MATH 556: MATHEMATICAL STATISTICS I

STOCHASTIC CONVERGENCE

The following definitions relate to a sequence $\{X_n\}$ of random variables defined on the same probability space $(\Omega, \mathcal{F}, P)$. The statements are given in terms of P for simplicity.

1. **Convergence Almost Surely:** $\{X_n\}$ **converges almost surely** to random variable X, denoted $X_n \xrightarrow{a.s.} X$, if for every $\epsilon > 0$

$$P\left[\lim_{n\longrightarrow\infty} |X_n - X| < \epsilon\right] = 1,$$

that is, if $A \equiv \{\omega : X_n(\omega) \longrightarrow X(\omega)\}$, then $P(A) = 1$. Equivalently, $X_n \xrightarrow{a.s.} X$ if for every $\epsilon > 0$

$$P\left[\lim_{n\longrightarrow\infty} |X_n - X| \geq \epsilon\right] = 0.$$

Equivalent terminology is

$X_n \longrightarrow X$ *almost everywhere*, $X_n \xrightarrow{a.e.} X$ $\qquad$ $X_n \longrightarrow X$ *with probability 1*, $X_n \xrightarrow{w.p.1} X$

Interpretation: The sequence of random variables $\{X_n\}$ corresponds to a sequence of functions defined on elements of Ω. Almost sure convergence requires that the sequence of real numbers $X_n(\omega)$ converges to $X(\omega)$ (as a real sequence) for all $\omega \in \Omega$, as $n \longrightarrow \infty$, except perhaps when ω is in a set having probability zero under the probability distribution of X. That is, for every $\omega \in \Omega$, except possibly those lying in a set of probability zero under P, we have

$$\lim_{n\longrightarrow\infty} X_n(\omega) = X(\omega).$$

Let $\epsilon > 0$, and for each $n \geq 1$, consider sets $A_n(\epsilon), B_n(\epsilon) \in \mathcal{F}$ defined by

$$A_n(\epsilon) \equiv \{\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\} \qquad\qquad B_n(\epsilon) \equiv \bigcup_{m=n}^{\infty} A_m(\epsilon).$$

Then we have $X_n \xrightarrow{a.s.} X$ if and only if $\lim_{n\longrightarrow\infty} P(B_n(\epsilon)) = 0$. Note that

$$A_n(\epsilon) \subseteq B_n(\epsilon) \qquad \Longrightarrow \qquad P(A_n(\epsilon)) \leq P(B_n(\epsilon))$$

so

$$\lim_{n\longrightarrow\infty} P(B_n(\epsilon)) = 0 \qquad \Longrightarrow \qquad \lim_{n\longrightarrow\infty} P(A_n(\epsilon)) = 0.$$

Note also that by continuity of probability,

$$\lim_{n\longrightarrow\infty} P(B_n(\epsilon)) = P\left(\lim_{n\longrightarrow\infty} B_n(\epsilon)\right) \equiv P\left(\bigcap_{n=1}^{\infty} B_n(\epsilon)\right) = P\left(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty} A_m(\epsilon)\right)$$

where, as $B_{n+1}(\epsilon) \subseteq B_n(\epsilon)$, $\{B_n(\epsilon)\}$ is a decreasing sequence of sets, we may define

$$\lim_{n\longrightarrow\infty} B_n(\epsilon) = \bigcap_{n=1}^{\infty} B_n(\epsilon).$$

- **Strong Law Of Large Numbers:** Suppose that $\{X_n\}$ is a sequence of random variables each with expectation μ. Let $\overline{X}_n$ be the sample mean. Then for all $\epsilon > 0$,

$$P\left[\lim_{n\longrightarrow\infty} \left|\overline{X}_n - \mu\right| < \epsilon\right] = 1,$$

that is, $\overline{X}_n \xrightarrow{a.s.} \mu$, and thus the mean of $X_1, \ldots, X_n$ converges almost surely to μ.

2. **Convergence in Probability:** The sequence $\{X_n\}$ **converges in probability** to random variable X, $X_n \stackrel{p}{\longrightarrow} X$, if, for all $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} P\left[|X_n - X| < \epsilon\right] = 1 \qquad \text{or equivalently} \qquad \lim_{n \longrightarrow \infty} P\left[|X_n - X| \geq \epsilon\right] = 0$$

Let $\epsilon > 0$, and consider $A_n(\epsilon)$ defined above. Then we have $X_n \stackrel{p}{\longrightarrow} X$ if

$$\lim_{n \longrightarrow \infty} P(A_n(\epsilon)) = 0$$

that is, if there exists an n such that for all $m \geq n$, $P(A_m(\epsilon))$ is arbitrarily small.

- As a special case, $\{X_n\}$ converges in probability to a **constant** c, denoted $X_n \stackrel{p}{\longrightarrow} c$, if for every $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} P\left[|X_n - c| < \epsilon\right] = 1 \qquad \text{or} \qquad \lim_{n \longrightarrow \infty} P\left[|X_n - c| \geq \epsilon\right] = 0$$

  that is, if the limiting distribution of $X_1, \ldots, X_n$ is **degenerate at** c.

- **Weak Law Of Large Numbers:** Suppose that $\{X_n\}$ is a sequence of i.i.d. random variables with expectation μ. Let $\overline{X}_n$ be the sample mean. Then for all $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} P\left[\left|\overline{X}_n - \mu\right| < \epsilon\right] = 1,$$

  that is, $\overline{X}_n \stackrel{p}{\longrightarrow} \mu$, and thus the mean of $X_1, \ldots, X_n$ converges in probability to μ. The Weak Law can be proved in a straightforward fashion using Chebychev's Inequality if the variables have finite variance σ^2; this inequality states that for any random variable X, and $\epsilon > 0$,

$$P_X\left[|X - \mu| < \epsilon\right] \geq 1 - \sigma^2/\epsilon^2.$$

  Applying this to $\overline{X}_n$ yields the result, as the variance converges to zero. However the result can be proved even without the finite variance assumption using characteristic functions.

3. **Convergence in Distribution:** Suppose $\{X_n\}$ have corresponding sequence of cdfs, $F_{X_1}, F_{X_2}, \ldots$ so that for $n = 1, 2, ..$ $F_{X_n}(x) = P\left[X_n \leq x\right]$. Suppose that there exists a cdf, F_X, such that **for all** x **at which** F_X **is continuous**,

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

Then $\{X_n\}$ **converges in distribution** to X with cdf F_X, denoted $X_n \stackrel{d}{\longrightarrow} X$, and F_X is the **limiting distribution**.

- Convergence of a sequence of mgfs or cfs also indicates convergence in distribution. For example, if for all t at which $M_X(t)$ is defined, as $n \longrightarrow \infty$, we have

$$M_{X_i}(t) \longrightarrow M_X(t) \qquad \Longleftrightarrow \qquad X_n \stackrel{d}{\longrightarrow} X.$$

- The sequence of random variables $X_1, \ldots, X_n$ converges in distribution to constant c if the limiting distribution of $X_1, \ldots, X_n$ is **degenerate at** c, that is,

$$X_n \stackrel{d}{\longrightarrow} X$$

  and $P\left[X = c\right] = 1$, so that

$$F_X(x) = \begin{cases} 0 & x < c \\ 1 & x \geq c \end{cases}$$

This special case occurs when the limiting distribution is discrete, with the probability mass function only being non-zero at a single value, that is, if the limiting random variable is X, then $P[X = c] = 1$ and zero otherwise. We say that the sequence of random variables $X_1, \ldots, X_n$ **converges in distribution** to c if and only if, for all $\epsilon > 0$,

$$\lim_{n \longrightarrow \infty} P\left[|X_n - c| < \epsilon\right] = 1$$

This definition indicates that convergence in distribution to a constant c occurs if and only if the probability becomes increasingly concentrated around c as $n \longrightarrow \infty$.

To show that we should ignore points of discontinuity of F_X in the definition of convergence in distribution, consider the following example: let

$$F_\epsilon(x) = \begin{cases} 0 & x < \epsilon \\ 1 & x \geq \epsilon \end{cases}$$

be the cdf of a degenerate distribution with probability mass 1 at $x = \epsilon$. Now consider a sequence $\{\epsilon_n\}$ of real values converging to ϵ from **below**. Then, as $\epsilon_n < \epsilon$, we have

$$F_{\epsilon_n}(x) = \begin{cases} 0 & x < \epsilon_n \\ 1 & x \geq \epsilon_n \end{cases}$$

which converges to $F_\epsilon(x)$ at all real values of x. However, if instead $\{\epsilon_n\}$ converges to ϵ from **above**, then $F_{\epsilon_n}(\epsilon) = 0$ for each finite n, as $\epsilon_n > \epsilon$, so $\lim_{n \longrightarrow \infty} F_{\epsilon_n}(\epsilon) = 0$. Hence, as $n \longrightarrow \infty$,

$$F_{\epsilon_n}(\epsilon) \longrightarrow 0 \neq 1 = F_\epsilon(\epsilon).$$

Thus the limiting function in this case is

$$F_\epsilon(x) = \begin{cases} 0 & x \leq \epsilon \\ 1 & x > \epsilon \end{cases}$$

which is not a cdf as it is not right-continuous. However, if $\{X_n\}$ and X are random variables with distributions $\{F_{\epsilon_n}\}$ and F_ϵ, then $P[X_n = \epsilon_n] = 1$ converges to $P[X = \epsilon] = 1$, however we take the limit, so F_ϵ does describe the limiting distribution of the sequence $\{F_{\epsilon_n}\}$. Thus, because of right-continuity, we ignore points of discontinuity in the limiting function.

4. **Convergence In rth Mean** The sequence of random variables $\{X_n\}$ **converges in rth mean** to random variable X, denoted $X_n \overset{r}{\longrightarrow} X$ if

$$\lim_{n \longrightarrow \infty} \mathbb{E}\left[|X_n - X|^r\right] = 0.$$

For example, if

$$\lim_{n \longrightarrow \infty} \mathbb{E}\left[(X_n - X)^2\right] = 0$$

then we write $X_n \overset{r=2}{\longrightarrow} X$. In this case, we say that $\{X_n\}$ converges to X *in mean-square* or *in quadratic mean*. For $r_1 > r_2 \geq 1$,

$$X_n \overset{r=r_1}{\longrightarrow} X \qquad \Longrightarrow \qquad X_n \overset{r=r_2}{\longrightarrow} X$$

as, by Lyapunov's inequality

$$\mathbb{E}[\,|X_n - X|^{r_2}\,]^{1/r_2} \leq \mathbb{E}[\,|X_n - X|^{r_1}\,]^{1/r_1} \qquad \therefore \qquad \mathbb{E}[\,|X_n - X|^{r_2}\,] \leq \mathbb{E}[\,|X_n - X|^{r_1}\,]^{r_2/r_1} \longrightarrow 0$$

as $n \longrightarrow \infty$, as $r_2 < r_1$. Thus

$$\mathbb{E}[\,|X_n - X|^{r_2}\,] \longrightarrow 0$$

and $X_n \overset{r=r_2}{\longrightarrow} X$. The converse does not hold in general.

Notes:

(a) **Relating The Modes Of Convergence** For sequence of random variables $X_1, \ldots, X_n$,

$$\left.\begin{array}{c} X_n \xrightarrow{a.s.} X \\ \text{or} \\ X_n \xrightarrow{r} X \end{array}\right\} \implies X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X$$

so almost sure convergence and convergence in rth mean for some r both imply convergence in probability, which in turn implies convergence in distribution to random variable X.

No other relationships hold in general, although there are some partial converse results.

(b) **Slutsky's Theorem:** Suppose that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ for some constant c. Then

(i) $X_n + Y_n \xrightarrow{d} X + c$

(ii) $X_n Y_n \xrightarrow{d} cX$

(iii) $X_n / Y_n \xrightarrow{d} X/c$ provided $c \neq 0$.

(c) **The Central Limit Theorem:** Suppose $X_1, \ldots, X_n$ are i.i.d. random variables with cf φ_X, with expectation μ and variance σ^2, both finite. Let the random variable Z_n be defined by

$$Z_n = \frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sqrt{n\sigma^2}} = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)$$

and denote by φ_{Z_n} the cf of Z_n. Then, as $n \longrightarrow \infty$,

$$\varphi_{Z_n}(t) \longrightarrow \exp\{-t^2/2\}$$

irrespective of the form of φ_X. Thus, as $n \longrightarrow \infty$, $Z_n \xrightarrow{d} Z \sim Normal(0,1)$.

Proof. First, let $Y_i = (X_i - \mu)/\sigma$ for $i = 1, \ldots, n$. Then $Y_1, \ldots, Y_n$ are i.i.d. with cf φ_Y say, and

$$\mathbb{E}_Y[Y_i] = 0 \qquad \text{Var}_Y[Y] = 1$$

for each i. By a previous result for cfs concerning moments, using a Taylor expansion for t in a neighbourhood of zero, we have

$$\varphi_Y(t) = 1 - \frac{t^2}{2} + \text{o}(t^3)$$

Re-writing Z_n as

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i$$

as $Y_1, \ldots, Y_n$ are independent, we have by a standard cf result that

$$\varphi_{Z_n}(t) = \prod_{i=1}^{n} \left\{ \varphi_Y \left(\frac{t}{\sqrt{n}} \right) \right\} = \left\{ 1 - \frac{t^2}{2n} + \text{o}(n^{-3/2}) \right\}^n = \left\{ 1 - \frac{t^2}{2n} + \text{o}(n^{-1}) \right\}^n.$$

so that, by the definition of the exponential function, as $n \longrightarrow \infty$

$$\varphi_{Z_n}(t) \longrightarrow \exp\{-t^2/2\} \qquad \therefore \qquad Z_n \xrightarrow{d} Z \sim Normal(0,1)$$

where no further assumptions on φ_X are required. ∎

Alternative statement: The theorem can also be stated in terms of

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}} = \sqrt{n}(\overline{X}_n - \mu)$$

so that

$$Z_n \xrightarrow{d} Z \sim Normal(0, \sigma^2).$$

and σ^2 is termed the **asymptotic variance** of Z_n.

Notes :

(i) The theorem holds for the i.i.d. case, but there are similar theorems for **non identically distributed**, and **dependent** random variables.

(ii) The theorem allows the construction of **asymptotic normal approximations**. For example, for **large but finite** n, by using the properties of the Normal distribution,

$$\begin{aligned} \overline{X}_n &\sim \mathcal{AN}(\mu, \sigma^2/n) \\ S_n = \sum_{i=1}^{n} X_i &\sim \mathcal{AN}(n\mu, n\sigma^2). \end{aligned}$$

where $\mathcal{AN}(\mu, \sigma^2)$ denotes an asymptotic normal distribution. The notation

$$\overline{X}_n \stackrel{.}{\sim} Normal(\mu, \sigma^2/n)$$

is sometimes used.

(iv) The **multivariate version** of this theorem can be stated as follows: Suppose $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are i.i.d. d-dimensional random variables with

$$\mathbb{E}_{\mathbf{X}}[\mathbf{X}_i] = \boldsymbol{\mu} \qquad \text{Var}_{\mathbf{X}}[\mathbf{X}_i] = \Sigma$$

where Σ is a positive definite, symmetric $d \times d$ matrix defining the variance-covariance matrix of the $\mathbf{X}_i$. Let the random variable $\mathbf{Z}_n$ be defined by

$$\mathbf{Z}_n = \sqrt{n}(\overline{\mathbf{X}}_n - \boldsymbol{\mu})$$

where

$$\overline{\mathbf{X}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i.$$

Then

$$\mathbf{Z}_n \xrightarrow{d} \mathbf{Z} \sim Normal_d(\mathbf{0}, \Sigma)$$

as $n \longrightarrow \infty$.

Appendix: Technical Details

Alternative characterizations of almost sure convergence:

(i) Let $\epsilon > 0$, and define the sets $A_n(\epsilon)$ and $B_n(\epsilon)$ be defined for $n \geq 1$ by

$$A_n(\epsilon) \equiv \{\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\} \qquad\qquad B_n(\epsilon) \equiv \bigcup_{m=n}^{\infty} A_m(\epsilon).$$

- $A_n(\epsilon)$ is the set of ω for which $X_n(\omega)$ is at least ϵ away from X.
- $B_n(\epsilon)$ is the set of ω for which $X_m(\omega)$ at least ϵ away from X, for **at least one** $m \geq n$.
- The event $B_n(\epsilon)$ occurs **if there exists** an $m \geq n$ such that $|X_m - X| \geq \epsilon$.
- $X_n \xrightarrow{a.s.} X$ if and only if $P(B_n(\epsilon)) \longrightarrow 0$.

(ii) $X_n \xrightarrow{a.s.} X$ if and only if

$$P[\,|X_n - X| \geq \epsilon \text{ infinitely often}\,] = 0$$

that is, $X_n \xrightarrow{a.s.} X$ if and only if there are **only finitely many** X_n for which $|X_n(\omega) - X(\omega)| \geq \epsilon$ if ω lies in a set of probability greater than zero.

Note that $X_n \xrightarrow{a.s.} X$ if and only if

$$\lim_{n \longrightarrow \infty} P(B_n(\epsilon)) = \lim_{n \longrightarrow \infty} P\left(\bigcup_{m=n}^{\infty} A_m(\epsilon)\right) = 0$$

in contrast with the definition of convergence in probability, where $X_n \xrightarrow{p} X$ if

$$\lim_{n \longrightarrow \infty} P(A_n(\epsilon)) = 0.$$

Clearly $A_n(\epsilon) \subseteq \bigcup\limits_{m=n}^{\infty} A_m(\epsilon)$ so therefore

$$\lim_{n \longrightarrow \infty} P\left(\bigcup_{m=n}^{\infty} A_m(\epsilon)\right) = 0 \qquad \Longrightarrow \qquad \lim_{n \longrightarrow \infty} P(A_n(\epsilon)) = 0$$

and hence almost sure convergence implies convergence in probability.

Proof. **Relating the modes of convergence.**

(a) $X_n \xrightarrow{a.s.} X \Longrightarrow X_n \xrightarrow{p} X$. Suppose $X_n \xrightarrow{a.s.} X$, and let $\epsilon > 0$. Then

$$P[\,|X_n - X| < \epsilon\,] \geq P[\,|X_m - X| < \epsilon,\ \forall m \geq n\,] \tag{1}$$

as, considering the original sample space,

$$\{\omega : |X_m(\omega) - X(\omega)| < \epsilon,\ \forall m \geq n\} \subseteq \{\omega : |X_n(\omega) - X(\omega)| < \epsilon\}$$

But, as $X_n \xrightarrow{a.s.} X$, $P[\,|X_m - X| < \epsilon,\ \forall m \geq n\,] \longrightarrow 1$, as $n \longrightarrow \infty$. So, after taking limits in equation (1), we have

$$\lim_{n \longrightarrow \infty} P[\,|X_n - X| < \epsilon\,] \geq \lim_{n \longrightarrow \infty} P[\,|X_m - X| < \epsilon,\ \forall m \geq n\,] = 1$$

and so

$$\lim_{n \longrightarrow \infty} P[\,|X_n - X| < \epsilon\,] = 1 \qquad \therefore \qquad X_n \xrightarrow{p} X.$$

(b) $X_n \xrightarrow{r} X \Longrightarrow X_n \xrightarrow{p} X$. Suppose $X_n \xrightarrow{r} X$, and let $\epsilon > 0$. Then, using an argument similar to Chebychev's Lemma,

$$\mathbb{E}[\,|X_n - X|^r\,] \geq \mathbb{E}[\,|X_n - X|^r \mathbb{1}\{|X_n - X| > \epsilon\}\,] \geq \epsilon^r P[|X_n - X| > \epsilon].$$

Taking limits as $n \longrightarrow \infty$, as $X_n \xrightarrow{r} X$, $\mathbb{E}[\,|X_n - X|^r\,] \longrightarrow 0$ as $n \longrightarrow \infty$, so therefore

$$P[|X_n - X| > \epsilon] \longrightarrow 0 \qquad \therefore \qquad X_n \xrightarrow{p} X.$$

(c) $X_n \xrightarrow{p} X \Longrightarrow X_n \xrightarrow{d} X$. Suppose $X_n \xrightarrow{p} X$, and let $\epsilon > 0$. Denote, in the usual way,

$$F_{X_n}(x) = P[X_n \leq x] \qquad \text{and} \qquad F_X(x) = P[X \leq x].$$

Then, by the theorem of total probability, we have two inequalities

$$F_{X_n}(x) = P[X_n \leq x] = P[X_n \leq x, X \leq x+\epsilon] + P[X_n \leq x, X > x+\epsilon] \leq F_X(x+\epsilon) + P[|X_n - X| > \epsilon]$$

$$F_X(x-\epsilon) = P[X \leq x-\epsilon] = P[X \leq x-\epsilon, X_n \leq x] + P[X \leq x-\epsilon, X_n > x] \leq F_{X_n}(x) + P[|X_n - X| > \epsilon].$$

as $A \subseteq B \Longrightarrow P(A) \leq P(B)$ yields

$$P[\,X_n \leq x, X \leq x+\epsilon\,] \leq F_X(x+\epsilon) \qquad \text{and} \qquad P[\,X \leq x-\epsilon, X_n \leq x\,] \leq F_{X_n}(x).$$

Thus

$$F_X(x-\epsilon) - P[\,|X_n - X| > \epsilon] \leq F_{X_n}(x) \leq F_X(x+\epsilon) + P[\,|X_n - X| > \epsilon]$$

and taking limits as $n \longrightarrow \infty$ (with care; we cannot yet write $\lim_{n\longrightarrow\infty} F_{X_n}(x)$ as we do not know that this limit exists) recalling that $X_n \xrightarrow{p} X$,

$$F_X(x-\epsilon) \leq \liminf_{n\longrightarrow\infty} F_{X_n}(x) \leq \limsup_{n\longrightarrow\infty} F_{X_n}(x) \leq F_X(x+\epsilon)$$

Then if F_X is continuous at x, $F_X(x-\epsilon) \longrightarrow F_X(x)$ and $F_X(x+\epsilon) \longrightarrow F_X(x)$ as $\epsilon \longrightarrow 0$, so

$$F_X(x) \leq \liminf_{n\longrightarrow\infty} F_{X_n}(x) \leq \limsup_{n\longrightarrow\infty} F_{X_n}(x) \leq F_X(x)$$

and thus $F_{X_n}(x) \longrightarrow F_X(x)$ as $n \longrightarrow \infty$.

Thus all results follow. ∎

Slutsky's Theorem: Suppose that $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$ for some constant c. Then

(a) $X_n + Y_n \xrightarrow{d} X + c$

(b) $X_n Y_n \xrightarrow{d} cX$

(c) $X_n/Y_n \xrightarrow{d} X/c$ provided $c \neq 0$.

Proof. For (a), let $x - c$ be a continuity point of F_X, some x, and choose $\epsilon > 0$ such that $x - c - \epsilon$ and $x - c + \epsilon$ are also continuity points. Let $Z_n = X_n + Y_n$. Then, as in the previous proof, by the theorem of total probability, we have the inequalities

$$\begin{aligned} F_{Z_n}(x) = P[X_n + Y_n \leq x] &= P[\,X_n + Y_n \leq x, |Y_n - c| < \epsilon\,] + P[\,X_n + Y_n \leq x, |Y_n - c| \geq \epsilon\,] \\ &\leq F_{X_n}(x - c + \epsilon) + P[\,|Y_n - c| \geq \epsilon] \end{aligned}$$

and similarly

$$\begin{aligned} F_{X_n}(x-c-\epsilon) = P[X_n \le x-c-\epsilon] =& P[\, X_n \le x-c-\epsilon, |Y_n - c| < \epsilon\,] \\ &+ P[\, X_n \le x-c-\epsilon, |Y_n - c| \ge \epsilon\,] \\ &\le F_{Z_n}(x) + P[\,|Y_n - c| \ge \epsilon] \end{aligned}$$

Thus

$$\begin{aligned} \limsup_{n\longrightarrow\infty} F_{Z_n}(x) &\le \limsup_{n\longrightarrow\infty} F_{X_n}(x-c+\epsilon) + \limsup_{n\longrightarrow\infty} P[\,|Y_n - c| \ge \epsilon] = F_X(x-c+\epsilon) \\ \liminf_{n\longrightarrow\infty} F_{Z_n}(x) &\ge \liminf_{n\longrightarrow\infty} F_{X_n}(x-c-\epsilon) + \liminf_{n\longrightarrow\infty} P[\,|Y_n - c| \ge \epsilon] = F_X(x-c-\epsilon) \end{aligned}$$

as $x-c-\epsilon$ and $x-c+\epsilon$ are continuity points of F_X. This holds for arbitrary $\epsilon > 0$, and thus

$$\lim_{n\longrightarrow\infty} F_{Z_n}(x) = F_X(x-c) = P[X \le x-c] = P[X + c \le x] = P[Z \le x] = F_Z(x)$$

Thus

$$\lim_{n\longrightarrow\infty} F_{Z_n}(x) = F_Z(x) \qquad \therefore \qquad Z \stackrel{d}{\longrightarrow} X + c$$

Results (b) and (c) follow in a similar fashion. ∎

Partial Converses

(a) If

$$\sum_{n=1}^{\infty} P[\,|X_n - X| > \epsilon\,] < \infty$$

for every $\epsilon > 0$, then $X_n \stackrel{a.s.}{\longrightarrow} X$.

(b) If, for some positive integer r,

$$\sum_{n=1}^{\infty} \mathbb{E}[\,|X_n - X|^r\,] < \infty$$

then $X_n \stackrel{a.s.}{\longrightarrow} X$.

Proof. The results follow from direct probability arguments.

(a) Let $\epsilon > 0$. Then for $n \ge 1$,

$$P[\,|X_n - X| > \epsilon, \text{ for some } m \ge n\,] \equiv P\left[\bigcup_{m=n}^{\infty} \{|X_m - X| > \epsilon\}\right] \le \sum_{m=n}^{\infty} P[\,|X_m - X| > \epsilon\,]$$

as, by elementary probability theory, $P(A \cup B) \le P(A) + P(B)$. But, as it is the tail sum of a convergent series (by assumption), it follows that

$$\lim_{n\longrightarrow\infty} \sum_{m=n}^{\infty} P[\,|X_m - X| > \epsilon\,] = 0.$$

Hence

$$\lim_{n\longrightarrow\infty} P[\,|X_n - X| > \epsilon, \text{ for some } m \ge n\,] = 0$$

and $X_n \stackrel{a.s.}{\longrightarrow} X$.

(b) Identical to part (a), and using part (b) of the previous theorem on relating the modes of convergence that $X_n \stackrel{r}{\longrightarrow} X \Longrightarrow X_n \stackrel{p}{\longrightarrow} X$.

Thus the partial converse results hold. ∎