M2RI Asymptotic Statistics Lecture 3: Delta Method and Method of Moments

Proof (of the Delta Method)

By Taylor applied at θ,

$$\phi(\hat{\theta}_n) = \phi(\theta) + (J\phi(\theta))(\hat{\theta}_n - \theta) + o(\|\hat{\theta}_n - \theta\|)$$

So $r_n(\phi(\hat{\theta}_n) - \phi(\theta)) - r_n(J\phi(\theta))(\hat{\theta}_n - \theta) = r_n o(\|\hat{\theta}_n - \theta\|)$.

Then, we can notice that $\hat{\theta}_n - \theta = o_{\mathbb{P}}(1)$.

Indeed, let $M > 0$ and $\varepsilon > 0$.

Since $r_n(\hat{\theta}_n - \theta) \longrightarrow X$, there exists $C_\varepsilon > 0$ such that

$\mathbb{P}(|r_n|\|\hat{\theta}_n - \theta\| \geq C_\varepsilon) < \varepsilon$ for n big enough.

then, $\mathbb{P}(|r_n|\|\hat{\theta}_n - \theta\| \geq M) \leq \mathbb{P}(|r_n|\|\hat{\theta}_n - \theta\| \geq |r_n| M)$

for n big enough ($r_n \to +\infty$)

and for n big enough, $\leq \mathbb{P}(|r_n|\|\hat{\theta} - \theta\| \geq C_\varepsilon)$

$\leq \varepsilon$ for n big enough. •

So $o(\|\hat{\theta}_n - \theta\|) = o_{\mathbb{P}}(\hat{\theta}_n - \theta)$.

Thus, $r_n o(\|\hat{\theta}_n - \theta\|) = o_{\mathbb{P}}(r_n(\hat{\theta}_n - \theta))$.

Furthermore, since $r_n(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} X$, $r_n(\hat{\theta}_n - \theta) = O_{\mathbb{P}}(1)$

(Prohorov)

Thus, $r_n o(\|\hat{\theta}_n - \theta\|) = o_{\mathbb{P}}(O_{\mathbb{P}}(1)) = o_{\mathbb{P}}(1)$.

So, $r_n(\phi(\hat{\theta}_n) - \phi(\theta)) - r_n(J\phi(\theta))(\hat{\theta}_n - \theta) \xrightarrow{\mathbb{P}} 0$

$n \to +\infty$

Furthermore, $r_n(\hat{\theta}_n - \theta) \longrightarrow X$, so by Slutsky

$$\left\{\left(r_n(\phi(\hat{\theta}_n) - \phi(\theta)) - r_n(J\phi(\theta))(\hat{\theta}_n - \theta),\; r_n(\hat{\theta}_n - \theta)\right)\right\} \xrightarrow{\mathcal{L}} (0, X)$$

and by the continuous mapping applied to $x, y \mapsto x + (J\phi(\theta))y$,

$$r_n\left(\phi(\hat{\theta}_n) - \phi(\theta)\right) \longrightarrow \left(J\phi(\theta)\right) X .$$

How to remember this result?

$$r_n\left(\phi(\hat{\theta}_n) - \phi(\theta)\right) \text{``}\approx\text{''} \underbrace{\left(r_n(\hat{\theta}_n - \theta)\right)}_{\approx X} \times \underbrace{\left(\frac{\phi(\hat{\theta}_n) - \phi(\theta)}{\hat{\theta}_n - \theta}\right)}_{\approx J\phi(\theta) \text{ when } n \text{ is big.}}$$

Delta Method for Variance Estimation

$X_1, X_2, \dots \overset{iid}{\sim} X$ with $\mathbb{E}(|X|^4) < +\infty$.

$\mu_1 = \mathbb{E}(X)$ $\quad \forall \ell \in \{2, 3, 4\}$, $\mu_\ell = \mathbb{E}\left((X - \mathbb{E}(X))^\ell\right)$.

$\hat{\mu}_{1,n} = \frac{1}{n} \sum_{i=1}^{n} X_i$; $\hat{\mu}_{2,n} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu}_{1,n})^2$

By formula for the variance and invariance of the variance by translation,

$$\hat{\mu}_{2,n} = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} X_i\right)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_1)^2 - \left(\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_1)\right)^2$$

Let's note $Y_i = \begin{pmatrix} X_i - \mu_1 \\ (X_i - \mu_1)^2 \end{pmatrix}$.

We have $Cov(Y_i) = \begin{pmatrix} \mu_2 & \mu_3 \\ \mu_3 & \mu_4 - \mu_2^2 \end{pmatrix}$,

By the central limit theorem,

$$\sqrt{n}\left(\overline{Y_n} - \begin{pmatrix} 0 \\ \mu_2 \end{pmatrix}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \begin{pmatrix} \mu_2 & \mu_3 \\ \mu_3 & \mu_4 - \mu_2^2 \end{pmatrix}\right)$$

Furthermore, by noting $\phi(x,y) = y - x^2$, we have

$$\hat{\mu}_{2,n} = \phi\left(\left(\overline{Y_n}\right)_1, \left(\overline{Y_n}\right)_2\right) \text{ and } \mu_2 = \phi(0, \mu_2)$$

Thus,

$$\sqrt{n}\left(\hat{\mu}_{2,n} - \mu_2\right) = \sqrt{n}\left(\phi\left(\overline{Y_n}\right) - \phi(0,\mu_2)\right)$$

$$\xrightarrow{\mathcal{L}} \underbrace{\underbrace{J\phi(0,\mu_2)}_{=(0,\ 1)} \mathcal{N}\left(0, \begin{pmatrix} \mu_2 & \mu_3 \\ \mu_3 & \mu_4 - \mu_2^2 \end{pmatrix}\right)}_{= \mathcal{N}(0,\ \mu_4 - \mu_2^2)}$$

Method of Moments - Gaussian Model

$\Theta = \mathbb{R} \times [0, +\infty)$, $\theta = (m, \sigma^2)$, $\mathcal{L}_\theta = \mathcal{N}(m, \sigma^2)$

$f_1(x) = x \qquad f_2(x) = x^2$

$$\mathbb{E}_{X \sim \mathcal{L}_\theta}\left(f_1(X)\right) = m$$
$$\mathbb{E}_{X \sim \mathcal{L}_\theta}\left(f_2(X)\right) = m^2 + \sigma^2$$

The estimator of moments solves

$$\boxed{\frac{1}{n}\sum_{i=1}^{n} X_i = \hat{m}_n \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \hat{m}_n^2 + \hat{\sigma}_n^2}$$

$$\Rightarrow \hat{m}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \hat{\sigma}_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - m_n^2$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - m_n\right)^2$$

<u>Proof - Method of moments</u>

$Je(\theta_0)$ is invertible.

Local inver. th.

$\Rightarrow$ $\exists\, U$ neighborhood of θ_0, V neighborhood of $e(\theta_0)$ st

$$e : U \longrightarrow V$$

is a C^1 diffeomorphism with, $\forall u \in U$ $(v = e(u))$

$$\left(Je^{-1}\right)(v) = \left(Je(u)\right)^{-1}.$$

<u>Law of large numbers</u> $\Rightarrow e_n := \begin{pmatrix} \frac{1}{n}\sum_i f_1(X_i) \\ \vdots \\ \frac{1}{n}\sum_i f_p(X_i) \end{pmatrix} \xrightarrow{\mathbb{P}} e(\theta_0).$

so $\mathbb{P}(e_n \in V) \xrightarrow[n \to +\infty]{} 1$

Let's define $\hat{\theta}_n = \begin{cases} e^{-1}(e_n) & \text{if } e_n \in V \\ \text{arbitrary} & \text{if } e_n \notin V. \end{cases}$

$\hat{\theta}$ is a solution of the MoM problem with probability $\to 1$.

Let's also define $\tilde{e}_n = \begin{cases} e_n & \text{if } e_n \in V \\ e(\theta_0) & \text{if } e_n \in V \end{cases}$

and observe that for $\varepsilon > 0$

$$\mathbb{P}\left(\left\| \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) - \sqrt{n}\left(e^{-1}(\tilde{e}_n) - e^{-1}(e(\theta_0))\right) \right\| \geq \varepsilon \right)$$

$$\leq \mathbb{P}(e_n \notin V) \xrightarrow[n \to +\infty]{} 0. \qquad *$$

<u>Finally</u>, $\sqrt{n}\left(e^{-1}(\tilde{e}_n) - e^{-1}(e(\theta_0))\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, (J_e(\theta_0))^{-1} \Sigma_f (J_e(\theta_0))^{-T}\right).$

(CLT + Approx CV in prob + Delta Method).

And thus, because of $*$ and the approximation by another sequence converging in proba,

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, (J_e(\theta_0))^{-1} \Sigma_f (J_e(\theta_0))^{-T}\right).$$