Delta Method and Method of Moment

Luo Fangzhi

Sun Yat-sen University

Content

Delta Method

If we know $T_n \to_w \theta$, then for any continuous map ϕ, $\phi(T_n) \to \phi(\theta)$.
If we know $T_n \to_w \theta$. Then for any differentiable map ϕ, with some regularization conditions, Delta method ~~allows~~shows,

$$\sqrt{n}(\phi(T_n) - \phi(\theta)) \to_w \phi'(\theta)\sqrt{n}(T_n - \theta).$$

Here's a small example :
$X_1, \cdots, X_n$ are i.i.d. samples from a random variable X, satisfying $E(X^4) < \infty$.
$S^2 \triangleq \frac{1}{n}\sum_i (X_i - \bar{X})^2 \triangleq \phi(\overline{X}, \overline{X^2})$, where $\phi(x, y) = y - x^2$. By CLT,

$$\sqrt{n}\left(\begin{pmatrix} \overline{X} \\ \overline{X^2} \end{pmatrix} - \vec{\boldsymbol{\mu}}\right) \to_w \mathcal{N}(\vec{0}, \Sigma) \triangleq \begin{pmatrix} T_1 \\ T_2 \end{pmatrix},$$

where $\vec{\boldsymbol{\mu}} = (\alpha_1, \alpha_2)$, then by the Delta method,

$$\sqrt{n}(\phi(\overline{X}, \overline{X^2}) - \phi(\vec{\boldsymbol{\mu}})) \to_w \sqrt{n}\phi'(\vec{\boldsymbol{\mu}}) \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = -2\alpha_1 T_1 + T_2.$$

Definition

$\phi : \mathbb{R}^k \mapsto \mathbb{R}^m$ is differentiable at $\boldsymbol{\theta} \in \boldsymbol{R}^k$ if there exists a $m \times k$ matrix (linear map) $\phi'(\theta) : \boldsymbol{R}^k \mapsto \boldsymbol{R}^m$ such that

$$\phi(\boldsymbol{\theta} + \boldsymbol{h}) - \phi(\boldsymbol{\theta}) = \phi'(\boldsymbol{\theta})\boldsymbol{h} + o(\|\boldsymbol{h}\|).$$

Thus we can write

$$\phi'(\theta) = \begin{pmatrix} \frac{\partial \phi_1}{\partial \theta_1}(\boldsymbol{\theta}) & \cdots & \frac{\partial \phi_1}{\partial \underline{x_k} \theta_k}(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ \frac{\partial \phi_m}{\partial \theta_1}(\boldsymbol{\theta}) & \cdots & \frac{\partial \phi_m}{\partial \theta_k}(\boldsymbol{\theta}) \end{pmatrix}, \phi(\boldsymbol{\theta} + \boldsymbol{h}) - \phi(\boldsymbol{\theta}) = \phi'(\boldsymbol{\theta})\boldsymbol{h} + o(\|\boldsymbol{h}\|).$$

where the elements in the matrix are the so-called partial derivatives. Note that here the definition of derivative is different from the ordinary one. $\phi'(\theta)$ is a matrix (linear mapping), recalled the defintion of directional derivative in multivariable calculus.

We are now ready to state the general Delta method.

Theorem

Conditions :

- *Let* $\phi : \mathbb{R}^k \mapsto \mathbb{R}^m$ *be a map defined on a subset of* $\mathbb{R}^k$.
- ϕ *is differentiable at* θ.
- *Let* T_n *be random vectors taking their values in the domain of* ϕ. *Suppose* $r_n\left(T_n-\theta\right) \rightarrow_w T$ *for numbers* $r_n \rightarrow \infty$.
- $T = O_P(1)$

then

- $r_n\left(\phi\left(T_n\right)-\phi(\theta)\right) \rightarrow_w \phi^{\prime}(\theta) T$.
- $\phi^{\prime}(\theta)\left(r_n\left(T_n-\theta\right)\right) \rightarrow_w \phi^{\prime}(\theta) T$.

Proof :

Delta method can be used in three way :

- As an approximation of variance.
- Bias correction for the expectation of a function of a random variable,
- Limiting distribution of a function of a random variable.

Let X be a random variable with $E(X) = \mu$ and variance $V(X) = \sigma^2$. Let $Y = f(X)$. Assume f is p^{th} order differentiable at μ, then by Taylor expansion,

$$f(X) = \sum_{i=0}^{p} \frac{f^{(i)}(\mu)(X-\mu)^i}{i!} + o_P((X-\mu)^p).$$

- Variance approximation : $f(X) - f(\mu) \approx f'(\mu)(X-\mu)$, $Var(Y) \approx [f'(\mu)]^2\sigma^2$,
- Bias reduction : $E(Y) \approx f(\mu) + \frac{1}{2}f''(\mu)\sigma^2$,
- Limiting distribution : If $\sqrt{n}(X_n - \mu) \to_w \mathcal{N}(0, \sigma^2)$, then

$$\sqrt{n}(f(X_n) - f(\mu)) \to_w \mathcal{N}(0, [f'(\mu)]^2\sigma^2).$$

Method of Moments

Let $X_1, \cdots, X_n$ be samples from a distribution P_θ,$\theta \in \Theta$. The method of moments consists of estimating θ by the solution of a system of equations :

$$\frac{1}{n}\sum_i f_j(X_i) = \mathbb{E}_\theta f_j(X). \quad j = 1, \cdots, k,$$

for given functions $f_1, \cdots, f_k$. The choices $f_j(x) = x^j$ lead to the method of moments in its simplest form. Usually, for k parameters, we use k moments.

Let $\mathbb{P}_n$ be the empirical measure of $X_1, \cdots, X_n$, P_θ be the population measure, which is defined as $\mathbb{P}_n f = \frac{1}{n}\sum_i f(X_i)$, $P_\theta(f) = \mathbb{E}_\theta(f(X))$. Let
$e : \Theta \mapsto \mathbb{R}^k, e(\theta) = P_\theta f$.
The moment estimator $\hat{\theta}_n$ solves the system of equations

$$\mathbb{P}_n f \equiv \frac{1}{n}\sum_i f(X_i) = e(\theta) \equiv P_\theta f.$$

MM

Intuitively, this is right because $\mathbb{P}_n$ converges to P_θ. Further, if e is somehow invertible, maybe also differentiable, then $\hat{\theta}_n = e^{-1}(\mathbb{P}_n f)$. What's more,

$$\sqrt{n}(\hat{\theta}_n - \theta) = \sqrt{n}(e^{-1}(\mathbb{P}_n f) - e^{-1}(P_\theta f)), \tag{1}$$

if $\mathbb{P}_n f$ is asymptotically normal, then by delta method, $\sqrt{n}(\hat{\theta}_n - \theta_n)$ is asymptotically normal.

Theorem

Conditions :

- *Suppose that* $e(\theta) = P_\theta f$ *is one-to-one on an open set* $\Theta \subset \mathbb{R}^k$.
- $e(\theta)$ *is continuously differentiable at* θ_0 *with nonsingular derivative* $e'(\theta_0)$. *Note that* $e'(\theta_0) : \mathbb{R}^k \mapsto \mathbb{R}^k$.
- $\mathbb{E}_{\theta_0}(f(X)^T f(X)) < \infty$.

Then

- *Then moment estimators* $\hat{\theta}_n = e^{-1}(\mathbb{P}_n f)$ *exists with probability tending to one.*
- *If* $\hat{\theta}_n$ *exists,*
  $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \to_w \mathcal{N}\left(0, (e^{-1})'(\theta_0)\mathbb{E}_{\theta_0}(f(X)f(X)^T)\left((e^{-1})'(\theta_0)\right)^T\right)$.

Proof :

GMM

Recall MM, consider a very simple example. Assume the population distribution has unknown mean μ and variance equal to one.

Assume we have i.i.d. samples $X_1, \cdots, X_n$ from X, then the simplest MM for μ is $\frac{1}{n}\sum_i X_i$.

Of course, we can define other f instead of $f(x) = x$ above. For any f satisfying the regularizaiton condition, we can obtain a MM by solving

$$\frac{1}{n}\sum_i f(x_i, \mu) = E(f(X, \mu)).$$

Suppose we believe to know that the samples are from a Poisson distribution with parameter λ. Then another moment condition would be $E[X_i^2] - \lambda^2 - \lambda = 0$. The MM should satisfied :

$$\begin{pmatrix} \frac{1}{n}\sum_i X_i - \hat{\lambda} \\ \frac{1}{n}\sum_i (X_i^2) - \hat{\lambda} - \hat{\lambda}^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{2}$$

GMM

Here, we use two equation to estimate one paramter (over-identify), thus we called this Generalized method of moment (GMM). By GMM, we can make use of more information of moments to obtain a better estimation.

Definition

Suppose that $e(\theta) = P_\theta f$ is one-to-one on an open set $\Theta \subset \mathbb{R}^k$ and continuously differentiable at θ_0 with nonsingular derivative $e'(\theta_0)$. For simplicity, we assume $E[f(x_i, \theta_0)] = 0$. $f_n(\theta)$ is the corresponding sample counterparts. We define the criterion function $Q_n(\theta)$ for GMM as :

$$Q_n(\theta) = f_n(\theta)^T W_n f_n(\theta), \tag{3}$$

where W_n, the weighting matrix, converges to a positive definite matrix W as n grows large. The the GMM estiamtor of θ_0 is given by

$$\hat{\theta} = \arg\min_{\theta \in \Theta} Q_n(\theta). \tag{4}$$

Remark that, W_n plays a role of precision matrix, used to adjust the correlation between different entries of f.

Suppose we are interested in the correlation of earning potentials and compulsory schooling laws and estimated the following equation :

$$\ln(w) = \beta_0 + \beta_1 ed + \text{ controls } + u,$$

where w denotes the earning potentials, ed denotes the length of compulsory schooling laws, and controls denote some covariates correlated to w.

Estimating this linear equation by OLS could be biased and inconsistent as ed_i is probably correlated with individual factors in the regression error term u_i such as individual costs and potential benefits of schooling or other options outside the schooling system, most of which are unobserved by the researcher.

Using the structure of compulsory school attendance laws at that time in the US they were able to argue that (in addition to the controls) dummy variables indicating the quarter of birth for each individual could be used to instrument for the years spent in education. Their exogeneity assumption implies that the following population moment conditions hold :

$$E\left[z_i\left(\ln\left(w_i\right)-\beta_0-\beta_1 ed_i-\text{ controls }\right)\right]=0$$

where the vector of instruments z_i contains the exogenous variables from the original model supplemented by the quarter of birth dummies. Note that there are more moment conditions then parameters and we could estimate the model by GMM.