropriate estimation procedure.

TABLE I

Dependency of μ, ρ and σ^2 on $x_i{=}(2i-1)/20$ and (β_0, β_1) under the beta-binomial model $\mathrm{betaB}(\mu_i/(1-\mu_i), 1)$.

(β_0, β_1)	$\mu(1)$	$\mu(x)$ range	$\rho(1)$	$\rho(x)$ range	$\sigma^2(1)$	$\sigma^2(x)$ range
$(-1, -1)$	0.12	$(0.05, 0.26)$	0.47	$(0.43, 0.49)$	5.22	$(4.83, 5.39)$
$(-1, +1)$	0.50	$(0.28, 0.72)$	0.33	$(0.22, 0.42)$	4.00	$(2.96, 4.77)$
$(-1, +3)$	0.88	$(0.30, 0.99)$	0.11	$(0.01, 0.41)$	1.96	$(1.07, 4.71)$

We specify $p_i \sim \mathrm{beta}(\phi\mu_i, \phi(1-\mu_i))$ with $\phi > 0$, yielding again $\mathrm{E}(p_i) = \mu_i$ but the variance structure $\mathrm{Var}(p_i) = (1+\phi)^{-1}\mu_i(1-\mu_i)$. In terms of the correlated-binomial model, this gives a common correlation coefficient $\rho = (1+\phi)^{-1}$. The global dispersion factor is $\sigma^2 = 1 + (m-1)\rho \geq 1$, which only allows the consideration of overdispersion.

6.2 INTERPRETATION

All simulation results are based on 1200 trials and should be regarded conditionally on the specific set $\{x_i\}$. Percentile intervals are obtained by using $B = 200$ replications to estimate the quantiles of the bootstrap distribution.

The following discussion is based on the deviations of the coverage probabilities of the proposed confidence intervals from the nominal level 95%. Moreover, the influence of sample size n, parameters (β_0, β_1) and overdispersion σ^2 will be examined. For the latter one, $\rho = 1/3$ is chosen, which is equivalent to the overdispersion factor $\sigma^2 = 4$, when generating beta-binomial variates.

We use the abbreviations of Table II to refer to the confidence intervals, which are based on normal quantiles and theoretical variances.

TABLE II

Theoretical confidence intervals considered in the simulation study.

Method	95% confidence interval	$\widehat{Var}(\hat{\beta})$
MLE(t)	$\hat{\beta}_j \pm 1.96\sqrt{\hat{F}_{jj}^{-1}}$	$(X'\hat{V}X)^{-1}$
MQE(t)	$\hat{\beta}_j \pm 1.96\sqrt{\hat{\sigma}^2\hat{F}_{jj}^{-1}}$	$\hat{\sigma}^2(X'\hat{V}X)^{-1}$
Res(t)	$\hat{\beta}_j \pm 1.96\sqrt{\hat{\sigma}_e^2\hat{F}_{jj}^{-1}}$	$\hat{\sigma}_e^2(X'\hat{V}X)^{-1}$
Sco_m(t)	$\hat{\beta}_j \pm 1.96\sqrt{\mathrm{Var}_{jj,F_t}^{-1}(s^*)}$	$(X'\hat{W}X)^{-1}$
Sco_r(t)	$\hat{\beta}_j \pm 1.96\sqrt{\left(\hat{F}^{-1}\mathrm{Var}_{F_t}(s^*)\hat{F}^{-1}\right)_{jj}}$	$(X'\hat{V}X)^{-1}(X'\hat{W}X)(X'\hat{V}X)^{-1}$

The matrices $\hat{V} = \mathrm{diag}(m_i\hat{\mu}_i(1-\hat{\mu}_i))$ and $\hat{W} = \mathrm{diag}(m_i^2(y_i - \hat{\mu}_i)^2)$ denote the model-based and robust variance estimates of $m_i y_i$. The bootstrap estimate $\hat{\sigma}_e^2$ is based on three different types of Pearson residuals, unscaled and scaled by $\sqrt{1-h_{ii}}$ or $\sqrt{1-p/n}$. The confidence intervals are Res_u(t), Res_h(t) and Res_p(t), respectively. The corresponding bootstrap percentile intervals are Res_u(p), Res_h(p) and Res_p(p) outlined in (12), MLE(p) described above, and Sco(p) represented in (19).

Theoretically, in the case of binomial variation the MLE(t) intervals are appropriate, whereas MQE(t) intervals are convenient for overdispersed data. Both statements are confirmed by our simulation (see Table III). Concerning small samples, it is interesting to note that the application of MQE(t) intervals in the binomial situation results in a slight change for the worse compared to the coverage of the MLE(t) intervals. This comes from the fact that in the quasi-likelihood approach $\hat{\sigma}^2$ has to estimate the parameter $\sigma^2 = 1$ whereas we are using this true value in the likelihood approach. On the other hand, the MLE(t) intervals are forced to use $\sigma^2 = 1$ even in the case of EBV $\sigma^2 = 4$. Therefore, the lengths of the intervals are just about one half of those described by MQE(t) yielding coverages less than 70%. Obviously the properties of the estimate $\hat{\sigma}^2$ depend on n. In the large sample case that appears to be the cause of hardly any difference between the coverages of the MLE(t) and MQE(t).

The residual resampling techniques differ only in using various estimates for dispersion. Especially in the Res_p(t) interval the estimate $\hat{\sigma}_e^2$ equals $\hat{\sigma}^2$ up to a correction by the squared sum of Pearson residuals divided by n^2, which is zero in OLM's if an intercept is included. In GLM's, however, the correction term does not vanish in general but is relatively small compared to the sum of squared residuals divided by n. Therefore, Res_p(t) and MQE(t) are roughly the same. The Res_u(t) interval includes a dispersion estimate $\hat{\sigma}_e^2$ which is of magnitude $(n-p)/n\hat{\sigma}^2$. Particular in the small sample case, the length of Res_u(t) is only 90% of MQE(t), whereas it is about 99% for large samples. Indeed it is more favourable to resample from scaled residuals than to use unscaled terms, but there are no essential differences in the results of the scaled versions Res_p(t) and Res_h(t). Regarded as a whole, the results of MQE(t) are very similar to those of scaled residual resampling.

When using binomial data without EBV, one of the most evident differences in the results of the score resampling techniques Sco_m(t) and Sco_r(t) is the fact that for $n = 20$ the interval based on robust variance estimation is too small whereas the model-based variance estimator yields intervals which seem to be a little bit too large. Even in the case of EBV the Sco_r(t) interval is of poor quality when used in small samples. But it is interesting to note that increasing the sample size seems to result in a better performance. In the EBV situation the length of the Sco_m(t) interval is only about a quarter of the MQE(t) interval and therefore useless. This is documented by coverage probabilities less than 50%.

A lot of percentile intervals show similar tendencies in their behaviour compared with the corresponding intervals based on normal quantiles. Of course, like the MLE(t) the MLE(p) cannot be applied in the case of EBV. The Res(p) intervals are often outperformed by the Res(t) intervals, especially the Res_u(p) intervals turn out badly. However, for $n = 80$ there is no essential difference between Res(p) and Res(t) intervals. The Sco(p) intervals, which are based on N(0,1) random variates t_i, behave in a manner like Sco_r(t). Both types seem to be of worse quality if $\beta_1 = -1$ (small values of μ only) or $\beta_1 = +3$ (large range of μ).

In all those situations mentioned in Table III the percentile intervals do not yield a substantial improvement in the results. Therefore, it seems that this computationally extensive method does not gain an advantage over confidence intervals based on normal quantiles. But the deviations may be caused by our choice $B = 200$, which was also used in Moulton and Zeger (1991). Moreover, we are aware of the fact that the small number of trials in our simulation can possibly lead to some results which are not representative.

TABLE III

Deviations of the coverage probabilities from 95% nominal level in percent.

	$(\beta_0, \beta_1) = (-1, -1)$				$(\beta_0, \beta_1) = (-1, +1)$				$(\beta_0, \beta_1) = (-1, +3)$			
	no EBV		EBV		no EBV		EBV		no EBV		EBV	
$n = 20$	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
MLE(t)	-0.9	-1.2	-27.2	-26.3	-0.2	0.1	-26.7	-29.4	1.1	0.3	-26.2	-27.2
MQE(t)	-2.7	-2.3	-1.8	-1.0	-1.5	-1.6	0.3	-0.2	-1.8	-2.7	-4.8	-3.7
MLE(p)	-2.1	-2.3	-30.6	-28.2	-1.2	-0.7	-27.6	-29.5	-0.5	-2.3	-28.1	-30.7
Res_u(t)	-4.8	-4.0	-3.2	-2.2	-2.7	-2.7	-0.5	-1.7	-3.2	-3.7	-6.6	-5.6
Res_h(t)	-2.7	-2.1	-1.7	-0.7	-1.5	-1.7	0.4	-0.3	-1.7	-2.7	-4.3	-3.5
Res_p(t)	-2.7	-2.3	-1.9	-1.0	-1.5	-1.6	0.3	-0.2	-1.8	-2.8	-4.8	-3.7
Res_u(p)	-6.1	-4.2	-5.8	-2.1	-3.0	-3.3	-1.1	-1.7	-3.8	-5.6	-7.3	-7.2
Res_h(p)	-4.5	-2.5	-4.5	-0.4	-1.6	-1.9	-0.4	-1.0	-2.0	-3.2	-4.2	-4.3
Res_p(p)	-4.4	-2.6	-4.7	-0.7	-1.5	-2.1	-0.3	-0.9	-2.2	-3.7	-5.1	-4.9
Sco_m(t)	0.7	1.5	-49.7	-45.4	1.4	2.1	-56.0	-53.3	2.2	2.3	-48.7	-45.5
Sco_r(t)	-7.6	-5.0	-6.6	-5.7	-3.7	-4.1	-2.4	-3.3	-6.1	-3.7	-5.5	-5.5
Sco(p)	-7.9	-5.7	-7.4	-6.6	-4.0	-4.2	-2.7	-3.5	-6.7	-4.8	-6.0	-6.4
$n = 80$	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1	β_0	β_1
MLE(t)	0.1	1.2	-27.2	-27.5	-0.1	0.6	-29.9	-30.7	0.2	0.3	-26.1	-27.9
MQE(t)	-0.2	0.5	-1.6	0.5	0.2	0.1	-0.2	-0.6	-0.9	0.1	-1.3	-1.0
MLE(p)	-0.2	0.0	-28.9	-28.9	-1.4	-0.6	-30.6	-31.7	-1.2	-1.2	-26.4	-27.8
Res_u(t)	-0.7	0.0	-2.0	0.0	-0.1	-0.1	-0.2	-1.1	-1.4	-0.2	-1.7	-1.3
Res_h(t)	-0.2	0.4	-1.5	0.5	0.2	0.1	-0.2	-0.6	-0.9	0.1	-1.2	-1.0
Res_p(t)	-0.2	0.5	-1.6	0.5	0.2	0.1	-0.2	-0.6	-0.9	0.0	-1.3	-1.0
Res_u(p)	-1.6	-0.4	-2.9	-0.1	-1.4	-0.3	-1.1	-1.8	-1.7	-1.3	-2.5	-2.0
Res_h(p)	-1.0	0.0	-2.7	0.1	-1.3	-0.2	-1.0	-1.3	-1.3	-1.2	-2.1	-1.7
Res_p(p)	-1.0	-0.2	-2.7	0.1	-1.3	-0.2	-0.9	-1.3	-1.2	-1.2	-1.9	-1.7
Sco_m(t)	-0.1	1.5	-54.8	-55.4	0.4	1.1	-58.1	-58.4	0.3	0.2	-56.1	-54.5
Sco_r(t)	-1.4	-0.8	-2.3	-1.4	-0.4	-0.1	-0.8	-1.7	-1.7	-0.5	-1.3	-1.0
Sco(p)	-2.1	-0.5	-2.2	-1.1	-0.9	-1.0	-1.6	-1.9	-1.7	-1.1	-1.7	-1.7

7. CONCLUDING REMARKS

Without exception, this paper is concerned with the investigation of the performance of various confidence intervals in the situation of misspecified variance models, namely overdispersion. Thus, the judgement of the quality of the estimates can be seen from this point of view only. But there exists a variety of interesting situations, which cannot be described by a standard logistic regression model. For example, if x denotes a random design vector with components $x_i \stackrel{iid}{\sim} F_x$ and the conditional distribution of y_i given x_i is the binomial one, then the unconditional moments of y_i are not covered by a simple dispersion model. Maybe, the observation vector resampling technique in Moulton and Zeger (1991) is quite a good choice to handle this situation. It is interesting to note that the "closed form" approximation of the variance of their one-step bootstrap coefficient based on vector resampling is the same as the $\mathrm{Sco}_r(\mathrm{t})$ variance estimator. That means the vector resampling method can be used to construct at least a variance estimator even if x is not random.

Especially in the context of MQE(t) intervals, the behaviour of alternative estimates for σ^2 (e.g. a deviance based estimator) needs further study. Both resampling techniques can be generalized onto the whole class of GLM's.

BIBLIOGRAPHY

Davison, A.C. and Snell, E.J. (1991). Residuals and diagnostics. In *Statistical Theory and Modelling: In honour of Sir David Cox, FRS* (D.V. Hinkley, N. Reid, E.J. Snell, eds), pp. 83-106. Chapman and Hall, London.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.*, **7**, 1-26.

Fahrmeir, L. and Kaufmann, H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Statist.*, **13**, 342-368.

Firth, D., Glosup, J. and Hinkley, D.V. (1991). Model checking with nonparametric curves. *Biometrika*, **78**, 245-252.

Freedman, D.A. (1981). Bootstrapping regression models. *Ann. Statist.*, **9**, 1218-1228.

Hinde, J. (1992). Choosing between non-nested models: a simulation approach. In *Lecture Notes in Statistics* (L. Fahrmeir *et al.* eds) pp. 119-124, **78**. Springer, New York.

Lele, S. (1992). Resampling using estimating equations. In *Estimating Functions* (V.P. Godambe ed) pp. 295-304, Oxford University Press, Oxford.

Liang, K.-Y. and McCullagh, P. (1993). Case studies in binary dispersion. *Biometrics*, **49**, 623-630.

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

Liu, R.Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.*, **16**, 1696-1708.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*, 2nd ed. Chapman and Hall, London.

Moulton, L.H. and Zeger, S.L. (1991). Bootstrapping generalized linear models. *Comp. Statist. & Data Analysis*, **11**, 53-63.

Wedderburn, R.W.M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439-447.

Williams, D.A. (1982). Extra-binomial variation in logistic linear models. *Appl. Statist.*, **31**, 144-148.

Wu, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.*, **14**, 1261-1350.