PH 712 Probability and Statistical Inference

Part VII: Evaluating Estimators II (for Large Samples)

Zhiyang Zhou (zhou67@uwm.edu, zhiyanggeezhou.github.io)

2024/11/06 14:21:31

Well-known (but NOT required) identities

- Laws of large numbers (LLN, CB Thm 5.5.2 & 5.5.9): if $X_1, \ldots, X_n$ are iid with finite mean μ, then $\bar{X} \approx \mu$ as $n \to \infty$.
  - The "$\approx$" notation is abused here and, is supposed to be "$\xrightarrow{p}$" (convergence in probability): $\bar{X} \xrightarrow{p} \mu \Leftrightarrow$ for each $\varepsilon > 0$, $\lim_{n\to\infty} \Pr(|\bar{X} - \mu| > \varepsilon) = 0$;
  - A sufficient condition for $\bar{X} \xrightarrow{p} \mu$: as $n \to \infty$, $\mathrm{E}(\bar{X}) \to \mu$ and $\mathrm{var}(\bar{X}) \to 0$.
- Central limit theorem (CLT, CB Thm 5.5.15): if $X_1, \ldots, X_n$ are iid with finite mean μ and finite variance σ^2, then as $n \to \infty$,
$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \approx \mathcal{N}(0, 1).$$
  - A normal approximation to the distribution of $\bar{X}$ (regardless the distribution of each X_i): $\bar{X} \approx \mathcal{N}(\mu, \sigma^2/n)$
  - The "$\approx$" notation is abused too and is supposed to be "$\xrightarrow{d}$" (convergence in distribution): $\sqrt{n}(\bar{X} - \mu)/\sigma \xrightarrow{d} \mathcal{N}(0, 1)$ means that the limiting distribution of $\sqrt{n}(\bar{X} - \mu)/\sigma$ is $\mathcal{N}(0, 1)$.

Consistency (or consistence, CB Sec 10.1.1)

- A statistic T_n is consistent for $g(\theta)$ if and only if $T_n \approx g(\theta)$ as $n \to \infty$.

Example Lec7.1

- Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with given μ and unknown σ^2. Please check the consistency of the following estimators of σ^2.
  1. $T_n = n^{-1} \sum_i (X_i - \mu)^2$
  2. $W_n = (n-1)^{-1} \sum_i (X_i - \mu)^2$

Asymptotic efficiency

- (CB Def 10.1.11) T_n is *asymptotically efficient* for $g(\theta)$ if and only if $\sqrt{n}\{T_n - g(\theta)\} \approx \mathcal{N}(0, I_1^{-1}(\theta)\{g'(\theta)\}^2)$
  - Where $I_1(\theta)$ is the Fisher information with $n = 1$
    - For an iid sample, $I_1(\theta) = n^{-1} I_n(\theta)$, no longer a function of n
  - Roughly speaking, when n is large enough, a asymptotically efficient T_n is expected to follow $\mathcal{N}(g(\theta), I_n^{-1}(\theta)\{g'(\theta)\}^2)$
- (CB Def 10.1.16 & HMC Def 6.2.3(c)) Denote by T_n and W_n two estimators for $g(\theta)$. Suppose that
$$\sqrt{n}\{T_n - g(\theta)\} \approx \mathcal{N}(0, \sigma_T^2) \quad \text{and} \quad \sqrt{n}\{W_n - g(\theta)\} \approx \mathcal{N}(0, \sigma_W^2).$$

The *asymptotic relative efficiency* (ARE) of T_n with respect to W_n is defined as

$$\mathrm{ARE}(T_n, W_n) = \sigma_W^2 / \sigma_T^2.$$

- T_n is asymptotically more efficient than W_n if and only if $\mathrm{ARE}(T_n, W_n) > 1$

Example Lec7.2

- Suppose $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with given μ and unknown σ^2. Please check the asymptotically efficiency of the following estimators of σ^2.
  1. $T_n = n^{-1} \sum_i (X_i - \mu)^2$
  2. $W_n = (n-1)^{-1} \sum_i (X_i - \mu)^2$