The Delta Method

John Duchi

Stats 300b – Winter Quarter 2021

Outline

- Delta-method
- Higher-order delta methods
- Examples

Reading: van der Vaart, Chapter 3.

The delta method

motivating question: suppose we know a sequence of statistics T_n estimates a parameter θ at some rate, say

$$r_n(T_n - \theta) \xrightarrow{d} T.$$

Can we say something about $\phi(T_n)$ for (smooth) mappings ϕ?

The delta method

Theorem

Let $r_n \to \infty$ be deterministic and $\phi : \mathbb{R}^d \to \mathbb{R}^k$ be differentiable at θ. Assume $r_n(T_n - \theta) \xrightarrow{d} T$ for some random vector T. Then

(1) $r_n(\phi(T_n) - \phi(\theta)) \xrightarrow{d} \phi'(\theta) T$

(2) $r_n\big(\phi(T_n) - \phi(\theta)\big) - r_n\phi'(\theta)(T_n - \theta) \xrightarrow{p} 0$

where $\phi'(\theta) \in \mathbb{R}^{k\times d}$ is the Jacobian matrix with entries $[\phi'(\theta)]_{ij} = \frac{\partial \phi_i(\theta)}{\partial \theta_j}$

Proof of Delta-method

The limiting distribution of a quadratic

Example (The delta method for quadratics)

Assume $X_i \stackrel{\text{iid}}{\sim} P$ with $\mathbb{E}[X] = \theta \neq 0$, $\text{Cov}(X) = \Sigma$, and set $\phi(h) = \frac{1}{2} \|h\|_2^2$. Then

$$\sqrt{n} \left(\frac{1}{2} \left\| \frac{1}{n} \sum_{i=1}^{n} X_i \right\|_2^2 - \frac{1}{2} \|\theta\|_2^2 \right) \stackrel{d}{\to} \mathcal{N}\left(0, \theta^T \Sigma \theta\right)$$

Sample variance

Example (Delta method for sample variance)

For X_i i.i.d. with $\mathrm{Var}(X_i) = \sigma^2$ and $\mathbb{E}[X_i^4] < \infty$, let

$$S_n^2 := \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}_n^2.$$

Then for $\phi(x, y) = y - x^2$ we have $S_n^2 = \phi(\overline{X}_n, \overline{X^2}_n)$, and

$$\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}[X^4] - \mathbb{E}[X^2]^2\right) \stackrel{\text{dist}}{=} \mathcal{N}\left(0, \mathrm{Var}(X^2)\right).$$

Higher-order delta methods

insight: the delta method is just a Taylor expansion, so if $\phi'(\theta) = 0$, we may consider higher-order terms. Consider $\phi : \mathbb{R}^d \to \mathbb{R}$ for simplicity (in notation)

Corollary

Let $r_n \to \infty$ be deterministic and assume $r_n(T_n - \theta) \stackrel{d}{\to} T$, and let ϕ be twice continuously differentiable at θ. Then

$$r_n^2(\phi(T_n) - \phi(\theta)) \stackrel{d}{\to} \frac{1}{2} T^\top \nabla^2 \phi(\theta) T$$

Proof of higher-order delta method

Example: KL divergences and log-likelihood ratios

recall KL-divergence between distributions

$$D_{\text{kl}}\left(P\|Q\right) := \int dP \log \frac{dP}{dQ} = \int p \log \frac{p}{q} d\mu$$

Example

Let $X_i \in \{0,1\}$, $X_i \sim P_\theta := \text{Bernoulli}(\theta)$ (i.e. $\mathbb{E}[X_i] = \theta$). For $\widehat{\theta}_n = \frac{1}{n}\sum_{i=1}^n X_i$,

$$nD_{\text{kl}}\left(P_{\widehat{\theta}_n}\|P_\theta\right) \stackrel{d}{\to} \frac{1}{2}W^2 \quad \text{and} \quad nD_{\text{kl}}\left(P_\theta\|P_{\widehat{\theta}_n}\right) \stackrel{d}{\to} \frac{1}{2}W^2$$

for $W \sim \mathcal{N}(0,1)$