STAT 620: Asymptotic Statistics **Spring 2022**

Lecture: Feb 1

Lecturer: Xianyang Zhang

1 Moment Estimator

1.1 Theorem

Suppose that $P_\theta f = e(\theta)$ is one-to-one on an open set $\Theta \subset \mathbb{R}^k$ and continuously differentiable at θ_0 with nonsingular derivative $De(\theta) \in \mathbb{R}^{k\times k}$. Assume $P_{\theta_0}\|f\|^2 < \infty$. Then the moment estimator $\hat{\theta}_n$ exists with probability tending to one and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to^d N(0, (De(\theta_0))^{-1}(P_{\theta_0}ff^\top - P_{\theta_0}fP_{\theta_0}f^\top)((De(\theta_0))^{-1})^\top).$$

1.2 Proof

Here we illustrate only the main steps of the proof.

Step 1: The continuity of $De(\theta)$ and nonsingularity at θ_0 imply nonsingularity in a neighborhood of θ_0 . There exist a neighborhood of θ_0 (say U) and a neighborhood of $P_{\theta_0}f$ (say V) such that $e : U \to V$ is differentiable, bijective with a differentiable inverse $e^{-1} : V \to U$. Note $P_nf \to^p P_{\theta_0}f$ and thus $P_{\theta_0}(P_nf \in V) \to 1$. Therefore $\hat{\theta}_n = e^{-1}(P_nf)$ exists with probability tending to one.

Step 2: $\sqrt{n}(P_nf - P_{\theta_0}f) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (f(x_i) - E(f(x_i)) \to^d N(0, P_{\theta_0}ff^\top - P_{\theta_0}fP_{\theta_0}f^\top)$

Step 3: Apply the delta method.

1.3 Example

Let $X_1, \ldots, X_n \sim \text{Gamma}(\alpha, \lambda)$. Then using the method of moments, we have for the first and the second moment

$$\alpha_1 = E(X_1) = \frac{\alpha}{\lambda},\ \alpha_2 - \alpha_1^2 = \text{var}(X) = \frac{\alpha}{\lambda^2}.$$

Thus we have

$$\lambda = \frac{\alpha_1}{\alpha_2 - \alpha_1^2},\ \alpha = \frac{\alpha_1^2}{\alpha_2 - \alpha_1^2}.$$

Therefore we have

$$\hat{\lambda} = \frac{\bar{X}}{\overline{X^2} - (\bar{X})^2},\ \hat{\alpha} = \frac{(\bar{X})^2}{\overline{X^2} - (\bar{X})^2}$$

and

$$\sqrt{n}\left(\begin{pmatrix}\hat{\lambda}\\ \hat{\alpha}\end{pmatrix} - \begin{pmatrix}\lambda\\ \alpha\end{pmatrix}\right) \xrightarrow{d} N(0, \Sigma).$$

2 Taylor Expansions

1. For $f : \mathbb{R}^d \to \mathbb{R}$ differentiable at $x \in \mathbb{R}^d$

$$f(y) = f(x) + \nabla f(x)^\top (y - x) + o(\|y - x\|)$$

and from the Mean Value Theorem

$$f(y) = f(x) + \nabla f(\tilde{x})^\top (y - x),$$

where $\tilde{x}$ is between x and y.

2. Let $f : \mathbb{R}^d \to \mathbb{R}^k$ with $f = (f_1, \ldots, f_k)^\top$. Denote

$$Df(x) = \begin{pmatrix} \nabla f_1(x)^\top \\ \vdots \\ \nabla f_k(x)^\top \end{pmatrix} \in \mathbb{R}^{k \times d}.$$

Then

$$f(y) = f(x) + Df(x)(y - x) + o(\|y - x\|).$$

2.1 Example

We illustrate why no mean value theorem (M.V.) holds for Case (2). Let $f : \mathbb{R} \to \mathbb{R}^k$ with $f(x) = (x, x^2, \ldots, x^k)^\top$. Then

$$Df(x) = \begin{pmatrix} 1 \\ 2x \\ \vdots \\ kx^{k-1} \end{pmatrix}.$$

Assuming M.V. holds, then we must have

$$f(y) = f(x) + Df(\tilde{x})(y - x).$$

Taking $y = 1, x = 0$, then we should have $f(y) - f(x) = f(1) - f(0) = Df(\tilde{x})$, but

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ k\tilde{x}^{k-1} \end{pmatrix}$$

does not have a solution.

3 A useful lemma

For $f : \mathbb{R}^d \to \mathbb{R}^k$ differentiable, assume that Df is L-Lipschitz and

$$f(y) = f(x) + Df(x)(y - x) + R(y - x),$$

where R is a remainder matrix. Then $\|R\|_{op} \leq \frac{L}{2}\|y - x\|$ and thus $\|R(y - x)\| \leq \frac{L}{2}\|y - x\|^2$

3.1 L-Lipschitz function

Recall that $g : \mathbb{R} \to \mathbb{R}$ is L-Lipschitz if $|g(x) - g(y)| \leq L|x - y|$ and $f : \mathbb{R}^d \to \mathbb{R}^{k \times d}$ is L-Lipschitz if

$$\|f(x) - f(y)\|_{op} \leq L\|x - y\|.$$

3.2 Operator Norm

For a matrix $A \in \mathbb{R}^{k \times d}$, define the operator norm

$$\|A\|_{op} = \sup_{\|u\|=1} \|Au\|.$$

Note that for $x \in \mathbb{R}^d$, we have $\|Ax\| \leq \|A\|_{op}\|X\|$.

3.3 Proof

Define $\phi_i(t) = f_i((1-t)x + ty)$ for $1 \leq i \leq k$. Note that $\phi_i(0) = f_i(x), \phi_i(1) = f_i(y)$ and $\phi_i'(t) = \nabla f_i((1-t)x + ty)^\top (y - x)$. Then

$$Df((1-t)x + ty)(y - x) = \begin{pmatrix} \nabla f_1((1-t)x + ty)^\top \\ \vdots \\ \nabla f_k((1-t)x + ty)^\top \end{pmatrix} (y - x) = \begin{pmatrix} \phi_1' \\ \vdots \\ \phi_k' \end{pmatrix}.$$

Let $\phi = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_k \end{pmatrix}$. Then we have

$$f(y) - f(x) = \int_0^1 d\phi(t) = \int_0^1 Df((1-t)x + ty)(y - x)dt = Df(x)(y - x) + \int_0^1 (Df((1-t)x + ty) - Df(x))(y - x)dt.$$

Thus

$$\begin{aligned} \|R(y - x)\| &= \left\| \int_0^1 (Df((1-t)x + ty) - Df(x))(y - x)dt \right\| \\ &\leq \int_0^1 \|Df((1-t)x + ty) - Df(x)\|_{op}\|y - x\|dt \leq \int_0^1 Lt\|y - x\|^2 dt = \frac{L}{2}\|y - x\|^2. \end{aligned}$$