MA 575 Linear Models:

Cedric E. Ginestet, Boston University

Delta Method

Week 10, Lecture 1

1 Delta Method

1.1 Motivation from Polynomial Regression

Consider the following quadratic regression mean function,

$$\mathbb{E}[Y|X=x] = \beta_0 + \beta_1 x + \beta_2 x^2.$$

Setting this function to zero and differentiating with respect to x, we obtain the formula for the critical value,

$$x_M := -\frac{\beta_1}{2\beta_2}.$$

This would be only relevant if we were expecting that predictor to exhibit a minimum or a maximum within the range of the data at hand. Note that this quantity is a **non-linear** combination of the two parameters of interest, β_1 and β_2. As a result, it is not clear whether we can apply normal theory to construct a confidence interval for that quantity.

1.2 Univariate Delta Method

The general idea of the delta method is to approximate the distribution of a non-linear transformation of an estimator, using the known distribution of that estimator. It pertains to any statistic T_n, which is well-behaved in the sense that the CLT for that statistic holds,

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \sigma^2).$$

Then, given a function of T_n, denoted $\phi(T_n)$, which is both continuous and differentiable at a point θ, we can approximate the distribution of that transformed statistic using,

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) \xrightarrow{d} N(0, [\phi'(\theta)]^2\sigma^2),$$

where $\phi'(\theta)$ is the derivative of ϕ with respect to T_n, estimated at θ. The delta method is a direct use of a **first-order** Taylor approximation.

1.3 Delta Method in Multiple Regression

Under the standard assumptions of a multiple linear model, the OLS estimators of interest have a known multivariate normal distribution, such that

$$\widehat{\boldsymbol{\beta}} \sim \text{MVN}_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}),$$

where recall that $(\mathbf{X}^T\mathbf{X})^{-1}$ is positive definite, since $\mathbf{X}$ is full-rank. The statistic of interest is the following transformation of the OLS estimators,

$$g(\widehat{\boldsymbol{\beta}}) := -\frac{\widehat{\beta}_1(Y_1,\dots,Y_n)}{2\beta_2(Y_1,\dots,Y_n)}. \tag{1}$$

Observe that $g(\widehat{\boldsymbol{\beta}})$ depends on the full random vector of Y_i's, and is therefore a random variable. We can then construct a Taylor series of $g(\widehat{\boldsymbol{\beta}})$ around $\boldsymbol{\beta}^*$, such that

$$g(\widehat{\boldsymbol{\beta}}) = g(\boldsymbol{\beta}^*) + \sum_{j=1}^{k} \frac{\partial g(\widehat{\boldsymbol{\beta}})}{\partial \beta_j}(\widehat{\beta}_j - \beta_j^*) + O(||\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*||^2).$$

In matrix notation, this gives the following **truncated** Taylor series,

$$g(\widehat{\boldsymbol{\beta}}) \doteq g(\boldsymbol{\beta}^*) + \dot{\mathbf{g}}(\boldsymbol{\beta}^*)^T(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*),$$

where we have defined the **score vector** as follows,

$$\dot{\mathbf{g}}(\boldsymbol{\beta}^*) := \left|\begin{matrix} \frac{\partial g(\widehat{\boldsymbol{\beta}})}{\partial \widehat{\beta}_1} \\ \vdots \\ \frac{\partial g(\widehat{\boldsymbol{\beta}})}{\partial \widehat{\beta}_p} \end{matrix}\right|_{\widehat{\boldsymbol{\beta}}=\boldsymbol{\beta}^*}$$

The variance of this estimator is given by

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[g(\widehat{\boldsymbol{\beta}})|\mathbf{X}] &= \mathbb{V}\mathrm{ar}[g(\boldsymbol{\beta}^*)|\mathbf{X}] + \mathbb{V}\mathrm{ar}\left[\dot{\mathbf{g}}(\boldsymbol{\beta}^*)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\,\middle|\,\mathbf{X}\right] \\
&= \dot{\mathbf{g}}(\boldsymbol{\beta}^*)^T\,\mathbb{V}\mathrm{ar}\left[\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\,\middle|\,\mathbf{X}\right]\dot{\mathbf{g}}(\boldsymbol{\beta}^*) \\
&= \sigma^2\dot{\mathbf{g}}(\boldsymbol{\beta}^*)^T(\mathbf{X}^T\mathbf{X})^{-1}\dot{\mathbf{g}}(\boldsymbol{\beta}^*).
\end{aligned}$$

The choice of $\boldsymbol{\beta}^*$ is here motivated by our goal to determine the coverage of the statistic x_M. Thus, we simply define $\boldsymbol{\beta}^*$ to be the estimate of that statistic for the data at hand, such that we choose

$$\boldsymbol{\beta}^* := \widehat{\boldsymbol{\beta}}(y_1,\dots,y_n).$$

Finally, observe that the derivation of the variance in the case of equation (1) would require the computation of the vector of partial derivatives of $g(\boldsymbol{\beta}) = -\beta_1/2\beta_2$ evaluated at $\widehat{\boldsymbol{\beta}}(y_1,\dots,y_n)$, which is

$$\dot{\mathbf{g}}(\widehat{\boldsymbol{\beta}}) = \left[0, -\frac{1}{2\widehat{\beta}_2}, \frac{\widehat{\beta}_1}{2\widehat{\beta}_2^2}\right]^T.$$