36-700: Probability and Mathematical Statistics **Spring 2019**

Lecture 7: WLLN and CLT

Lecturer: Jing Lei

7.1 Review and Outline

Last class we saw:

- Convergence of random variables
- Some limiting theorems

This lecture we will give a more detailed treatment of LLN and CLT.

7.2 Law of Large Numbers

Suppose we toss a fair coin n times and let $\bar{X}_n$ be the frequency of head. We shall expect $\bar{X}_n$ to converge to $1/2$ in an appropriate sense.

Theorem 7.1 (Weak Law of Large Numbers) *Let X_n be an iid sequence of random variables with finite mean μ and variance σ^2. Then*

$$\bar{X}_n \xrightarrow{P} \mu\,.$$

Proof: It is straightforward to verify that $\mathbb{E}(\bar{X}_n) = \mu$ and $\text{Var}(\bar{X}_n) = \sigma^2/n$. By Chebyshev's inequality

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \epsilon) \leq \epsilon^{-2}\text{Var}(\bar{X}_n) = \frac{\sigma^2}{n\epsilon^2} \to 0\,,$$

for all $\epsilon > 0$. ■

The proof of WLLN provides a simple way to bound the probability of $\bar{X}_n$ deviates from μ by a certain amount.

Example 7.2 *Let $\bar{X}_n$ be the head frequency in n tosses of a fair coin. We can calculate a lower bound of the probability of $0.4 < \bar{X}_n < 0.6$.*

$$\mathbb{P}(0.4 < \bar{X}_n < 0.6) = \mathbb{P}(|\bar{X}_n - \mu| < 0.1) = 1 - \mathbb{P}(|\bar{X}_n - \mu| \geq 0.1) \geq 1 - \frac{25}{n}\,.$$

In particular, when $n = 100$ we have $\mathbb{P}(0.4 < \bar{X}_n < 0.6) \geq 0.75$.

The result of Theorem 7.1 can be strengthened as follows. Let $0 \le r < 1/2$ be a constant. Under the setting of Theorem 7.1, using the same reasoning with Chebyshev inequality we have

$$\mathbb{P}(|n^r(\bar{X}_n - \mu)| \ge \epsilon) \le \frac{\sigma^2}{n^{1-2r}\epsilon^2} \to 0\,.$$

Therefore, $n^r(\bar{X}_n - \mu) \xrightarrow{P} 0$ for all $0 \le r < 1/2$.

7.3 The Central Limit Theorem

What happens if $r = 1/2$? The answer is given by the Central Limit Theorem.

Theorem 7.3 (Central Limit Theorem) *Let X_n be an iid sequence of random variables with finite mean μ and variance σ^2, then*

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \rightsquigarrow N(0,1)\,.$$

Example 7.4 *In Theorem 7.2, we can also approximate the probability $\mathbb{P}(0.4 < \bar{X}_n < 0.6)$ using the CLT. Recall that $\mu = 0.5$ and $\sigma^2 = 1/4$, so*

$$\begin{aligned}
&\mathbb{P}(0.4 < \bar{X}_n < 0.6)\\
=&\mathbb{P}(-0.1 < \bar{X}_n - \mu < 0.1)\\
=&\mathbb{P}\left(\frac{-0.1\times\sqrt{n}}{\sigma} < \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} < \frac{0.1\times\sqrt{n}}{\sigma}\right)\\
\approx&\Phi(0.2\sqrt{n}) - \Phi(-0.2\sqrt{n})\,,
\end{aligned}$$

where Φ is the cdf of $N(0,1)$. In the special case of $n = 100$, we have $\mathbb{P}(0.4 < \bar{X}_n < 0.6) \approx 0.9545$.

This is a much sharper approximation than the bound obtained in Theorem 7.2 using Chebyshev's inequality.

In Theorem 7.4, a key ingredient in the application of CLT is that we know the value of σ. In many applications, σ^2 is not known and needs to be estimated. Given an iid sample of size n, a popular estimator of σ^2 is

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2\,.$$

One can show that $S_n^2 \xrightarrow{P} \sigma^2$. Then using Slutsky's theorem we obtain the following version of CLT.

Theorem 7.5 *Under the same setting as Theorem 7.3, we have*

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \rightsquigarrow N(0,1)\,.$$

The CLT can be generalized to the multivariate case.

Theorem 7.6 (Multivariate CLT) *Let X_n be a sequence of iid random vectors*

$$X_n = \begin{pmatrix} X_{1n} \\ X_{2n} \\ \vdots \\ X_{kn} \end{pmatrix}$$

with mean

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix}$$

where $\mu_j = \mathbb{E}(X_{jn})$, and variance matrix Σ where $\Sigma_{jl} = \mathsf{Cov}(X_{jn}, X_{ln})$, then

$$\sqrt{n}(\bar{X}_n - \mu) \rightsquigarrow N(0,\Sigma)\,.$$

7.4 The Delta Method

A combination of CLT and Slutsky's theorem gives the very useful Delta Method.

Theorem 7.7 *If*

$$\frac{\sqrt{n}(Y_n - \mu)}{\sigma} \rightsquigarrow N(0,1)$$

and g is a differentiable function such that $g'(\mu) \neq 0$. Then

$$\frac{\sqrt{n}(g(Y_n) - g(\mu))}{g'(\mu)\sigma} \rightsquigarrow N(0,1)\,.$$

The formal proof, though quite elementary, is a bit tricky and is given in the Appendix. Here we provide a quick heuristic argument. Using Taylor expansion, $g(Y_n) - g(\mu) = g'(\mu)(Y_n - \mu) + Rem$, where Rem is the remainder term. Thus

$$\frac{\sqrt{n}(g(Y_n) - g(\mu))}{g'(\mu)\sigma} \approx \frac{\sqrt{n}(Y_n - \mu)g'(\mu)}{g'(\mu)\sigma} \rightsquigarrow N(0,1)\,.$$

Example 7.8 *Let $X_1, ...X_n, ...$ be iid with finite mean μ and variance σ^2. By CLT we have $\sqrt{n}(\bar{X}_n - \mu) \rightsquigarrow N(0, \sigma^2)$. Let $W_n = e^{\bar{X}_n}$, then the Delta Method implies that $\sqrt{n}(W_n - e^\mu) \rightsquigarrow N(0, \sigma^2 e^{2\mu})$.*

The Delta Method can be extended to the multivariate case.

Theorem 7.9 (Multivariate Delta Method) *Let $Y_n = (Y_{n1}, ..., Y_{nk})^T$ be a sequence of random vectors and $g : \mathbb{R}^k \mapsto \mathbb{R}$ be a differentiable function with derivative*

$$\nabla g(y) = \begin{pmatrix} \frac{\partial g}{\partial y_1} \\ \frac{\partial g}{\partial y_2} \\ \vdots \\ \frac{\partial g}{\partial y_k} \end{pmatrix}.$$

If

$$\sqrt{n}(Y_n - \mu) \rightsquigarrow N(0, \Sigma)$$

then

$$\sqrt{n}(g(Y_n) - g(\mu)) \rightsquigarrow N\left(0, \nabla_\mu^T \Sigma \nabla_\mu\right),$$

where ∇_μ is $\nabla g(y)$ evaluated at $y = \mu$.

Example 7.10 *Let $X_n = \begin{pmatrix} X_{n1} \\ X_{n2} \end{pmatrix}$ be iid bivariate random variables with mean $\mu = (\mu_1, \mu_2)^T$ and covariance Σ. Let $\hat{\mu}_{1,n} = n^{-1} \sum_{i=1}^n X_{i1}$, and $\hat{\mu}_{2,n} = n^{-1} \sum_{i=1}^n X_{i2}$. By multivariate CLT,*

$$\sqrt{n} \begin{pmatrix} \hat{\mu}_{1,n} - \mu_1 \\ \hat{\mu}_{2,n} - \mu_2 \end{pmatrix} \rightsquigarrow N(0, \Sigma).$$

Let $Y_n = \hat{\mu}_{1,n}\hat{\mu}_{2,n}$, then $Y_n = g(\hat{\mu}_{1,n}, \hat{\mu}_{2,n})$ with $g(s_1, s_2) = s_1 s_2$, and

$$\nabla g(s) = \begin{pmatrix} \frac{\partial g}{\partial s_1} \\ \frac{\partial g}{\partial s_2} \end{pmatrix} = \begin{pmatrix} s_2 \\ s_1 \end{pmatrix}.$$

Thus

$$\nabla_\mu^T \Sigma \nabla_\mu = (\mu_2, \mu_1) \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} \mu_2 \\ \mu_1 \end{pmatrix} = \mu_2^2 \sigma_{11} + 2\mu_1\mu_2\sigma_{12} + \mu_1^2 \sigma_{22}.$$

By Multivariate Delta Method,

$$\sqrt{n}(\hat{\mu}_{1,n}\hat{\mu}_{2,n} - \mu_1\mu_2) \rightsquigarrow N(0, \mu_2^2\sigma_{11} + 2\mu_1\mu_2\sigma_{12} + \mu_1^2\sigma_{22}).$$

Appendix

Sketch of proof of CLT.

Assume further that the mgf of X_1 exists.

$$\begin{aligned}
M_{\sqrt{n}(\bar{X}_n-\mu)}(t) =& M_{\sum_{i=1}^n (X_i-\mu)/\sqrt{n}}(t) = \prod_{i=1}^n M_{(X_i-\mu)/\sqrt{n}}(t) \\
=& \prod_{i=1}^n \mathbb{E}e^{(X_i-\mu)t/\sqrt{n}} = \left[\mathbb{E}e^{(X_1-\mu)t/\sqrt{n}}\right]^n \\
=& \left[\mathbb{E}\left(1+(X_1-\mu)t/\sqrt{n}+\frac{t^2}{2n}(X_1-\mu)^2+...\right)\right]^n \\
=& \left(1+0+\frac{t^2}{2n}\mathbb{E}(X_1-\mu)^2+...\right)^n \\
\approx& (1+t^2\sigma^2/(2n))^n \to e^{t^2\sigma^2/2} .
\end{aligned}$$

Sketch of proof of the Delta method.

Let $h(y) = \frac{g(y)-g(\mu)}{y-\mu} - g'(\mu)$. When $y \to \mu$, $h(y) \to 0$ by definition of derivative. Define $h(\mu) = 0$, then h is continuous at μ.

Now we have $g(y) - g(\mu) = (y-\mu)g'(\mu) + (y-\mu)h(y)$. So,

$$\sqrt{n}\,[g(Y_n) - g(\mu)] = \sqrt{n}(Y_n-\mu)g'(\mu) + \sqrt{n}(Y_n-\mu)h(Y_n) := A_{1,n} + A_{2,n} .$$

By assumption the first term $A_{1,n} \rightsquigarrow N(0, g'(\mu)^2\sigma^2)$. Now the proof concludes if we show that $A_{2,n} \xrightarrow{P} 0$.

First the assumption $\sqrt{n}(Y_n-\mu) \rightsquigarrow N(0,\sigma^2)$ implies that $Y_n \xrightarrow{P} \mu$. Thus the continuous mapping theorem implies that $h(Y_n) \xrightarrow{P} 0$. Then Slutsky's theorem implies that $A_{2,n} \rightsquigarrow 0$.