Asymptotics II: Limiting Distributions

MIT 18.655

Dr. Kempthorne

Spring 2016

Outline

1. Asymptotics II
   - Delta Method: Multivariate Case
   - Asymptotic Normality of Exponential Family MLE
   - Asymptotic Normality of M-Estimators
   - Asymptotic Normality of MLE
   - Super-Efficiency

Multivariate Case

Lemma 5.3.3 Suppose

- $\{U_n\}$ are d-dimensional random vectors
- $\{a_n\}$ constants with $a_n \to \infty$.
- $a_n(U_n - u) \xrightarrow{\mathcal{L}} V$, for $(d \times 1)$ vector u.
- $g : R^d \to R^p$ has differential $g^{(1)}(u)$ $(p \times d)$ at u,
$$||g^{(1)}(u)||_{i,j} = \frac{\partial g(u)_i}{\partial u_j},$$
$i = 1, \dots, p$ and $j = 1, \dots, d$

Then
$$a_n[g(U_n) - g(u)] \xrightarrow{\mathcal{L}} g^{(1)}(u)V.$$

Proof: Multivariate delta method (Theorem 5.3.2).

Multivariate Case

Theorem 5.3.4

- $Y_1, \ldots, Y_n$ iid d-vectors
- $E[|Y_1|^2] < \infty$.
- $E[Y_1] = m \in R^d$.
- $Var[Y_1] = \Sigma$ $(d \times d)$ positive definite.
- $h : \mathcal{O} \to R^p$ where $\mathcal{O}$ is an open subset of R^d.
- $h = (h_1, \ldots, h_p)$ and has a total differential
$$h^{(1)}(m) = ||\frac{\partial h_i}{\partial x_j}(m)||_{p \times d}.$$

Then:

$$h(\overline{Y}) = h(m) + h^{(1)}(m)(\overline{Y} - m) + o_P(n^{-1/2})$$
$$\sqrt{n}[h(\overline{Y}) - h(m)] \xrightarrow{\mathcal{L}} N_p(\mathbf{0}_p, h^{(1)}(m)\Sigma[h^{(1)}(m)]^T)$$

Outline

Asymptotic Normality of Exponential Family MLE

Theorem 5.3.5 Suppose:

- $\mathcal{P}$: canonical exponential family
- $T(X_1) = (T_1(X_1), \ldots, T_d(X_1))^T$.
- $\mathcal{E}$ open
- $X_1, \ldots, X_n$ iid $P_\eta \in \mathcal{P}$
- $\hat{\eta}$: MLE (if it exists, otherwise constant c)

Then

(i) $\hat{\eta} = \eta + \frac{1}{n}\sum_{i=1}^{n} \ddot{A}^{-1}(T(X_i) - \dot{A}(\eta)) + o_{P_\eta}(n^{-\frac{1}{2}})$.

(ii) $\sqrt{n}(\hat{\eta} - \eta) \xrightarrow{\mathcal{L};} N_d(\mathbf{0}_d, \ddot{A}^{-1}(\eta))$

Proof:

- Applying the multivariate CLT to $\overline{T}$:
  with
  - $E[\overline{T} \mid \eta] = \dot{A}(\eta) = E[T(X_1) \mid \eta]$.
  - $Var[\overline{T} \mid \eta] = \ddot{A}(\eta) = Var[T(X_1) \mid \eta]$.

  gives
  $$\sqrt{n}(\overline{T} - \dot{A}(\eta)) \xrightarrow{\mathcal{L}} N_d(0, \ddot{A}(\eta)).$$
- $\hat{\eta}$ solves $\dot{A}(\eta) = \overline{T} = \frac{1}{n}\sum_{i=1}^{n} T(X_i)$, so
  $$\hat{\eta} = h(\overline{T}), \text{ where } h(t) = \dot{A}^{-1}(t).$$
- If $t = \dot{A}(\eta)$, then
  $$h^{(1)}(t) = ||\frac{\partial[\dot{A}^{-1}(t)]_i}{\partial t_j}|| = D\dot{A}^{-1}(t) = [D\dot{A}(\eta)]^{-1} = [\ddot{A}(\eta)]^{-1}$$
- Apply Theorem 5.3.4 to $\overline{T}$, using $h(\overline{T})$, noting that
  $$m = E[T \mid \eta],\ h(t) = \dot{A}^{-1}(t).$$

Outline

Asymptotic Normality of Minimum-Contrast Estimators

Minimum Contrast Estimator:

- $X_1, \ldots, X_n$ iid $P_\theta, \theta \in \Theta$, where $\Theta(open) \subset R$.
- Contrast function:
  $$\rho : \mathcal{X} \times \Theta \to R,$$
- Discrepancy function:
  $$D(\theta_0, \theta) = E[\rho(X_1, \theta) - \rho(X_1, \theta_0) \mid \theta_0].$$
  Uniquely minimized at $\theta = \theta_0$.
- Minimum-contrast estimate:
  $$\overline{\theta}_n \text{ minimizes } \overline{\rho}_n(\theta) = \frac{1}{n}\sum_{i=1}^n \rho(X_i, \theta).$$
- Assumption $A_0 : \psi = \frac{\partial \rho}{d\theta}$ is well defined:
  $\overline{\theta}_n$ solves
  $$\frac{1}{n}\sum_{i=1}^n \psi(X_i, \theta) = \overline{\psi}_n(\theta) = 0.$$

Asymptotic Normality of Minimum-Contrast Estimators

- For a distribution P, define the parameter:

  $\theta(P)$: unique solution of

  $$\int \psi(x,\theta)dP(x) = 0.$$

  Note:
  - For $\theta(P)$ to be well-defined, assume that
    $$\int |\psi(x,\theta)|dP(x) < \infty,\ \theta \in \Theta,\ P \in \mathcal{P}.$$
  - P may come from larger class than $\{P_\theta, \theta \in \Theta\}$. The original parameter of interest can be extended to larger class of distributions.

- Consider Taylor expansion of $\overline{\psi}_n(\theta)$ at $\theta = \overline{\theta}_n$ centered at $\theta = \theta(P)$:

  $$0 = \overline{\psi}_n(\overline{\theta}_n) = \overline{\psi}_n(\theta(P)) + (\overline{\theta}_n - \theta(P)) \times \frac{\partial}{\partial\theta}[\overline{\psi}_n(\theta^*)]$$

  where $\theta^* : |\theta^* - \theta(P)| < |\overline{\theta}_n - \theta(P)|$.

  $$\frac{\partial}{\partial\theta}[\overline{\psi}_n(\theta)] = \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta}[\psi(X_i,\theta)]$$

Asymptotic Normality of Minimum-Contrast Estimators

- Consider Taylor expansion of $\overline{\psi}_n(\theta)$ at $\theta = \overline{\theta}_n$ centered at $\theta = \theta(P)$:
$$0 = \overline{\psi}_n(\overline{\theta}_n) = \overline{\psi}_n(\theta(P)) + (\overline{\theta}_n - \theta(P)) \times \frac{\partial}{\partial\theta}[\overline{\psi}_n(\theta^*)]$$
where
$$\theta^* : |\theta^* - \theta(P)| < |\overline{\theta}_n - \theta(P)|.$$
$$\frac{\partial}{\partial\theta}[\overline{\psi}_n(\theta)] = \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta}[\psi(X_i, \theta)]$$
- Note components of Taylor expansion:
  - $\overline{\psi}_n(\theta(P)) = \frac{1}{n}\sum_1^n \psi(X_i, \theta(P))$, a sample mean; CLT applies, so long as:
$$A2 : E_P[\psi^2(X_1, \theta(P))] < \infty, \text{ for all } P \in \mathcal{P}.$$
  - If $A5 : \overline{\theta}_n \xrightarrow{P} \theta(P)$, then $\theta^* \xrightarrow{P} \theta(P)$
  - If $A4$:
$$sup_t\{|\frac{1}{n}\sum_{i=1}^n(\frac{\partial\psi}{\partial\theta}(X_i, t) - \frac{\partial\psi}{\partial\theta}(X_i, \theta(P))| : |t - \theta(P)| \le \epsilon_n\}$$
$$\xrightarrow{P} 0, \text{ if } \epsilon_n \to 0.$$
then
$$\frac{\partial}{\partial\theta}[\overline{\psi}_n(\theta^*)] \xrightarrow{P} E_P(\frac{\partial}{\partial\theta}[\psi(X_i, \theta(P)])$$

- Define

$$J(\theta(P)) = -E_P(\tfrac{\partial}{\partial\theta}[\psi(X_i,\theta)])|_{\theta=\theta(P)}$$

Suppose $A3: J(\theta(P)) \neq 0$, then we can rewrite:

$$0 = \overline{\psi}_n(\overline{\theta}_n) = \overline{\psi}_n(\theta(P)) + (\overline{\theta}_n - \theta(P)) \times \tfrac{\partial}{\partial\theta}[\overline{\psi}_n(\theta^*)]$$

as

$$\begin{aligned}
\overline{\theta}_n - \theta(P) &= \frac{\overline{\psi}_n(\theta(P))}{-\frac{\partial}{\partial\theta}[\overline{\psi}_n(\theta^*)]} \\
&= \frac{\overline{\psi}_n(\theta(P))}{J(\theta(P)) + o_P(1)} \\
\implies \sqrt{n}(\overline{\theta}_n - \theta(P)) &= [J(\theta(P) + o_P(1)]^{-1} \times \sqrt{n}[\overline{\psi}_n(\theta(P))] \\
&= [J(\theta(P)]^{-1} \times [1 + o_P(1)] \times \sqrt{n}[\overline{\psi}_n(\theta(P))] \\
&\xrightarrow{\mathcal{L}} N(0, \sigma^2(\psi, P))
\end{aligned}$$

where $\sigma^2(\psi, P) = E_P[\psi^2(X_1, \theta(P))]/[J(\theta(P))]^2$.

Remarks

- The asymptotic normal distribution applies to solutions of the estimating equations; these equations can be motivated by M-Estimators (distinct from minimum-contrast estimators).
- The limiting distribution results apply to sampling distribution $P \notin \mathcal{P}$ so long as

  $\theta(P)$ is unique minimum of $E_P(\rho(X_1, \theta))$ or

  $\theta(P)$ uniquely solves: $E_P[\psi(X_1, \theta)] = 0$.

Outline

1. Asymptotics II
   - Delta Method: Multivariate Case
   - Asymptotic Normality of Exponential Family MLE
   - Asymptotic Normality of M-Estimators
   - Asymptotic Normality of MLE
   - Super-Efficiency

Asymptotic Normality of MLE

Maximum-Likelihood Estimators

- $\rho(X,\theta) = -log[p(x \mid \theta)] = -l(x,\theta)$
- $\psi(X,\theta) \equiv \frac{\partial l}{\partial \theta}(x,\theta)$
- Note that:

$$\begin{aligned} J(\theta) &= -E_P(\tfrac{\partial}{\partial\theta}\psi(X,\theta)) \\ &= E_\theta[(\tfrac{\partial l}{\partial \theta}(X_1,\theta))^2] = Var_\theta[\tfrac{\partial l}{\partial\theta}(X_1,\theta)] = I(\theta) \end{aligned}$$

$$\begin{aligned} \Longrightarrow \quad \sqrt{n}(\bar{\theta}_n - \theta(P)) &= [J(\theta(P)]^{-1} \times \sqrt{n}[\overline{\psi}_n(\theta(P))] + o_P(1) \\ &\xrightarrow{\mathcal{L}} N(0, \sigma^2(\psi, P)) \end{aligned}$$

where $\sigma^2(\psi, P) = E_P[\psi^2(X_1, \theta(P))]/[J(\theta(P))]^2 = 1/I(\theta)$, and

$$\begin{aligned} \Longrightarrow \quad \bar{\theta}_n &= \theta(P) + \frac{\overline{\psi}_n(\theta(P))}{J(\theta(P))} + o_P(\overline{\psi}_n(\theta(P))) \\ &= \theta(P) + \frac{1}{n}\sum_1^n \frac{1}{I(\theta(P))}\frac{\partial}{\partial\theta}[l(X_i,\theta)] + o_P(n^{-1/2}) \end{aligned}$$

Theorem 5.4.3 If $\bar{\theta}_n$ is a minimum contrast estimator corresponding to $\rho(x,\theta)$ and $\psi(x,\theta)$, which satistfy assumptions $A0 - A6$, then

$$\sigma^2(\psi, P_\theta) \geq \frac{1}{I(\theta)}$$

with equality if and only if $\psi(x,\theta) = a(\theta)\frac{\partial l(x,\theta)}{\partial\theta}$, for some $a(\theta) \neq 0$.

Proof: Assuming that $l(x,\theta) = -log(p(x \mid \theta))$ is differentiable:

$\int \psi(x,\theta)p(x \mid \theta)dx = 0$ implies (by differentiating),

$$\begin{aligned} E_\theta[\tfrac{\partial\psi}{\partial\theta}(X_1,\theta(P))] &= -E_\theta[\tfrac{\partial l}{\partial\theta}(X_1,\theta)\psi(X_1,\theta)] \\ &= -Cov_\theta[\tfrac{\partial l}{\partial\theta}(X_1,\theta),\psi(X_1,\theta)] = -J(\theta(P)) \end{aligned}$$

The covariance inequality

$$[J(\theta(P))]^2 \leq Var[\tfrac{\partial l}{\partial\theta}(X_1,\theta)] \times E_\theta([\psi(X_1,\theta)]^2) \text{ gives}$$

$$\sigma^2(\psi,P_\theta) = \frac{E_\theta([\psi(X_1,\theta)]^2)}{[J(\theta(P))]^2} \geq (Var[\frac{\partial l}{\partial\theta}(X_1,\theta)])^{-1} = \frac{1}{I(\theta)}$$

with equality iff $\psi(X_1,\theta)$ is a linear multiple of $\partial l(x,\theta)/\partial\theta$.

Outline

Hodges' Super-Efficiency Example

Hodges' Example:

- $X_1, \dots, X_n$ iid $N(\theta, 1)$
- $\overline{X}_n$ is the MLE of θ
- $I(\theta) = 1$.

Consider the estimator:

$$\tilde{\theta}_n = \begin{cases} 0, & \textit{if} \quad |\overline{X}_n| \leq n^{-1/4} \\ \overline{X}_n, & \textit{if} \quad |\overline{X}_n| > n^{-1/4} \end{cases}$$

$\tilde{\theta}_n$ is a "Pre-Test" Estimator:

- Test $H_0 : \theta = 0$ vs $H_1 : \theta \neq 0$.
- Reject H_0 if $\overline{X}_n > n^{-1/4}$.
- Use $\overline{X}_n$ if H_0 rejected, otherwise 0.

Hodges' Super-Efficiency Example

Limiting Distribution: $\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{\mathcal{L}} ?$

- Consider:

$$\begin{aligned} P_\theta[\tilde{\theta}_n = 0] &= P[|\overline{X}_n| \le n^{-1/4}] \\ &= P[|N(\theta, \tfrac{1}{n})| \le n^{-1/4}] \\ &= \Phi(z_n^{**}) - \Phi(z_n^{*}) \end{aligned}$$

where

$$\begin{aligned} z_n^{**} &= \frac{n^{-1/4} - \theta}{n^{-1/2}} = n^{1/4} - n^{1/2}\theta \\ z_n^{*} &= \frac{-n^{-1/4} - \theta}{n^{-1/2}} = -n^{1/4} - n^{1/2}\theta \end{aligned}$$

- Suppose $\theta \neq 0$. Then since $z_n^{**} \to -\infty$ and $z_n^{*} \to -\infty$,

$$P_\theta[\tilde{\theta}_n = 0] \xrightarrow{P} 0 \text{ and } P_\theta[\tilde{\theta}_n = \overline{X}_n] \to 1.$$

- Suppose $\theta = 0$. Then since $z_n^{**} \to +\infty$ and $z_n^{*} \to -\infty$,

$$P_\theta[\tilde{\theta}_n = 0] \xrightarrow{P} 1 \text{ and } P_\theta[\tilde{\theta}_n = \overline{X}_n] \to 0.$$

Hodges' Super-Efficiency Example

Limiting Distribution: $\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{\mathcal{L}} ?$

$$\sqrt{n}(\tilde{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, \sigma^2(\theta))$$

where

$$\sigma^2(\theta) = \begin{cases} 1/I(\theta), & \textit{if} \quad \theta \neq 0 \\ 0, & \textit{if} \quad \theta = 0 \end{cases}$$

The estimator $\tilde{\theta}_n$ is

- **efficient** for all $\theta \neq 0$, and
- **super-efficient** at $\theta = 0$.

18.655 Mathematical Statistics
Spring 2016