Notes on Asymptotic Analysis

Jungbin Jun

1 Information Inequality

Theorem 1. *Let $\eta_n := \eta_n(X_1, \dots, X_n)$ be estimator for $\tau := \tau(\theta)$. Then, under R0-R5,*

$$Var_\theta(\eta_n) \geq \nabla \mathbb{E}_\theta(\eta_n)^\top \{nI(\theta)\}^{-1} \nabla \mathbb{E}_\theta(\eta_n)$$

Proof. Note that we have

$$\begin{aligned} Cov_\theta(\eta_n, S_n(\theta)) &= \mathbb{E}_\theta[\eta_n S_n(\theta)^\top] \\ &= \int \eta_n (\nabla \ell_n(\theta))^\top \prod_{i \in [n]} f_\theta(x_i) dx_1 \dots dx_n \\ &= \int \eta_n (\nabla \ell_n(\theta))^\top \exp(\ell_n(\theta)) dx_1 \dots dx_n \\ &= \int \eta_n \{\nabla \exp(\ell_n(\theta))\}^\top dx_1 \dots dx_n \\ &= \nabla \{ \int \eta_n \exp(\ell_n(\theta) dx_1 \dots dx_n \} \\ &= \nabla \mathbb{E}_\theta[\eta_n]^\top \end{aligned}$$

The best linear predictor of η_n in terms of $S_n(\theta)$ is

$$\begin{aligned} &\mathbb{E}_\theta(\eta_n) + Cov_\theta(\eta_n, S(\theta)) \{Var_\theta[S(\theta)]\}^{-1} \{S_n(\theta) - \mathbb{E}_\theta[S_n(\theta)]\} \\ &= \mathbb{E}_\theta(\eta_n) + \nabla \mathbb{E}_\theta[\eta_n]^\top \{nI(\theta)\}^{-1} S_n(\theta) \\ &= a + B\, S_n(\theta) \end{aligned}$$

The variance of the difference of η_n and its best linear predictor is then

$$\begin{aligned} &Var_\theta[\eta_n - \mathbb{E}_\theta(\eta_n) - \nabla \mathbb{E}_\theta[\eta_n]^\top \{nI(\theta)\}^{-1} S_n(\theta)] \\ &= Var_\theta(\eta_n) + B\, Var_\theta(S_n(\theta)) B^\top - Cov_\theta(\eta_n, S_n(\theta)) B^\top - B\, Cov_\theta(S_n(\theta), \eta_n) \\ &= Var_\theta(\eta_n) - \nabla \mathbb{E}_\theta[\eta_n]^\top \{nI(\theta)\}^{-1} \nabla \mathbb{E}_\theta[\eta_n] \geq 0 \end{aligned}$$

□

Corollary 1. *If η_n is an unbiased estimator for $\tau := \tau(\theta)$,*

$$Var_\theta(\eta_n) \geq \nabla \tau(\theta)^\top \{nI(\theta)\}^{-1} \nabla \tau(\theta)$$

Proof. It is straightforward by the fact that $\mathbb{E}_\theta[\eta_n] = \tau(\theta)$ □

Corollary 2 (Cramer-Rao Inequality)**.** *If $\hat{\theta}_n$ is an unbiased estimator for θ,*

$$Var_\theta(\theta_n) \succeq \{nI(\theta)\}^{-1}$$

Proof. Say $\eta := c^\top \theta$ for any c, and consider $\eta_n = c^\top \hat{\theta}_n$. Then, by the Theorem 1, we have

$$Var_\theta(\eta_n) = c^\top Var_\theta(\hat{\theta}_n) c \geq c^\top \{nI(\theta)\}^{-1} c,$$

which gives us the desired result. □

2 Big/Little-Oh in a Probability Language

Let $\{X_n\}$ and $\{R_n\}$ be sequences of random variables, and $\{a_n\}$ be a corresponding sequences of constants.

Definition 1 (Big-O-p Notations)**.** We say X_n is bounded in probability and denoted $X_n := O_p(R_n)$ if

$$\forall \varepsilon > 0,\ \exists M < \infty,\ \exists N \in \mathbb{N} \text{ s.t. } \forall n \geq N,\ \mathbb{P}\big(|X_n/R_n| > M\big) < \varepsilon.$$

Below statement is equivalent.

$$\lim_{k\to\infty} \sup_{n} \mathbb{P}(|X_n/R_n| > k) = 0$$

Definition 2 (Little-O-p Notations)**.** We denote $X_n := o_p(R_n)$ if

$$\left|\frac{X_n}{R_n}\right| \xrightarrow{P} 0$$

It suffices to work with the orders $O_p(1)$ and $o_p(1)$. Indeed, to compare two sequences of random variables (X_n) and (R_n), one can reduce the problem to the ratio X_n/R_n (whenever $R_n \neq 0$ with high probability) and then study whether this ratio is $O_p(1)$ or $o_p(1)$.

Theorem 2. *Let X_n, Y_n and Z_n be random variables such that*

$$X_n = O_p(1), \qquad Y_n = O_p(1), \qquad Z_n = o_p(1).$$

Then, as $n \to \infty$,

1. $X_n + Y_n = O_p(1)$,
2. $X_n Y_n = O_p(1)$,
3. $Z_n X_n = o_p(1)$,
4. $Z_n = O_p(1)$,
5. $Z_n + X_n = O_p(1)$. *(This follows immediately from 4.)*

Theorem 3. *If $f(x) = o(g(x))$ as $x \to a$ and $X_n \xrightarrow{P} a$, then $f(X_n) = o_p(g(X_n))$*

Proof. Let $\epsilon_1, \epsilon_2 > 0$ be given. By the assumption, we have

$$\exists \delta > 0 \text{ s.t } |x - a| < \delta \Rightarrow |f(x)/g(x)| < \epsilon_1$$
$$\exists N \in \mathbb{N} \text{ s.t. } n \geq N \Rightarrow \mathbb{P}(|X_n - a| \geq \delta) \leq \epsilon_2.$$

Thus,

$$\frac{f(x)}{g(x)} \geq \epsilon_1 \Rightarrow |x - a| \geq \delta$$

Substitute X_n for x and take probability on both side. Then,

$$\mathbb{P}\left(\frac{f(X_n)}{g(X_n)} \geq \epsilon_1\right) \leq \mathbb{P}\left(|X_n - a| \geq \delta\right) \leq \epsilon_2$$

□