Asymptotics: Consistency and Delta Method

MIT 18.655

Dr. Kempthorne

Spring 2016

Outline

Consistency

Statistical Estimation Problem

- $X_1, \ldots, X_n$ iid $P_\theta, \theta \in \Theta$.
- $q(\theta)$: target of estimation.
- $\hat{q}(X_1, \ldots, X_n)$: estimator of $q(\theta)$.

Definition: $\hat{q}_n$ is a **consistent** estimator of $q(\theta)$, i.e.,

$$\hat{q}_n(X_1, \ldots, X_n) \xrightarrow{P_\theta} q(\theta)$$

if for every $\epsilon > 0$,

$$\lim_{n\to\infty} P_\theta(|q_n(X_1, \ldots, X_n) - q(\theta)| > \epsilon) = 0.$$

Example: Consider P_θ such that:

- $E[X_1 \mid \theta] = \theta$
- $q(\theta) = \theta$
- $\hat{q}_n = \dfrac{\sum_{i=1}^n X_i}{n} = \overline{X}$

When is $\hat{q}_n$ consistent for θ?

Consistency: Example

Example: Consistency of sample mean $\hat{q}_n(X_1, \ldots, X_n) = \overline{X} = \overline{X}_n$.

- If $Var(X_1 \mid \theta) = \sigma^2(\theta) < \infty$, apply Chebychev's Inequality. For any $\epsilon > 0$:
$$P_\theta(|\overline{X}_n| \geq \epsilon) \leq \frac{Var(\overline{X}_n \mid \theta)}{\epsilon^2} = \frac{\sigma^2(\theta)/\epsilon^2}{n} \xrightarrow{n\to\infty} 0$$
- If $Var(X_1 \mid \theta) = \infty$, $\overline{X}_n$ is consistent if
$$E[|X_1| \mid \theta] < \infty.$$

Proof: Levy Continuity Theorem.

Consistency: A Stronger Definition

Definition: $\hat{q}_n$ is a **uniformly consistent** estimator of $q(\theta)$, if for every $\epsilon > 0$,

$$\lim_{n\to\infty}\left(\sup_{\theta\in\Theta}[P_\theta(|q_n(X_1,\ldots,X_n) - q(\theta)| > \epsilon)]\right) = 0.$$

Example: Consider the sample mean $\hat{q}_n = \overline{X}_n$ for which

- $E[\overline{X}_n \mid \theta] = \theta$
- $Var[\overline{X}_n \mid \theta] = \sigma^2(\theta)$.

Proof of consistency of $\hat{q}_n = \overline{X}_n$ extends to uniform consistency if

$$\sup_{\theta\in\Theta}\sigma^2(\theta) \leq M < \infty \qquad (*).$$

Examples Satisfying $(*)$

- X_i i.i.d. $Bernoulli(\theta)$.
- X_i i.i.d. $Normal(\mu,\sigma^2)$, where $\theta = (\mu,\sigma^2)$
  $\Theta = \{\theta\} = (-\infty,+\infty)\times[0,M]$, for finite $M < \infty$.

Consistency: The Strongest Definition

Definition: $\hat{q}_n$ is a **strongly consistent** estimator of $q(\theta)$, if

$$P_\theta\left(\lim_{n\to\infty} |q_n(X_1,\ldots,X_n) - q(\theta)| \leq \epsilon)]\right) = 1, \text{ for every } \epsilon > 0.$$

$$\hat{q}_n \xrightarrow{a.s.} q(\theta). \qquad (\text{a.s.} \equiv \text{"almost surely"})$$

Compare to:

Definition: $\hat{q}_n$ is a **(weakly) consistent** estimator of $q(\theta)$, if for every $\epsilon > 0$,

$$\lim_{n\to\infty}[P_\theta(|q_n(X_1,\ldots,X_n) - q(\theta)| > \epsilon)] = 0.$$

Consistency of Plug-In Estimators

Plug-In Estimators: Discrete Case

- Discrete outcome space of size K:
  $$\mathcal{X} = \{x_1, \dots, x_K\}$$
- $X_1, \dots, X_n$ iid $P_\theta, \theta \in \Theta$, where
  $$\theta = (p_1, \dots, p_K)$$
  $$P(X_1 = x_k \mid \theta) = p_k,\ k = 1, \dots, K$$
  $$p_k \geq 0 \text{ for } k = 1, \dots, K \text{ and } \quad \sum_1^K p_k = 1.$$
- $\Theta = \mathcal{S}_K$ (K-dimensional simplex).
- Define the **empirical distribution**:
  $$\hat{\theta}_n = (\hat{p}_1, \dots, \hat{p}_K)$$
  where
  $$\hat{p}_k = \frac{\sum_{i=1}^n \mathbf{1}(X_i = x_k)}{n} \equiv \frac{N_k}{n}$$
- $\hat{\theta}_n \in \mathcal{S}_K$.

Consistency of Plug-In Estimators

Proposition/Theorem (5.2.1) Suppose $\mathbf{X}_n = (X_1, \ldots, X_n)$ is a random sample of size n from a discrete distribution $\theta \in \mathcal{S}$. Then:

- $\hat{\theta}_n$ is uniformly consistent for $\theta \in \mathcal{S}$.
- For any continuous function $q : \mathcal{S} \to R^d$,
  $$\hat{q} = q(\hat{\theta}_n) \text{ is uniformly consistent for } q(\theta).$$

Proof:

- For any $\epsilon > 0$, $\quad P_\theta(|\hat{\theta}_n - \theta| \geq \epsilon) \longrightarrow 0.$
  This follows upon noting that:
  $$\{\mathbf{x}_n : |\hat{\theta}_n(\mathbf{x}_n) - \theta|^2 < \epsilon^2\} \supset \cap_{k=1}^K \{\mathbf{x}_n : |(\hat{\theta}_n(\mathbf{x}_n) - \theta)_k|^2 < \epsilon^2/K\}$$
  So
  $$\begin{aligned} P(\{\mathbf{x}_n : |\hat{\theta}_n(\mathbf{x}_n) - \theta|^2 \geq \epsilon^2\} &\leq \textstyle\sum_{k=1}^K P(\{\mathbf{x}_n : |(\hat{\theta}_n(\mathbf{x}_n) - \theta)_k|^2 \geq \epsilon^2/K\} \\ &\leq \textstyle\sum_{k=1}^K \frac{1}{4n}/(\epsilon^2/K) = \frac{K^2}{4n\epsilon^2} \end{aligned}$$

Consistency of Plug-In Estimators

Proof (continued):

- $q(\cdot)$: continuous on compact $\mathcal{S}$
  $\Longrightarrow q(\cdot)$ uniformly continuous on $\mathcal{S}$.
  For every $\epsilon > 0$, there exists $\delta(\epsilon) > 0$ such that:
  $$|\theta_1 - \theta_0| < \delta(\epsilon) \Longrightarrow |q(\theta_1) - q(\theta_0)| < \epsilon,$$
  uniformly for all $\theta_0, \theta_1 \in \Theta$.
- It follows that
  $$\{\mathbf{x} : |q(\hat{\theta}_n(\mathbf{x})) - q(\theta)| < \epsilon\}^c \subseteq \{\mathbf{x} : |\hat{\theta}_n - \theta| < \delta(\epsilon)\}^c$$
  $$\Longrightarrow \quad P_\theta[|\hat{q}_n - q(\theta)| \geq \epsilon] \leq P_\theta[|\hat{\theta}_n - \theta| \geq \delta(\epsilon)]$$

Note: uniform consistency can be shown; see B&D.

Consistency of Plug-In Estimators

Proposition 5.2.1 Suppose:

- $\mathbf{g} = (g_1, \ldots, g_d) : \mathcal{X} \to \mathcal{Y} \subset R^d$.
- $E[|g_j(X_1)| \mid \theta] < \infty$, for $j = 1, \ldots, d$, for all $\theta \in \Theta$.
- $m_j(\theta) \equiv E[g_j(X_1) \mid \theta]$, for $j = 1, \ldots, d$.

Define:

$$q(\theta) = h(\mathbf{m}(\theta)),$$

where

$$\mathbf{m}(\theta) = (m_1(\theta), \ldots, m_d(\theta))$$
$$h : \mathcal{Y} \to R^p, \text{ is continuous.}$$

Then:

$$\hat{q}_n = h(\bar{\mathbf{g}}) = h\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{g}(X_i)\right)$$

is a consistent estimator of $q(\theta)$.

Consistency of Plug-In Estimators

Proposition 5.2.1 Applied to Non-Parametric Models

$\mathcal{P} = \{P : E_P(|\mathbf{g}(X_1)|) < \infty\}$ and $\nu(P) = h(E_P\mathbf{g}(X_1))$

$\nu(\hat{P}_n) \xrightarrow{P} \nu(P)$ where $\hat{P}_n$ is empirical distribution.

Consistency of MLEs in Exponential Family

Theorem 5.2.2 Suppose:

- $\mathcal{P}$ is a canonical exponential family of rank d generated by $\mathbf{T} = (T_1(X), \ldots, T_d(X))^T$.
- $p(x \mid \eta) = h(x) exp\{\mathbf{T}(x)\eta - A(\eta)\}$
- $\mathcal{E} = \{\eta\}$ is open.
- $X_1, \ldots, X_n$ are i.i.d $P_\eta \in \mathcal{P}$

Then:

- $P_\eta[$ the MLE $\hat{\eta}$ exists$] \xrightarrow{n \to \infty} 1$.
- $\hat{\eta}$ is consistent.

Consistency of MLEs in Exponential Family

Proof:

- $\hat{\eta}(X_1, \ldots, X_n)$ exists iff $\overline{\mathbf{T}}_n = \sum_1^n \mathbf{T}(X_i)/n \in C^o_{\mathbf{T}_n}$.
- If η_0 is true, then $\mathbf{t}_0 = E[\mathbf{T}(X_1) \mid \eta_0] \in C^o_{\mathbf{T}_n}$ and $\dot{A}(\boldsymbol{\eta}_0) = \mathbf{t}_0$.
- By definition of the interior of the convex support, there exists $\delta > 0$: $S_\delta = \{\mathbf{t} : |\mathbf{t} - E_{\eta_0}[\mathbf{T}(X_1)| < \delta\} \subset C^o_{\mathbf{T}_n}$.
- By the WLLN:
$$\overline{\mathbf{T}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{T}(X_i) \xrightarrow{P_{\eta_0}} E_{\eta_0}[\mathbf{T}(X_1)]$$
$$\Longrightarrow P_{\eta_0}\left[\overline{\mathbf{T}}_n \in C^o_{\mathbf{T}_n}\right] \xrightarrow{n\to\infty} 1.$$
- $\hat{\eta}$ exists if it solves $\dot{A}(\boldsymbol{\eta}) = \overline{\mathbf{T}}_n$, i.e., if $\overline{\mathbf{T}}_n \in C^o_{\mathbf{T}_n}$,
- The map $\eta \to \dot{A}(\boldsymbol{\eta})$ is 1-to-1 on $C^0_{\mathbf{T}}$ and continuous on $\mathcal{E}$, so the inverse function $\dot{A}^{-1}$ is continuous, and Prop. 5.2.1 applies.

Consistency of Minimum-Contrast Estimates

Minimum-Contrast Estimates

- $X_1, \ldots, X_n$ iid P_θ, $\theta \in \Theta \subset R^d$.
- $\rho(X, \theta) : \mathcal{X} \times \Theta \to R$, a contrast function
- $D(\theta_0, \theta) = E[\rho(X, \theta) \mid \theta_0]$: the discrepancy function
  $\theta = \theta_0$ uniquely minimizes $D(\theta_0, \theta)$.
- The Minimum-Contrast Estimate $\hat{\theta}$ minimizes:

$$\rho_n(\mathbf{X}, \theta) = \frac{1}{n} \sum_{i=1}^{n} \rho(X_i, \theta).$$

Theorem 5.2.3 Suppose

- $\sup_{\theta \in \Theta} \{ \frac{1}{n} \sum_{i=1}^{n} \rho(X_i, \theta) - D(\theta_0, \theta) \} \xrightarrow{P_{\theta_0}} 0$
- $\inf_{|\theta - \theta_0| \geq \epsilon} \{ D(\theta_0, \theta) \} > D(\theta_0, \theta_0)$, for all $\epsilon > 0$.

Then $\hat{\theta}$ is consistent.

Outline

1. Asymptotics
   - Asymptotics: Consistency
   - Delta Method: Approximating Moments
   - Delta Method: Approximating Distributions

Delta Method

Theorem 5.3.1 (Applying Taylor Expansions) Suppose that:

- $X_1, \ldots, X_n$ iid with outcome space $\mathcal{X} = R$.
- $E[X_1] = \mu$, and $Var[X_1] = \sigma^2 < \infty$.
- $E[|X_1|^m] < \infty$.
- $h : R \to R$, m-times differentiable on R, with $m \geq 2$.
$$h^{(j)} = \frac{d^j h(x)}{dx^j}, \; j = 1, 2, \ldots, m.$$
- $||h^{(m)}||_\infty \equiv \sup_{x \in \mathcal{X}} |h^{(m)}(x)| \leq M < \infty$.

Then
$$E(h(\overline{X})) = h(\mu) + \sum_{j=1}^{m-1} \frac{h^{(j)}(\mu)}{j!} E[(\overline{X} - \mu)^j] + R_m$$
where
$$|R_m| \leq M \frac{E[|X_1|^m]}{m!} n^{-m/2}$$

Proof:

- Apply Taylor Expansion to $h(\overline{X})$:

$$h(\overline{X}) = h(\mu) + \sum_{j=1}^{m-1} \frac{h^{(j)}(\mu)}{j!}(\overline{X} - \mu)^j + h^{(m)}(X^*)(\overline{X} - \mu)^m,$$

where $|X^* - \mu| \leq |\overline{X} - \mu|$

- Take expectations and apply Lemma 5.3.1: If $E|X_1|^j < \infty, j \geq 2$, then there exist constants $C_j > 0$ and $D_j > 0$ such that

$$E|\overline{X} - \mu|^j \leq C_j E|X_1|^j n^{-j/2}$$
$$|E[(\overline{X} - \mu)^j]| \leq D_j E|X_1|^j n^{-(j+1)/2} \text{ for } j \text{ odd}$$

Applying Taylor Expansions

Corollary 5.3.1 (a).
If $E|X_1|^3 < \infty$ and $||h^{(3)}||_\infty < \infty$, then

$$E[h(\overline{X})] = h(\mu) + 0 + [\frac{h^{(2)}(\mu)}{2}]\frac{\sigma^2}{n} + O(n^{-3/2})$$

Corollary 5.3.1 (b).
If $E|X_1|^4 < \infty$ and $||h^{(4)}||_\infty < \infty$, then

$$E[h(\overline{X})] = h(\mu) + 0 + [\frac{h^{(2)}(\mu)}{2}]\frac{\sigma^2}{n} + O(n^{-2})$$

For (b), use Lemma 5.3.2 with $j = 3$ (odd) gives

$$|E(\overline{x} - \mu)^3| \leq D_3 E[|X_1|^3] \times \frac{1}{n^2} = O(n^{-2})$$

Note: Asymptotic bias of $h(\overline{X})$ for $h(\mu)$:

- If $h^{(2)}(\mu) \neq 0$, then $O(n^{-1})$
- If $h^{(2)}(\mu) = 0$, then $O(n^{-3/2})$ if third-moment finite and $O(n^{-2})$ if fourth-moment finite.

Applying Taylor Expansions: Asymptotic Variance

Corollary 5.3.2 (a).
If $||h^{(j)}||_\infty < \infty$, for j=1,2,3, and $E|Z_1|^3 < \infty$, then

$$Var[h(\overline{X})] = \frac{\sigma^2[h^{(1)}(\mu)]^2}{n} + O(n^{-3/2})$$

Proof: Evaluate

$$Var[h(\overline{X})] = E[(h(\overline{X})^2] - (E[h(\overline{X})])^2$$

From Corollary 5.3.1 (a):

$$\begin{aligned} E[h(\overline{X})] &= h(\mu) + 0 + [\tfrac{h^{(2)}(\mu)}{2}]\tfrac{\sigma^2}{n} + O(n^{-3/2}) \\ \Longrightarrow (E[h(\overline{X})])^2 &= (h(\mu) + [\tfrac{h^{(2)}(\mu)}{2}]\tfrac{\sigma^2}{n})^2 + O(n^{-3/2}) \\ &= (h(\mu))^2 + [h(\mu)h^{(2)}(\mu)]\tfrac{\sigma^2}{n} + O(n^{-3/2}) \end{aligned}$$

Taking Expectation of the Taylor Expansion:

$$\begin{aligned} E([h(\overline{X})]^2) = \; & [h(\mu)]^2 + E[\overline{X} - \mu]\left(2[h(\mu)]h^{(1)}(\mu)\right) \\ & + \tfrac{1}{2}E[(\overline{X} - \mu)^2]\left(2[h^{(1)}(\mu)]^2 + 2[h(\mu)]h^{(2)}(\mu)\right) \\ & + \tfrac{1}{6}E[(\overline{X} - \mu)^3]\left([h^2(\mu)]^{(3)}(X^*)\right) \end{aligned}$$

Difference gives result.

Note:

- Asymptotic bias of $h(\overline{X})$ for $h(\mu)$ is $O(\frac{1}{n})$.
- Asymptotic standard deviation of $h(\overline{X})$ is $O(\frac{1}{\sqrt{n}})$ unless (!) $h^{(1)}(\mu) = 0$.
- More terms in aTaylor Series with finite expectations of $E[|\overline{X} - \theta|^j]$ yields finer approximation to order $O(n^{-j/2})$
- Taylor Series Expansions apply to functions of vector-valued statistics (See Theorem 5.3.2).

Outline

1. Asymptotics
   - Asymptotics: Consistency
   - Delta Method: Approximating Moments
   - Delta Method: Approximating Distributions

Theorem 5.3.3 Suppose

- $X_1, \ldots, X_n$ iid with $\mathcal{X} = R$.
- $E[X_1^2] < \infty$.
- $\mu = E[X_1]$ and $\sigma^2 = Var(X_1)$.
- $h : R \to R$ is differentiable at μ.

Then

$$\mathcal{L}\left(\sqrt{n}(h(\overline{X}) - h(\mu))\right) \to N(0, \sigma^2(h))$$

where $$\sigma^2(h) = [h^{(1)}(\mu)]^2\sigma^2.$$

Proof: Apply Taylor expansion of $h(\overline{X})$ about μ :

$$\begin{aligned} h(\overline{X}) &= h(\mu) + (\overline{X} - \mu)[h^{(1)}(\mu) + R_n] \\ \Longrightarrow \sqrt{n}(h(\overline{X}) - h(\mu)) &= [\sqrt{n}(\overline{X} - \mu)][h^{(1)}(\mu) + R_n] \\ &\xrightarrow{\mathcal{L}} [N(0, \sigma^2)] \times h^{(1)}(\mu) \end{aligned}$$

Limiting Distributions of t Statistics

Example 5.3.3 One-sample t-statistic

- $X_1, \ldots, X_n$ iid $P \in \mathcal{P}$
- $E_P[X_1] = \mu$
- $Var_P(X_1) = \sigma^2 < \infty$
- For a given μ_0, define t-statistic for testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$.

$$T_n = \sqrt{n}\frac{(\overline{X} - \mu_0)}{s_n} \quad \text{where}$$
$$s_n^2 = \sum_{1}^{n}(X_i - \overline{X})^2/(n-1).$$

If H is true then $T_n \xrightarrow{\mathcal{L}} N(0,1)$.

Proof: Apply Slutsky's theorem for limit of $\{U_n/v_n\}$ where

- $U_n = \sqrt{n}\frac{(\overline{X} - \mu_0)}{\sigma} \xrightarrow{\mathcal{L}} N(0,1)$.
- $v_n = s_n/\sigma \xrightarrow{P} 1$.

Limiting Distributions of t Statistics

Example 5.3.3 Two-Sample t-statistic

- $X_1, \ldots, X_{n_1}$ iid with $E[X_1] = \mu_1$ and $Var[X_1] = \sigma_1^2 < \infty$.
- $Y_1, \ldots, X_{n_2}$ iid with $E[Y_1] = \mu_2$ and $Var[Y_1] = \sigma_2^2 < \infty$.
- Define t-statistic for testing $H_0 : \mu_2 = \mu_1$ versus $H_1 : \mu_2 > \mu_1$.

$$T_n = \frac{\overline{Y} - \overline{X}}{\sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}}} = \sqrt{\frac{n_1 n_2}{n}}\left(\frac{\overline{Y} - \overline{X}}{s}\right)$$

where $$s^2 = \frac{1}{n-2}\left[\sum_{i=1}^{n_1}(X_i - \overline{X})^2 + \sum_{j=1}^{n_2}(Y_i - \overline{Y})^2\right]$$

If H is true, $\sigma_1^2 = \sigma_2^2$, and all distributions are Gaussian, then

$$T_n \sim t_{n-2}, \text{ (a } t\text{-distribution)}$$

In general, if H is true, and $n_1 \to \infty$ and $n_2 \to \infty$, with $n_1/n \to \lambda$, $(0 < \lambda < 1)$, then

$$T_n \xrightarrow{\mathcal{L}} N(0, \tau^2), \text{ where } \tau^2 = \frac{(1-\lambda)\sigma_1^2 + \lambda\sigma_2^2}{\lambda\sigma_1^2 + (1-\lambda)\sigma_2^2} \quad (\approx 1, \text{ when?})$$

Additional Topics

- Monte Carlo simulations/studies: evaluating asymptotic distribution approximations.
- Variance-stabilizing transformations: When $E[\overline{X}] = \mu$, but $Var[\overline{X}] = \sigma^2(\mu)$, consider $h(\overline{X})$ such that
  $$\sigma^2(\mu)[h^{(1)}(\mu)]^2 = constant$$
  Asymptotic distribution approximation for $h(\overline{X})$ will have a constant variance.
- Edgeworth Approximations: Refining the Central Limit Theorem to match nonzero skewness and non-Gaussian kurtosis.

Taylor Series Review

Power Series Representation of function $f(x)$

$$f(x) = \sum_{j=0}^{\infty} c_j (x-a)^j, \text{ for } x : |x-a| < d$$

$$a = \text{center; and } d = \text{radius of convergence}$$

Theorem: If $f(x)$ has a power series representation, then

- $c_j = \dfrac{f^{(j)}(a)}{j!} = \dfrac{\frac{d^j}{dx^j}[f(x)]}{j!}\Big|_{x=a}$.
- Define $T_m(x) = \sum_{j=1}^{m} c_j (x-a)^j$, and $R_m(x) = f(x) - T_m(x)$.
  $\lim\limits_{m\to\infty} T_m(x) = f(x)$ and $(\Leftrightarrow)$ $\lim\limits_{m\to\infty} R_m(x) = 0$.

Power Series Approximation of $f(x)$ where

- $f^{(j)}(x)$: finite for $1 \leq j \leq m$
- $\sup_x ||f^{(m)}(x)|| \leq M$.

For $m = 2$:

$$
\begin{array}{lrcl}
 & f^{(2)}(x) & \leq & M \\
\Longrightarrow & \int_a^x f^{(2)}(t)dt & \leq & \int_a^x Mdt \\
\Longleftrightarrow & f^{(1)}(x) - f^{(1)}(a) & \leq & M(x-a) \\
\Longleftrightarrow & f^{(1)}(x) & \leq & f^{(1)}(a) + M(x-a)
\end{array}
$$

Integrate again:

$$
\begin{array}{lrcl}
 & \int_a^x f^{(1)}(t)dt & \leq & \int_a^x [f^{(1)}(a) + M(t-a)]dt \\
\Longleftrightarrow & f(x) - f(a) & \leq & f^{(1)}(a)(x-a) + M\frac{(x-a)^2}{2} \\
\Longleftrightarrow & f(x) & \leq & f(a) + f^{(1)}(a)(x-a) + M\frac{(x-a)^2}{2}
\end{array}
$$

Reverse inequality and use $-M$:

$$
\begin{array}{lrcl}
\Longrightarrow & f(x) & \geq & f(a) + f^{(1)}(a)(x-a) - M\frac{(x-a)^2}{2} \\
\Longrightarrow & f(x) & = & f(a) + f^{(1)}(a)(x-a) + R_2(x) \\
 & \textit{where} & & |R_2(x)| \leq M\frac{(t-a)^2}{2}
\end{array}
$$

Delta Method for Function of a Random Variable

- X a r.v. with $\mu = E[x]$
- $h(\cdot)$ function with m derivatives
- $h(X) = h(\mu) + (X-\mu)h^{(1)}(\mu) + R_2(X)$
  where $|R_2(X)| \leq \frac{M}{2}(X-\mu)^2$.
- $h(X) = h(\mu) + (X-\mu)h^{(1)}(\mu) + (X-\mu)^2\frac{h^{(2)}(\mu)}{2} + R_3(X)$
  where $|R_3(X)| \leq \frac{M}{3!}|X-\mu|^3$.
- $h(X) = h(\mu) + (X-\mu)h^{(1)}(\mu) + (X-\mu)^2\frac{h^{(2)}(\mu)}{2}$
  $+(X-\mu)^3\frac{h^{(3)}(\mu)}{3!} + R_4(X)$
  where $|R_4(X)| \leq \frac{M}{4!}|X-\mu|^4$.

Key Example: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ **, for i.i.d.** X_i

$$\begin{array}{rclrcl} E[X_1] & = & \theta, & E[(X_1-\theta)^2] & = & \sigma^2 \\ E[(X_1-\mu)^3] & = & \mu_3 & E[|X_1-\mu|^3] & = & \kappa_3 \end{array}$$

With $\bar{X}$ **for a given sample size** n

$$\begin{array}{rclrcl} E[\bar{X}] & = & \theta, & E[(\bar{X}-\theta)^2] & = & \frac{\sigma^2}{n} \\ E[(\bar{X}-\theta)^3] & = & \frac{\mu_3}{n^2} & E[|\bar{X}-\mu|^3] & = & 0_p[(\frac{1}{\sqrt{n}})^3] \end{array}$$

Taking Expectations of the Delta Formulas (cases $m = 2, 3$**)**

$$\begin{array}{rcl} E[h(\bar{X})] & = & E[h(\theta) + (\bar{X}-\theta)h^{(1)}(\theta) + R_2(\bar{X})] \\ & = & h(\theta) + E[(\bar{X}-\theta)]h^{(1)}(\theta) + E[R_2(\bar{X})] \\ & = & h(\theta) + 0 + E[R_2(\bar{X})] \end{array}$$

where $|E[R_2(\bar{X})]| \leq E[|R_2(\bar{X})|] \leq \frac{M}{2}E[(\bar{X}-\theta)^2] = \frac{M}{2}\frac{\sigma^2}{n}$

$$\begin{array}{rcl} E[h(\bar{X})] & = & E[h(\theta) + (\bar{X}-\theta)h^{(1)}(\theta) + (\bar{X}-\theta)^2\frac{h^{(2)}(\theta)}{2} + R_3(X)] \\ & = & h(\theta) + \frac{\sigma^2}{n}\frac{h^{(2)}(\theta)}{2} + E[R_3(\bar{X})] \end{array}$$

where

$$|E[R_3(\bar{X})]| \leq [E|R_3(\bar{X})|] \leq \frac{M}{3!}E[|\bar{X}-\theta|^3] = \frac{M}{3!}O_p(|\frac{1}{\sqrt{n}}|^3).$$

Taking Expectations of the Delta Formula (case $m = 4$)

$$\begin{aligned}
E[h(\bar{X})] &= E[h(\theta) + (\bar{X} - \theta)h^{(1)}(\theta) + (\bar{X} - \theta)^2 \frac{h^{(2)}(\theta)}{2}] + \\
&\quad + E[(\bar{X} - \theta)^3 \frac{h^{(3)}}{3!} + R_4(X)] \\
&= h(\theta) + \frac{\sigma^2}{n}\frac{h^{(2)}(\theta)}{2} + E[(\bar{X} - \theta)^3]\frac{h^{(3)}}{3!} + E[R_4(X)] \\
&= h(\theta) + \frac{\sigma^2}{n}\frac{h^{(2)}(\theta)}{2} + \frac{\mu_3}{n^2}\frac{h^{(3)}}{3!} + E[R_4(X)] \\
&= h(\theta) + \frac{\sigma^2}{n}\frac{h^{(2)}(\theta)}{2} + O_p(\frac{1}{n^2})
\end{aligned}$$

because $|E[R_4(\bar{X})]| \leq [E|R_4(\bar{X})|] \leq \frac{M}{4!}E[|X - \theta|^4] = \frac{M}{4!}O_p[|\frac{1}{n^2}|]$.

Taking Expectations of the Delta Formula (case $m = 4$)

$$
\begin{aligned}
E[h(\bar{X})] &= E[h(\theta) + (\bar{X}-\theta)h^{(1)}(\theta) + (\bar{X}-\theta)^2 \frac{h^{(2)}(\theta)}{2}] + \\
&\quad + E[(\bar{X}-\theta)^3 \frac{h^{(3)}}{3!} + R_4(X)] \\
&= h(\theta) + \frac{\sigma^2}{n}\frac{h^{(2)}(\theta)}{2} + E[(\bar{X}-\theta)^3]\frac{h^{(3)}}{3!} + E[R_4(X)] \\
&= h(\theta) + \frac{\sigma^2}{n}\frac{h^{(2)}(\theta)}{2} + \frac{\mu_3}{n^2}\frac{h^{(3)}}{3!} + E[R_4(X)] \\
&= h(\theta) + \frac{\sigma^2}{n}\frac{h^{(2)}(\theta)}{2} + O_p(\frac{1}{n^2})
\end{aligned}
$$

because $|E[R_4(\bar{X})]| \leq [E|R_4(\bar{X})|] \leq \frac{M}{4!}E[|X-\theta|^4] = \frac{M}{4!}O_p[|\frac{1}{n^2}|]$.

Key Example: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ **, for i.i.d.** X_i

$$\begin{array}{rclcrcl} E[X_1] & = & \theta, & & E[(X_1-\theta)^2] & = & \sigma^2 \\ E[(X_1-\mu)^3] & = & \mu_3 & & E[|X_1-\mu|^3] & = & \kappa_3 \end{array}$$

With $\bar{X}$ **for a given sample size** n

$$\begin{array}{rclcrcl} E[\bar{X}] & = & \theta, & & E[(\bar{X}-\theta)^2] & = & \frac{\sigma^2}{n} \\ E[(\bar{X}-\theta)^3] & = & \frac{\mu_3}{n^2} & & E[|\bar{X}-\mu|^3] & = & 0_p[(\frac{1}{\sqrt{n}})^3] \end{array}$$

Limit Laws from the Delta Formula (case $m=2$**)**

$$\begin{array}{rcl} h(\bar{X}) & = & h(\theta) + (\bar{X}-\theta)h^{(1)}(\theta) + R_2(\bar{X}) \\ \Longrightarrow \sqrt{n}[h(\bar{X}) - h(\theta)] & = & \sqrt{n}(\bar{X}-\theta)h^{(1)}(\theta) + \sqrt{n}R_2(\bar{X}) \\ & = & \sqrt{n}(\bar{X}-\theta)h^{(1)}(\theta) + O_p(\frac{1}{\sqrt{n}}) \\ & \xrightarrow{\mathcal{L}} & N(0, \frac{\sigma^2}{n}[h^{(1)}(\theta)]^2) \end{array}$$

since $\sqrt{n}|E[R_2(\bar{X})]| \leq \sqrt{n}\frac{M}{2}\frac{\sigma^2}{n}$

Note: if $h^{(1)}(\theta) = 0$, then

- $\sqrt{n}[h(\bar{X}) - h(\theta)] \xrightarrow{Pr} 0.$
- Consider increasing scaling to $n[h(\bar{X}) - h(\theta)]$