Stat210A: Theoretical Statistics — **Lecture Date: October 9, 2014**

Central Limit Theorem, MLEs, Asymptotic Relative Efficiency

Lecturer: Michael I. Jordan — *Scribe: Lucia Petito*

1 Central Limit Theorem and Other Asymptotic Results

Definition 1 (Convergence in Distribution)**.** Let Y_n be a sequence of random variables, H_n be the CDF of Y_n, and H be the CDF of Y. $Y_n \xrightarrow{d} Y$ if $H_n(y) \to H(y)$ if H is continuous at y.

Lemma 2 (Two Implications of Convergence in Distribution)**.** .

1. *$Y_n \xrightarrow{d} Y$ iff $Ef(Y_n) \to Ef(Y)$ $\forall$ bounded, continuous functions f.*
2. *$Y_n \xrightarrow{d} Y$ and g is continuous $\implies$ $g(Y_n) \xrightarrow{d} g(Y)$.*

Theorem 3 (Central Limit Theorem)**.** *Given X_i i.i.d. with mean μ, variance σ^2,*

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

Proof. Can be found in Stat 205A, follows from basic Fourier analysis □

Example. Let H_n be the CDF of $\sqrt{n}(\bar{X}_n - \mu)$. What is $P(\mu - a/\sqrt{n} \leq \bar{X}_n \leq \mu + a/\sqrt{n})$?

$$\begin{aligned} P(\mu - a/\sqrt{n} \leq \bar{X}_n \leq \mu + a/\sqrt{n}) &= P(-a \leq \sqrt{n}(\bar{X}_n - \mu) \leq a) \\ &= H_n(a) - H_n(-a) \xrightarrow{d} \Phi(a/\sigma) - \Phi(-a/\sigma) \\ &\text{(By the CLT)} \end{aligned}$$

Theorem 4 (Slutsky's Theorem)**.** *Suppose $Y_n \xrightarrow{d} Y$, $A_n \xrightarrow{P} a$, $B_n \xrightarrow{P} b$. Then*

$$A_n + B_n Y_n \xrightarrow{d} a + bY$$

Delta Method Given assumptions underlying the CLT (3) and f is differentiable at μ,

$$\sqrt{n}(f(\bar{X}_n) - f(\mu)) \xrightarrow{d} N(0, (f'(\mu))^2\sigma^2)$$

Definition 5 (Uniform Integrability)**.**

$$\sup_{n \geq 1} E[|X_n| 1_{|X_n| \geq t}] \to 0 \text{ as } t \to \infty$$

Theorem 6 (8.16)**.** *$X_n \xrightarrow{d} X$ & uniform integrability* $\implies$ $E[X_n] \to E[X]$

2 Maximum Likelihood Estimator (MLE)

If X has density p_θ and $L(\theta) = p_\theta(x)$, then $\hat{\theta}_{MLE} = \text{argmax}_\theta L(\theta)$.

What is the MLE of $g(\theta)$? $g(\hat{\theta}_{MLE})$

Example. Exponential Family

$$\begin{aligned}
p_\eta(x) &= \exp(\eta T(x) - A(\eta))h(x) \\
\log p_\eta(x) &= \eta T(x) - A(\eta) + \log h(x) \\
\frac{\partial}{\partial \eta} \log p_\eta(x) &= T(x) - A'(\eta) \equiv 0 \\
T(x) &= A'(\eta) \\
\hat{\eta}_{MLE} &= \psi(T) \text{ where } \psi \text{ is the inverse of } A' \\
&\quad \text{(Show the second derivative is always } < 0) \\
\frac{\partial^2}{\partial \eta^2} \log p_\eta(x) &= -A''(\eta) = -Var(T) < 0
\end{aligned}$$

If we look at this in a random sample, $\{X_i\}$

$$0 = \sum_{i=1}^{n} T(x_i) - nA'(\eta) \implies \hat{\eta}_{MLE} = \psi(\bar{T})$$

What about estimating $\mu = E_\eta[T]$?

$$E_\eta[T] = A'(\eta) \implies \hat{\mu}_{MLE} = A'(\hat{\eta}_{MLE}) = A'(A'^{-1}(\bar{T})) = \bar{T} \text{ which is UMVU}$$

Now for the derivative of ψ. Given $\psi(T) = \eta$, $A'(\eta) = T$, then $A''(\eta) = \frac{dT}{d\eta}$. We want to calculate $\frac{d\eta}{dT}$.

$$\frac{d\eta}{dT} = \frac{1}{A''(\eta)} = \frac{1}{(A'' \circ \psi)(T)} \qquad \text{(setting up the delta method)}$$

Evaluate at $\mu = E_\eta[T] = A'(\eta) = \frac{1}{A''(\psi(A'(\eta)))} = \frac{1}{A''(\eta)} = \frac{1}{Var(T)}$

By the Delta Method,

$$\sqrt{n}(\hat{\eta}_{MLE} - \eta) \xrightarrow{d} N\left(0, \frac{1}{A''(\eta)}\right)$$

[Note: we've seemingly achieved the Cramer-Rao bound]

How does this relate to Cramer-Rao?

Fisher Info (Exponential Family): $A''(\eta)$ [proved previously] For unbiased $\tilde{\eta}$,

$$Var_\eta(\tilde{\eta}) \geq \frac{1}{nA''(\eta)} \implies Var_\eta(\tilde{\eta} - \eta) \geq \frac{1}{A''(\eta)}$$

$Var_\eta \sqrt{n}(\tilde{\eta} - \eta) = nVar_\eta \tilde{\eta}$

- MLE achieves Cramer-Rao in an asymptotic sense
- This is an heuristic argument! Super efficiency throws a wrench in the works (some points can break CR bound)
- Martin will discuss in 210B
- Keener in Ch 16 is harder to read, vander Vaart presents it nicely

3 Asymptotic Relative Efficiency (ARE)

Example. Compare the mean and median: which has the smaller limiting variance?

Mean follows CLT, limiting distribution is $N(0, \sigma^2)$.

Median (actually all quantiles)

$$\tilde{X}_n = \begin{cases} X_{(m)} & n = 2m - 1 \text{ (odd)} \\ 1/2[X_{(m)} + X_{(m+1)}] & n = 2m \text{ (even)} \end{cases}$$

Let $\{X_i\}$ be i.i.d. F, and let F have a unique median θ. $F(\theta) = 1/2$, and assume $0 < F'(\theta) < \infty$.

We want to approximate

$$P(\sqrt{n}(\tilde{X}_n - \theta) \leq a) = P(\tilde{X}_n \leq \theta + a/\sqrt{n})$$

"Trick" step. Let $S_n = \#\{i \leq n : X_i \leq \theta + a/\sqrt{n}\}$. These can be thought of as Bernoulli dots, do they fall before or after $\theta + a/\sqrt{n}$?

We have $\tilde{X}_n \leq \theta + a/\sqrt{n} \iff S_n \geq m$.

$S_n \sim Bin(n, F(\theta + a/\sqrt{n}))$ CLT for Binomials: say $Y_n \sim Bin(n, p)$. Then, CLT implies

$$\sqrt{n}(Y_n/n - p) \xrightarrow{d} N(0, p(1-p))$$

So,

$$P\left(\frac{Y_n - np}{\sqrt{n}} > y\right) = 1 - P\left(\frac{Y_n - np}{\sqrt{n}} \leq y\right) \xrightarrow{n\to\infty} 1 - \Phi\left(\frac{y}{\sqrt{p(1-p)}}\right) = \Phi\left(\frac{-y}{\sqrt{p(1-p)}}\right)$$

Now we can continue with our approximation from above:

$$P(\sqrt{n}(\tilde{X}_n - \theta) \leq a) = P(S_n > m + 1) = P\left(\frac{S_n - nF(\theta + a/\sqrt{n}}{\sqrt{n}} \geq \frac{m + 1 - nF(\theta + a/\sqrt{n}}{\sqrt{n}}\right)$$

$$\to \Phi\left(\frac{(nF(\theta + a/\sqrt{n}) - m + 1)/\sqrt{n}}{\sqrt{F(\theta + a/\sqrt{n})(1 - F(\theta + a/\sqrt{n}))}}\right)$$

$$\text{Numerator} = a\frac{F(\theta + a/\sqrt{n}) - F(\theta)}{a/\sqrt{n}} + \frac{nF(\theta) - m + 1}{\sqrt{n}} \to aF'(\theta)$$

(First term is deriv of F at θ, second goes to 0), denominator of this approaches 1/2

In conclusion:

$$P(\sqrt{n}(\tilde{X}_n - \theta) \leq a) \to \Phi(2aF'(\theta)) \implies \sqrt{(}\tilde{X}_n - \theta) \xrightarrow{d} N\left(0, \frac{1}{4(F'(\theta))^2}\right)$$