The Delta Method

Case I : Univariate (θ is a scalar)

Let $\hat{\theta}$ be an estimator for θ such that $\hat{\theta}$ satisfies a CLT :

$$\sqrt{n}\,(\hat{\theta} - \theta) \xrightarrow{d} W \sim N(0, \sigma^2)$$

or

$$\hat{\theta} \overset{A}{\sim} N\left(\theta, \frac{\sigma^2}{n}\right)$$

Let $\eta = g(\theta)$ denote the "parameter of interest" where g is a continuous function independent of the sample size n. Then

$$\sqrt{n}(\hat{\eta} - \eta) = \sqrt{n}\,(g(\hat{\theta}) - g(\theta)) \xrightarrow{d} \frac{dg(\theta)}{d\theta} \cdot W$$

$$\sim N\left(0, \left(\frac{dg(\theta)}{d\theta}\right)^2 \cdot \sigma^2\right)$$

or

$$\hat{\eta} = g(\hat{\theta}) \overset{A}{\sim} N\left(g(\theta), \left(\frac{dg(\theta)}{d\theta}\right)^2 \cdot \sigma^2\right)$$

If θ is unknown, a practically useful formula is based on estimating the asymptotic variance:

$$g(\hat{\theta}) \overset{A}{\sim} N\Big(g(\theta), \underbrace{\Big(\frac{d(\hat{\theta})}{d\theta}\Big)^2 \cdot \hat{\sigma}^2}_{\substack{\text{estimated} \\ \text{asymptotic} \\ \text{variance.}}}\Big)$$

<u>Example</u>

Let $X_1, .., X_n$ be an iid sample with $E[X_i] = \mu$ and $var(X_i) = \sigma^2$. Let $\eta = g(\mu) = \mu^{-1}$ be the parameter of interest. The CLT gives

$$\sqrt{n}(\bar{X} - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Now $\dfrac{d\, g(\mu)}{d\mu} = \dfrac{d}{d\mu}\mu^{-1} = -\mu^{-2}$

$$\Big(\frac{d\, g(\mu)}{d\mu}\Big)^2 = +\mu^{-4}$$

Then the delta method gives

$$\sqrt{n}\left(g(\bar{x}) - g(\mu)\right) \xrightarrow{d} N\left(0, \left(\frac{dg(\mu)}{d\mu}\right)^2 \cdot \sigma^2\right.$$

or

$$\sqrt{n}\left(\frac{1}{\bar{X}} - \frac{1}{\mu}\right) \xrightarrow{d} N\left(0, \mu^{-4}\sigma^2\right)$$

Alternatively,

$$\frac{1}{\bar{X}} \overset{A}{\sim} N\left(\frac{1}{\mu}, \frac{1}{n}\mu^{-4}\cdot\sigma^2\right)$$

A practically useful formula uses consistent estimates for μ and σ^2 in the asymptotic variance:

$$\bar{X} \xrightarrow{p} \mu$$

$$\hat{\sigma}^2 = \frac{1}{n}\Sigma_i^n (X_i - \bar{x})^2 \xrightarrow{p} \sigma^2$$

$$\frac{1}{\bar{X}} \overset{A}{\sim} N\left(\frac{1}{\mu}, \frac{1}{n}(\bar{x})^{-4}\hat{\sigma}^2\right)$$

Remark: The method is called the "delta method" because ~~then~~ it is based on a first order Taylor series approximation to $\sqrt{n}\, g(\hat{\theta})$:

$$g(\hat{\theta}_n) = g(\theta) + \frac{d}{d\theta} g(\theta)(\hat{\theta} - \theta) + \text{error}$$

$$\Rightarrow \quad \sqrt{n}\left(g(\hat{\theta}) - g(\theta)\right) = \frac{d}{d\theta} g(\theta) \cdot \sqrt{n}\,(\hat{\theta} - \theta) + \sqrt{n}\,\text{error}$$

$$\xrightarrow{d} \quad \frac{d}{d\theta} g(\theta) * N(0, \sigma^2)$$

$$= N\left(0, \left(\frac{d}{d\theta} g(\theta)\right)^2 \cdot \sigma^2\right)$$

The hard part is proving that $\sqrt{n}\cdot\text{error}^* \xrightarrow{P} 0$.

The Delta Method Continued

Case II: $\underset{\sim}{\theta}$ is a vector

Let $\underset{k\times 1}{\hat{\underset{\sim}{\theta}}}$ be an estimator for $\underset{\sim}{\theta}$ s.t.

$$\sqrt{n}\,(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}) \xrightarrow{d} N(0, \underset{k\times k}{\Sigma})$$

or

$$\hat{\underset{\sim}{\theta}} \overset{A}{\sim} N(\theta, \tfrac{1}{n}\Sigma)$$

Let $\underset{1\times 1}{\eta} = g(\underset{\sim}{\theta})$ denote the "parameter of interest" where g is a continuous function independent of n. Then

$$\sqrt{n}\,(\hat{\eta} - \eta) = \sqrt{n}\,(g(\hat{\underset{\sim}{\theta}}) - g(\theta))$$

$$\xrightarrow{d} N\left(0, \underset{1\times k}{\frac{\partial g(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}'}} \underset{k\times k}{\Sigma} \underset{k\times 1}{\left(\frac{\partial g(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}'}\right)'}\right)$$

or

$$\hat{\eta} = g(\hat{\underset{\sim}{\theta}}) \overset{A}{\sim} N\left(g(\theta), \frac{1}{n}\frac{\partial g(\underset{\sim}{\theta})}{\partial \theta'} \Sigma \left(\frac{\partial g(\theta)}{\partial \theta'}\right)'\right)$$

<u>Note</u>: $$\underset{1\times k}{\frac{\partial g(\theta)}{\partial \theta'}} = \left(\frac{\partial g(\theta)}{\partial \theta_1}, \frac{\partial g(\theta)}{\partial \theta_2}, \ldots, \frac{\partial g(\theta)}{\partial \theta_k}\right)$$

If $\underset{\sim}{\theta}$ is unknown, a practically useful formula is based on consistent estimates of $\hat{\underset{\sim}{\theta}}$ and Σ:

$$\hat{\eta} = g(\hat{\underset{\sim}{\theta}}) \overset{A}{\sim} N\left(g(\theta),\ \frac{1}{n} \frac{\partial g(\hat{\underset{\sim}{\theta}})}{\partial \theta'} \hat{\Sigma} \left(\frac{\partial g(\hat{\theta})}{\partial \theta'} \right)' \right)$$

Example: Generalized learning curve (Berndt Ch. 2)

$$C_t = C_1 \cdot N_t^{\alpha_c/R} \cdot Y_t^{(1-R)/R} \exp(u_t)$$

~~total~~ unit real cost at time t

cumulative production up to time t

production at time t

multiplicative error term

α_c = elasticity of unit cost w.r.t. cumulative production

= learning curve parameter (typically negative)

R = returns to scale parameter

$R = 1$: CRS

$R < 1$: DRS

$R > 1$: IRS.

u_t = iid $(0, \sigma^2)$ error term.

Intuition: learning is proxied by the cumulative production of a variable. If the learning curve effect is present, then as cumulative production (learning) increases real unit costs should fall. If production technology exhibits constant returns to scale, then real unit costs should not vary with the level of output.

If returns to scale are increasing then real unit costs should decline as the level of production increases.

<u>Linearized learning curve regression</u> (Valid if multiplicative errors

$$\ln C_t = \ln C_1 + \left(\frac{\alpha_c}{R}\right) \ln n_t + \left(\frac{1-R}{R}\right) \ln Y_t + u$$

or

$$= \beta_0 + \beta_1 \ln n_t + \beta_2 \ln Y_t + u_t \quad (*$$

where

$$\left.\begin{aligned} \beta_0 &= \ln C_1 \\ \beta_1 &= \frac{\alpha_c}{R} \\ \beta_2 &= \frac{1-R}{R} \end{aligned}\right\} \quad \begin{aligned} \alpha_c &= \frac{\beta_1}{1+\beta_2} = g_1(\beta_0, \beta_1, \beta \\ &= g_1(\underset{\sim}{\beta}) \\ R &= \frac{1}{1+\beta_2} = g_2(\beta_0, \beta_1, \beta_2 \\ &= g_2(\underset{\sim}{\beta}) \end{aligned}$$

We can estimate β_0, β_1 & β_2 by OLS on $(*)$

but the parameters of interest are α_c and R

OLS applied to $(*)$ gives

$$\underset{3\times 1}{\hat{\underset{\sim}{\beta}}} = (X'X)^{-1}(X'y) \quad , \; x_t = (1, \ln n_t, \ln Y_t) \quad y_t = \ln c_t$$

and we know that

$$\hat{\beta} \xrightarrow{p} \beta$$

$$\hat{\beta} \overset{A}{\sim} N(\beta, \hat{\sigma}^2 (X'X)^{-1})$$

Then from Slutsky's theorem

$$\hat{\alpha}_c = \frac{\hat{\beta}_1}{1+\hat{\beta}_2} \xrightarrow{p} \frac{\beta_1}{1+\beta_2} = \alpha_c$$

provided $\beta_2 \neq -1$

$$\hat{R} = \frac{1}{1+\hat{\beta}_2} \xrightarrow{p} \frac{1}{1+\beta_2} = R$$

provided $\beta_2 \neq -1$.

We can use the delta method to get the asymptotic distributions of $\hat{\alpha}_c$ and $\hat{R}$:

(i) $\hat{\alpha}_c \overset{A}{\sim} N\left(\alpha_c ,\ \underset{1\times 3}{\frac{\partial g_1(\hat{\beta})}{\partial \beta'}} \underset{3\times 3}{\left[\hat{\sigma}^2 (X'X)^{-1}\right]} \underset{3\times 1}{\left(\frac{\partial g_1(\hat{\beta})}{\partial \beta'}\right)'} \right)$

(ii) $\hat{R} \overset{A}{\sim} N\left(R ,\ \frac{\partial g_2(\hat{\beta})}{\partial \beta'} \left[\hat{\sigma}^2 (X'X)^{-1}\right] \left(\frac{\partial g_2(\hat{\beta})}{\partial \beta'}\right)' \right)$

Now

$$\underset{1\times 3}{\frac{\partial g_1(\beta)}{\partial \beta'}} = \left[\frac{\partial g_1(\beta)}{\partial \beta_0} \quad \frac{\partial g_1(\beta)}{\partial \beta_1} \quad \frac{\partial g_1(\beta)}{\partial \beta_2} \right]$$

$$\frac{\partial g_1(\beta)}{\partial \beta_0} = \frac{\partial}{\partial \beta_0}\left(\frac{\beta_1}{1+\beta_2} \right) = 0$$

$$\frac{\partial g_1(\beta)}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\frac{\beta_1}{1+\beta_2} \right) = \frac{1}{1+\beta_2}$$

$$\frac{\partial g_1(\beta)}{\partial \beta_2} = \frac{\partial}{\partial \beta_2}\left(\frac{\beta_1}{1+\beta_2} \right) = -\frac{\beta_1}{(1+\beta_2)^2}$$

A consistent estimate of the asymptotic variance of $\hat{\alpha}_c$ is then

$$\hat{\text{Var}}(\hat{\alpha}_c) = \left[0 \quad \frac{1}{1+\hat{\beta}_2} \quad \frac{-\hat{\beta}_1}{(1+\hat{\beta}_2)^2} \right] \left[\hat{\sigma}^2 (X'X)^{-1} \right] * \begin{bmatrix} 0 \\ \frac{1}{1+\hat{\beta}_2} \\ \frac{-\hat{\beta}_1}{(1+\hat{\beta}_2)^2} \end{bmatrix}$$

Similarly,

$$\frac{\partial g_2(\beta)}{\partial \beta'} = \left[\frac{\partial g_2(\beta)}{\partial \beta_0} \quad \frac{\partial g_2(\beta)}{\partial \beta_1} \quad \frac{\partial g_2(\beta)}{\partial \beta_2} \right]$$

and

$$\frac{\partial g_2(\beta)}{\partial \beta_0} = \frac{\partial}{\partial \beta_0}\left(\frac{1}{1+\beta_2} \right) = 0$$

$$\frac{\partial g_2(\beta)}{\partial \beta_1} = \frac{\partial}{\partial \beta_1}\left(\frac{1}{1+\beta_2} \right) = 0$$

$$\frac{\partial g_2(\beta)}{\partial \beta_2} = \frac{\partial}{\partial \beta_2}\left(\frac{1}{1+\beta_2} \right) = \frac{-1}{(1+\beta_2)^2}$$

A consistent estimate of the asymptotic variance of $\hat{R}$ is then

$$\hat{\text{var}}(\hat{R}) = \begin{bmatrix} 0 & 0 & \frac{-1}{(1+\hat{\beta}_2)^2} \end{bmatrix} \left[\hat{\sigma}^2 (X'X)^{-1} \right] \begin{bmatrix} 0 \\ 0 \\ -\frac{1}{(1+\hat{\beta}_2)^2} \end{bmatrix}$$

These formulas can be used to construct asymptotic valid confidence intervals for α_c and R.

Case III: $\underset{\sim}{\theta}_{K\times 1}$,

$$\underset{\sim}{\eta}_{J\times 1} = \underset{\sim}{g}(\underset{\sim}{\theta}) = \begin{pmatrix} g_1(\underset{\sim}{\theta}) \\ g_2(\underset{\sim}{\theta}) \\ \vdots \\ g_J(\underset{\sim}{\theta}) \end{pmatrix}$$

scalar function

= J parameters of interest are functions of the K parameters in $\underset{\sim}{\theta}$.

Let $\underset{\sim}{\hat{\theta}}_{K\times 1}$ be an estimator for $\underset{\sim}{\theta}$ s.t.

$$\sqrt{n}(\underset{\sim}{\hat{\theta}} - \underset{\sim}{\theta}) \xrightarrow{d} N(0, \underset{K\times K}{\Sigma})$$

or

$$\underset{\sim}{\hat{\theta}}_{K\times 1} \overset{\wedge}{\sim} N(0, \tfrac{1}{n}\Sigma)$$

Then

$$\sqrt{n}(\underset{\sim}{\hat{\eta}}_{(J\times 1)} - \underset{\sim}{\eta}) = \sqrt{n}(\underset{\sim}{g}(\underset{\sim}{\hat{\theta}}) - \underset{\sim}{g}(\underset{\sim}{\theta}))$$

$$\xrightarrow{d} N\Big(\underset{\sim}{0}_{J\times 1}, \underbrace{\underset{J\times K}{\frac{\partial \underset{\sim}{g}(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}'}} \underset{K\times K}{\Sigma} \underset{K\times J}{\Big(\frac{\partial \underset{\sim}{g}(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}'}\Big)'}}_{J\times J}\Big)$$

over

↓

$$\underset{K\times 1}{\underset{\sim}{\theta}} = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_K \end{pmatrix}, \quad \underset{J\times 1}{\underset{\sim}{g}}(\underset{\sim}{\theta}) = \begin{bmatrix} g_1(\underset{\sim}{\theta}) \\ \vdots \\ g_J(\underset{\sim}{\theta}) \end{bmatrix}$$

Note: $$\underset{J\times K}{\frac{\partial \underset{\sim}{g}(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}'}} = \begin{bmatrix} \frac{\partial g_1(\underset{\sim}{\theta})}{\partial \theta_1} & \frac{\partial g_1(\underset{\sim}{\theta})}{\partial \theta_2} & \cdots & \frac{\partial g_1(\underset{\sim}{\theta})}{\partial \theta_K} \\ \frac{\partial g_2(\underset{\sim}{\theta})}{\partial \theta_1} & \frac{\partial g_2(\underset{\sim}{\theta})}{\partial \theta_2}, & \cdots, & \frac{\partial g_2(\underset{\sim}{\theta})}{\partial \theta_K} \\ \vdots & & & \\ \frac{\partial g_J(\underset{\sim}{\theta})}{\partial \theta_1} & \frac{\partial g_J(\underset{\sim}{\theta})}{\partial \theta_2}, & \cdots, & \frac{\partial g_J(\underset{\sim}{\theta})}{\partial \theta_K} \end{bmatrix}$$

Cost e Learning Curve example continued

Suppose we are interested in the asymptotic distribution of $\underset{2\times 1}{\underset{\sim}{\eta}} = \begin{pmatrix} \alpha_c \\ R \end{pmatrix} = \begin{pmatrix} g_1(p) \\ g_2(p) \end{pmatrix} = \underset{\sim}{g}(\underset{\sim}{p})$. This would be the case if we wanted to ~~make~~ test joint hypotheses about α_c and R. For example,

$$H_0: \alpha_c = 0 \text{ and } R = 1$$

vs

$$H_a: \alpha_c \neq 0 \text{ or } R \neq 1 \text{ or both.}$$

The delta method tells us that

$$\begin{pmatrix} \hat{\alpha}_c \\ \hat{R} \end{pmatrix} = \underset{\sim}{g}(\hat{\underset{\sim}{\beta}}) \overset{A}{\sim} N\left(\underset{\sim}{g}(\underset{\sim}{\beta}),\ \frac{\partial \underset{\sim}{g}(\underset{\sim}{\beta})}{\partial \beta'} \left[\sigma^2 (X'X)^{-1}\right] \left(\frac{\partial \underset{\sim}{g}(\beta)}{\partial \beta'} \right)' \right)$$

Now

$$\underset{2\times 3}{\frac{\partial g(\beta)}{\partial \beta'}} = \begin{bmatrix} \frac{\partial g_1(\beta)}{\partial \beta_0} & \frac{\partial g_1(\beta)}{\partial \beta_1} & \frac{\partial g_1(\beta)}{\partial \beta_2} \\ \frac{\partial g_2(\beta)}{\partial \beta_0} & \frac{\partial g_2(\beta)}{\partial \beta_1} & \frac{\partial g_2(\beta)}{\partial \beta_2} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & \frac{1}{1+\beta_2} & \frac{-\beta_1}{(1+\beta_2)^2} \\ 0 & 0 & -\frac{1}{(1+\beta_2)^2} \end{bmatrix}$$

A consistent estimate of the asymptotic variance of $\begin{pmatrix} \hat{\alpha}_c \\ \hat{R} \end{pmatrix}$ is then

$$\begin{bmatrix} 0 & \frac{1}{1+\hat{\beta}_2} & -\frac{\hat{\beta}_1}{(1+\hat{\beta}_2)^2} \\ 0 & 0 & -\frac{1}{(1+\hat{\beta}_2)^2} \end{bmatrix} \left[\hat{\sigma}^2 (X'X)^{-1}\right] \begin{bmatrix} 0 & 0 \\ \frac{1}{1+\hat{\beta}_2} & 0 \\ -\frac{\hat{\beta}_1}{(1+\hat{\beta}_2)^2} & -\frac{1}{(1+\hat{\beta}_2)^2} \end{bmatrix}$$