

Bootstrapping generalized linear models

Lawrence H. Moulton

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

Scott L. Zeger

Department of Biostatistics, The Johns Hopkins University, Baltimore, MD 21205, USA

Received July 1989

Revised February 1990

Abstract: Several methods for bootstrapping generalized linear regression models are introduced. One-step techniques, both conditional and unconditional on the covariates, are examined with respect to robustness and coverage properties.

Keywords: Bootstrap, Generalized linear models, Robustness, Logistic regression, Overdispersion.

1. Introduction

The popularity of the use of the class of generalized linear models (GLMs; McCullagh and Nelder [12]) has given rise to a variety of applications in which functions of the estimated parameters of GLMs need to be evaluated. This paper describes the use of two bootstrap methods for GLMs analogous to those used for ordinary linear models (OLMs). These methods may be used for assessing the variability of the estimated functions of interest, and are particularly useful in situations with small sample sizes or very complicated functions. The proposed methods are efficient with respect to computation time, since only one iteration is performed for each bootstrap replication, instead of carrying out the iteratively weighted least squares algorithm to convergence. By allowing nonparametric (within the context of the chosen GLM) evaluation of variability, bootstrap methods can achieve varying degrees of robustness to assumptions commonly made about data. In small-to-moderate sized samples, it may be difficult or impossible to verify distributional assumptions, making robust approaches desirable.

Methods for bootstrapping OLMs are reviewed in section 2. Although most of the work in this field has been concerned with residual resampling, we put equal emphasis on complete observation vector resampling. The analogy between

ordinary and generalized linear models is exploited in section 3 to arrive at two one-step bootstrap techniques, some properties of which are explored in section 4. Section 5 illustrates the bootstrap techniques with a well-known data set (Finney [6]), and extensive simulation results are presented in section 6.

2. An overview of bootstrapping ordinary linear models

In order to gain insight into the problem of bootstrapping GLMs, we first examine some aspects of what is already known about the bootstrapping of OLMs of the form:

$$Y = X \beta + \epsilon, \quad (2.1)$$

$N \times 1$ $N \times p$ $p \times 1$ $N \times 1$

where ϵ is a vector of independent and identically distributed random variables with mean 0 and variance $\sigma^2 I$. Suppose we are interested in obtaining a bootstrap distribution of some function $\hat{\psi} \equiv \psi(\hat{\beta})$, where $\hat{\beta}$ is the least squares estimator $\hat{\beta} = (X'X)^{-1}X'Y$. We may proceed by calculating Monte Carlo approximations to either of two bootstrap distributions for $\hat{\psi}$, based on either the residuals of the least squares fit or the complete observation vectors. These two methods were first considered by Efron [4], and have been examined in greater depth by Freedman [7] and Wu [20], among others.

2.1. Residual resampling

The most common method of bootstrapping an OLM relies upon the assumed exchangeability of the error terms. This has been the focus of recent work by Hinkley [10], Freedman and Peters [8], and Wu [20]. It is based on the empirical distribution function

$$\hat{F}_r: \text{mass } 1/N \text{ at } r_i, \quad i = 1, \dots, N,$$

where the r_i are the elements of the residual vector $r = [I - X(X'X)^{-1}X']Y$. The Monte Carlo approximation is calculated by randomly drawing B bootstrap samples of N residuals from \hat{F}_r , yielding the vectors r^{*b} , $b = 1, \dots, B$. For each r^{*b} , we calculate $Y^{*b} = X\hat{\beta} + r^{*b}$, and then obtain

$$\begin{aligned} \beta^{*b} &= (X'X)^{-1}X'Y^{*b} \\ &= \hat{\beta} + (X'X)^{-1}X'r^{*b}, \end{aligned}$$

from which we calculate the $\psi^{*b} \equiv \psi(\beta^{*b})$ to approximate the distribution of $\hat{\psi}$.

Now (dropping the replication superscript) $E_r r_i^* = N^{-1} \sum r_i = 0$, and therefore $E_r \beta^* = \hat{\beta}$. Here, E_r denotes expectation taken with respect to \hat{F}_r . Also,

$$\begin{aligned} \text{Var}_r \beta^* &= (X'X)^{-1}X' \text{diag}[E_r(r^{*b})^2]X(X'X)^{-1} \\ &= \hat{\sigma}^2(X'X)^{-1}, \end{aligned}$$

where $\hat{\sigma}^2 = N^{-1} \sum r_i^2$. This is the usual maximum likelihood estimator, which is biased, since $E_Y[\text{Var}_r \beta^*] = (N-p)N^{-1}\sigma^2(X'X)^{-1}$. The bias results from the fact that $\text{Var}_Y(r_i) = \sigma^2(1-h_i)$, where h_i is the i th diagonal element of $X(X'X)^{-1}X'$. A bootstrap method which uses the standardized residuals $r_i(1-h_i)^{-1/2}$, each of which has mean 0 and variance σ^2 , will eliminate this bias.

2.2. Vector resampling

A conceptually simpler bootstrap distribution arises from employing the empirical distribution function of the observation vectors $(y_i; x_i)$, where x_i is the i th row vector of X . The bootstrap distributions of $\hat{\beta}$ and $\hat{\psi}$ may be approximated by drawing B samples of size N from

$$\hat{F}_r: \text{mass } 1/N \text{ at } (y_i; x_i), \quad i = 1, \dots, N.$$

each time creating a pseudo-data set (Y^{*b}, X^{*b}) from which the $\beta^{*b} = (X^{*b'}X^{*b})^{-1}X^{*b'}Y^{*b}$ and then the ψ^{*b} are calculated.

This bootstrapping approach corresponds to a correlation model in which X is random, not fixed as in a regression model (Freedman [7]). As a consequence, we get a quite different variance result (Efron [5]; Moulton [13]):

$$\text{Var}_r \beta^* = (X'X)^{-1}X'\text{diag}(r_i^2)X(X'X)^{-1} + O(N^{-2}).$$

Under the homoscedastic model where $E(\epsilon_i^2 | x_i) = E(\epsilon_i^2) = \sigma^2$, this variance is asymptotically equivalent to $\text{Var}_r \beta^*$ (Freedman [7]), but such will not be the case when there is heteroscedasticity. In the latter situation, $N[\text{Var}_r \beta^* - \text{Var}_r \hat{\beta}] \rightarrow 0$, while $\text{Var}_r \beta^* = \hat{\sigma}^2(X'X)^{-1}$ will no longer consistently estimate $\text{Var}_Y \hat{\beta}$. Note that, apart from the $O(N^{-2})$ term, $\text{Var}_r \beta^*$ is identical to the weighted jackknife estimator of Hinkley [9].

3. Bootstrapping generalized linear models

A generalized linear model is characterized by three components (McCullagh and Nelder [12]): an error distribution from the exponential family $f_\phi(y; \theta) = \exp\{[y\theta - a(\theta) + b(y; \phi)]/\phi\}$, a systematic component $\eta = X\beta$, and a monotonic, differentiable link function $g(\mu) = \eta$, where $\mu = E(Y)$. The log-likelihood function $\sum [y_i\theta_i - a(\theta_i) + b(y_i; \phi)]$ is maximized by finding the solution of the system of equations

$$\sum (y_i - \hat{y}_i)\Delta_i x_{ij} = 0, \quad j = 1, \dots, p,$$

where $\hat{y}_i = g^{-1}(x_i\hat{\beta})$, $\Delta_i = \partial\theta_i/\partial\eta_i | \hat{\beta}$. These estimating equations may be written in matrix form as $X'\Delta s = 0$, where $\Delta = \text{diag}(\Delta_i)$, $s = y - \hat{y}$. The maximum likelihood estimator $\hat{\beta}$ may be found via iteratively weighted least squares (IWLS), in which the $(t+1)$ st update is given by:

$$\beta^{(t+1)} = (X'\Delta V \Delta X)^{-1} X'\Delta V \Delta z,$$

where Δ , V , and z are evaluated at $\beta^{(t)}$, $z = X\beta^{(t)} + (\Delta V)^{-1}s$, and $V = \text{diag}(\hat{\sigma}^2 a(\theta_i)/\partial\theta_i^2 | \beta^{(t)})$. This scheme is continued until convergence of the log-likelihood is attained.

3.1. Standardized Pearson residual resampling

We may bootstrap a GLM in a manner analogous to the residual resampling method for OLMs in section 2.1, utilizing a one-step procedure to obtain bootstrap coefficients. Although there are several different types of residuals defined for GLMs, the Pearson residual is most amenable (both conceptually and computationally) to a process of bootstrapping to obtain the coefficients β^{*b} . Ideally, we would want to resample i.i.d. quantities, but no such residuals are readily available for the class of GLMs. The Pearson residual is $r_i = (y_i - \hat{y}_i)v_i^{-1/2}$, with vector $r = V^{-1/2}s$. We will use the more nearly exchangeable quantities $e_i = s_i/[v_i(1 - h_i)]^{1/2}$, $i = 1, \dots, N$, where now h_i is the i th diagonal element of $G(G'G)^{-1}G'$, with $G = V^{1/2}\Delta X$. In our simulations, bootstrapping with these standardized residuals, e_i , consistently outperformed use of the raw residuals, r_i , and consequently are the only ones discussed in the remainder of this paper. The residual bootstrap thus is based on \hat{F}_e : mass $1/N$ at e_i , $i = 1, \dots, N$, and bootstrap coefficients may be calculated by reattaching resampled residual vectors as was done for OLMs:

$$\beta^{*b} = \hat{\beta} + (G'G)^{-1}G'e^{*b}, \quad b = 1, \dots, B. \quad (3.1)$$

Note that this is a one-step procedure; iteration to convergence cannot in general be expected. Such an effort would represent an attempt to solve the equations $X'\Delta V^{1/2}e^* = 0$ using the expected Hessian, $X'\Delta V\Delta X$, that is used to solve $X'\Delta s = 0$.

Of particular interest is the variance of these bootstrap quantities. The following proposition tells us the variance using the standardized residual bootstrap is the inverse of the information matrix multiplied by an estimate of scale.

Proposition 3.1. *The variance of the bootstrap coefficient given in (3.1) for a GLM is:*

$$\text{Var}_e \beta^* = \hat{\phi}_e (G'G)^{-1},$$

where $\hat{\phi}_e = N^{-1}\sum e_i^2 - N^{-2}[\sum e_i]^2$.

This proposition follows directly from the fact that $\text{Var}_e \beta^* = (G'G)^{-1}G' \text{Var}_e(e^*)G(G'G)^{-1}G'$, and that $\text{Var}_e e^* = E_e(e_i^*)^2 - E_e^2(e_i^*) = N^{-1}\sum e_i^2 - N^{-2}[\sum e_i]^2$. The standardized residual bootstrap estimate of $\text{Var} \hat{\beta}$ may therefore be calculated by multiplying the inverse of the estimated information matrix by the scale estimate $\hat{\phi}_e$.

3.2. Observation vector resampling

Proceeding as in section 2.2., we can create pseudo-data sets (Y^{*b}, X^{*b}) by resampling from \hat{F}_e . We could then carry out the IWLS scheme until convergence

is attained, and then calculate the desired ψ^{*b} from the resulting β^{*b} . However, there are two drawbacks to this approach. The first is that if N and/or p is large, computational costs may be quite high, as several iterations may be needed for each bootstrap replication, even when the MLE $\hat{\beta}$ is used as the starting value.

The second and more serious problem is that we would risk obtaining “extreme” data replications for which either the likelihood fails to converge (rare) or the parameter estimates fail to converge (much more likely). We have encountered the latter situation in several instances when using a logit link with binomial error, both with simulated data and with the bioassay data analyzed by Finney [6], as published in Pregibon [16]. In the latter instance, 21 of 100 replications resulted in the $\beta^{(i)}$ vector diverging, although the likelihood converged. This phenomenon is discussed more thoroughly in McCullagh and Nelder [12]. A bootstrap procedure which attempts to iterate to convergence will face this difficulty. Ignoring the divergent bootstrap replicates can introduce bias. However, if all we require is a bootstrap *percentile* interval, only a high proportion (at least as great as the level of the interval) of the bootstrap replications must converge, since the divergent ones result in coefficients with known values of plus or minus infinity.

In response to the foregoing difficulties, we propose employing a one-step approximation to the bootstrap distribution of $\hat{\beta}$. We start with V , Δ , and z evaluated at $\hat{\beta}$ and take one Newton–Raphson step in the direction of what would be the MLE for (Y^{*b}, X^{*b}) . This scheme is best represented through use of a resampling matrix $D = \text{diag}(d_i)$, where $d = (d_1, \dots, d_N)$ is distributed as a Multinomial($N; 1/N, \dots, 1/N$) random vector. The bootstrap coefficients can then be obtained from

$$\begin{aligned} \beta^{*b} &= (X' \Delta V^{1/2} D_b V^{1/2} \Delta X)^{-1} X' \Delta V^{1/2} D_b V^{1/2} \Delta [X \hat{\beta} + (\Delta V)^{-1} s] \\ &= \hat{\beta} + (G' D_b G)^{-1} G' D_b r, \quad b = 1, \dots, B. \end{aligned} \quad (3.2)$$

A Taylor series expansion of β^{*b} about $D_b = I_N$ permits calculation of the variance of the β^{*b} (see the Appendix of Moulton and Zeger [14]), as given in the following proposition.

Proposition 3.2. *The variance of the one-step bootstrap coefficient given in (3.2) for a GLM is*

$$\text{Var}_r \beta^* = (G'G)^{-1} G' \text{diag}(r_i^2) G (G'G)^{-1} + O(N^{-2}).$$

Thus $\text{Var}_r \beta^*$ cannot be calculated in closed form as can $\text{Var}_e \beta^*$, but it has a closed form approximation that results from dropping the $O(N^{-2})$ term, which we shall denote by $\hat{\text{Var}}_{r,\cdot}$.

4. Robustness of the one-step bootstrap methods

We use the term “robust” in the classical sense indicating that an estimator remains consistent in the face of failure of assumptions underlying the method of

analysis (Box [1]). In this section we briefly describe robustness properties we can expect of the one-step bootstrap techniques of section 3.

The residual (or conditional) bootstrap coefficients of (3.1) have variance $\hat{\phi}_e(G'G)^{-1}$, automatically incorporating an estimate of scale. For Gaussian response models, the scale is always estimated; for binomial or Poisson response models, $\phi = 1$ under the assumption of independence of observations and no estimate is required. However, when there is extra-binomial or extra-Poisson variation, one possible remedy is to use such an estimate of scale. For binomial response, this corresponds to a situation where $\text{Var}(Y_i) = \phi n p_i (1 - p_i)$, where $Y_i | p_i \sim \text{Binomial}(n, p_i)$. It is also equivalent to a quasilielihood estimate of variance (McCullagh and Nelder [12]). Thus, using these residual-based bootstrap coefficients in situations close to these will reflect the appropriate amount of variability.

Robustness to a broader range of departures from the posited model can be gained through use of the vector-based (or unconditional) bootstrap coefficient (3.2). A consistent estimator of the variance of these coefficients, $\hat{\text{V}}\hat{a}r_e$, is the same as that employed by Cox [2], Huber [11], White [18], Pregibon [15], and Royall [17] for handling quite general model misspecification. Virtually all that is required for consistency of this variance estimator is that our estimator of β be consistent. Its main drawback is that it is not a resistant estimator of variance, as extreme residuals may have undue influence.

5. Application to logistic regression analysis of Finney's data

For purposes of illustration we apply the different bootstrap methods to the bioassay data of D.J. Finney [6], as reported in Pregibon [16]. The dichotomous response "dilation of the blood vessels" is expressed as a function of the rate and volume of inspired air. Table 1 gives the following variance estimates of the estimated coefficients, using: Var_e , Var_r , and the closed form approximation $\hat{\text{V}}\hat{a}r_e$, as well as the standard information-based estimate for comparison. Because of the small sample size ($N = 39$) and the presence of two observations with large Pearson residuals, $B = 10,000$ bootstrap replications were necessary to achieve stability of the Var_r estimates.

Table 1
Estimated coefficient variances for the Finney data.

| Variable | Coefficient estimate | Variance estimates | | | |
|------------|----------------------|--------------------|----------------|----------------|----------------------------|
| | | Information based | Var_e | Var_r | $\hat{\text{V}}\hat{a}r_e$ |
| log Rate | 5.2 | 3.5 | 3.3 | 5.9 | 5.6 |
| log Volume | 4.6 | 3.4 | 3.2 | 5.4 | 5.0 |

We see that the residual-based variance estimator yields slightly lower values than the standard information-based estimate, since the estimate $\hat{\phi} = 0.96$ is less than one. The difference is negligible; note that multiplying by the less biased (although non-bootstrap) scale factor $\hat{\phi}N/(N-p) = 1.04$ results in standard errors slightly greater than the standard ones. The vector-based estimators are much larger because of two “outliers” in the data, which have large Pearson residuals; without these observations the data are completely separable, producing infinite MLEs, thus precluding fully iterative bootstrap variance estimates.

Which is the correct variance estimate in this case is debatable. The simulations in the following section are designed to give us a basis for comparison.

6. Comparison of bootstrap approaches through simulation

6.1. Design of the simulations

The simulations evaluate the different variance assessment techniques for coefficients of logistic regression models. Three factors were used:

- (1) Sample sizes of $N = 20, 40,$ and 80 ;
- (2) Response probabilities of about 0.12 and 0.50 ;
- (3) Extra-binomial variation (EBV) absent or present.

Each simulated data set was composed of N binomial responses $Y_i = \sum_{j=1}^{10} W_{ij}$, where the W_{ij} are generated from the relation: $\text{logit}\{\text{Pr}[W_{ij} = 1]\} = \beta_0 + \beta_1 x_i$, for the trials in which there is no EBV, and from

$$\text{logit}\{\text{Pr}[W_{ij} = 1]\} = U_{ij}^{1/v_i}, \quad \text{where } v_i = \exp(\beta_0 + \beta_1 x_i)$$

for the trials with EBV, where the U_{ij} are drawn from a Uniform $(0, 1)$, distribution. This latter method yields beta-binomial responses Y_i with means $10v_i/(1+v_i)$ and variances $10v_i(12+v_i)/[(1+v_i)^2(2+v_i)]$.

All simulations were performed conditional on a set of 20 values x_i drawn from a Uniform $(0, 2)$ distribution, copied once and three times, respectively, for the sample sizes 40 and 80. For all trials the intercept was $\beta_0 = -1$; for half the trials, the value of the slope parameter was $\beta_1 = -1$, with $\beta_1 = 1$ for the other half.

The simulations consisted of generating 1600 data sets for each combination of the above three factors. For each data set, $B = 200$ bootstrap replications were performed using the Pearson residual, standardized Pearson residual, and one-step vector techniques. In ordinary data analysis situations, a larger number, say $B = 1000$, would be recommended for the construction of the percentile intervals. In our simulation a much smaller number suffices, since by far the largest component of variances is between data set variation. All of the pseudo-random numbers were produced by the generator of Wichmann and Hill [19], and all calculations were done using double-precision FORTRAN-77 routines.

The focus of the simulation was the coverage of confidence intervals for the functions ψ : β_0 , β_1 , $\beta_0\beta_1$, and β_0/β_1 . For a general, thorough discussion of bootstrap confidence intervals, see DiCiccio and Romano [3]. The following are abbreviations for and descriptions of the 95% nominal level confidence intervals we calculated:

- II: $\hat{\psi} \pm 1.96 \hat{\sigma}_\psi$, where $\hat{\sigma}_\psi$ is the estimate of standard error derived from the inverse of the Fisher information matrix; for the product and quotient functions, delta-method variance estimates are used.
- RB: $[\hat{F}_{\psi_e^*}^{-1}(0.025), \hat{F}_{\psi_e^*}^{-1}(0.975)]$, where $\hat{F}_{\psi_e^*}^{-1}$ is the empirical distribution function of the $\psi(\beta_e^*)$; the bootstrap percentile interval based on the standardized Pearson residuals.
- RC: $\hat{\psi} \pm 1.96 \hat{\phi}_e \hat{\sigma}_\psi$; the closed form interval based on the variance of the standardized residual bootstrap.
- VB: $[\hat{F}_{\psi_e^*}^{-1}(0.025), \hat{F}_{\psi_e^*}^{-1}(0.975)]$, where $\hat{F}_{\psi_e^*}^{-1}$ is the empirical distribution function of the $\psi(\beta_e^*)$; the bootstrap percentile interval based on the one-step vector bootstrap method.
- VC: $\hat{\psi} \pm 1.96 \text{V}\hat{\text{a}}\text{r}_e(\hat{\psi})$; the interval based on the closed form approximation to the one-step vector bootstrap variance estimator.

In summary, RB and VB intervals are bootstrap percentile intervals calculated from the sets of 200 bootstrap quantities; II, RC and VC are calculated in closed form, using delta-method variance approximations for the product and quotient of the coefficients.

6.2. Simulation results

Table 2 shows deviations from the nominal level of the coverages of the various confidence intervals for each of the simulation trial configurations. The most striking departures from nominal levels are seen when the usual inverse information variance estimator is used on data with extra-binomial variation, as was to be expected. Simple sums of absolute deviations (or of deviations from coverages transformed to achieve symmetry) may be used to compare the methods. Note also that when the observed coverage based on 1600 observations is 95%, a 95% confidence interval for the true coverage will have a half-width of $1.96 (0.95 - 0.05)^{1/2} / 40 = 0.01$. Thus, rough general comparisons may be made by counting the number of times the absolute value of the deviation is two or more percentage points. Applying this criterion to all the entries of Table 2, the RB bootstrap percentile interval using the standardized residuals edged out the closed form version RC, 18 to 21. The worst performance (not counting the standard II approach) was given by VC, the closed form approximation to the vector-based bootstrap variance, with 31 large deviations. RC fared better than VB for the trials without extra-binomial variation, while VB did a bit better than RC in the face of extra-binomial variation.

For the 80 observation sample size, there was relatively little difference between the methods (except for II in the presence of EBV). This is not surprising

Table 2
Coverage, in percentage points difference from the nominal 95%, of the bootstrap methods as applied to simulated logistic regressions. For each configuration, 1600 data sets were simulated, and 200 bootstrap replications performed for each data set.

| Method | EBV | | | | | | | | | | | | | | | |
|---------------|------------|-----|-----|-----------|-----|-----|------------|-----|-----|-----------|-----|-----|-----|-----|-----|-----|
| | No EBV | | | | | | EBV | | | | | | | | | |
| | Slope = -1 | | | Slope = 1 | | | Slope = -1 | | | Slope = 1 | | | | | | |
| | Int | Slo | Pro | Quo | Int | Slo | Pro | Quo | Int | Slo | Pro | Quo | | | | |
| <i>N</i> = 20 | | | | | | | | | | | | | | | | |
| II | 1 | 1 | 3 | -5 | 0 | 1 | -3 | 3 | -31 | -34 | -16 | -31 | -29 | -27 | -29 | -16 |
| RB | -1 | 0 | 1 | -5 | -1 | -1 | -1 | -1 | -4 | -2 | -14 | -3 | -2 | -1 | 1 | 1 |
| RC | 0 | -1 | 1 | -5 | -1 | -1 | -4 | 2 | -1 | -2 | 3 | -18 | -2 | -1 | -8 | 3 |
| VB | -3 | -1 | -7 | 1 | -3 | -2 | -3 | -2 | -4 | -5 | -11 | -2 | -2 | -2 | 0 | 1 |
| VC | -5 | -3 | -1 | -9 | -4 | -4 | -7 | 0 | -9 | -7 | 1 | -19 | -5 | -5 | -10 | 1 |
| <i>N</i> = 40 | | | | | | | | | | | | | | | | |
| II | 0 | 0 | 2 | -3 | 0 | 0 | -1 | 0 | -32 | -33 | -23 | -25 | -30 | -28 | -29 | -21 |
| RB | -1 | -2 | -4 | 1 | -2 | -1 | -1 | -2 | -2 | -1 | -8 | 0 | -2 | -1 | -1 | 1 |
| RC | -1 | -1 | 2 | -3 | -1 | 0 | -2 | 0 | 0 | 0 | 3 | -9 | -1 | 0 | -6 | 3 |
| VB | -3 | -2 | -5 | 1 | -2 | -2 | -2 | -3 | -2 | -4 | -8 | 0 | 0 | 0 | 0 | 1 |
| VC | -3 | -3 | 1 | -5 | -4 | -3 | -4 | -1 | -3 | -4 | 3 | -10 | -1 | -1 | -7 | 2 |
| <i>N</i> = 80 | | | | | | | | | | | | | | | | |
| II | 1 | 0 | 1 | -1 | 0 | 1 | 0 | 0 | -32 | -33 | -27 | -21 | -31 | -28 | -30 | -27 |
| RB | -1 | -1 | -3 | 0 | -1 | 0 | -1 | -1 | -2 | -1 | -7 | 1 | -3 | -1 | -2 | -1 |
| RC | 0 | -1 | 1 | -2 | 0 | 1 | 0 | 0 | -1 | -1 | 3 | -5 | -3 | 0 | -3 | 2 |
| VB | -1 | -2 | -4 | -1 | -1 | -1 | -1 | -1 | -1 | -3 | -5 | 1 | -2 | -2 | -2 | -1 |
| VC | -1 | -1 | 0 | -2 | -1 | 0 | -1 | 0 | -2 | -3 | 2 | -7 | -2 | -1 | -4 | 1 |

in view of the fact that the methods are asymptotically equivalent for the situations studied, where we have introduced at most a scale parameter that reflects overdispersion. In situations of different or greater heterogeneity of variance, we may expect the vector-based methods to outperform the residual-based ones. For all sample sizes, we see uniformly better performance when the mean response probability is 0.5 ($\beta_1 = 1$) than when it is 0.12 ($\beta_1 = -1$), due to the lower variability of $\hat{\beta}_1$ in the former case.

Because of their greater variability, the product and quotient of the coefficients proved more difficult to estimate than the coefficients themselves. The percentile intervals had more difficulty with the product, while the closed-form methods, relying on the accuracy of the delta-method approximation, were further off in coverages of the quotient. The delta-method approximation used for the closed-form methods was less accurate for the more highly nonlinear quotient.

7. Final comments

The simulations of the preceding section compared the performances of the bootstrap techniques with respect to confidence interval construction. Numerous robust closed-form competitors exist for this purpose. However, the bootstrap can provide much more than interval estimation, and the results of section 6 encourage the use of these one-step methods in other situations. For example, they could be useful in such computationally intensive applications as bootstrapping step-

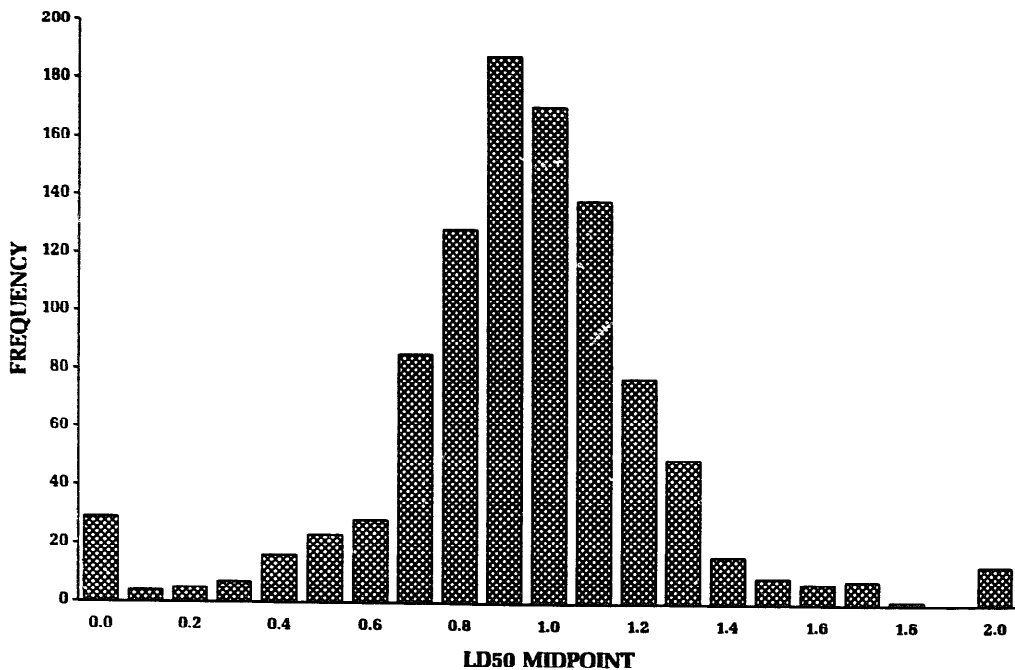


Fig. 1. Estimated bootstrap distribution of $-B_0/B_1$ for a simulated data set.

wise variable selection procedures for a logistic regression problem with many variables. Even simple examination of an estimated bootstrap distribution of a quantity of interest can be more informative than a mere confidence interval. Figure 1 shows the distribution of 1000 bootstrap estimates of $-\beta_0/\beta_1$ obtained from the one-step standardized residual method. We see that for this data set, although the distribution may be close enough to Gaussian to calculate an adequate delta-method interval, the distribution is slightly skewed left, with a mean value of 0.872 as compared to the median 0.945 and the MLE 0.959.

References

- [1] G.E.P. Box, Non-normality and tests on variances. *Biometrika* **40** (1953) 318–335.
- [2] D.R. Cox, Tests of separate families of hypotheses. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1961) 105–123.
- [3] T.J. DiCiccio and J.P. Romano, A review of bootstrap confidence intervals (with discussion). *J. Roy. Statist. Soc. Ser. B* **50** (1988) 338–370.
- [4] B. Efron, Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** (1979) 1–26.
- [5] B. Efron, The Jackknife, the bootstrap and other resampling plans. In *Regional Conference Series in Applied Mathematics* (38) (SIAM, Philadelphia, 1982).
- [6] D.J. Finney, The estimation from individual records of the relationship between dose and quantal response. *Biometrika* **34** (1947) 320–334.
- [7] D.A. Freedman, Bootstrapping regression models. *Ann. Statist.* **9** (1981) 1218–1228.
- [8] D.A. Freedman and S.C. Peters, Bootstrapping a regression equation: some empirical results. *JASA* **79** (1984) 9–16.
- [9] D.V. Hinkley, Jackknifing in unbalanced situations. *Technometrics* **19** (1977) 285–292.
- [10] D.V. Hinkley, Bootstrap methods. *J. Roy. Statist. Soc. Ser. B* **50** (1988) 321–337.
- [11] P.J. Huber, The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1967) 221–233.
- [12] P. McCullagh and J.A. Nelder, *Generalized Linear Models* (2nd ed.) (Chapman and Hall, London, 1989).
- [13] L.H. Moulton, *Bootstrapping Generalized Linear Models With Application to Longitudinal Data*. Unpublished Ph.D. thesis (The Johns Hopkins University, Baltimore, 1986).
- [14] L.H. Moulton and S.L. Zeger, Analyzing repeated measures on generalized linear models via the bootstrap. *Biometrics* **45** (1989) 381–394.
- [15] D. Pregibon, An alternative covariance estimated for generalized linear models. *The GLIM Newsletter* **6** (1983) 51–55.
- [16] D. Pregibon, Logistic regression diagnostics. *Ann. Statist.* **4** (1981) 705–724.
- [17] R.M. Royall, Model robust confidence intervals using maximum likelihood estimators. *Int. Statist. Rev.* **54** (1986) 221–226.
- [18] H. White, Maximum likelihood estimation of misspecified models. *Econometrica* **50** (1982) 1–25.
- [19] B.A. Wichmann and I.D. Hill, An efficient and portable pseudo-random number generator. Algorithm AS 182. *Applied Statistics* **31** (1982) 188–192.
- [20] C.F.J. Wu, Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** (1986) 1261–1350.