Delta Method

The delta method is really a theorem which states that a smooth function of an asymptotically normal estimator is also asymptotically normally distributed (*see* **Large-sample Theory**). The result is applied in numerous contexts for the computation of large-sample tests and confidence limits for nonlinear functions of parameters which have already been estimated. Typically, the method of **estimation** is a standard large-sample technique which cannot be directly applied to the problem of interest.

Let $\hat{\theta}_n$ denote a sequence of estimates of some parameter θ, such that

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathrm{L}} \mathrm{N}[0, \sigma^2(\theta)] \qquad (1)$$

(*see* **Convergence in Distribution and in Probability**). For example, $\hat{\theta}_n = \theta(X_1, X_2, \ldots, X_n)$, where $X_1, \ldots, X_n$ is a random sample from a distribution F_θ, the simplest cases being $\hat{\theta}_n = \overline{X}_n$, the sample **mean**, with $\theta = \mu_x$, or $\hat{\theta}_n = s_n^2$, the sample **variance**, with $\theta = \sigma_x^2$. Let g be a function which is differentiable in a neighborhood of the true value θ with $g'(\theta) \neq 0$. Then the delta method states that

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{\mathrm{L}} \mathrm{N}\{0, \sigma^2(\theta)[g'(\theta)]^2\}. \qquad (2)$$

The result follows from (1) using standard convergence theorems [14, 2c.4]. Briefly, substitution of $\hat{\theta}_n$ into a Taylor expansion for g about θ yields

$$g(\hat{\theta}_n) = g(\theta) + (\hat{\theta}_n - \theta)g'(\theta) + (\hat{\theta}_n - \theta)o_p(1),$$

where $o_p(1) \xrightarrow{\mathrm{p}} 0$. Then

$$\begin{aligned}\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) - \sqrt{n}(\hat{\theta}_n - \theta)g'(\theta)\\ = \sqrt{n}(\hat{\theta}_n - \theta)o_p(1) = o_p(1).\end{aligned}$$

In fact, a slightly stronger result is needed, in which we assume that g' and σ^2 are continuous at $\theta, \sigma(\hat{\theta}_n)|g'(\hat{\theta}_n)|$ is used to standardize (2), and convergence is to a standard normal distribution [14, 6a.2; 2, 12.1.2]. In practice, the **standard error** (se) of $\hat{\theta}_n$ is taken to be $\sigma/\sqrt{n}$, and the se of $g(\hat{\theta}_n)$ is $|g'(\theta)|\sigma(\theta)/\sqrt{n}$, with θ estimated by $\hat{\theta}_n$. Thus the delta method can also be viewed as a technique for approximating the mean and variance of a function of a random variable, $g(T)$ [10]. For this approximation to be valid we must have $\mathrm{var}(T) = O(1/n)$. The Taylor expansion can also be used to provide a bias correction [8, 8.4(iii)].

In applications, a multivariate version of the theorem is typically needed. Suppose that $\hat{\theta}_n = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ is an asymptotically [in the sense of (1)] multivariate normal random vector with asymptotic mean θ and variance matrix $\Sigma(\theta)$. Let $\mathbf{g}(\theta) = [g_1(\theta), \ldots, g_q(\theta)]'$ and $\partial\mathbf{g}/\partial\theta$ denote the $q \times k$ matrix of partial derivatives. Then, if for each i, $\partial g_i/\partial\theta_j \neq 0$ for some j, we have

$$\sqrt{n}[\mathbf{g}(\hat{\theta}_n) - \mathbf{g}(\theta)] \xrightarrow{\mathrm{L}} \mathrm{N}\left(\mathbf{0}, \frac{\boldsymbol{\partial}\mathbf{g}}{\boldsymbol{\partial\theta}} \sum \frac{\boldsymbol{\partial}\mathbf{g}'}{\boldsymbol{\partial\theta}}\right). \qquad (3)$$

The proof involves either a multivariate Taylor expansion [1, Appendix C; [4], 14.6.3] or, alternatively, an application of (2) using an arbitrary linear combination of the coordinates of $\mathbf{g}(\hat{\theta}_n)$ and well-known characterizations of the **multivariate normal distribution** and of convergence in distribution of random vectors [14, 6a.2]. In most applications $q = 1$.

The delta method is closely related to the method of **maximum likelihood**. As is well known, if $\hat{\theta}$ is the maximum likelihood estimator (MLE) of θ and $\mathbf{g}$ is a one-to-one differentiable transformation, then $\hat{\phi} = \mathbf{g}(\hat{\theta})$ is the MLE of $\phi = \mathbf{g}(\theta)$. A Taylor expansion of the score vector (*see* **Likelihood**) may also be used to show that the **information matrix** becomes [6; [8], Exercise 4.15]

$$I(\phi) = \frac{\partial\theta'}{\partial\phi} I(\theta) \frac{\partial\theta}{\partial\phi}.$$

Thus the asymptotic distribution of $\hat{\phi}$ obtained from the delta method is the same as that of the MLE. Indeed, in single parameter **exponential families**, where there is a **sufficient statistic** for the natural parameter (such as the **multinomial distribution**) the method of maximum likelihood can be regarded as an application of the delta method using the transformation implicitly defined by the score equations [5, 3, 2, 12.2.1]

Two classes of applications may be distinguished. The first involves choosing a transformation $g(\theta)$ so that var[$g(\hat{\theta})$] = constant, in which case g is known as a variance-stabilizing transformation. From the delta method,

$$g(x) = c \int \frac{\mathrm{d}\theta}{\sigma(\theta)}.$$

Encyclopedia of Biostatistics, Online © 2005 John Wiley & Sons, Ltd.

This article was published in the *Encyclopedia of Biostatistics* in 2005 by John Wiley & Sons, Ltd.
DOI: 10.1002/0470011815.b2a15029

Examples include the angular (arc sin square root) transformation for a binomial proportion, and Fisher's z transformation for the sample correlation coefficient, r, [14, 6g.4]. The multivariate delta method can be used to establish the asymptotic normality of r [15, Chapter 3]; a similar technique can be used to show the asymptotic normality of the sample variance, s^2. Another general class of examples is given by the ladder of powers, $\mathrm{var}(X) = c^2\mu^{2\alpha}$ for $\alpha > 0$ (*see* **Power Transformations**). In this case

$$g(x) = \frac{1}{c(1-\alpha)}\mu^{1-\alpha},$$

with the understanding that $\alpha = 1$ (constant coefficient of variation) gives $g(x) = \log x$. Other special cases include the square root transformation ($\alpha = 1/2$), which is used with the **Poisson distribution**. Strictly speaking, if $X_n \sim$ Poisson $(n\theta)$, then the delta method must be applied to the sequence $\hat{\theta}_n = X_n/n$ [4, Example 14.6-3; [15], p. 121].

The second class of applications involves nonlinear functions of parameters for which large-sample estimates can easily be obtained, often from the **central limit theorem**, or from a **generalized linear model**. The simplest example is provided by the asymptotic normality of the sample **standard deviation** $s = \sqrt{s^2}$. The standard example is the variance of a ratio:

$$\mathrm{var}\left(\frac{T_1}{T_2}\right) \approx \left[\frac{\mathrm{E}(T_1)}{\mathrm{E}(T_2)}\right]^2 \left\{\frac{\mathrm{var}(T_1)}{[\mathrm{E}(T_1)]^2} - \frac{2\mathrm{cov}(T_1, T_2)}{\mathrm{E}(T_1)\mathrm{E}(T_2)} + \frac{\mathrm{var}(T_2)}{[\mathrm{E}(T_2)]^2}\right\}.$$

An application is the estimation of the dose corresponding to a given frequency of response in the analysis of **quantal response** data in toxicology [13, 2.7.1]. Related applications include the variance of the log of the **relative risk**, $rr = p_1/p_2$, and the log **odds ratio** (the logarithms of these quantities being more nearly normally distributed) in epidemiology [11, 15.5]. For example,

$$\mathrm{var}[\log(rr)] \approx \frac{1-p_1}{n_1 p_1} + \frac{1-p_2}{n_2 p_2}.$$

Another interesting class of examples involves the calculation of large-sample standard deviations for various measures of **association** in two-way contingency tables. These have the form $\zeta = \nu(\pi_{ij})/\delta(\pi_{ij})$, where $\nu(\pi_{ij})$ and $\delta(\pi_{ij})$ are known functions of the population proportions [1, 10.3; [4], 11.3]. A special case is the measurement of **agreement** between two raters on a categorical scale. The standard measure of interrater reliability is the observed proportion of agreement corrected for chance, known as **kappa**. A large sample standard deviation may be computed using the delta method [4, 11.4; [9], 13.1].

In multinomial regression models, the delta method can be used to establish the asymptotic distribution of the predicted cell probabilities and residuals, typically standardized cell **residuals** [2, 12.2–12.3]. Cox & Ma [7] used a similar application to develop confidence bands for generalized **nonlinear regression** models. For extensions of the basic result ($g'(\theta) = 0$ and $\sigma_n \to 0$ instead of $\sigma/\sqrt{(n)} \to 0$), see [15, Chapter 3].

Lehman [15, Section 6.3] considers the extension to limit distributions of statistical functionals.

References

[1] Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. Wiley, New York.

[2] Agresti, A. (1990). *Categorical Data Analysis*. Wiley, New York.

[3] Benichou, J. & Gail, M.H. (1989). A delta method for implicitly defined random variables, *American Statistician* **43**, 41–44.

[4] Bishop, Y.M.M., Fienberg, S.E. & Holland, P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge, Mass.

[5] Cox, C. (1984). An elementary introduction to maximum likelihood estimation for multinomial models: Birch's theorem and the delta method, *American Statistician* **38**, 283–287.

[6] Cox, C. (1990). Fieller's theorem, the likelihood, and the delta method, *Biometrics* **46**, 709–718.

[7] Cox, C. & Ma, G. (1995). Asymptotic confidence bands for generalized nonlinear regression models, *Biometrics* **51**, 142–150.

[8] Cox, D.R. & Hinkley, D.V. (1974). *Theoretical Statistics*. Chapman & Hall, London.

[9] Fleiss, J.L. (1981). *Statistical Methods for Rates and Proportions*, 2nd Ed. Wiley, New York.

[10] Johnson, N.L. & Kotz, S. (1988). *Encyclopedia of Statistical Sciences*, Vol. 8. Wiley, New York, pp. 646–647.

[11] Kleinbaum, D.G., Kupper, L.L. & Morgenstern, H. (1982). *Epidemiologic Research*. Lifetime Learning Publications, Belmont.

[12] Lehman, E.L. (1999). *Elements of Large Sample Theory*. Springer-Verlag, New York.

This article was published in the *Encyclopedia of Biostatistics* in 2005 by John Wiley & Sons, Ltd.
DOI: 10.1002/0470011815.b2a15029

[13] Morgan, B.J.T. (1992). *Analysis of Quantal Response Data*. Chapman & Hall, London.
[14] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, 2nd Ed. Wiley, New York.
[15] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

C. Cox

DOI: 10.1002/0470011815.b2a15029