The Delta Method in Arc

Sanford Weisberg

School of Statistics, University of Minnesota, St. Paul, MN 55108-6042.

Supported by National Science Foundation Grant DMS 0103983.

September 26, 2001

Abstract

This paper describes an **Arc** add-in for the delta method for regression models. The method works for linear models, generalized linear models, nonlinear models and also for generalized nonlinear models.

The *delta method* is a procedure for finding estimates and approximate standard errors of arbitrary functions of normally distributed random variables. As implemented in **Arc**, the delta method is used to find estimates and standard errors of nonlinear functions of parameter estimates.

1 Getting the delta method add-on

Download the file http://www.stat.umn.edu/arc/lindelta.lsp. Place it in the Extras directory in your Arc directory (if you don't have such a directory, create one). The file will be automatically loaded every time you start **Arc**. An additional item called "Delta method" will be added to regression menus, and this item is used to access the delta method code.

2 Usage

As an example of the usage of the delta method, consider the Lake Mary data discussed in Cook and Weisberg (1999, Chapter 1), relating $y = \mathit{Length}$ of fish to $x = \mathit{Age}$. It is suggested in the text that a reasonable approximation to the mean length at age over the range of ages in the data is a quadratic curve, $\mathrm{E}(\mathit{Length}|\mathit{Age}) = \eta_0 + \eta_1 \mathit{Age} + \eta_2 \mathit{Age}^2$ An important characteristic of the quadratic curve is the age at which the maximum length is attained (at least according to the quadratic fit). We can find the value of *Age* that maximizes *Length* by differentiating the mean function (2) with respect to *Age*, and set the result to zero. Straightforward calculus, given later in this report, gives $\mathit{AgeMax} = -\eta_1/2\eta_2$ as the *Age* with maximum expected length.

To estimate *AgeMax* using **Arc**, first fit the quadratic regression, given by

```
Data set = LakeMary, Name of Fit = L1
Normal Regression
Kernel mean function = Identity
Response      = Length
Terms         = (Age Age^2)
Coefficient Estimates
Label      Estimate        Std. Error    t-value    p-value
Constant   13.6224         11.0164        1.237     0.2201
Age        54.0493         6.48884        8.330     0.0000
Age^2     -4.71866         0.943959      -4.999     0.0000

R Squared:              0.801138
Sigma hat:               10.9061
Number of cases:              78
Degrees of freedom:           75

Summary Analysis of Variance Table
Source          df       SS             MS           F    p-value
Regression       2    35938.        17969.     151.07    0.0000
Residual        75    8920.7       118.943
 Lack of fit     3   108.012        36.004       0.29    0.8295
 Pure Error     72   8812.68       122.398
```

The parameter of interest is $-\hat{\eta}_1/(2 * \hat{\eta}_2) = 54.0493/(2 * -4.71866) = 5.727$ years of age. This value and the corresponding standard error can be obtained using the "Delta method" item in the regression menu. This item gives a dialog box that shows the names of the parameters that correspond to each of the terms in the regression model. In the text area in the dialog, enter a mathematical expression for the function of interest, which in this case is `Eta1/(2*Eta2)`. The resulting output is

```
Data set = LakeMary, Name of Fit = L1
Normal Regression
Kernel mean function = Identity
Response      = Length
Terms         = (Age Age^2)
Function of parameters: -Eta1/(2*Eta2)
Estimate = 5.72718, with se = 0.493826
```

which gives the same estimate presented before, but also gives the approximate standard error of about 0.49 years. Assuming that the sample size is large enough, this standard error can be the basis of inference statements concerning the function of interest. For normal linear and nonlinear models, the t distribution can be used for inference, while in other regression problems, the normal distribution should be used.

In the remainder of this report, we give some details on how the delta method works.

3 Details

Let θ be a $k \times 1$ parameter vector, with estimator $\hat{\theta}$ such that

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, n\sigma^2 D)$$

where D is positive definite. This distributional result can be exact or asymptotically valid, as in nonlinear models. In some cases, like many generalized linear models, σ^2 may be known, usually equal to one. Estimating a function $g(\theta)$ is easy, as long as g is a *linear* function of θ. In some problems, we may wish to obtain an estimate for a nonlinear function of the parameters. This can be done asymptotically using a general method called the δ-method.

Suppose $g(\theta)$ is a nonlinear continuous function of θ. Expanding in Taylor series about the true value θ^*,

$$\begin{aligned} g(\hat{\theta}) &= g(\theta^*) + \sum_{j=1}^{p} \frac{\partial g}{\partial \theta_j}(\hat{\theta}_j - \theta_j^*) + o(\| \hat{\theta} - \theta^* \|) \\ &= g(\hat{\theta}) + g'(\theta^*)^T(\hat{\theta} - \theta^*) + o(\| \hat{\theta} - \theta^* \|) \end{aligned}$$

where we have defined

$$g'(\theta^*) = \frac{\partial g}{\partial \theta} = \left(\frac{\partial g}{\partial \theta_1}, \ldots, \frac{\partial g}{\partial \theta_k} \right)^T$$

evaluated at θ^*. The vector g' has dimension $k \times 1$. Rearranging terms,

$$g(\hat{\theta}) - g(\theta^*) = g'(\theta^*)^T(\hat{\theta} - \theta^*) + o(\| \hat{\theta} - \theta^* \|)$$

We will estimate $g(\theta)$ by $g(\hat{\theta})$. If $\hat{\theta}$ is a maximum likelihood estimate, then $g(\hat{\theta})$ is the mle of $\widehat{g(\theta)}$. Taking the variance of both sides,

$$\text{var}(g(\hat{\theta}) - g(\theta^*)) = g'(\theta^*)^T \text{var}(\hat{\theta}) g'(\theta^*) \tag{1}$$

where $\text{var}(\hat{\theta}) = \sigma^2 D$. This equation is the heart of the delta method, so I'll write it out again as a scalar equation. Let g'_i be the i-th element of $g'(\hat{\theta})$, and let v_{ij} be the ij-element of the matrix $\hat{\sigma}^2 D$. Then the estimated variance of $g(\hat{\theta})$ is

$$\text{var}(g(\hat{\theta})) = \sum_{i=1}^{k} \sum_{j=1}^{k} g'_i g'_j v_{ij} \tag{2}$$

In large samples and under regularity conditions, $g(\hat{\theta})$ will converge to $g(\theta)$, and

$$\sqrt{n}(g(\hat{\theta}) - g(\theta^*)) \to N(0, n\sigma^2[g'(\theta^*)^T D g'(\theta^*)])$$

In practice, all derivatives are evaluated at $\hat{\theta}$.

3.1 Example: Maximum of a quadratic function.

Consider fitting a quadratic regression

$$y = \eta_0 + \eta_1 x + \eta_2 x^2 + \varepsilon$$

The value of x at the maximum of the quadratic curve is the solution to

$$\frac{dy}{dx} = \eta_1 + 2\eta_2 x = 0$$

so the maximum occurs at $-\eta_1/2\eta_2$. We may well be interested in finding this particular value of x, so define

$$g(\eta) = -\eta_1/2\eta_2$$

which is estimated by $g(\hat{\eta})$. To compute the asymptotic variance of the estimate, we need the partial derivative, evaluated at $\hat{\eta}$,

$$\left(\frac{\partial g}{\partial \eta}\right)^T = (0, -\frac{1}{2\hat{\eta}_2}, \frac{\hat{\eta}_1}{2\hat{\eta}_2^2})$$

Using (1), straightforward calculation gives

$$\text{var}\widehat{(g(\eta))} = \frac{1}{4\hat{\eta}_2^2}\left[\text{var}(\hat{\eta}_1) + \frac{\hat{\eta}_1^2}{\hat{\eta}_2^2}\text{var}(\hat{\eta}_2) - \frac{2\hat{\eta}_1}{\hat{\eta}_2}\text{cov}(\hat{\eta}_1, \hat{\eta}_2)\right]$$

In practice, the variances and covariances are estimated from the corresponding elements of $\hat{\sigma}^2 D = \hat{\sigma}^2 (X^T X)^{-1}$, and X has columns 1, x, and x^2.

3.2 The LD50 for logistic regression

In logistic regression with logit$(\theta(x)) = \eta_0 + \eta_1 x$, the LD50, the point where the probability of success is 0.5, is easily shown to be estimated by $\hat{\eta}_0/\hat{\eta}_1$. If the estimated variance covariance matrix of the estimates has elements $v_{ij}, i, j = 1, 2$, then it is easy to show that

$$\text{var}\widehat{(\hat{\eta}_0/}\hat{\eta}_1) = \frac{1}{\hat{\eta}_1^2}v_{11} + 2\frac{\hat{\eta}_0}{\hat{\eta}_1}v_{12} + \hat{\eta}_0^2 v_{22}$$

4 References

R. D. Cook and S. Weisberg (1999). *Applied Regression Including Computing and Graphics*. New York: Wiley.

C. R. Rao (1965). *Linear Statistical Inference and Its Applications*. New York: Wiley, p. 321, 327.