Large Sample Theory

References:
1. Bickel, D. and Doksum, K. *Mathematical Statistics*
2. Lehman, E. and Casella, G. *Theory of Point Estimation*

Convergence and Limit Theorems – Review

Let $X_1, X_2, \ldots$ be a sequence of random variables. Then there are different modes of convergence that apply (unlike sequences of real or complex numbers).

Definition. $X_n \xrightarrow{P} X$ as $n \to \infty$ (i.e., X_n converges to X in probability) if $P(|X_n - X| \geq \epsilon) \to 0$ as $n \to \infty$ for each $\epsilon > 0$.

Note that $P(|X_n - X| \geq \epsilon) \equiv P(\{\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\})$.

Example. By the Weak Law of Large Numbers (WLLN), in the i.i.d case,
(i) $\bar{X} \xrightarrow[n\to\infty]{P} \mu$;
(ii) $\hat{p} \xrightarrow[n\to\infty]{P} p$.

Definition. $X_n \xrightarrow{a.s.} X$ as $n \to \infty$ (i.e., X_n converges to X almost surely or almost everywhere) if $P\left(\lim_{n\to\infty} X_n = X\right) = 1$.

Note again that $P\left(\lim_{n\to\infty} X_n = X\right) = P\left(\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right)$.

Example. By the Strong Law of Large Numbers (SLLN), in the i.i.d case,
(i) $\bar{X} \xrightarrow[n\to\infty]{a.s.} \mu$;
(ii) $\hat{p} \xrightarrow[n\to\infty]{a.s.} p$.

Definition. $X_n \xrightarrow{d} X$ as $n \to \infty$ (i.e., X_n converges to X in distribution) if $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$ for all x where F_X is continuous.

Note that X_n and X need not be on the same space, or have a joint distribution.

Example. By the Central Limit Theorem (CLT), $\sqrt{n}(\bar{X} - \mu) \xrightarrow[n\to\infty]{d} N(0, \sigma^2)$.

Result. We have that

$$X_n \xrightarrow[n\to\infty]{a.s.} X \implies X_n \xrightarrow[n\to\infty]{P} X \implies X_n \xrightarrow[n\to\infty]{d} X.$$

Example. Consider independent $X_n \sim \text{Bernoulli}(\frac{1}{n})$. Then

$$P(|X_n - 0| > \epsilon) = P(X_n > \epsilon) = P(X_n = 1) = \frac{1}{n} \quad \underset{n\to\infty}{\longrightarrow} 0,$$

so that $X_n \underset{n\to\infty}{\overset{P}{\longrightarrow}} 0$, but does $X_n \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} 0$?

Result. We have

$$X_n \underset{n\to\infty}{\overset{d}{\longrightarrow}} X, \quad X \equiv c \text{ (a constant) then } X_n \underset{n\to\infty}{\overset{P}{\longrightarrow}} X.$$

Proof. We have $F_{X_n}(x) \to F_X(x)$ for all $x \neq c$. Thus, $F_{X_n}(c+\epsilon) \to 1$ and $F_{X_n}(c-\epsilon) \to 0$. Therefore,

$$\begin{aligned} P(|X_n - c| \geq \epsilon) &= P(X_n \leq c - \epsilon \text{ or } X_n \geq c + \epsilon) \\ &\leq F_{X_n}(c-\epsilon) + (1 - F_{X_n}(c+\epsilon/2)) \to 0. \end{aligned}$$

Result. Let g be a continuous function. Then

$$X_n \longrightarrow X \implies g(X_n) \longrightarrow g(X)$$

for all three modes of convergence.

This is easy to see for a.s. convergence.

Theorem (Slutsky). Suppose $X_n \underset{n\to\infty}{\overset{d}{\longrightarrow}} X$ and $U_n \underset{n\to\infty}{\overset{P}{\longrightarrow}} u_0$. Then

(a) $X_n + U_n \underset{n\to\infty}{\overset{d}{\longrightarrow}} X + u_0$;

(b) $U_n X_n \underset{n\to\infty}{\overset{d}{\longrightarrow}} u_0 X$.

Proof. Since $(X_n, U_n) \underset{n\to\infty}{\overset{d}{\longrightarrow}} (X, u_0)$, and $g(x,y) = x + y$ and $g(x,y) = xy$ are continuous functions, the result follows from the previous result.

Result. If $X_n - Y_n \underset{n\to\infty}{\overset{P}{\longrightarrow}} 0$ and $X_n \underset{n\to\infty}{\overset{d}{\longrightarrow}} X$, then $Y_n \underset{n\to\infty}{\overset{d}{\longrightarrow}} X$.

Proof. Note that $(X_n, X_n - Y_n) \underset{n\to\infty}{\overset{d}{\longrightarrow}} (X, 0)$. Therefore, $Y_n = X_n - (X_n - Y_n) \underset{n\to\infty}{\overset{d}{\longrightarrow}} X$, by Slutsky.

Chebychev's Inequality. For any random variable X, and $a > 0$,

$$P\left(|X| \geq a\right) \leq \frac{E(X^2)}{a^2}.$$

Proof. If $Y > 0$ and $a > 0$, we have

$$\begin{aligned} E(Y) &= \int_0^a y f_Y(y)\, dy + \int_a^\infty y f_Y(y)\, dy \\ &\geq \int_a^\infty y f_Y(y)\, dy \geq a P(Y \geq a), \end{aligned}$$

so that

$$P(Y \geq a) \leq \frac{E(Y)}{a}.$$

Therefore,

$$P(|X| \geq a) = P\left(X^2 \geq a^2\right) \leq \frac{E(X^2)}{a^2}.$$

The familiar form of this inequality is

$$P(|X-\mu| \geq \epsilon) = P\left((X-\mu)^2 \geq \epsilon^2\right) \leq \frac{E((X-\mu)^2)}{\epsilon^2} = \frac{Var(X)}{\epsilon^2}.$$

WLLN (Khinchine). If $X_1, X_2, \ldots$ are i.i.d such that $E(X)$ exists, then $\bar{X} \xrightarrow[n\to\infty]{P} E(X)$.

WLLN (Chebychev). If $X_1, X_2, \ldots$ is a sequence of random variables with $E(X_i) = \mu_i$, $Var(X_i) = \sigma_i^2$, $Cov(X_i, X_j) = 0$ for $i \neq j$, then

$$\frac{1}{n^2}\sum_{i=1}^n \sigma_i^2 \xrightarrow[n\to\infty]{} 0 \implies \bar{X}_n - \bar{\mu}_n \xrightarrow[n\to\infty]{P} 0.$$

Proof. Applying Chebychev's inequality,

$$\begin{aligned} P\left(|\bar{X}_n - \bar{\mu}_n| \geq \epsilon\right) &\leq \frac{1}{\epsilon^2} E\left(\bar{X}_n - \bar{\mu}_n\right)^2 = \frac{1}{\epsilon^2} E\left(\frac{1}{n}\sum_{i=1}^n (X_i - \mu_i)\right)^2 \\ &= \frac{1}{\epsilon^2}\frac{1}{n^2}\sum_{i=1}^n Var(X_i) \end{aligned}$$

If $\mu_i \equiv \mu$ then we obtain $\bar{X}_n \xrightarrow[n\to\infty]{P} \mu$ subject to conditions on σ_i^2 such as $\sigma_i^2 \equiv \sigma^2$. WLLN for an i.i.d sequence then is a special case.

SLLN (Kolmogorov). If $X_1, X_2, \ldots$ is an i.i.d sequence such that $E(X) = \mu$ exists and is finite, then $\bar{X}_n \xrightarrow[n\to\infty]{a.s.} \mu$.

CLT (Lindberg-Levy). Let $X_1, X_2, \ldots$ is an i.i.d sequence such that $E(X) = \mu$ and $Var(X) = \sigma^2$, $0 < \sigma^2 < \infty$. Then

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \xrightarrow[n\to\infty]{d} N(0,1).$$