Chapter 10: Asymptotic Evaluation

Samples $X_1, \ldots, X_n$ i.i.d. $f(x|\theta)$, n large. We will see what happens if $n \to \infty$.

- This assumption $n \to \infty$ generally makes life easier.
- Because limit theorems become available, distributions can be found approximately. Limiting distributions are much simpler than actual distributions

The behaviors (including its distribution, bias, variance) of an estimator under $n \to \infty$ are known as its *asymptotic* properties.

1 Consistency

Does the estimator converge/get closer to the parameter when n gets larger and larger?

Definition: Let $T_n = T_n(X_1, \ldots, X_n)$ be a sequence of estimators for $\tau(\theta)$. We say that T_n is **consistent** for estimating $\tau(\theta)$ if

$$T_n \to_p \tau(\theta) \text{ under } \mathrm{P}_\theta, \quad \forall \theta.$$

That is, given any $\epsilon > 0$, $\lim_{n\to\infty} P(|T_n - \tau(\theta)| > \epsilon) = 0$.

How to prove consistency of $\hat{\theta}$ for estimating θ?

- by definition (often complicated)
- Chebychev's Inequality.

$$P(|T_n - \tau(\theta)| > \epsilon) \leq \frac{E[T_n - \tau(\theta)]^2}{\epsilon^2}.$$

Theorem 10.1.3

If T_n is a sequence of estimators of $\tau(\theta)$ satisfying

$$\begin{aligned} \lim_{n\to\infty} \text{Bias}_\theta(T_n) &= 0 \quad \text{(asymptotically unbiased)}, \\ \lim_{n\to\infty} \text{Var}_\theta(T_n) &= 0, \end{aligned}$$

for all θ, then T_n is consistent for $\tau(\theta)$. Here $\text{Bias}_\theta(T_n) = E(T_n) - \theta$.

Theorem 10.1.5

Let T_n be a consistent sequence of estimators of $\tau(\theta)$. Let a_n and b_n be a sequence constant satisfying $a_n \to 1, b_n \to 0$. Then the sequence $U_n = a_n T_n + b_n$ is a consistent estimator of $\tau(\theta)$.

Invariance Principle of Consistency

- If T_n is consistent for θ and g is a continuous function, then $g(T_n)$ is consistent for $g(\theta)$.
- MME (method of moment estimator) is generally consistent.
- MLE (maximum likelihood estimator) is consistent.
  Let $X_1, \ldots, X_n$ be i.i.d. $f(x|\theta)$ and let $\hat{\theta}_n$ be the MLE of θ. Then $\hat{\theta}_n$ is consistent for θ.
- UMVUE (unbiased minimal variance estimator) is consistent.
  Let $X_1, \ldots, X_n$ be i.i.d. $f(x|\theta)$ and let T_n be the UMVUE of $\tau(\theta)$. Then T_n is consistent for $\tau(\theta)$.

Example: Let $X_1, \dots, X_n$ be i.i.d. with $E(X) = \mu$ and $\text{Var}(X) = \sigma^2$, then

(1) $\bar{X}_n$ is consistent for μ;

(2) $S_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2/(n-1)$ is consistent for σ^2.

(3) $\sum_{i=1}^n (X_i - \bar{X}_n)^2/n$ is consistent for σ^2.

2 Asymptotic normality

A statistic T_n is *asymptotically normal* if

$$\sqrt{n}\{T_n - \tau(\theta)\} \to_d N(0, v(\theta)) \quad \text{for all} \quad \theta.$$

- $\tau(\theta)$ is called the *asymptotic mean*;
- $v(\theta)$ is called the *asymptotic variance*.

We write T_n is $AN(\tau(\theta), v(\theta)/n)$,

Central Limit Theorem:

Assume $X_1, \ldots, X_n$ is i.i.d. $f(x|\theta)$, with finite mean $\mu = \mu(\theta)$ and variance $\sigma^2 = \sigma^2(\theta)$. Then

$$\bar{X}_n \quad \text{is} \quad AN(\mu(\theta), \sigma^2(\theta)/n),$$

Delta method:

Assume T_n is $AN(\theta, v(\theta)/n)$. If a function g satisfies that $g'(\theta) \neq 0$, then

$$g(T_n) \quad \text{is} \quad AN(g(\theta), [g'(\theta)]^2 v(\theta)/n).$$

Example: Let $X_1, \ldots, X_n$ be iid $N(\mu, \sigma^2)$, $\mu \neq 0$. Show the MLE of μ^2 is $AN(a, b)$, and specify its asymptotic mean a and variance b.

Example: Let $X_1, \ldots, X_n$ be iid from Bernoulli(p), where $p \neq \frac{1}{2}$. Find the asymptotic mean and variance of $\bar{X}(1 - \bar{X})$.

3 Asymptotic Relative Efficiency (ARE)

If two estimators are both asymptotically unbiased and normal. Which is better? Can compare asymptotic variances.

Definition: If two estimators T_n and S_n satisfy

$$\sqrt{n}[T_n - \tau(\theta)] \to_d N(0, \sigma_T^2),$$
$$\sqrt{n}[S_n - \tau(\theta)] \to_d N(0, \sigma_S^2).$$

The *asymptotic relative efficiency* (ARE) of T_n with respect to S_n is

$$\text{ARE}(T_n, S_n) = \frac{\sigma_S^2}{\sigma_T^2}.$$

If $ARE(T_n, S_n) \leq 1, \forall\theta$, then S_n is *asymptotically more efficient* than T_n.

Example: (ARE of two Poisson Estimators)

$X_1, \ldots, X_n$ iid Poisson (λ). Consider two estimator of $P_\lambda(X = 0) = e^{-\lambda}$. One estimator is $Y_n = \frac{1}{n}\sum_{i=1}^n I(X_i = 0)$, and the other is the MLE.

4 Asymptotic Efficiency

Definition: A sequence T_n is *asymptotically efficient* for $\tau(\theta)$ if for all $\theta \in \Theta$,

$$\sqrt{n}(T_n - \tau(\theta)) \to_d N\left(0, \frac{[\tau'(\theta)]^2}{I(\theta)}\right).$$

- The asymptotic variance of T_n achieves the Cramér-Rao lower bound.

Theorem: Asymptotic efficiency of MLEs

Let $X_1, \ldots, X_n$ be iid $f(x|\theta)$, and let $\hat{\theta}$ be the MLE of θ. Under some regularity conditions, $\hat{\theta}$ is $AN(\theta, 1/(nI(\theta)))$ or $AN\left(\theta, I_n^{-1}(\theta)\right)$, i.e.

$$\sqrt{n}(\hat{\theta} - \theta) \to_d N\left(0, \frac{1}{I(\theta)}\right),$$

for all $\theta \in \Theta$. Assume $\tau(\theta)$ is continuous and differentiable in θ. then

$$\sqrt{n}[\tau(\hat{\theta}) - \tau(\theta)] \to_d N\left(0, \frac{[\tau'(\theta)]^2}{I(\theta)}\right).$$

That is, $\tau(\hat{\theta})$ is a consistent and asymptotic efficient estimator of $\tau(\theta)$.

Approximating the variance

For finite sample size n, let $\hat{\theta}$ be the MLE. The variance of $\tau(\hat{\theta})$ can be approximated as

$$\begin{aligned}
\mathrm{Var}\Big(\tau(\hat{\theta})\Big) &\approx \tfrac{[\tau'(\theta)]^2}{I_n(\theta)} \quad \text{(asymptotic variance)} \\
&\approx \tfrac{[\tau'(\hat{\theta})]^2}{I_n(\hat{\theta})} \quad \text{(plug in the estimate).} \\
\mathrm{Var}\Big(\tau(\hat{\theta})\Big) &\approx \frac{[\tau'(\theta)]^2}{\mathrm{E}_\theta\left(-\frac{\partial^2}{\partial\theta^2}\log L(\theta|\mathbf{X})\right)} \\
&\approx \frac{[\tau'(\hat{\theta})]^2}{-\frac{\partial^2}{\partial\theta^2}\log L(\theta|\mathbf{X})\big|_{\theta=\hat{\theta}}}.
\end{aligned}$$

The denominator is called *observed information number.*

Example: (Approximate Binomial Variance)

$X_1, \ldots, X_n$ iid from $\mathrm{Bin}(1, p)$. Calculate the variance of the MLE of p.

(ii) Calculate the variance of the MLE of the odds $p/(1-p)$.

5 Highlights of Chapter 5

1) W_n **converges in probability** to a constant c, i.e. $W_n \to_p c$

$$P(|W_n - c| > \epsilon) \to 0 \quad \text{as} \quad n \to \infty, \qquad \text{for all} \;\; \epsilon > 0.$$

2) **Weak Law of Large Numbers**:
If $X_1, \ldots, X_n$ are iid with $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2 < \infty$. Then

$$\bar{X}_n \to_p \mu.$$

3) **Central Limit Theorem**:
Let $X_1, \ldots, X_n$ be iid with $\mathrm{E}(X) = \mu$ and $0 < \mathrm{Var}(X) = \sigma^2 < \infty$. Then

$$\sqrt{n}(\bar{X}_n - \mu) \to_d N(0, \sigma^2).$$

Thus the central limit theorem says that

$$P\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \le x\right) \to \Phi(x), \quad \forall x,$$

where Φ stands for the standard normal cdf.

4) **Slutsky's theorem**:
If $W_n \to_d W$ and $Z_n \to_p c$, then

$$W_n + Z_n \to_d W + c; \quad W_n Z_n \to_d cW; \quad W_n/Z_n \to_d W/c \text{ if } c \neq 0.$$

5) Let $A_n \to a, B_n \to b, X_n \to_p x$, then $A_n + B_n X_n \to_p a + bx$.

6) Suppose g is continuous, then

(i) $W_n \to_d W$ implies that $g(W_n) \to_d g(W)$.

(ii) $W_n \to_p W$ implies that $g(W_n) \to_p g(W)$.

7) **First-Order Delta method**:
Let W_n be a sequence of random variables satisfying $\sqrt{n}(W_n - \theta) \to_d N(0, \sigma^2)$. For a given function g, suppose $g'(\theta)$ exists and not equal to zero. Then

$$\sqrt{n}[g(W_n) - g(\theta)] \to_d N(0, \sigma^2[g'(\theta)]^2)$$

8) **Second-order Delta method**:
Let W_n be a sequence of random variables satisfying $\sqrt{n}(W_n - \theta) \to_d N(0, \sigma^2)$. For a given function g, suppose $g'(\theta) = 0$ and $g''(\theta) \neq 0$. Then

$$n[g(W_n) - g(\theta)] \to_d \sigma^2 \, \frac{g''(\theta)}{2} \chi_1^2.$$

9) **Approximate Mean and Variance**:
Suppose $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$. We want to estimate mean and variance of the function $g(X)$.

Using the first-order Taylor expansion

$$g(X) \approx g(\mu) + g'(\mu)(X - \mu).$$

Then we approximately have

$$\begin{aligned} \mathrm{E}_\mu[(g(X)] &\approx g(\mu), \\ \mathrm{Var}_\mu[(g(X)] &\approx [g'(\mu)]^2 \sigma^2. \end{aligned}$$