Chapter 5: Micellaneous Topics

In this chapter we explore the large sample properties of an estimator focusing on the concept of consistency for point estimators. We then consider the bootstrap method for precision estimation such as estimating the standard errors and bootstrap c.i.s

1 Consistency of Estimators

Reading Assignment: Sections 5.5.1 and 5.5.2, Handout 6, and Section 10.1.1 of the textbook.

So far we have focused on finite-sample criteria or properties of an estimator. There is a limitation to what can be developed for a point estimator in the finite-sample paradigm as exact inference can be untractable. Luckily, we may rely on approximations if the sample size is large as often is the case in the era of big data. There are several large sample properties, such as consistency, rate of convergence, asymptotic distributions, efficient etc. We will only have time to explore the most fundamental large sample concept of an estimator, its consistency.

What happens to a point estimator $\hat{\theta}$ as the sample size $n \to \infty$?

Definition 5.1.1 Let $\hat{\theta}_n = W(X_1, \dots, X_n)$ be a statistic, $\hat{\theta}_n$ is called a consistent estimator of θ
$\Leftrightarrow \hat{\theta}_n \xrightarrow{P} \theta$, where $\xrightarrow{P}$ means convergence in probability,
i.e. $\forall \varepsilon$, $P(|\hat{\theta}_n - \theta| > \varepsilon) \xrightarrow{P} 0$. ■

The consistency defined above is also called weak consistency versus “strong consistency” which means $P(\hat{\theta}_n \to \theta) = 1$.

Obviously, strong consistency implies weak consistency (Try to prove it!).

Example 5.1.1 By the Law of Large Number (LLN) $\bar{X} \xrightarrow{P} \mu$, if $\mu < \infty$.
So $\bar{X}$ is a consistent estimator of μ.
Likewise, $m_k(\theta)$, the sample kth moment is a consistent estimator of $\mu_k = E(X^k)$.
$\Rightarrow$ any MoM estimator $q(m_1, \dots, m_k)$ is consistent if q is continuous at θ.
This follows from Theorem 3.8.1 (b) below.

A stronger version by invoking the SLLN implies that any moment estimator is also a strong consistent of θ. ■

For the rest of the section we focus on weak consistency due to its simplicity. Additional convergence concepts and properties are provided in Handout 6 for those interested to learn more.

Theorem 5.1.1 (Convergence in probability is preserved under a continuous transformation)

(a) If a sequence of r.v.s Z_n converges in probability to another r.v. Z (i.e. $Z_n \xrightarrow{P} Z$) and g is continuous $\Rightarrow g(Z_n) \xrightarrow{P} g(c)$.

(b) $Z_n \xrightarrow{P} c$ and g is continuous at $c \Rightarrow g(Z_n) \xrightarrow{P} g(c)$.

(c) $\text{MSE}(\hat{\theta}) = E(\hat{\theta}_n - \theta)^2 \to 0 \Rightarrow \hat{\theta}_n \xrightarrow{P} \theta$.

Here $E(\hat{\theta}_n - \theta)^2 \to 0$ is called convergence in quadratic mean, so convergence in quadratic mean $\Rightarrow$ convergency in probability.

Proof: The proof of (a) is not difficult and is left as a homework problem. Part (b) follows from (a) when the limit $Z = c$ a constant. Below we show the proof of (c).

By Markov's inequality, $P(|\hat{\theta}_n - \theta| > \varepsilon) \le \frac{E(\hat{\theta}-\theta)^2}{\varepsilon^2} \to 0$. ■

HW 5.1 Prove Theorem 5.1.1 (a). A hint for the proof is provided in Exercise 5.39 (a) of the textbook. ■

Remark: To show consistency, it often works by showing $\text{MSE}(\hat{\theta}) = E(\hat{\theta}_n - \theta)^2 \to 0$.

That is, show convergence in quadratic mean first.

To show convergence in q.m., use $E(\hat{\theta}_n - \theta)^2 = \text{bias}^2(\hat{\theta}_n) + \text{Var}(\hat{\theta}_n)$, then show bias $\to 0$ and Var $\to 0 \Rightarrow \text{MSE}(\hat{\theta}_n) \to 0$.

Example 5.1.2 Let $X_1, \ldots, X_n \sim \text{Pois}(\theta)$, $\theta \sim \text{Gamma}(\alpha, \beta) \Rightarrow E(\theta) = \frac{\alpha}{\beta}$.

It can be shown that the Posterior $\pi(\theta \mid \mathbf{x}) \sim \text{Gamma}(\alpha + \sum_i x_i, \beta + n)$.

$\Rightarrow$ Bayes estimator for θ w.r.t sq error loss is $\hat{\theta}$ = mean of $\text{Gamma}(\alpha + \sum_i x_i, \beta + n) = \frac{\alpha + \sum_i x_i}{\beta + n}$.

Note $\hat{\theta} = \frac{\beta}{\beta+n}\frac{\alpha}{\beta} + \frac{n}{\beta+n}\bar{X} \Rightarrow$ Bayes estimator shrinks MLE towards posterior mean.

$$\hat{\theta} = \frac{\alpha}{\beta+n} + \frac{\sum_i X_i}{\beta+n} = \underbrace{\frac{\alpha}{\beta+n}}_{\substack{\downarrow \\ 0}} + \underbrace{\frac{\bar{X}}{\beta/n+1}}_{\substack{\downarrow \\ 1}} \underset{?}{\xrightarrow{P}} \bar{X} \xrightarrow{P} \theta \qquad \text{as } n \to \infty$$

The above is a sketch of the proof but we need to provide a rigorous proof.

[Method 1] Use $\text{MSE}_\theta\,(\hat\theta)$, i.e. convergence in q.m.

$$\begin{aligned}
E_\theta(\hat\theta) &= E(\hat\theta \mid \theta) = \frac{\alpha + nE(X_i)}{\beta+n} = \frac{\alpha+n\theta}{\beta+n}.\\
\Rightarrow \text{bias}_\theta(\hat\theta) &= \frac{\alpha+n\theta}{\beta+n} - \theta = \frac{\alpha-\beta\theta}{\beta+n} \to 0,\ \forall\theta.\\
\text{Var}_\theta(\hat\theta) &= \text{Var}\left(\frac{\bar X}{\beta/n+1} \mid \theta\right) = \left(\frac{1}{\beta/n+1}\right)^2 \text{Var}(\bar X \mid \theta) = \left(\frac{1}{\beta/n+1}\right)^2 \frac{\theta}{n} \to 0,\ \forall\theta.
\end{aligned}$$

$\Rightarrow \text{MSE}(\hat\theta \mid \theta) = \text{MSE}_\theta(\hat\theta) \to 0,\ \forall\theta.$

$\Rightarrow$ For each θ, $\hat\theta \xrightarrow{P} \theta$.

[Method 2] Use properties of convergence in probability as listed in Handout 6.

For each θ fixed, $\bar X \xrightarrow{P} \theta$ by LLN $\Rightarrow \frac{1}{\beta/n+1}\bar X \xrightarrow{P} \theta$, since $c_n = \frac{1}{\beta/n+1} \to 1$.
(See Handout 6, property 5)

$$\begin{aligned}
X_n \xrightarrow{P} X, Y_n \xrightarrow{P} Y \quad &\Rightarrow \quad X_n \pm Y_n \xrightarrow{P} X \pm Y \ \ (Y_n \ \&\ Y \ \text{ could be constant})\\
&\ \ \& \quad X_nY_n \xrightarrow{P} XY\\
&\ \ \& \quad \frac{X_n}{Y_n} \xrightarrow{P} \frac{X}{Y},\ \text{ if } \ P(Y=0)=0.
\end{aligned}$$

Since $\frac{\alpha}{\beta+n} \to 0, \Rightarrow \frac{\alpha}{\beta+n} + \frac{1}{\beta/n+1}\bar X \xrightarrow{P} 0 + \theta = \theta$, for each fixed θ.
$\Rightarrow \hat\theta_n \xrightarrow{P} \theta,\ \forall\theta.$ ■

Example 5.1.3 Let $X_1, \ldots, X_n \sim Exp(\beta) \Rightarrow E(X_i) = \frac{1}{\beta}$. MLE of β is $\frac{1}{\bar X}$.

Since $\bar X \xrightarrow{P} \mu = \frac{1}{\beta}$, $g(\bar X) = \frac{1}{\bar X} \xrightarrow{P} \beta$. ■

Sometimes it is easier to use the definition to prove consistency as the next example shows.

Example 5.1.4 $X_1, \ldots, X_n \sim U(0, \theta)$. Show that the MLE=$X_{(n)}$ is consistent.

Proof:

$$
\begin{aligned}
P(|\hat{\theta} - \theta| > \varepsilon) &= P(\hat{\theta} - \theta > \varepsilon \text{ or } \hat{\theta} - \theta < -\varepsilon) \\
&= P(\underbrace{\hat{\theta} > \theta + \varepsilon}_{\text{impossible}} \text{ or } \hat{\theta} - \theta < -\varepsilon) \\
&= P(\hat{\theta} - \theta < -\varepsilon) \\
&= P(\hat{\theta} < \theta - \varepsilon) \\
&= P(X_i < \theta - \varepsilon, i = 1, \dots, n) \\
&= \prod_i P(X_i < \theta - \varepsilon), \text{ since } X_i \text{ are independent} \\
&= \left(\frac{\theta - \varepsilon}{\theta}\right)^n \\
&\to 0
\end{aligned}
$$

■

Remark. MLE and Bayes estimators are usually consistent (under mild assumptions).

Summary: Methods to show consistency (sometimes you need several of them) :

1. LLN
2. MSE $\to 0$. (This is a higher standard as convergency in q.m $\Rightarrow$ convergence in probability)
3. Use properties of convergence in probability in Handout 6.
4. Use the definition. ■

2 Bootstrap Methods

Reading Assignment: Singh and Xie (2014)