Delta method

From Wikipedia, the free encyclopedia

In statistics, the **delta method** is a result concerning the approximate probability distribution for a function of an asymptotically normal statistical estimator from knowledge of the limiting variance of that estimator.

Contents

Univariate delta method

While the delta method generalizes easily to a multivariate setting, careful motivation of the technique is more easily demonstrated in univariate terms. Roughly, if there is a sequence of random variables X_n satisfying

$$\sqrt{n}[X_n - \theta] \xrightarrow{D} \mathcal{N}(0, \sigma^2),$$

where θ and σ^2 are finite valued constants and $\xrightarrow{D}$ denotes convergence in distribution, then

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{D} \mathcal{N}(0, \sigma^2 [g'(\theta)]^2)$$

for any function g satisfying the property that $g'(\theta)$ exists and is non-zero valued.

Proof in the univariate case

Demonstration of this result is fairly straightforward under the assumption that $g'(\theta)$ is continuous. To begin, we use the mean value theorem (i.e.: the first order approximation of a Taylor series using Taylor's theorem):

$$g(X_n) = g(\theta) + g'(\tilde{\theta})(X_n - \theta),$$

where $\tilde{\theta}$ lies between X_n and θ. Note that since $X_n \xrightarrow{P} \theta$ and $X_n < \tilde{\theta} < \theta$, it must be that $\tilde{\theta} \xrightarrow{P} \theta$ and since $g'(\theta)$ is continuous, applying the continuous mapping theorem yields

$$g'(\tilde{\theta}) \xrightarrow{P} g'(\theta),$$

where $\xrightarrow{P}$ denotes convergence in probability.

Rearranging the terms and multiplying by $\sqrt{n}$ gives

$$\sqrt{n}[g(X_n) - g(\theta)] = g'\left(\tilde{\theta}\right)\sqrt{n}[X_n - \theta].$$

Since

$$\sqrt{n}[X_n - \theta] \xrightarrow{D} \mathcal{N}(0, \sigma^2)$$

by assumption, it follows immediately from appeal to Slutsky's theorem that

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{D} \mathcal{N}(0, \sigma^2[g'(\theta)]^2).$$

This concludes the proof.

Proof with an explicit order of approximation

Alternatively, one can add one more step at the end, to obtain the order of approximation:

$$\sqrt{n}[g(X_n) - g(\theta)] = g'\left(\tilde{\theta}\right)\sqrt{n}[X_n - \theta] = \sqrt{n}[X_n - \theta]\left[g'(\tilde{\theta}) + g'(\theta) - g'(\theta)\right]$$

$$= \sqrt{n}[X_n - \theta]\left[g'(\theta)\right] + \sqrt{n}[X_n - \theta]\left[g'(\tilde{\theta}) - g'(\theta)\right]$$

$$= \sqrt{n}[X_n - \theta]\left[g'(\theta)\right] + O_p(1) \cdot o_p(1)$$

$$= \sqrt{n}[X_n - \theta]\left[g'(\theta)\right] + o_p(1)$$

This suggests that the error in the approximation converges to 0 in probability.

Multivariate delta method

By definition, a consistent estimator B converges in probability to its true value β, and often a central limit theorem can be applied to obtain asymptotic normality:

$$\sqrt{n}\left(B - \beta\right) \xrightarrow{D} N\left(0, \Sigma\right),$$

where n is the number of observations and Σ is a (symmetric positive semi-definite) covariance matrix. Suppose we want to estimate the variance of a function h of the estimator B. Keeping only the first two terms of the Taylor series, and using vector notation for the gradient, we can estimate $h(B)$ as

$$h(B) \approx h(\beta) + \nabla h(\beta)^T \cdot (B - \beta)$$

which implies the variance of $h(B)$ is approximately

$$\begin{aligned}
\mathbf{Var}(h(B)) &\approx \mathbf{Var}\left(h(\beta) + \nabla h(\beta)^T \cdot (B - \beta)\right) \\
&= \mathbf{Var}\left(h(\beta) + \nabla h(\beta)^T \cdot B - \nabla h(\beta)^T \cdot \beta\right) \\
&= \mathbf{Var}\left(\nabla h(\beta)^T \cdot B\right) \\
&= \nabla h(\beta)^T \cdot \mathbf{Cov}(B) \cdot \nabla h(\beta) \\
&= \nabla h(\beta)^T \cdot (\Sigma/n) \cdot \nabla h(\beta)
\end{aligned}$$

One can use the mean value theorem (for real-valued functions of many variables) to see that this does not rely on taking first order approximation.

The delta method therefore implies that

$$\sqrt{n}\left(h(B) - h(\beta)\right) \xrightarrow{D} N\left(0, \nabla h(\beta)^T \cdot \Sigma \cdot \nabla h(\beta)\right)$$

or in univariate terms,

$$\sqrt{n}\left(h(B) - h(\beta)\right) \xrightarrow{D} N\left(0, \sigma^2 \cdot (h'(\beta))^2\right).$$

Example

Suppose X_n is Binomial with parameters $p \in (0,1]$ and n. Since

$$\sqrt{n}\left[\frac{X_n}{n} - p\right] \xrightarrow{D} N(0, p(1-p)),$$

we can apply the Delta method with $g(\theta) = \log(\theta)$ to see

$$\sqrt{n}\left[\log\left(\frac{X_n}{n}\right) - \log(p)\right] \xrightarrow{D} N(0, p(1-p)[1/p]^2)$$

Hence, even though for any finite n, the variance of $\log\left(\frac{X_n}{n}\right)$ does not actually exist (since X_n can be zero), the asymptotic variance of $\log\left(\frac{X_n}{n}\right)$ does exist and is equal to

$$\frac{1-p}{pn}.$$

Note that since $p > 0$, $\mathbf{Pr}\left(\frac{X_n}{n} > 0\right) \to 1$ as $n \to \infty$, so with probability converging to one, $\log\left(\frac{X_n}{n}\right)$ is finite for large n.

Moreover, if $\hat{p}$ and $\hat{q}$ are estimates of different group rates from independent samples of sizes n and m respectively, then the logarithm of the estimated relative risk $\frac{\hat{p}}{\hat{q}}$ has asymptotic variance equal to

$$\frac{1-p}{pn} + \frac{1-q}{qm}.$$

This is useful to construct a hypothesis test or to make a confidence interval for the relative risk.

Note

The delta method is often used in a form that is essentially identical to that above, but without the assumption that X_n or B is asymptotically normal. Often the only context is that the variance is "small". The results then just give approximations to the means and covariances of the transformed quantities. For example, the formulae presented in Klein (1953, p. 258) are:

$$\mathbf{Var}(h_r) = \sum_i \left(\frac{\partial h_r}{\partial B_i}\right)^2 \mathbf{Var}(B_i) + \sum_i \sum_{j \neq i} \left(\frac{\partial h_r}{\partial B_i}\right)\left(\frac{\partial h_r}{\partial B_j}\right) \mathbf{Cov}(B_i, B_j)$$

$$\mathbf{Cov}(h_r, h_s) = \sum_i \left(\frac{\partial h_r}{\partial B_i}\right)\left(\frac{\partial h_s}{\partial B_i}\right) \mathbf{Var}(B_i) + \sum_i \sum_{j \neq i} \left(\frac{\partial h_r}{\partial B_i}\right)\left(\frac{\partial h_s}{\partial B_j}\right) \mathbf{Cov}(B_i, B_j)$$

where h_r is the rth element of $h(B)$ and B_i is the ith element of B. The only difference is that Klein stated these as identities, whereas they are actually approximations.

See also

- Taylor expansions for the moments of functions of random variables
- Variance-stabilizing transformation

References

- Casella, G.; Berger, R. L. (2002). *Statistical Inference* (2nd ed.).
- Cramér, H. (1946). *Mathematical Methods of Statistics*. p. 353.
- Davison, A. C. (2003). *Statistical Models*. pp. 33–35.
- Greene, W. H. (2003). *Econometric Analysis* (5th ed.). pp. 913f.
- Klein, L. R. (1953). *A Textbook of Econometrics*. p. 258.
- Oehlert, G. W. (1992). "A Note on the Delta Method". *The American Statistician*. **46** (1): 27–29. JSTOR 2684406.

External links

- Asmussen, Søren. "Some Applications of the Delta Method" (PDF). *Lecture notes*. Aarhus University.
- Feiveson, Alan H. "Explanation of the delta method". Stata Corp.
- Xu, Jun; Long, J. Scott (August 22, 2005). "Using the Delta Method to Construct Confidence Intervals for Predicted Probabilities, Rates, and Discrete Changes" (PDF). *Lecture notes*. Indiana University.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Delta_method&oldid=750239657"

Categories: Estimation methods | Statistical approximations

- This page was last modified on 18 November 2016, at 15:25.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.