The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: https://www.tandfonline.com/loi/utas20

Who Invented the Delta Method?

Jay M. Ver Hoef

To cite this article: Jay M. Ver Hoef (2012) Who Invented the Delta Method?, The American Statistician, 66:2, 124-127, DOI: 10.1080/00031305.2012.687494

To link to this article: https://doi.org/10.1080/00031305.2012.687494

Accepted author version posted online: 01 Jun 2012.
Published online: 15 Aug 2012.

Who Invented the Delta Method?

Jay M. Ver Hoef

Many statisticians and other scientists use what is commonly called the "delta method." However, few people know who proposed it. The earliest article was found in an obscure journal, and the author is rarely cited for his contribution. This article briefly reviews three modern versions of the delta method and how they are used. Then, some history on the author and the journal of the first known article on the delta method is given. The original author's specific contribution is reproduced, along with a discussion on possible reasons that it has been overlooked.

KEY WORDS: Approximate variance; Bias correction; Limiting distribution; Taylor expansion.

1. INTRODUCTION

The "delta method" is commonly used by statisticians and other scientists. While many people use the delta method, and it is described in many textbooks, few people know who invented it. The method is often used without citation, or it is cited to a secondhand author, such as the author of a textbook on mathematical statistics. It is common to drop citations to authors when methods become so pervasive that everyone knows them. For example, Newton and Leibniz are not cited every time that a derivative or integral is used; it is common knowledge that they "invented" calculus. Similarly, Fisher is not cited every time that maximum likelihood estimation is used. On the other hand, despite the widespread use of the delta method, it is almost impossible to find who first proposed it. According to Oehlert (1992), Cramèr (1946) first offered a rigorous treatment of the delta method, but Cramèr did not invent it. The purpose of this note is to give a quick review of various definitions of the delta method, and then clear up some of the mystery about who invented it. Some history on the author and the journal of the first known article on the delta method is given, and the original author's specific contribution is reproduced. The discussion and conclusions give some possible reasons that the original author's contribution has been overlooked.

Jay M. Ver Hoef is Statistician, National Marine Mammal Laboratory, Alaska Fisheries Science Center, National Oceanic and Atmospheric Administration's (NOAA) National Marine Fisheries Service, Seattle, WA 98115-6349 (E-mail: *jay.verhoef@noaa.gov*). This project received financial support from the NOAA's National Marine Fisheries Service to the Alaska Fisheries Science Center. Reference to trade names does not imply endorsement by the National Marine Fisheries Service, NOAA.

2. WHAT IS THE DELTA METHOD?

There are three fairly distinct meanings for the delta method: (1) as an approximation for the variance of a function of a random variable, (2) as a bias correction for the expectation of a function of a random variable, and (3) as the limiting distribution of a function of a random variable. For the following, let X be a random variable with expectation $\mathrm{E}(X) = \mu$ and variance $\mathrm{var}(X) = \sigma^2$. As an example, consider the antilogit, $Y = f(X) = \exp(X)/(1 + \exp(X))$, which transforms X to the probability scale, between 0 and 1. This example will illustrate the three meanings of the delta method: as an approximation to $\mathrm{var}(Y)$, a bias-corrected $\mathrm{E}(Y)$, and the limiting distribution of Y.

2.1 An Approximate Variance

The delta method uses a Taylor expansion, yielding an approximate variance for a nonlinear function of a random variable. A Taylor series expansion around the mean gives

$$Y = f(X) = f(\mu) + f'(\mu)(X-\mu) + \frac{1}{2} f''(\mu)(X-\mu)^2 + \cdots,$$

where $f'(\mu)$ and $f''(\mu)$ are the first and second derivatives (respectively) of f evaluated at μ. A Taylor expansion can be expressed up to any order, with a remainder term, as

$$Y = f(X) = \sum_{i=0}^{p} \frac{f^{(i)}(\mu)(X-\mu)^i}{i!} + R, \tag{1}$$

where $f^{(i)}(\mu)$ is the ith derivative evaluated at μ and

$$R = \frac{f^{(p+1)}(\xi)(X-\mu)^{p+1}}{(p+1)!}$$

for ξ between X and μ. The Taylor expansion up to the first order, $i = 1$ in (1) and dropping R, is

$$f(X) - f(\mu) \approx f'(\mu)(X-\mu),$$

and after squaring and taking expectation,

$$\mathrm{var}(Y) \approx [f'(\mu)]^2 \sigma^2. \tag{2}$$

For the antilogit example, because $f'(X) = \exp(X)/(1 + \exp(X))^2$, using (2) yields

$$\operatorname{var}(Y) \approx \frac{[\exp(\mu)]^2}{(1+\exp(\mu))^4}\sigma^2. \tag{3}$$

A simulation illustrates (3). Assuming $X \sim N(2, 0.2^2)$, the estimated variance from one million simulations of $f(X)$ was 0.00247, and using (3), it was 0.00220, that is, about 10% error. The accuracy of the delta method depends on the functional form of f and the precision of X; Oehlert (1992) gave several other caveats.

2.2 Bias Correction

The delta method for bias correction also starts with the Taylor expansion (1). Up to the second-order term, $i = 2$ in (1) and dropping R, taking expectations gives

$$\mathrm{E}(Y) \approx f(\mu) + \frac{1}{2} f''(\mu)\sigma^2. \tag{4}$$

Continuing the antilogit example, because $f''(X) = \exp(X)(1 - \exp(X))/(1 + \exp(X))^3$,

$$\mathrm{E}(Y) \approx \frac{\exp(\mu)}{1+\exp(\mu)} + \frac{\exp(\mu)(1-(\exp(\mu))}{2(1+(\exp(\mu))^3}\sigma^2. \tag{5}$$

Again assuming $X \sim N(2, 0.2^2)$, the estimated mean from one million simulations of $f(X)$ was 0.862, and using (5), it was 0.861, while the naive back-transformation $\exp(\mu)/(1 + \exp(\mu))$ was 0.881.

2.3 A Limiting Distribution

The first item on a Google search of "delta method" is a wikipedia.com entry, where the delta method is defined as a limiting probability distribution for a function of an asymptotically normal statistical estimator. If

$$\sqrt{n}(X_n - \mu) \xrightarrow{D} N(0, \sigma^2)$$

for a sequence of random variables X_n depending on n, then

$$\sqrt{n}(f(X_n) - f(\mu)) \xrightarrow{D} N(0, [f'(\mu)]^2\sigma^2); \tag{6}$$

see Agresti (1990, p. 429) for a proof. The convergence as sample sizes increase is illustrated in Figure 1.

2.4 Using the Delta Method

The three meanings of the delta method are mostly used in two ways. Suppose X is an estimator that depends on data. The first use of the delta method depends on known values of μ and σ^2, which are used in a theoretical analysis of the properties of $f(X)$; see Agresti (1990, p. 419) for many examples. The second primary use of the delta method is when μ and σ^2 are unknown. Then, sample estimators, $\hat{\mu}$ and $\hat{\sigma}^2$, are used in place of μ and σ^2, which is called the "plug-in" method. This method is commonly used in practice; for example, see Cox (1990). Abatih et al. (2008) gave an example using the antilogit f with plug-in values for (3).

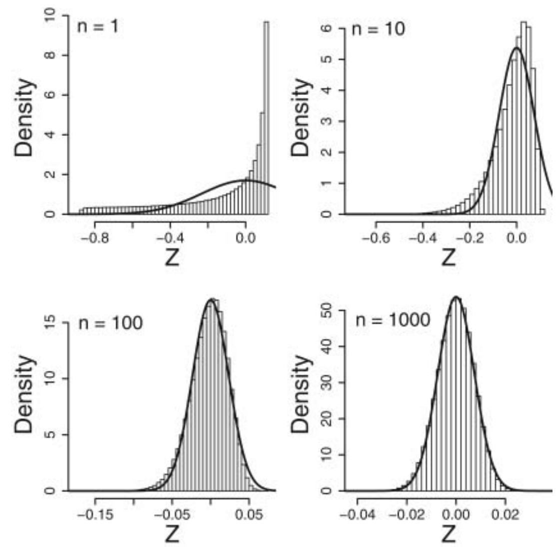

Figure 1. One million values were simulated from $X \sim N(\mu, \sigma^2/n)$ for $\mu = 2$ and $\sigma^2 = 5$, and histograms created of $f(X) - f(\mu) = \exp(X)/(1 + \exp(X)) - \exp(\mu)/(1 + \exp(\mu))$ for sample sizes of n = 1, 10, 100, and 1000. The normal distribution $N(0, [f'(\mu)]^2\sigma^2/n)$ is shown as a solid curve. Note that the histogram approaches the normal distribution as sample size increases.

3. DORFMAN'S δ-METHOD

An article by Robert Dorfman (1938) in the *Biometric Bulletin* originally proposed a "δ-method" as an approximate variance for a nonlinear function of one or more random variables. A brief biography of Dorfman and a history of the *Biometric Bulletin* precede a summary of his original contribution.

3.1 Robert Dorfman, 1916–2002

Dorfman was a man of great achievements, and he is well known for many contributions. Remarkably, Dorfman published his work on the delta method directly after earning his B.A. in mathematical statistics from Columbia College, New York, in 1936. After that, he worked for the federal government, and then he was an operations analyst during World War II. After the war, Dorfman entered a Ph.D. program in economics at the University of California, Berkeley, where he graduated in 1950. He held an associate professor position there until he moved to Harvard in 1955, where he retired in 1987. As an economist, Dorfman was one of the pioneers in linear programming and collaborated with economists and Nobel laureates Robert Solow and Paul Samuelson. He wrote pioneering textbooks such as *Application of Linear Programming to the Theory of the Firm* (Dorfman 1951), and *Linear Programming and Economic Analysis* (Dorfman, Samuelson, and Solow 1958). Later, Dorfman made important contributions in environmental economics. For example, he collaborated with engineers and hydrologists to study the conservation and distribution of scarce natural resources, such as water, in places ranging from Pakistan to the Middle East. More on his work is presented in his obituary (President and Fellows of Harvard College 2002).

3.2 Biometric Bulletin

Electronic searching for the journal *Biometric Bulletin* can be confusing. The *Biometric Bulletin* of Dorfman's delta-method article was published by Worcester State Hospital, the first public asylum for the insane in New England, which opened in 1833. The *Biometric Bulletin* was started by E.M. Jellinek (Roizen 2011), who was heading up a new biometric group after arriving at the hospital in 1931. The *Biometric Bulletin* was supposed to be published quarterly (subscription: $2 per year!), yet only four issues of a single volume went to press. The first volume spanned 3 years, appearing in June and December of 1936, December of 1937, and September of 1938. Dorfman published his work in the fourth issue in 1938. It is somewhat unclear how to cite the date of his work, as the date of the first issue (1936) of the one and only volume, or the actual date (I chose the latter). There may be confusion about the journal because *Biometrics* began as the *Biometrics Bulletin* in 1945–1946 with volumes 1–2 before becoming *Biometrics*. Adding to this confusion about the title is the "Biometric Bulletin" newsletter currently published by the Biometric Society.

3.3 Dorfman's Original Contribution

Here, Dorfman's original contribution is reproduced and explained, using Dorfman's δ notation, with a slight variation on some of his other notation to make it more modern and compact. In fact, Dorfman's result was a multivariate version of Section 2.1. Let

$$Y = f(X_1, X_2, \ldots, X_k) = f(\mathbf{x}). \tag{7}$$

The multivariate version of a Taylor expansion in (1), for our purposes, can be written as

$$Y = f(\mathbf{x}) = f(\boldsymbol{\mu}) + \mathbf{d}'(\mathbf{x}-\boldsymbol{\mu}) + (\mathbf{x}-\boldsymbol{\mu})'\mathbf{H}(\mathbf{x}-\boldsymbol{\mu})/2, \tag{8}$$

where

$$\mathbf{d} \equiv [f_1(\boldsymbol{\mu}), f_2(\boldsymbol{\mu}), \ldots, f_k(\boldsymbol{\mu})]',$$

has

$$f_i(\boldsymbol{\mu}) \equiv \left.\frac{\partial f(\mathbf{x})}{\partial X_i}\right|_{\mathbf{x}=\boldsymbol{\mu}},$$

and where

$$\mathbf{H} = \begin{pmatrix} f_{1,1}(\boldsymbol{\xi}) & f_{1,2}(\boldsymbol{\xi}) & \ldots & f_{1,k}(\boldsymbol{\xi}) \\ f_{2,1}(\boldsymbol{\xi}) & f_{2,2}(\boldsymbol{\xi}) & \ldots & f_{2,k}(\boldsymbol{\xi}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{k,1}(\boldsymbol{\xi}) & f_{k,2}(\boldsymbol{\xi}) & \ldots & f_{k,k}(\boldsymbol{\xi}) \end{pmatrix},$$

has

$$f_{i,j}(\boldsymbol{\xi}) \equiv \left.\frac{\partial^2 f(\mathbf{x})}{\partial X_i \partial X_j}\right|_{\mathbf{x}=\boldsymbol{\xi}},$$

for $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_k)'$, with ξ_i between X_i and μ_i.

Let Y and $\mathbf{x} = (X_1, X_2, \ldots, X_k)'$ be random with expectations $\mathrm{E}(Y) = \mu_y$, $\mathrm{E}(X_1) = \mu_1, \ldots, \mathrm{E}(X_k) = \mu_k$, and let $\boldsymbol{\mu} \equiv (\mu_1, \mu_2, \ldots, \mu_k)'$. Also, let the variance-covariance matrix be $\mathrm{var}(\mathbf{x}) = \boldsymbol{\Sigma} = (\boldsymbol{\sigma}\boldsymbol{\sigma}') \odot \mathbf{R}$, where $\mathrm{var}(X_i) = \sigma_i^2$, $\boldsymbol{\sigma} \equiv (\sigma_1, \sigma_2, \ldots, \sigma_k)'$, the i, jth element of $\mathbf{R}$ is the correlation between X_i and X_j, denoted as $\mathbf{R}[i, j] = \rho_{i,j}$, and $\mathbf{A} \odot \mathbf{B}$ is the Hadamard (elementwise) product of matrices $\mathbf{A}$ and $\mathbf{B}$.

Dorfman wrote (7) as

$$\mu_y + \delta_y = f(\mu_1 + \delta_1, \mu_2 + \delta_2, \ldots, \mu_k + \delta_k), \tag{9}$$

where δ_y and $\boldsymbol{\delta}' \equiv (\delta_1, \ldots, \delta_k)$ are zero-mean random variables, that is, they are random "deviations" or "errors" from the mean, and $\boldsymbol{\delta}$ has the same variance structure as $\mathbf{x}$, that is, $\mathrm{var}(\boldsymbol{\delta}) = \boldsymbol{\Sigma} = (\boldsymbol{\sigma}\boldsymbol{\sigma}') \odot \mathbf{R}$. Dorfman then wrote (9) as a multivariate Taylor expansion using (8),

$$\begin{aligned} \mu_y + \delta_y = {} & f(\boldsymbol{\mu}) + f_1(\boldsymbol{\mu})\delta_1 + f_2(\boldsymbol{\mu})\delta_2 + \cdots + f_k(\boldsymbol{\mu})\delta_k \\ & + \frac{1}{2}\{f_{1,1}(\boldsymbol{\mu} + \boldsymbol{\delta} \odot \boldsymbol{\theta})\delta_1^2 + \cdots + f_{k,k}(\boldsymbol{\mu} + \boldsymbol{\delta} \odot \boldsymbol{\theta})\delta_k^2 \\ & + 2 f_{1,2}(\boldsymbol{\mu} + \boldsymbol{\delta} \odot \boldsymbol{\theta})\delta_1\delta_2 + \cdots + 2(\boldsymbol{\mu} + \boldsymbol{\delta} \odot \boldsymbol{\theta}) \\ & \times \delta_{k-1}\delta_k f_{k-1,k}\}, \end{aligned}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ is a vector with $0 < \theta_i < 1$. Note that because $0 < \theta_i < 1$, and δ_i is a deviation from μ_i, the use of $\boldsymbol{\mu} + \boldsymbol{\delta} \odot \boldsymbol{\theta}$ is equivalent to $\boldsymbol{\xi}$ in (8). Ignoring the last terms in (8) containing the $\mathbf{H}$ and $\boldsymbol{\xi}$, and hence the terms containing $\boldsymbol{\theta}$ in Dorfman's formulation, and letting $\mu_y = f(\boldsymbol{\mu})$,

$$\delta_y \approx \delta_1 f_1(\boldsymbol{\mu}) + \delta_2 f_2(\boldsymbol{\mu}) + \cdots + \delta_k f_k(\boldsymbol{\mu}).$$

Note that Dorfman's use of the δ's was to isolate the zero-mean random components. Squaring each side,

$$\begin{aligned} \delta_y^2 \approx {} & \delta_1^2 f_1^2(\boldsymbol{\mu}) + \delta_2^2 f_2^2(\boldsymbol{\mu}) + \cdots + \delta_k^2 f_k^2(\boldsymbol{\mu}) \\ & + 2\delta_1\delta_2 f_1(\boldsymbol{\mu}) f_2(\boldsymbol{\mu}) + \cdots + 2\delta_{k-1}\delta_k f_{k-1}(\boldsymbol{\mu}) f_k(\boldsymbol{\mu}), \end{aligned} \tag{10}$$

and then taking expectations yields

$$\begin{aligned} \sigma_y^2 \approx {} & \sigma_1^2 f_1^2(\boldsymbol{\mu}) + \sigma_2^2 f_2^2(\boldsymbol{\mu}) + \cdots + \sigma_k^2 f_k^2(\boldsymbol{\mu}) \\ & + 2\sigma_1\sigma_2\rho_{1,2} f_1(\boldsymbol{\mu}) f_2(\boldsymbol{\mu}) + \cdots \\ & + 2\sigma_{k-1}\sigma_k\rho_{k-1,k} f_{k-1}(\boldsymbol{\mu}) f_k(\boldsymbol{\mu}), \end{aligned} \tag{11}$$

(Dorfman 1938), which agrees with the modern result; see (12).

He made several comments after his result: (1) often the X's are independent so that cross products disappear in (11), simplifying it considerably, (2) the method is exact if f is linear, and (3) the magnitude of $\boldsymbol{\theta}$ will determine the reliability of the approximation. Dorfman (1938) gave two examples. The first was an approximation to the t-distribution. The t is a ratio of two random variables: the sample mean over the sample standard deviation, which are independent of each other (if the original sample random variables are normal and independent) so that the cross-product term in (11) disappears. His second example was again a ratio, but this time a sample of blood volumes to a sample of body weights. In this example, the two variables were correlated so that the cross-product term in (11) was retained using the sample correlation between the two variables. Note that in both of the examples, Dorfman used plug-in values (as described in Section 2.4). Dorfman finished his article with some discussion on the limitations of the delta method. He noted that, first, if f does not deviate sharply from linear (a relatively small second derivative), then the approximation will be better (a justification for dropping the f'' term). Second, because he used the plug-in method (Section 2.4), the variances of the X's "must be known with some accuracy if the method is to be of much use."

Note that more modern treatments of the delta method skip the delta notation completely, as was done in Section 2.1. The multivariate version can be described compactly by noting that, from (8),

$$Y - f(\boldsymbol{\mu}) \approx \mathbf{d}'(\mathbf{x} - \boldsymbol{\mu}).$$

After squaring,

$$(Y - \mu_y)^2 \approx \mathbf{d}'(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'\mathbf{d},$$

where $\mu_y = f(\boldsymbol{\mu})$, and taking expectations yields

$$\sigma_y^2 \approx \mathbf{d}'\boldsymbol{\Sigma}\mathbf{d}. \tag{12}$$

There is no need to invoke the δ's, yet the term "delta method" has remained.

4. CONCLUSIONS

As of November 2011, a search for "Dorfman" and "delta method" in Google Scholar finds neither his 1938 article, nor any references to it. One may never know if Dorfman "invented" the delta method. A similar idea may be found in earlier literature. However, it seems very likely that Dorfman coined the term "delta method." He calls it the "δ-method" in the title of his article and uses the δ notation for the derivation within. The name has persisted but the association to Dorfman has not.

It is interesting to speculate on the reasons that Dorfman is not known for the delta method. Perhaps it was because his main career was as an economist, perhaps it was the obscurity of the *Biometric Bulletin*, or perhaps it was because he used the symbol δ in his title, rather than the word "delta," making modern electronic searches miss his original contribution. For whatever reason, it is time for Dorfman to get credit for his contribution.

[Received January 2012. Revised July 2012.]

REFERENCES

Abatih, E., Van Oyen, H., Bossuyt, N., and Bruckers, L. (2008), "Variance Estimation Methods for Health Expectancy by Relative Socio-Economic Status," *European Journal of Epidemiology*, 23, 243–249. [125]

Agresti, A. (1990), *Categorical Data Analysis*, New York: Wiley. [125]

Cox, C. (1990), "Fieller's Theorem, the Likelihood and the Delta Method," *Biometrics*, 46, 709–718. [125]

Cramèr, H. (1946), *Mathematical Methods of Statistics*, Princeton, NJ: Princeton University Press. [124]

Dorfman, R. (1938), "A Note on the δ-Method for Finding Variance Formulae," *The Biometric Bulletin*, 1, 129–137. [125,126]

——— (1951), *Application of Linear Programming to the Theory of the Firm: Including an Analysis of Monopolistic Firms by Non-Linear Programming*, Berkeley: University of California Press. [125]

Dorfman, R., Samuelson, P., and Solow, R. (1958), *Linear Programming and Economic Analysis*, New York: McGraw-Hill. [125]

Oehlert, G. (1992), "A Note on the Delta Method," *The American Statistician*, 46, 27–29. [124,125]

President and (2002), "Economist Dorfman dies at 85," [online]. Available at *http://news.harvard.edu/gazette/2002/07.18/25-dorfman.html*. [125]

Roizen, R. (2011), "E.M. Jellinek at Worcester: A Bare Beginning" [online]. Available at *http://pointsadhsblog.wordpress.com/2011/06/05/e-m-jellinek-at-worcester-a-bare-beginning/*. [125]