A Note on the Delta Method
Author(s): Gary W. Oehlert
Source: *The American Statistician*, Vol. 46, No. 1 (Feb., 1992), pp. 27-29
Published by: American Statistical Association
Stable URL: http://www.jstor.org/stable/2684406
Accessed: 24/09/2009 13:26

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=astata.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

American Statistical Association is collaborating with JSTOR to digitize, preserve and extend access to *The American Statistician.*

http://www.jstor.org

A Note on the Delta Method

GARY W. OEHLERT*

The delta method is an intuitive technique for approximating the moments of functions of random variables. This note reviews the delta method and conditions under which delta-method approximate moments are accurate.

KEY WORDS: Approximate moments; Asymptotic approximation; Taylor series.

For linear functions g of a random variable X, we have that

$$E[g(X)] = g(E[X]) \tag{1}$$

(assuming that X has a finite expectation). A common mistake for beginning statistics students is to assume that Equation (1) holds for all functions g. This is not true, of course, but students quickly learn that exact computation of $E[g(X)]$ is often possible only for the simplest functions g and random variables X. The *delta method* is a technique for approximating expected values of functions of random variables when direct evaluation of the expectation is not feasible.

The delta method approximates the expectation of $g(X)$ by taking the expectation of a polynomial approximation to $g(X)$. The polynomial is usually a truncated Taylor series centered at the mean of X. This kind of approximation is intuitively appealing and often gives good results. However, good results are not guaranteed. Section 1 of this note reviews some of the conditions under which the delta method will give accurate results. We conclude with some examples in Section 2.

We note that the term *delta method* is also used for a related technique, wherein we compute the moments of an approximating asymptotic distribution. See, for example, Rao (1965 p. 319) or Bishop, Fienberg, and Holland (1975 p. 486). In this note, however, we are concerned with approximating the moments of a random variable. This is not the same as finding the moments of an approximating distribution.

The delta method approximates the expected value of a function by the expected value of an approximation to the function. Alternatively, we could approximate the expected value by approximating the distribution with respect to which the expectation is taken. This second approach is the subject of asymptotic expansions [see Bhattacharya and Rao (1976) for a thorough treatment of this subject]. The results given here can be proved using asymptotic expansion techiques (see Oehlert 1981), but the proofs are not as straightforward as those given here.

*Gary W. Oehlert is Associate Professor, Department of Applied Statistics, University of Minnesota, St. Paul, 55108. The author thanks two anonymous referees for comments that improved the readability of this article.

1. THE DELTA METHOD

We begin with some notation for observations and moments. We have X, a random variable with distribution F, and $x_1, x_2, \ldots, x_n$, an iid sample from F. Let μ_j be the jth population moment of X, $\mu_j = E[X^j]$, where the mean may be denoted by μ_1 or just μ. Let $\overline{\mu}_j$ be the jth population central moment of X, $\overline{\mu}_j = E[(X - \mu)^j]$. Denote the jth sample moment by $m_{j,n}$, $m_{j,n} = \Sigma_{i=1}^n x_i^j/n$, and the jth sample central moment by $\overline{m}_{j,n}$, $\overline{m}_{j,n} = \Sigma_{i=1}^n (x_i - \bar{x})^j/n$. The second subscript showing the sample size may be dropped from time to time when no confusion will result.

The first fairly rigorous delta method appeared in Cramér (1946 p. 353). Cramér stated the result for $g(\overline{m}_{i,n}, \overline{m}_{j,n})$, a function of two sample central moments that depends on the sample size n only through the sample moments themselves, but Cramér noted that the same proof can be used for functions of any number of central moments or of $\bar{x}$. Cramér assumed that g is twice continuously differentiable in a neighborhood of the population moments $\overline{\mu}_i$ and $\overline{\mu}_j$ and bounded by Cn^p for positive constants C and p. Cramér did not list explicit assumptions about X, but earlier in the chapter he stated that X will be assumed to have enough finite moments so that the formulas are correct. Cramér's conclusion was that

$$E[g(\overline{m}_{i,n}, \overline{m}_{j,n})] = g(\overline{\mu}_i, \overline{\mu}_j) + O(n^{-1}). \tag{2}$$

Thus, up to an error of order $1/n$, the order of the function g and the expectation operator E may be interchanged.

Hurt (1976) extended Cramér's results by (a) allowing more general random variables as arguments to g, (b) allowing the function g to depend on n, and (c) taking more terms in the Taylor series approximation of g. Hurt considered a sequence of random variables W_n consistent for w such that

$$E|W_n - w|^{2(q+1)} = O(n^{-(q+1)}).$$

Note, in particular, that this is true for sample moments $m_{j,n}$ provided that X has $2j(q + 1)$ finite moments (Loève 1977 p. 276):

$$E|m_{j,n} - \mu_j|^{2(q+1)} = O(n^{-(q+1)}).$$

The function g is allowed to depend on the sample size n but must still be smooth and bounded. Hurt assumed that $g(n, W_n)$ is $q + 1$ times differentiable with respect to W_n in an interval around w, that g is bounded, and that the first $q + 1$ derivatives of g are bounded in a neighborhood of w. Hurt's conclusion was that

$$E[g(n, W_n)] = g(n, w) + \sum_{j=1}^{q} \frac{g^{(j)}(n, w)}{j!} E(W_n - w)^j + O(n^{-(q+1)/2}). \tag{3}$$

© 1992 American Statistical Association

Cramér's result corresponds to the case where $q = 1$ and g does not depend on n.

Lehmann (1983 p. 106, Theorem 5.1) gave the special case of Hurt's result where $W_n = \bar{x}$, $q = 3$, and g does not depend on n. (Theorem 5.1b on page 109 allows a limited form of dependence on n.)

All the results to this point are limited in that they are restricted to bounded functions. While some bounded functions are of interest, there are many unbounded functions, for example, squared error loss functions, for which we would like to be able to compute approximate expectations. Lehmann (1983 p. 109, Theorem 5.1b) showed that

$$E[g(\bar{x})] = g(\mu) + g^{(2)}(\mu)\sigma^2/(2n) + O(n^{-2}) \qquad (4)$$

when g has k derivatives ($k \geq 3$), the kth derivative is bounded, and X has at least k finite moments. Such a g need not be bounded, but it will be bounded by a polynomial in its argument. This polynomial bounding is the key. [Withers (1987, App. E) gives a result analogous to (4). However, his theorem as stated is incorrect, because it omits any bounding conditions on the function g or its derivatives. See the examples.]

The following theorem extends the previous delta-method theorems. The extensions are that the approximating polynomial need not be a truncated Taylor series (though it generally will be) and that the polynomial bounding is put directly on the approximation error. We prove the theorem for functions in the normalized sample moments $u_{j,n} = \Sigma_{i=1}^{n} (x_i^j - \mu_j)/\sqrt{n}$, but it can be extended to other random variables as well.

We will use the following notation for polynomials in the first J normalized sample moments. Let $\mathbf{p} = (p_1, p_2, \ldots, p_J)'$ be a vector of powers, and let $\mathbf{u}^{\mathbf{p}} = \mathbf{u}_n^{\mathbf{p}} = u_{1,n}^{p_1} u_{2,n}^{p_2} \cdots u_{J,n}^{p_J}$. A polynomial in the u's is a linear combination of a finite number of the $\mathbf{u}^{\mathbf{p}}$. The sets P_A and P_B are finite sets of powers that define the approximating and bounding polynomials.

Theorem. Let the random variables $u_{i,n}$ be the normalized sample moments of an iid sample of size n from a distribution with finite tth moment. Suppose that there are approximating and bounding polynomials

$$A_n(\mathbf{u}_n) = \sum_{\mathbf{p} \in P_A} a_{n,\mathbf{p}} \mathbf{u}^{\mathbf{p}},$$

and

$$B(\mathbf{u}_n) = \sum_{\mathbf{p} \in P_B} b_{\mathbf{p}} \mathbf{u}^{\mathbf{p}}$$

such that

$$n^{\beta}|g(n, \mathbf{u}_n) - A_n(\mathbf{u}_n)| \xrightarrow{P} 0$$

and

$$n^{\beta}|g(n, \mathbf{u}_n) - A_n(\mathbf{u}_n)| \leq B(\mathbf{u}_n)$$

for all n sufficiently large. If $t > 2J$ and $t > \max_{\mathbf{p} \in P_B \cup P_A} \Sigma_{j=1}^{J} jp_j$, then

$$n^{\beta}E|g(n, \mathbf{u}_n) - A_n(\mathbf{u}_n)| \rightarrow 0,$$

and consequently,

$$E[g(n, \mathbf{u}_n)] = E[A_n(\mathbf{u}_n)] + o(n^{-\beta}).$$

Proof. The proof is straightforward, given the fact that if a sequence of random variables converges to zero in probability and the sequence of $(1 + \epsilon)$th absolute moments is bounded for some ϵ greater than zero, then the sequence of random variables converges to zero in L_1 (see Loève 1977, p. 166). By the Minkowski inequality, we only need to show that $E|\mathbf{u}^{\mathbf{p}}|^{1+\epsilon}$ is bounded over n for some positive ϵ and all $\mathbf{p} \in P_B$. Taking the absolute values inside the power, it suffices to bound $E[|u_{1,n}|^{p_1(1+\epsilon)}|u_{2,n}|^{p_2(1+\epsilon)} \cdots |u_{J,n}|^{p_J(1+\epsilon)}]$. This reduces (by Holder's inequality) to showing that $E[|u_j|^{(1+\epsilon)\Sigma_{k=1}^{J} kp_k/j}]$ is bounded over n for $1 \leq j \leq J$. By our second assumption on t, it suffices to show that $E[|u_j|^{t/j}]$ is bounded over n. Loève (1977, p. 276) showed that if $y_1, y_2, \ldots, y_n$ are iid with $E[y_i] = 0$ and $E[|y_i|^s]$ finite for $s > 2$, then $E|\Sigma_{i=1}^{n} y_i|^s \leq cn^{s/2}E|y_i|^s$, for some constant c. Identify y_i with $(x_i^j - \mu_j)$ and s with t/j; then, by our first assumption on t and the Loève result, we have that $E[|u_j|^{t/j}]$ is bounded over n and the L_1 convergence is proved. Our second assumption on t insures that the expected value of $A_n(\mathbf{u}_n)$ exists, and the theorem is proved.

The basic point of the theorem is that when the function we are approximating is polynomially bounded in the random variables, then the naive Taylor series approximation will yield the correct asymptotic approximation to the expected value of the function provided only that the underlying sequences of random variables have enough bounded moments.

2. EXAMPLES

Example 1: Computation of the mean squared error of an adaptive shrinkage estimator of the mean. Let $\hat{\mu} = \bar{x} \times \bar{x}^2/(\bar{x}^2 + s^2/n)$, where $\bar{x}$ and s^2 are the sample mean and variance. This estimator was introduced by Thompson (1968); variants and improvements have been studied by a number of authors, including Mehta and Srinivasan (1971) and Oehlert (1981). The error $\hat{\mu} - \mu$ may be expanded, after some algebra, as

$$\begin{aligned}\hat{\mu} - \mu = {} & \frac{u_1}{n^{1/2}} - \frac{n^{1/2}\sigma^2 - \sigma^2 u_1/\mu + u_2 - 2\mu u_1}{n^{3/2}\mu} \\ & + \frac{u_1^2(\mu - u_1/n^{1/2}) + u_1(u_2 - 2\mu u_1)}{n^2\mu^2} \\ & - \frac{\bar{x}s^2}{n(n\bar{x}^2 + s^2)}(u_1^2 + s^2/\mu^2) + \frac{2s^4}{\mu^3 n^2} \\ & - \frac{2s^4\left(\dfrac{\bar{x} + \mu}{\mu^2} u_1 + \dfrac{s^2}{n^{1/2}\mu^2}\right)}{n^{3/2}\mu(n\bar{x}^2 + s^2)},\end{aligned} \qquad (5)$$

where u_1 and u_2 are the normalized sample moments defined in the preceding section. We are interested in the mean squared error of $\hat{\mu}$, so our g function is the squared error (SE) $(\hat{\mu} - \mu)^2$, or equivalently, the right side of (5) squared. Let T be the first two terms on the right side of (5); our approximating polynomial A_n will be T^2. This approximate squared error satisfies $n^2|\text{SE} - A_n| \xrightarrow{P} 0$. The difference $n^2|\text{SE} - T^2|$ is equal to $n^{1/2}(\hat{\mu} - \mu + T)n^{3/2}(\hat{\mu}$

$- \mu - T)$; each factor in this product can be polynomially bounded by noting that terms of the form $ab/(a^2 + b)$ (where b is positive) are bounded. The resulting bounding polynomial has largest order term u_2^4, so if $E|X|^{8+\delta}$ is finite for some positive δ, then the mean squared error of $\hat{\mu}$ can be expressed as

$$\mathrm{MSE}(\hat{\mu}) = \frac{\sigma^2}{n} + \frac{1}{n^2\mu^2}(3\sigma^4 - 2\mu\bar{\mu}_3) + o(n^{-2}).$$

The remaining examples give some idea of what can happen when the conditions of the theorem are not met.

Example 2: Convergence in probability is not sufficient to get convergence in mean without further conditions. Let $x_1, x_2, \ldots, x_n$ be iid unit normal random variables, take $g(n, \mathbf{u}_n)$ to be $\exp(n\bar{x}^2/2)/[\sqrt{n}(1 + \bar{x}^2)]$, and let $A_n = 1/\sqrt{n}$ (the first term in the Taylor series expansion of g) be the approximating polynomial. Now, both g and $|g - A_n|$ converge in probability to zero, yet g has constant expectation equal to $\sqrt{\pi/2}$. Thus the polynomial boundedness condition on g cannot be disregarded with impunity.

Example 3: The polynomial boundedness condition is not necessary. For the iid unit normals in Example 2, let $g(n, \mathbf{u}_n) = \exp(\bar{x}) = \exp(u_1/\sqrt{n})$. Using facts about the lognormal distribution, we have that

$$E[\exp(\bar{x})] = \exp(1/(2n)) = 1 + \frac{1}{2n} + o(n^{-1}),$$

which agrees with the delta-method approximation based on the truncated Taylor series $A_n = 1 + u_1/\sqrt{n} + u_1^2/(2n)$, even though the function g is not polynomially bounded.

Example 4: The delta method approximation can fail to hold when the moment conditions are not met. Suppose that $x_1, x_2, \ldots, x_n$ are iid observations from the distribution with density $f(x) = (3/x^4)I[x \geq 1]$ (which has mean 1.5 and variance .75 but no moments of order three or higher). Apply approximation (4) to the function $g(\bar{x}) = \bar{x}^3$ without checking the moment conditions (which are not met!). We obtain the approximate expectation $\mu^3 + 3\sigma^2\mu/n$, but the exact expectation is infinite.

[*Received June 1990. Revised November 1990.*]

REFERENCES

Bhattacharya, R. N., and Rao, R. R. (1976), *Normal Approximation and Asymptotic Expansions*, New York: John Wiley.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975), *Discrete Multivariate Analysis*, Cambridge, MA: MIT Press.

Cramér, H. (1946), *Mathematical Methods of Statistics*, Princeton, NJ: Princeton University Press.

Hurt, J. (1976), "Asymptotic Expansions of Functions of Statistics," *Aplikace Matematiky*, 21, 444–456.

Lehmann, E. L. (1983), *Theory of Point Estimation*, New York: John Wiley.

Loève, M. (1977), *Probability Theory I* (4th ed.), New York: Springer-Verlag.

Mehta, J. S., and Srinivasan, R. (1971), "Estimation of the Mean by Shrinkage to a Point," *Journal of the American Statistical Association*, 66, 86–90.

Oehlert, G. W. (1981), "Estimating the Mean of A Positive Random Variable," unpublished Ph.D. dissertation, Yale University, Dept. of Statistics.

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

Thompson, J. R. (1968), "Some Shrinkage Techniques for Estimating the Mean," *Journal of the American Statistical Association*, 63, 113–122.

Withers, C. S. (1987), "Bias Reduction by Taylor Series," *Communications in Statistics—Theory and Methods*, 16, 2369–2383.

Exponential Families and Variance Component Models

DENNIS L. CLASON and LEIGH W. MURRAY*

This article shows that two slightly differential exponential families are needed for commonly used linear models. The first family is the usual textbook definition. The second family, due to a lemma by Gautschi, is needed for some cross-classified variance component models.

KEY WORDS: Completeness; Lehmann-Scheffé theorem; Random components; Sufficiency.

Uniformly Minimum Variance Unbiased Estimators (UMVUE) are usually obtained by appeals to the Lehmann-Scheffé theorem and the properties of exponential families. Graybill (1976) defines the k-parameter exponential family in definition 2.6.3 as

$$f(\mathbf{Y}:\mathbf{\Theta}) = H(\mathbf{\Theta})G(\mathbf{Y}) \exp\left[\sum_{j=1}^{k} T_j(\mathbf{Y})Q_j(\mathbf{\Theta})\right], \quad (1)$$

with some regularity conditions on H, G, $\mathbf{T}$, and $\mathbf{Q}$. (This definition is similar to that of Rohatgi 1976, Roussas 1973, and Lehmann 1959). The form of the exponential family (1) is not, however, met by some cross-classified variance component models. These models are members of a slightly different exponential family, whose completeness was first proven in a lemma by Gautschi (1959) and subsequently reproven for the normal, twofold crossed model with an interaction component by Arnold (1981). Gautschi's lemma is cited by Hocking (1985, app. B).

*Dennis L. Clason is Assistant Professor and Leigh W. Murray is Associate Professor, both in the Department of Experimental Statistics, New Mexico State University, Las Cruces, 88003. This is Journal Article 1617, New Mexico Agricultural Experiment Station. The authors thank the referees and the editors for helpful comments.

© 1992 American Statistical Association