Convergence Concepts

Feng-Chang Lin

Department of Biostatistics
University of North Carolina at Chapel Hill

flin@bios.unc.edu

(C&B §5.5)

Introduction

- For random samples from the normal distribution, we derived the *exact* joint distribution of $(\bar{X}, S^2)$.
- For other distributions, the *exact* distribution may be too complicated to be of practical use.
- Instead, *approximate* distribution may be easier to derive or be computed.
- In this section, we will study the behavior of sample statistics in large samples, or say, $n \to \infty$.
- The term *large sample theory* or *asymptotic theory* refers to this approach.

Two Basic Tools

- Law of large numbers (LLN) and central limit theorem (CLT)
- That says, loosely, when sample size is large, the sample mean is close to the population mean (LLN) and the sample mean is approximately normally distributed (CLT).
- We will need Taylor's expansion from calculus, Slutsky's theorem, and delta method.
- All these tools rely on the mathematical notion of convergence.

Convergent Non-Random Sequences

- Sequences will be denoted by either $a_1, a_2, \cdots$ or by $\{a_n\}$.
- A sequence $\{a_n\}$ of real numbers is said to *converge* if there is a point a with the following property:
- For every $\epsilon > 0$, there is an integer N such that $n \geq N$ implies that $|a_n - a| < \epsilon$.
- In this case we say that $\{a_n\}$ converges to a, or that a is the limit of $\{a_n\}$, and we write $\lim_{n\to\infty} a_n = a$, or $a_n \to a$ as $n \to \infty$.
- If $\{a_n\}$ does not converge, it is said to *diverge*.
- The above definitions apply as well to sequences in R^k (finite k), with $|\cdot|$ replaced by Euclidean distance $\|\cdot\|$.

Convergent Random Sequences

- Does a sequence $\{X_n\}$ of random variables converge to a limit random variable X?
- Is there a meaningful way to say that "$X_n \to X$ as $n \to \infty$"?
- Remember $\{X_n\}$ is a "random" sequence, so whether $\{X_n\}$ converges to X or not is a "random" event.
- That means, some sequences converge, others do not.
- Since $\{X_n \to X \;\; as \;\; n \to \infty\}$ is a random event, we can put

$$P(X_n \to X \;\; as \;\; n \to \infty) = 1.$$

- We claim "the sequence of random variables $X_1, X_2, \cdots$, *converges almost surely* to a random variable X.
- Written as $P(\lim_{n\to\infty} |X_n - X| = 0) = 1$.

Converge Almost Surely

- Recall, a random variable is a real-value function defined on the sample space S.
- One may also write almost sure convergence as

$$P(\{s : \lim_{n\to\infty} |X_n(s) - X(s)| = 0\}) = 1$$

- Notation: $X_n \rightarrow_{a.s.} X$ as $n \to \infty$.
- Almost sure convergence means that $X_n(s) - X(s)$ for all $s \in S$, except possibly for a subset of S that has zero probability.
- **Example** S is uniform on $[0, 1]$, and define $X_n(s) = s + s^n$. For every $s \in [0, 1)$, $X_n(s) \to s$. But for $s = 1$, $s^n \to 1$, and $X_n(1) \to 2 \neq 1$.
- One can still claim $X_n \rightarrow_{a.s.} s = X(s)$ as $n \to \infty$ since $P(S = 1) = 0$.

Strong Law of Large Numbers (SLLN)

- Let $X_1, \cdots, X_n$ be iid random variables with $EX_i = \mu$ and $VarX_i = \sigma^2 < \infty$. Let $\bar{X}_n = \sum_{i=1}^n X_i/n$. Then, for every $\epsilon > 0$,

$$P(\lim_{n\to\infty} |\bar{X}_n - \mu| < \epsilon) = 1.$$

- That is, $\bar{X}_n$ converges almost surely to μ.
- The property $\bar{X}_n \to_{a.s.} \mu$ is called *strong consistency* of $\bar{X}_n$ as an estimator of μ.
- One may also say that $\bar{X}_n$ is a *strongly consistent estimator* of μ.

Converge in Probability

- A weaker form of convergence.
- A sequence of random variables X_n converges in probability to a random variable X if, for every $\epsilon > 0$,

$$\lim_{n \to \infty} P(|X_n - X| < \epsilon) = 1.$$

- One may say, for $\epsilon > 0$, define $a_n(\epsilon) = P(|X_n - X| < \epsilon)$.
- Convergence in probability means that $a_n(\epsilon) \to 1$ as $n \to \infty$, for every $\epsilon > 0$.
- Notation: $X_n \to_p X$ as $n \to \infty$.
- **Convergence in probability, not almost surely** see example 5.5.8 in C&B.

Weak Law of Large Numbers (WLLN)

- Let $X_1, \cdots, X_n$ be iid random variables with $EX_i = \mu$ and $VarX_i = \sigma^2 < \infty$. Let $\bar{X}_n = \sum_{i=1}^n X_i/n$. Then, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \epsilon) = 1.$$

- That is, $\bar{X}_n$ converges in probability to μ.
- The property $\bar{X}_n \to_p \mu$ is called *consistency* of $\bar{X}_n$.
- Comment: The condition that EX_i exists and is finite is *sufficient* in both WLLN and SLLN.

Converge in Distribution

- Let F_{X_n} be the cdf of X_n.
- A sequence of random variables X_n converges in distribution to a random variable X if,

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

  at all points x where $F_X(x)$ is continuous.
- Notation: $X_n \to_d X$ as $n \to \infty$.
- Convergence in distribution does not imply that X_n and X approximate each other.
- It only says that, for large n, the cdf of X_n becomes close to the cdf of X.

Central Limit Theorem (CLT)

- Let $X_1, \cdots, X_n$ be iid random variables with $EX_i = \mu$ and $VarX_i = \sigma^2 < \infty$. Define $\bar{X}_n = \sum_{i=1}^{n} X_i/n$, $Z_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$, and let G_n denote the cdf of Z_n. For any $-\infty < z < \infty$,

$$lim_{n\to\infty} G_n(z) = \Phi(z).$$

- That is, Z_n has a limiting standard normal distribution, $Z_n \to_d N(0,1)$ as $n \to \infty$.

Central Limit Theorem (cont'd)

- **Example** Suppose that $X_1, \cdots, X_n$ are iid *Bernoulli*(p), and define $Y = \sum_{i=1}^{n} X_i$. The CLT states that $Z_n = \sqrt{n}(\bar{X}_n - p)/\sqrt{p(1-p)}$ is approximately $N(0,1)$ for large n.
- Since $Z_n = (Y - np)/\sqrt{np(1-p)}$, that shows one can use a normal approximation to the binomial distribution of Y.
- Suppose $n = 100$ and $p = 0.5$. One can calculate $P(Y \leq 57) = 0.933$, which is close to $\Phi(z) = \Phi(1.4) = 0.919$.

Relationships between Modes of Convergence

- $X_n \rightarrow_{a.s.} X \Rightarrow X_n \rightarrow_p X \Rightarrow X_n \rightarrow_d X$.
- The converse statements are "generally" not true.
- **Example** A special case for $X_n \rightarrow_d X \Rightarrow X_n \rightarrow_p X$: If c is a non-random constant, $P(X = c) = 1$ then $X_n \rightarrow_d X$ implies that $X_n \rightarrow_p c$ (proofs in C&B).
- That is, convergence in distribution to a degenerate one-point distribution implies convergence in probability.

Slutsky's Theorem

- If $X_n \to_d X$ and $Y_n \to_p a$, where a is a finite constant, then
  1. $Y_n X_n \to_d aX$;
  2. $Y_n + X_n \to_d a + X$;
  3. $X_n / Y_n \to_d X/a$ if $a \neq 0$;
- Slutsky's theorem allows substituting consistent estimators when proving convergence in distribution.
- X_n and Y_n need not be independent.
- **Example** Suppose that $X \sim N(0, \sigma^2)$ and $T_n \to_d X$ as $n \to \infty$. By Slutsky's theorem, $T_n/\sigma \to_d X/\sigma$. Since $X/\sigma \sim N(0, 1)$, we conclude that $T_n/\sigma \to_d N(0, 1)$.

Convergence of Transformed Sequences

- Suppose that h is a continuous function.
- One has
  1. If $X_n \to_{a.s.} X$ then $h(X_n) \to_{a.s.} h(X)$.
  2. If $X_n \to_p X$ then $h(X_n) \to_p h(X)$.
  3. If $X_n \to_d X$ then $h(X_n) \to_d h(X)$.
- h needs be continuous only on the range of X. For example, if X is non-negative, the behavior of $h(x)$ for $x < 0$ does not matter.
- **Example** Let $X_1, \cdots, X_n$ be iid random variables with mean μ and variance $\sigma^2 < \infty$. Does the sample variance S_n^2 converge to σ^2 in some sense? Write

$$S_n^2 = \frac{1}{n-1}\left\{\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2\right\} = \frac{n}{n-1}\frac{\sum_{i=1}^{n} X_i^2}{n}. - \frac{n}{n-1}\bar{X}_n^2$$

Convergence of Transformed Sequences (cont'd)

- As $n \to \infty$, $n/(n-1) \to 1$,

$$\frac{\sum_{i=1}^{n} X_i^2}{n} \to_{a.s.} EX_1^2 = \mu^2 + \sigma^2,$$

  and

$$\bar{X}_n^2 \to_{a.s.} \mu^2.$$

- Slutsky's theorem and convergence of transformed random sequences lead to the result that $S_n^2 \to_{a.s.} \sigma^2$ as $n \to \infty$.
- **Example** Suppose that $\{T_n\}$ is a random sequence with $\sqrt{n}(T_n - \theta) \to_d N(0, \sigma^2)$. The asymptotic distribution of T_n is centered about θ. But, does T_n converge to θ in some sense? That is, is $T_n \to_p \theta$?

Convergence of Transformed Sequences (cont'd)

- Let $Z_n = \sqrt{n}(T_n - \theta)/\sigma \to_d N(0, 1)$
- Given $\epsilon > 0$, one has

$$
\begin{aligned}
P(|T_n - \theta| < \epsilon) &= P(-\sqrt{n}\epsilon/\sigma < Z_n < \sqrt{n}\epsilon/\sigma) \\
&< P(-\sqrt{n}\epsilon/\sigma < Z_n \leq \sqrt{n}\epsilon/\sigma) \\
&= P(Z_n \leq \sqrt{n}\epsilon/\sigma) - P(Z_n \leq -\sqrt{n}\epsilon/\sigma).
\end{aligned}
$$

- Since Z_n converges in distribution, $P(Z_n \leq \sqrt{n}\epsilon/\sigma) \to 1$ and $P(Z_n \leq -\sqrt{n}\epsilon/\sigma) \to 0$.
- Hence $P(|T_n - \theta| < \epsilon) \to 1$ as $n \to \infty$. That means, $T_n \to_p \theta$.

Delta Method - Univariate

- Suppose that $\{T_n\}$ is a random sequence with $\sqrt{n}(T_n - \theta) \to_d N(0, \sigma^2)$, and g is a function with $g'(\theta)$ exists and is not 0. Then

$$\sqrt{n}\{g(T_n) - g(\theta)\} \to_d N(0, \{g'(\theta)\}^2\sigma^2).$$

- We say that θ is the *asymptotic mean* of T_n. However θ may or may not be the mean of T_n. In fact, the mean of T_n may not even exist (example below).
- **Example** Suppose that $X_1, \cdots, X_n$ are iid Bernoulli(θ), $0 < \theta < 1$, and we want to make statistical inferences about the log-odds, which is defined by

$$\psi = \log\left(\frac{\theta}{1-\theta}\right).$$

Delta Method - Univariate (cont'd)

- Define $g(u) = \log\{u/(1-u)\}$ for $u \in (0,1)$, so $\psi = g(\theta)$.
- By SLLN, $\bar{X}_n \rightarrow_{a.s.} \theta$. Since g is continuous at $\theta \in (0,1)$, one has that $g(X_n) \rightarrow_{a.s.} g(\theta)$.
- Since $g'(\theta) = 1/\{\theta(1-\theta)\} \neq 0$ for $\theta \in (0,1)$, the delta method gives

$$\sqrt{n}(g(\bar{X}_n) - g(\theta)) \rightarrow_d N(0, \{g'(\theta)\}^2\theta(1-\theta)),$$

  or, equivalently,

$$\sqrt{n}(g(\bar{X}_n) - \psi) \rightarrow_d N\left(0, \frac{1}{\theta(1-\theta)}\right).$$

- The *asymptotic mean* of $g(\bar{X}_n)$ is ψ.
- The *exact mean* $Eg(\bar{X}_n)$ does not exist because $g(0) = -\infty$, $P(\bar{X}_n = 0) > 0$, $g(1) = \infty$, $P(\bar{X}_n = 1) > 0$, $Eg(\bar{X}_n) = \infty - \infty$.

Delta Method - Univariate (cont'd)

- Can the distribution above be used in practice? Why?
- We know that if a random variable Z follows a $N(0, 1/\{\theta(1-\theta)\})$, then $\sqrt{\theta(1-\theta)}Z$ follows a $N(0, 1)$.
- Is the following statement true?

$$\sqrt{\theta(1-\theta)}\sqrt{n}(g(\bar{X}_n) - \psi) \to_d N(0, 1),$$

  and

$$\sqrt{\bar{X}(1-\bar{X})}\sqrt{n}(g(\bar{X}_n) - \psi) \to_d N(0, 1).$$

- To construct a 95% CI for log-odds ψ, which one to use?

Second-order Delta Method

- Suppose that T_n is a random sequence with $\sqrt{n}(T_n - \theta) \to_d N(0, \sigma^2)$, and g is a function with $g'(\theta) = 0$ and $g''(\theta)$ exists and is not 0. Then

$$n\{g(T_n) - g(\theta)\} \to_d \sigma^2 \frac{g''(\theta)}{2} \chi_1^2$$

- **Example** $g(T_n) = \bar{X}_n(1 - \bar{X}_n)$, $g(\theta) = \theta(1 - \theta)$, $g'(\theta) = 1 - 2\theta$, $g''(\theta) = -2$. If $\theta = 1/2$, one can have

$$n\left\{\bar{X}_n(1 - \bar{X}_n) - \frac{1}{4}\right\} \to_d -\frac{1}{4}\chi_1^2.$$

Delta Method - multivariate

- Let the p-dimensional random vectors $X_1, \cdots, X_n$ be a random sample with $EX_{ij} = \mu_j$ ($j = 1, \cdots, p$) and $Cov(X_{ij}, X_{ik}) = \sigma^2_{jk}$.
- The population mean vector will be denoted by $\mu = (\mu_1, \cdots, \mu_p)$.
- If a function g maps R^p into R and has continuous first partial derivatives, $\partial g(t)/\partial t_j$, then

$$\sqrt{n}\{g(\bar{X}_1, \cdots, \bar{X}_p) - g(\mu_1, \cdots, \mu_p)\} \to_d N(0, \tau^2),$$

where

$$\tau^2 = \sum_{j=1}^{p} \sum_{k=1}^{p} \sigma^2_{jk} \frac{\partial g(\mu)}{\partial \mu_j} \frac{\partial g(\mu)}{\partial \mu_k},$$

provided that $\tau^2 > 0$.

Pair-Matched Case-Control Study

- A case (i.e., a diseased person, denoted D) is ”matched" (on covariates such as age, race, and sex) to a control (i.e., non-diseased person, denoted $\bar{D}$).
- Each member of the pairs is then interviewed as to the presence (E) or absence ($\bar{E}$) of a history of exposure to some harmful substance (e.g., cigarette smoke, asbestos, benzene, etc.)
- The data from such study involving n case-control pairs can be presented in tabular form as follows:

		$\bar{D}$		
		E	$\bar{E}$	
D	E	Y_{11}	Y_{10}	
	$\bar{E}$	Y_{01}	Y_{00}	
				n

Pair-Matched Case-Control Study (cont'd)

- Y_{11} is the number of pairs where *both* case *and* control are exposed (i.e., both have a history of exposure).
- Y_{10} is the number of pairs where case is exposed but the control is not, and so on.
- Clearly $\sum_{j=0}^{1}\sum_{k=0}^{1} Y_{jk} = n$.
- Assume that $\{Y_{ij}\}$ have a multinomial distribution with sample size n and associated cell probabilities $\{\pi_{ij}\}$, where

$$\sum_{j=0}^{1}\sum_{k=0}^{1} \pi_{jk} = 1.$$

- The interpretation is that π_{10} is the probability of obtaining a pair in which the case is exposed and its matched control is not.

Pair-Matched Case-Control Study (cont'd)

- In such study, the parameter measuring the association between exposure and disease is the odds ratio $\psi = \pi_{10}/\pi_{01}$. Intuitively, the estimator for ψ is $\hat{\psi} = Y_{10}/Y_{01}$.
- To derive the large sample distribution of $\hat{\psi}$, it will be easier to work on $\log(\hat{\psi})$, instead of $\hat{\psi}$.
- By the delta method in the multivariate case, think about $g(\pi_{10}, \pi_{01}) = \log(\pi_{01}/\pi_{01})$.
- Then, one has

$$\sqrt{n}\{\log(Y_{10}/Y_{01}) - \log(\pi_{10}/\pi_{01})\} \to_d N(0, \tau^2),$$

where

$$\tau^2 = A + B + C,$$

Pair-Matched Case-Control Study (cont'd)

- With

$$A = \left\{ \frac{\partial g(\pi_{10}, \pi_{01})}{\partial \pi_{10}} \right\}^2 \sigma_{10}^2 = \frac{1}{\pi_{10}^2} \pi_{10}(1 - \pi_{10}),$$

$$B = \left\{ \frac{\partial g(\pi_{10}, \pi_{01})}{\partial \pi_{01}} \right\}^2 \sigma_{01}^2 = \frac{1}{\pi_{01}^2} \pi_{01}(1 - \pi_{01})$$

and

$$C = 2 \frac{\partial g(\pi_{10}, \pi_{01})}{\partial \pi_{01}} \frac{\partial g(\pi_{10}, \pi_{01})}{\partial \pi_{10}} \sigma_{10,01}^2 = \frac{-2}{\pi_{10}\pi_{01}} (-\pi_{10}\pi_{01}).$$

- That concludes,

$$\tau^2 = \frac{1}{\pi_{10}}(1 - \pi_{10}) + \frac{1}{\pi_{01}}(1 - \pi_{01}) + 2 = \frac{1}{\pi_{10}} + \frac{1}{\pi_{01}}.$$

Pair-Matched Case-Control Study (cont'd)

- A common way to express $Var\{\log(\hat{\psi})\}$ is

$$Var\{\log(\hat{\psi})\} \approx \frac{1}{n\pi_{10}} + \frac{1}{n\pi_{01}}.$$

- That gives a common estimator for the variance of $\log(\hat{\psi})$ as

$$\widehat{Var}\{\log(\hat{\psi})\} \approx \frac{1}{Y_{10}} + \frac{1}{Y_{01}}.$$

- And, the large sample distribution of $\log(\hat{\psi})$ is

$$\frac{\log(\hat{\psi}) - \log(\psi)}{\sqrt{\widehat{Var}\{\log(\hat{\psi})\}}} \sim N(0, 1).$$