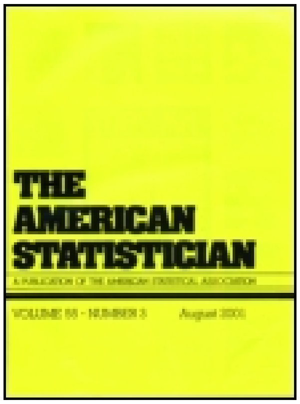

The American Statistician
Publication details, including instructions for authors and subscription information:
http://amstat.tandfonline.com/loi/utas20

Simulation-Based Confidence Intervals for Functions With Complicated Derivatives

Micha Mandel [a]

[a] Department of Statistics , The Hebrew University of Jerusalem , Jerusalem , Israel

Accepted author version posted online: 25 Mar 2013.Published online: 29 May 2013.

To cite this article: Micha Mandel (2013): Simulation-Based Confidence Intervals for Functions With Complicated Derivatives, The American Statistician, 67:2, 76-81

To link to this article: http://dx.doi.org/10.1080/00031305.2013.783880

Simulation-Based Confidence Intervals for Functions With Complicated Derivatives

Micha MANDEL

In many scientific problems, the quantity of interest is a function of parameters that index the model, and confidence intervals are constructed by applying the delta method. However, when the function of interest has complicated derivatives, this standard approach is unattractive and alternative algorithms are required. This article discusses a simple simulation-based algorithm for estimating the variance of a transformation, and demonstrates its simplicity and accuracy by applying it to several statistical problems.

KEY WORDS: Asymptotic normal estimator; Delta method; Multiple sclerosis; Parametric bootstrap.

1. INTRODUCTION

Many scientific questions concern functions of parameters indexing the probabilistic model. Examples include the correlation coefficient, which is a function of the first two moments; the odds ratio, which is a function of two probabilities; the probability of 1 year survival after a certain medical treatment, which is a function of the parameters indexing the underlying distributional law; and so forth. In such cases, confidence intervals for the parameter of interest are usually constructed by first estimating the basic parameters of the problem and the corresponding variance matrix, and then applying the delta method; see Ver Hoef (2012) for the use and origin of the method.

The delta method requires computation of the derivative of the function of interest. This is often a simple task that poses no special difficulties. However, in certain cases, the computation can be burdensome if not impossible, and alternative methods are required. The aim of this article is to present one such alternative and to demonstrate its simplicity and usefulness in several theoretical and practical situations.

To state the problem more formally, let $\theta \in \mathbb{R}^m$ be an unknown parameter indexing the assumed model, and let θ_0 denote the true value. Suppose that the estimator $\hat{\theta}_n$ of θ satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N_m(0, \Sigma), \qquad (1)$$

where the asymptotic variance $\Sigma = \Sigma(\theta_0)$ often depends on θ_0. The standard approximate $1-\alpha$ confidence region for θ is given by $A = \{\theta : n(\hat{\theta}_n - \theta)^t \hat{\Sigma}_n^{-1}(\hat{\theta}_n - \theta) \le q_{1-\alpha}\}$, where q_α is the α quantile of the chi-squared distribution with m degrees of freedom, and $\hat{\Sigma}_n$ is a consistent estimator for Σ (often, but not always, $\hat{\Sigma}_n = \Sigma(\hat{\theta}_n)$).

In this article, we consider the problem of constructing a confidence interval for a function $g(\theta) : \mathbb{R}^m \to \mathbb{R}$, that for ease of presentation we take to be one-dimensional (extension to the multivariate case $g(\theta) : \mathbb{R}^m \to \mathbb{R}^k$ is straightforward). A natural confidence region for $g(\theta)$ is $g(A) = \{g(\theta) : \theta \in A\}$. However, calculating $g(A)$ is often complicated and, furthermore, $g(A)$ may have a coverage level larger than the desired $1-\alpha$ level (as a simple example, suppose that $\theta \in \mathbb{R}$ and let $g(\theta) = \theta^2$, then $P(g(\theta) \in g(A)) = P(\{\theta \in A\} \cup \{\theta \in -A\}) > P(\theta \in A) = 1-\alpha$). Instead, the delta method is commonly used when g is differentiable at θ_0: Denote the gradient of g by $D(\theta) = (\partial g/\partial\theta_1, \ldots, \partial g/\partial\theta_m)^t$, then

$$\sqrt{n}\{g(\hat{\theta}_n) - g(\theta_0)\} \xrightarrow{D} N(0, D^t(\theta_0)\Sigma D(\theta_0)). \qquad (2)$$

When $D(\theta)$ is continuous in a neighborhood of θ_0, an approximate confidence interval for θ is constructed by $g(\hat{\theta}_n) \pm z_{1-\alpha/2} n^{-1/2}\{D^t(\hat{\theta}_n)\hat{\Sigma}_n D(\hat{\theta}_n)\}^{1/2}$, where z_α is the α quantile of the standard normal distribution.

This article considers the construction of confidence intervals in situations where g is differentiable but the calculation of the derivative, and hence the approximation of the asymptotic variance of $g(\hat{\theta}_n)$ by (2), is complicated (e.g., Aalen et al. 1997; see also the references below and Section 3.4). One popular alternative for constructing confidence intervals is the bootstrap, an algorithm that requires: (i) the simulation of many datasets, (ii) the estimation of the parameter θ in each replication, and (iii) the evaluation of the function $g(\theta)$. The variance or the quantiles of the simulated $g(\theta)$ values are used to construct the confidence region. The first two steps, that is, generating the datasets and estimating θ, are usually the most demanding ones in terms of programming and computer time. Here, we demonstrate the use of an alternative parametric bootstrap-like algorithm that replaces these two steps with repeated sampling from the m-variate normal distribution. The algorithm is not new and has been applied to several problems before (Aalen et al. 1997; Singh, Stukel, and Pfeffermann 1998; Drton et al. 2003; Whitaker and Farrington 2004; Mandel et al. 2007; Mandel and Betensky 2008; van den Hout, Jagger, and Matthews 2009; Mandel 2010; van den Hout and Matthews 2010; Titman et al. 2011). The aim of the current article is two-fold. First, to explain the algorithm and to discuss why it may provide valid

Micha Mandel is Senior Lecturer, Department of Statistics, The Hebrew University of Jerusalem, Jerusalem, Israel (E-mail: *msmic@huji.ac.il*). The work was partially supported by The Israel Science Foundation (Grant No. 774/11). The author thanks Yosi Rinott, an associate editor, and a referee for their helpful comments. I am grateful to the editor, Ronald Christensen, whose comments and suggestions greatly helped to sharpen and improve the article.

inference (Section 2 and the Appendix), and second, to demonstrate its use and to compare its performance to that of the delta method under different models (Section 3). Section 4 concludes the article with a short summary.

2. THE ALGORITHM

An informal argument leading to the algorithm is as follows. Consider a parametric bootstrap where the estimated value $\hat{\theta}_n$ is used to generate new datasets and the parameter θ is estimated at each. If (1) holds, one might expect that the distribution of the bootstrap estimates be approximately normal with moments $\hat{\theta}_n$ and $\mathrm{var}(\hat{\theta}_n)$, and therefore, one could hope that replacing these, typically demanding steps, with the much simpler task of sampling normal random variables would yield asymptotically correct variance estimators for $g(\hat{\theta}_n)$. This is the basic idea underlying the suggested algorithm that can be summarized as follows:

1. Sample B-independent realizations, $Y_{n1}, \ldots, Y_{nB}$, from the $N(0, \hat{\Sigma}_n)$ law, and define $\theta^*_{nb} = \hat{\theta}_n + n^{-1/2} Y_{nb}$.
2. For each realization, calculate $g^*_{nb} = g(\theta^*_{nb})$, and calculate the sample mean squared error (MSE): $\widehat{\mathrm{MSE}}^*_n = B^{-1}\sum_{b=1}^{B}(g^*_{nb} - g(\hat{\theta}_n))^2$, which is an approximation of $\mathrm{MSE}^*_n = E[\{g^*_{nb} - g(\hat{\theta}_n)\}^2 \mid \hat{\theta}_n, \hat{\Sigma}_n]$.
3. Construct the confidence interval $g(\hat{\theta}_n) \pm z_{1-\alpha/2}\sqrt{\widehat{\mathrm{MSE}}^*_n}$.

Confidence intervals constructed by both the algorithm above and the delta method rely on the asymptotic normal property (1). However, the former approach replaces the need of calculating analytically the derivative with evaluation of g many times (Step 2).

In the Appendix, we discuss the validity of the algorithm. We emphasize here that this requires not only the differentiability of g (needed also for the validity of the delta method) but also the technical condition that MSE^*_n is finite, which fails in some important cases, for example, the univariate case with $g(\theta) = 1/\theta$. Thus, the advantage of the method is not in providing valid intervals when the delta method fails, but rather in suggesting a simple algorithm for problems in which calculation of derivatives is complicated.

3. EXAMPLES

Given the simplicity of the algorithm, it is of interest to compare its performance to that of the widely used delta method. Section 3.1 compares the two approaches theoretically in a one-parameter model, and Sections 3.2 and 3.3 compare them by simulation in the important case of confidence intervals for the correlation coefficient and in a problem of estimating time to progression in multiple sclerosis. Section 3.4 discusses an example where the computation of the derivative of g is difficult but the simulation method can be easily applied.

3.1 Confidence Intervals for exp(μ)

Let $X_1, \ldots, X_n$ be iid $N(\mu, 1)$ and consider interval estimation of $g(\mu) = \exp(\mu)$. This simple example enables exact calculations of the intervals and their coverage probabilities and demonstrates well the differences between the two methods. The example can be used in statistical inference courses to explore the properties of the delta method and the simulation approach that, in this example, is identical to parametric bootstrap.

Let $\bar{X}_n = n^{-1}\sum X_i$, then $[\bar{X}_n - z_{1-\alpha/2}/\sqrt{n}\ ,\ \bar{X}_n + z_{1-\alpha/2}/\sqrt{n}]$ is the standard $1-\alpha$ confidence interval for μ, and the resulting exact $1-\alpha$ confidence interval for $g(\mu)$ is

$$[\exp(\bar{X}_n - z_{1-\alpha/2}/\sqrt{n}), \exp(\bar{X}_n + z_{1-\alpha/2}/\sqrt{n})]. \tag{3}$$

Applying the delta method, we find that $\sqrt{n}(e^{\bar{X}_n} - e^{\mu}) \xrightarrow{D} N(0, e^{2\mu})$, which yields the delta method interval $\exp(\bar{X}_n) \pm z_{1-\alpha/2}\exp(\bar{X}_n)/\sqrt{n}$, with probability of covering the true parameter:

$$\begin{aligned} &P_\mu(\exp(\bar{X}_n)[1 - z_{1-\alpha/2}/\sqrt{n}] < \exp(\mu) \\ &\quad < \exp(\bar{X}_n)\ [1 + z_{1-\alpha/2}/\sqrt{n}]) \\ &= \Phi(\sqrt{n}\log(1 + z_{1-\alpha/2}/\sqrt{n})) - \Phi(\sqrt{n}\log(1 - z_{1-\alpha/2}/\sqrt{n})). \end{aligned} \tag{4}$$

The simulation method described in Section 2 uses $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{D} N(0,1)$, so in the first and second steps of the algorithm, $Y_{n1}, \ldots, Y_{nB}$ are sampled from the $N(0,1)$ distribution and $g^*_{nb} = \exp(\bar{X}_n + n^{-1/2}Y_{nb})$ $(b = 1, \ldots, B)$ are calculated. To calculate MSE^*_n, note that, $g^*_{nb} \mid \bar{X}_n$ has a log-normal distribution with parameters $\bar{X}_n$ and $1/n$, so that $\mathrm{E}(g^*_{nb}|\bar{X}_n) = \exp(\bar{X}_n + 1/2n)$ and $\mathrm{var}(g^*_{nb}|\bar{X}_n) = \exp(2\bar{X}_n + 1/n)\{\exp(1/n) - 1\}$; therefore,

$$\begin{aligned} \mathrm{MSE}^*_n &= \{\mathrm{E}(g^*_{nb}|\bar{X}_n) - \exp(\bar{X}_n)\}^2 + \mathrm{var}(g^*_{nb}|\bar{X}_n) \\ &= \exp(2\bar{X})\{1 + \exp(2/n) - 2\exp(1/2n)\}. \end{aligned}$$

In practice, MSE^*_n is approximated using simulation, which for large B is close to the real value. The resulting confidence interval is $\exp(\bar{X}_n) \pm z_{1-\alpha/2}\exp(\bar{X}_n)\{1 + \exp(2/n) - 2\exp(1/2n)\}^{1/2}$ with corresponding coverage probabilities of

$$\begin{aligned} &\Phi(\sqrt{n}\log(1 + z_{1-\alpha/2}\sqrt{1 + e^{2/n} - 2e^{1/2n}})) \\ &\quad - \Phi(\sqrt{n}\log(1 - z_{1-\alpha/2}\sqrt{1 + e^{2/n} - 2e^{1/2n}})). \end{aligned} \tag{5}$$

One can check that $1 + e^{2/n} - 2e^{1/2n} > 1/n$, showing that the simulation interval is wider than the delta method interval in the current example.

Table 1 compares the coverage of the delta method given in Equation (4) to that of the simulation method provided in Equation (5) for several values of α and n. As expected, the coverage becomes more accurate as the sample size grows. The

Table 1. Coverage of intervals for exp(μ)

α	Method	$n = 10$	$n = 15$	$n = 20$	$n = 25$	$n = 50$	$n = 100$
0.01	Delta	0.970	0.976	0.979	0.981	0.985	0.988
	Simulation	0.978	0.981	0.982	0.984	0.987	0.988
0.05	Delta	0.935	0.940	0.943	0.944	0.947	0.949
	Simulation	0.949	0.950	0.951	0.951	0.951	0.951
0.10	Delta	0.897	0.899	0.899	0.899	0.900	0.900
	Simulation	0.918	0.915	0.912	0.910	0.905	0.903

delta method interval is less accurate when the confidence level required is larger. For example, if 99% level is required, then the probability of not covering the true parameter is almost twice than declared when the sample size is 25. The simulation interval is always wider than the delta method interval. For the most common 95% level it gives very good results, but for the 90% level the delta method performs better, unless one puts more weight on conservativeness than on accuracy.

Instead of calculating the variance in Step 3 of the algorithm, the quantiles of $g^*_{n1}, \dots, g^*_{nB}$ can be used as the limits of the confidence interval. Thus, the limits of this simulation quantile interval are the $[B\alpha/2]$ and $[B(1-\alpha/2)]$ order statistics of $g^*_{n1}, \dots, g^*_{nB}$, where $[a]$ is an integer value close to a. To explore the properties of this interval, assume that B is large so that the left limit is (approximately) the value C_l that satisfies

$$\alpha/2 = P(g^*_{n1} \le C_l \mid \hat{\theta}_n) = P(\exp(\bar{X}_n + n^{-1/2} Y_{n1}) \le C_l \mid \bar{X}_n).$$

Recalling that $Y_{n1} \sim N(0,1)$ is independent of $\bar{X}_n$, it is easy to solve this equation to obtain $C_l = \exp(\bar{X}_n - z_{1-\alpha/2}/\sqrt{n})$. Similar calculations for the right limit give $C_r = \exp(\bar{X}_n + z_{1-\alpha/2}/\sqrt{n})$. Thus, for large B, the quantile simulation interval $[C_l, C_r]$ equals (approximately) to the interval (3), which has an exact coverage of $1-\alpha$.

The normality assumption enables exact calculations of the confidence levels, but the discussion in this section holds, approximately, for any distribution with expectation μ and a unit standard deviation. As a concrete example, consider the case $X_i \sim \exp(\theta)$, with $\mu = 1/\theta$. Here, $\bar{X}_n \sim \text{Gamma}(n, \mu/n)$ having expectation μ and variance μ^2/n. The delta approximation gives $\sqrt{n}(e^{\bar{X}_n} - e^{\mu}) \xrightarrow{D} N(0, e^{2\mu}\mu^2)$, with corresponding interval $\exp(\bar{X}_n) \pm z_{1-\alpha/2}\bar{X}_n \exp(\bar{X}_n)/\sqrt{n}$. The simulation method samples $Y_{nb}|\bar{X}_n \sim N(0, \bar{X}_n^2)$, and g^*_{nb} has, conditionally on $\bar{X}_n$, a log-normal distribution with parameters $\bar{X}_n$ and $\bar{X}_n^2/n$, yielding the confidence interval $\exp(\bar{X}_n) \pm z_{1-\alpha/2}\exp(\bar{X}_n)\{1 + \exp(2\bar{X}_n^2/n) - 2\exp(\bar{X}_n/2n)\}^{1/2}$. Using 10^6 simulated averages for the case $\mu = 1$ and $n = 100$ and calculating 95% confidence intervals for e^{μ} in each, we obtained a coverage of 0.941 and 0.949 for the delta and simulation intervals, respectively.

3.2 Confidence Interval for the Correlation Coefficient

As a second example, consider the estimation of the correlation coefficient ρ using independent pairs (X_{1i}, X_{2i}), $i = 1, \dots, n$ from some bivariate distribution. This problem has been intensively discussed in the literature and we do not aim at studying it thoroughly. Instead, our aim is of demonstrating the close relationship between the delta method and the simulation algorithm in a way that may be relevant to other less familiar but similar examples.

Let $M = (\frac{1}{n}\sum X_{1i}, \frac{1}{n}\sum X_{2i}, \frac{1}{n}\sum X_{1i}^2, \frac{1}{n}\sum X_{2i}^2, \frac{1}{n}\sum X_{1i} X_{2i})$ denote the vector of first and second empirical moments, then $\sqrt{n}(M - E(M))$ has an asymptotic normal distribution and the empirical correlation coefficient $\hat{\rho}_n$ is a smooth function of M having asymptotic distribution $\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{D} N(0, (1-\rho^2)^2)$ (e.g., van der Vaart 1998). This result can be used to calculate the approximate confidence interval

$$\left[\hat{\rho}_n - z_{1-\alpha/2}\left(1-\hat{\rho}_n^2\right)/\sqrt{n},\ \hat{\rho}_n + z_{1-\alpha/2}\left(1-\hat{\rho}_n^2\right)/\sqrt{n}\right]. \quad (6)$$

However, (6) performs poorly for small datasets, and instead, the Fisher's z-transformation of the correlation coefficient, $z(\rho) = 0.5\log\frac{1+\rho}{1-\rho}$ is often used (see discussion in van der Vaart 1998). This transformation stabilizes the variance so that $\sqrt{n}(z(\hat{\rho}_n) - z(\rho)) \xrightarrow{D} N(0,1)$. A confidence interval for $z(\rho)$ is easily constructed by $[z^l, z^r] = [z(\hat{\rho}_n) - z_{1-\alpha/2}/\sqrt{n}\ ,\ z(\hat{\rho}_n) + z_{1-\alpha/2}/\sqrt{n}]$ and then inverted to provide confidence interval for ρ:

$$[(\exp(2z^l)-1)/(\exp(2z^l)+1), (\exp(2z^r)-1)/(\exp(2z^r)+1)]. \quad (7)$$

The simulation method in this problem reduces to sampling B realizations from the asymptotic law of M, calculating the correlation coefficient in each and estimating the variance using the simulated values. Table 2 compares the coverage probability of 95% intervals of the simulation method (using $B = 1000$) to that of the intervals (6) and (7). The table also reports the coverage of a simulation interval that uses the quantiles of the simulated values, as described in the previous section. The coverage probability is approximated by simulating 10,000 samples from a bivariate normal distribution with correlation ρ and calculating the proportion of intervals containing the true parameter.

For $n = 50$, the simulation method has slightly better coverage than the delta method before transformation, and slightly worse than the intervals based on Fisher's z-statistic; all intervals are somewhat anticonservative. For $n = 15$, the delta method intervals are anticonservative and the simulation intervals are conservative. For all sample sizes, the quantile simulation method seems to outperform all other methods. We remark that various corrections have been suggested for improving the small-sample accuracy of confidence intervals for this special, important case.

Table 2. Coverage of intervals for the correlation coefficient. Delta: delta method, Fisher: Fisher method, sim.MSE: simulation method, sim.q: simulation method based on quantiles

ρ	n	Delta	Fisher	sim.MSE	sim.q
0	15	0.885	0.925	0.964	0.945
	25	0.905	0.931	0.946	0.948
	50	0.931	0.944	0.938	0.948
0.2	15	0.876	0.919	0.962	0.944
	25	0.910	0.934	0.947	0.948
	50	0.933	0.945	0.941	0.949
0.4	15	0.878	0.919	0.964	0.946
	25	0.907	0.932	0.947	0.949
	50	0.927	0.941	0.938	0.947
0.6	15	0.879	0.920	0.975	0.946
	25	0.905	0.929	0.949	0.946
	50	0.927	0.939	0.942	0.948
0.8	15	0.874	0.920	0.985	0.944
	25	0.907	0.936	0.958	0.949
	50	0.927	0.939	0.940	0.948

3.3 Prediction in Markov Chains Models

The following example motivated our use of the simulation method. Let $Y_{it} \in \{1, \ldots, J\}$ be a categorical variable measured for subject i at time t, and let X_{it} be the corresponding vector of covariates. Consider a regression model of the form

$$P(Y_{it} = k | Y_{it-1} = j, X_{it} = x) = p_{j,k}(\theta|x),$$

where $\sum_{k=1}^{J} p_{j,k}(\theta|x) = 1$. This is a Markov transition model often used for analyzing longitudinal categorical data, see Diggle et al. (2002).

Let (Y_i, X_i), $i = 1, \ldots, n$, be a random sample of Markov chains with transition matrices $P_t(\theta|X_i)$ between time $t-1$ to t, where $Y_i = (Y_{i0}, \ldots, Y_{iT})$ is a vector containing the states of the ith chain at times $0, 1, \ldots, T$, X_i is a $p \times T$ matrix of covariates associated with the ith chain whose jth column contains the covariates for the jth transition, and θ is an unknown parameter vector indexing the model. Thus, each observation is a nonhomogeneous Markov chain with transitions determined by time-dependent covariates. The maximizer of the conditional likelihood (conditional on the covariate vectors and the initial states), $\mathcal{L}(\theta; Y_1, \ldots, Y_n | Y_{10}, \ldots, Y_{n0}, X_1, \ldots, X_n)$, has an asymptotic normal law with variance that can be easily estimated (Diggle et al. 2002).

Often, the parameter θ is not of main interest, but instead one is interested in derived quantities. For example, consider the parameter $g(\theta|x) = P_\theta(Y_{is} = k | Y_{i0} = j, X_i = x)$, which is the probability of occupying state k at discrete time s, conditionally on being in state j at time 0 and a matrix of covariates x. For example, if Y_{is} is an ordinal variable measuring the stage of a disease then $g(\theta|x) = P_\theta(Y_{is} = j+1 | Y_{i0} = j, X_i = x)$ is the probability of progression between times 0 and s, which is of clear importance for patients and doctors.

The parameter $g(\theta|x)$ is naturally estimated by the (j, k) entry of $\prod_{t=1}^{s} P_t(\hat{\theta}|x)$. If $P_t(\theta|x)$ is smooth in θ, which is most often the case, so is $g(\theta|x)$, and to apply the delta method, one should calculate

$$\frac{\partial}{\partial \theta_k} g(\theta|x) = \sum_{j=1}^{s} \left(\prod_{t=1}^{j-1} P_t(\theta|x) \right) \times \frac{\partial}{\partial \theta_k} P_j(\theta|x) \times \left(\prod_{t=j+1}^{s} P_t(\theta|x) \right).$$

This derivative requires a special, somewhat complicated, programming that can be eliminated if using the simulation method. The latter requires only an algorithm for generating Gaussian random variables and an algorithm for calculating $g(\theta|x)$, which is needed also for computation of the point estimate.

Here, we compare the delta and simulation methods using longitudinal semiannual disability data on multiple sclerosis patients with three states {no disability}, {minimal disability}, and {moderate disability} (Mandel et al. 2007). We use the same Markov partial proportional odds model as in the original article and calculate the time to sustained progression defined by being in a higher than baseline disability state for at least two consecutive visits. Figure 1 compares confidence intervals obtained by the standard delta method to those generated by the simulation method. The X-axis describes the visit number and the Y-axis is the probability of sustained worsening for a patient having no disability at baseline. The figure is based on the fitted model of Mandel et al. (2007) for the baseline covariates of an

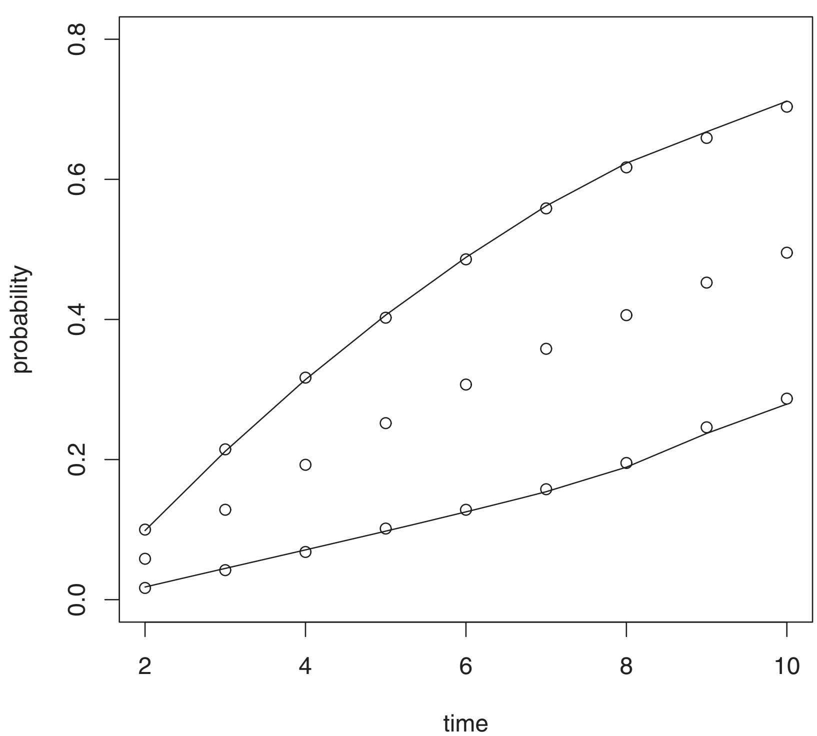

Figure 1. Probability of sustained progression with 95% confidence intervals based on the delta method (lines) and the simulation method (circles).

"average person": sex = female, brain parenchymal fraction = 0.88, lesion volume = 4.23, age = 35, and disease duration = 3 years. The last two covariates change in time making the model time nonhomogeneous.

The intervals obtained in the two methods are very similar with negligible differences, supporting the use of the simulation method, which is far more simple for programming. Typical to multiple sclerosis is the relatively slow rate of progression, where after 5 years, only about 50% of the patients deteriorate to a sustained minimal disability. It reflects the study population that comprises relapsing-remitting multiple sclerosis patients, including the subgroup of benign patients. The confidence intervals, however, become wider with time, reflecting an increase in uncertainty due to sampling.

3.4 Distribution Function of a Sum

The simulation method is most powerful in cases where the parameter of interest is not given in a simple closed form or when analytical calculation of it is difficult. As an example, consider a regression of a binary variable Y on a set of covariates $\mathbf{x} = (x_1, \ldots, x_m)$ according to the logistic model

$$P_\theta(Y = 1|\mathbf{x}) = \frac{\exp(\theta_1 x_1 + \cdots + \theta_m x_m)}{1 + \exp(\theta_1 x_1 + \cdots + \theta_m x_m)}. \tag{8}$$

Standard asymptotic arguments show that the maximum likelihood estimator satisfies (1) under mild regularity conditions. The maximum likelihood estimator and its covariance matrix are provided by most statistical software. The logistic regression (8) is one of the most popular statistical models and the delta method is sometimes used to estimate the variance of many derived functions of θ, such as $P_\theta(Y = 1|\mathbf{x})$ or $\exp(\theta_j)$.

Suppose that interest lies in the distribution function of the number of successes ($Y = 1$) in a certain population, for example, patients in a hospital where $\{Y = 1\}$ indicates mortality, or clients of an insurance company where $\{Y = 1\}$ indicates claim. Specifically, suppose that k-independent subjects with covariate vectors $\mathbf{x_1}, \ldots, \mathbf{x_k}$ are put on test and we are interested in estimating the distribution function of $\sum_{i=1}^k Y_i$, where $P(Y_i = 1|\mathbf{x}_i)$ is given by (8). This model is known as the Poisson–Binomial distribution (Chen and Liu 1997) and it is an example of a sum of independent but not identically distributed random variables. A similar, more complicated, problem where the parameter of interest is the distribution function of a random sum is studied by Mandel (2010).

Let $g_r(\theta) = P_\theta(\sum_{i=1}^k Y_i = r)$ be the function of interest:

$$g_r(\theta) = \sum_{S_r} \prod_{i=1}^{k} P_\theta(Y_i = y_i|\mathbf{x}_i),$$

where $S_r = \{(y_1, \ldots, y_k) : y_i = 0, 1;\ \sum y_i = r\}$. The function $g_r(\theta)$ is clearly differentiable and the delta method can be theoretically used to estimate its variance. However, it is difficult to calculate $g_r(\theta)$ by a direct sum as the set S_r may be very large. Special recursive algorithms have been suggested to calculate $g_r(\theta)$ (see Chen and Liu 1997 for more details), or standard Monte Carlo methods can be used to approximate it. These approaches can be applied in conjunction with the simulation method to obtain estimates for the variance without the need of evaluating the derivative.

4. SUMMARY

This article presents a simple simulation method for the estimation of the asymptotic variance of a transformation when the derivative exists but its computation is complicated. The standard bootstrap approach requires sampling many datasets and estimating θ for each, whereas the simulation method studied here replaces these steps by the simple generation of normal vectors, hence it is computationally less demanding.

The asymptotic approximation (1) is not guaranteed to be better than (2). In fact, one can apply the delta method from a vector version of (2) to (1) using g^{-1}, showing that the numbering of the equations is merely a matter of parameterization. For the simulation method to work well, one should start with an estimator that converges fast to the normal distribution (e.g., an average).

The quantile simulation method seems to perform very well in practice, though it typically requires a larger number of simulations than the variance-based approach. More study is required to evaluate the performance of quantile-based confidence intervals.

APPENDIX: JUSTIFICATION OF THE ALGORITHM

Let $\theta \in \mathbb{R}^m$ be a parameter that indexes a model, and denote by θ_0 the true parameter. Let $\hat\theta_n$ and $\hat\Sigma_n$ be sequences of estimators satisfying $\sqrt{n}(\hat\theta_n - \theta_0) \xrightarrow{D} N(0, \Sigma)$ and $\hat\Sigma_n \xrightarrow{p} \Sigma$, and let $Y_n \sim N(0, \hat\Sigma_n)$. Suppose that $g : \mathbb{R}^m \to \mathbb{R}$ has a continuous derivative in a neighborhood of θ_0. Expanding both $g(\hat\theta_n)$ and $g(\hat\theta_n + n^{-1/2}Y_n)$ around θ_0 gives

$$\begin{aligned} &\sqrt{n}[g(\hat\theta_n + n^{-1/2}Y_n) - g(\hat\theta_n)] \\ &\quad = \dot g(\theta_0)^T \sqrt{n}(\hat\theta_n + n^{-1/2}Y_n - \theta_0) + o_p(1) \\ &\qquad - [\dot g(\theta_0)^T \sqrt{n}(\hat\theta_n - \theta_0) + o_p(1)] \\ &\quad = \dot g(\theta_0)^T Y_n + o_p(1). \end{aligned} \tag{A.1}$$

Thus, the sequence on the left has the same limiting distribution as that of $\dot g(\theta_0)^T Y_n$, with a variance given by $\dot g(\theta_0)^T \Sigma \dot g(\theta_0)$. The suggested algorithm estimates this variance by calculating the second moment of $\sqrt{n}[g(\hat\theta_n + n^{-1/2}Y_n) - g(\hat\theta_n)]$, plugging-in $\hat\theta_n$ and $\hat\Sigma_n$.

In Step 1, the algorithm generates $Y_{n1}, \ldots, Y_{nB}$ from a zero mean normal distribution. Simulating normal variables is convenient, but other distributions can also be used. The key idea is to generate variables with zero expectation and covariance $\hat\Sigma_n$. For example, Step 1 can be replaced by generating vectors $Z_b = (Z_{b1}, \ldots, Z_{bm})^T$ of length m whose components are iid satisfying $P(Z_{bi} = -1) = P(Z_{bi} = 1) = 1/2$, and defining $Y_{nb} = \hat\Sigma_n^{1/2} Z_b$. Such an approach can be used to estimate the variance of $g(\theta) = 1/\theta$ or similar functions for which MSE_n^* is infinite when sampling normal vectors (see discussion at the end of Section 2).

[Received December 2011. Revised February 2013.]

REFERENCES

Aalen, O. O., Farewell, V. T., De Angelis, D., Day, N. E., and Gill, O. N. (1997), "A Markov Model for HIV Disease Progression Including the Effect of HIV Diagnosis and Treatment: Application to AIDS Prediction in England and Wales," *Statistics in Medicine*, 16, 2191–2210. [76]

Chen, S. X., and Liu, J. S. (1997), "Statistical Applications of the Poisson-Binomial and Conditional Bernoulli Distributions," *Statistica Sinica*, 7, 875–892. [80]

Diggle, P. J., Heagerty, P., Linag, K. Y., and Zeger, S. L. (2002), *Analysis of Longitudinal Data* (2nd ed.), Oxford: Oxford University Press. [79]

Drton, M., Caren, M., Guttorp, P., and Schaefer, J. T. (2003), "A Markov Chain Model of Tornadic Activity," *Monthly Weather Review*, 131, 2941–2953. [76]

Mandel, M. (2010), "Estimating Disease Progression Using Panel Data," *Biostatistics*, 11, 304–316. [76,80]

Mandel, M., and Betensky, R. A. (2008), "Estimating Time-to-Event From Longitudinal Ordinal Data Using Random-Effects Markov Models: Application to Multiple Sclerosis Progression," *Biostatistics*, 9, 750–764. [76]

Mandel, M., Gauthier, S. A., Guttmann, C. R. G., Weiner, H. L., and Betensky, R. A. (2007), "Estimating Time to Event From Longitudinal Categorical Data: an Analysis of Multiple Sclerosis Progression," *Journal of the American Statistical Association*, 102, 1254–1266. [76,79]

Singh, A. C., Stukel, D., and Pfeffermann, D. (1998), "Bayesian Versus Frequentist Measures of Error in Small Area Estimation," *Journal of the Royal Statistical Society,* Series B, 60, 377–396. [76]

Titman, A. C., Lancaster, G. A., Carmichael, K., and Scutt, D. (2011), "Accounting for Bias Due to a Non-Ignorable Tracing Mechanism in a Retrospective Breast Cancer Cohort Study," *Statistics in Medicine*, 30, 324–334. [76]

van den Hout, A., Jagger, C., and Matthews, F. E. (2009), "Estimating Life Expectancy in Health and Ill Health by Using a Hidden Markov Model," *Applied Statistics*, 58, 449–465. [76]

van den Hout, A., and Matthews, F. E. (2010), "Estimating Stroke-Free and Total Life Expectancy in the Presence of Non-Ignorable Missing Values," *Journal of the Royal Statistical Society,* Series A, 173, 331–349. [76]

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge: Cambridge University Press. [78]

Ver Hoef, J. M. (2012), "Who Invented the Delta Method?," *The American Statistician*, 66, 124–127. [76]

Whitaker, H. J., and Farrington, C. P. (2004), "Infections with Varying Contact Rates: Application to Varicella," *Biometrics*, 60, 615–623. [76]