556: MATHEMATICAL STATISTICS I

CHAPTER 5: STOCHASTIC CONVERGENCE

The following definitions are stated in terms of scalar random variables, but extend naturally to vector random variables defined on the same probability space with measure P. For example, some results are stated in terms of the Euclidean distance in one dimension $|X_n - X| = \sqrt{(X_n - X)^2}$, or for sequences of k-dimensional random variables $\mathbf{X_n} = (X_{n1}, \ldots, X_{nk})^\top$,

$$\|\mathbf{X_n} - \mathbf{X}\| = \left(\sum_{j=1}^{k}(X_{nj} - X_j)^2\right)^{1/2}.$$

5.1 Convergence in Distribution

Consider a sequence of random variables $X_1, X_2, \ldots$ and a corresponding sequence of cdfs, $F_{X_1}, F_{X_2}, \ldots$ so that for $n = 1, 2, ..$ $F_{X_n}(x) = \mathrm{P}[X_n \leq x]$. 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 $X_1, \ldots, X_n$ **converges in distribution** to random variable X with cdf F_X, denoted

$$X_n \xrightarrow{d} X$$

and F_X is the **limiting distribution**. Convergence of a sequence of mgfs or cfs also indicates convergence in distribution, that is, if for all t at which $M_X(t)$ is defined, if as $n \longrightarrow \infty$, we have

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

Definition : DEGENERATE DISTRIBUTIONS
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, $X_n \xrightarrow{d} X$ and $P[X = c] = 1$, so that

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

Interpretation: A special case of convergence in distribution 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 $\epsilon > 0$,

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

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

Note: Points of Discontinuity
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.

5.2 Convergence in Probability

Definition : CONVERGENCE IN PROBABILITY TO A CONSTANT
The sequence of random variables $X_1, \ldots, X_n$ **converges in probability** to constant c, denoted $X_n \xrightarrow{p} c$, if

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

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

Interpretation : Convergence in probability to a constant is precisely equivalent to convergence in distribution to a constant.

THEOREM (WEAK LAW OF LARGE NUMBERS)
Suppose that $X_1, \ldots, X_n$ is a sequence of i.i.d. random variables with expectation μ and finite variance σ^2. Let Y_n be defined by

$$Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

then, for all $\epsilon > 0$,

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

that is, $Y_n \xrightarrow{p} \mu$, and thus the mean of $X_1, \ldots, X_n$ converges in probability to μ.

Proof. Using the properties of expectation, it can be shown that Y_n has expectation μ and variance σ^2/n, and hence by the Chebychev Inequality,

$$P\left[|Y_n - \mu| \geq \epsilon\right] \leq \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0 \qquad \text{as } n \longrightarrow \infty$$

for all $\epsilon > 0$. Hence

$$P\left[|Y_n - \mu| < \epsilon\right] \longrightarrow 1 \qquad \text{as } n \longrightarrow \infty$$

and $Y_n \xrightarrow{p} \mu$. ∎

Definition : CONVERGENCE IN PROBABILITY TO A RANDOM VARIABLE
The sequence of random variables $X_1, \ldots, X_n$ **converges in probability** to random variable X, denoted $X_n \stackrel{p}{\longrightarrow} X$, if, for all $\epsilon > 0$,

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

To understand this definition, let $\epsilon > 0$, and consider

$$A_n(\epsilon) \equiv \{\omega : |X_n(\omega) - X(\omega)| \geq \epsilon\}$$

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 $m \geq n$, $P(A_m(\epsilon)) < \epsilon$.

5.3 Convergence Almost Surely

The sequence of random variables $X_1, \ldots, X_n$ **converges almost surely** to random variable X, denoted $X_n \stackrel{a.s.}{\longrightarrow} X$ 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, $X_n \stackrel{a.s.}{\longrightarrow} X$ if for every $\epsilon > 0$

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

This can also be written

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

for every $\omega \in \Omega$, except possibly those lying in a set of probability zero under P.

Alternative characterization:

- Let $\epsilon > 0$, and the sets $A_n(\epsilon)$ and $B_m(\epsilon)$ be defined for $n, m \geq 0$ by

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

  Then $X_n \stackrel{a.s.}{\longrightarrow} X$ if and only if $P(B_m(\epsilon)) \longrightarrow 0$ as $m \longrightarrow \infty$.
  Interpretation:
  - The event $A_n(\epsilon)$ corresponds to the set of ω for which $X_n(\omega)$ is more than ϵ away from X.
  - The event $B_m(\epsilon)$ corresponds to the set of ω for which $X_n(\omega)$ is more than ϵ away from X, for **at least one** $n \geq m$.
  - The event $B_m(\epsilon)$ occurs **if there exists** an $n \geq m$ such that $|X_n - X| > \epsilon$.
  - $X_n \stackrel{a.s.}{\longrightarrow} X$ if and only if and only if $P(B_m(\epsilon)) \longrightarrow 0$.

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

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

  that is, $X_n \stackrel{a.s.}{\longrightarrow} X$ if and only if there are **only finitely many** X_n for which

  $$|X_n(\omega) - X(\omega)| > \epsilon$$

  if ω lies in a set of probability greater than zero.

- Note that $X_n \stackrel{a.s.}{\longrightarrow} X$ if and only if

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

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

$$\lim_{m \longrightarrow \infty} P(A_m(\epsilon)) = 0.$$

Clearly

$$A_m(\epsilon) \subseteq \bigcup_{n=m}^{\infty} A_n(\epsilon)$$

and hence almost sure convergence is a stronger form.

Alternative terminology:

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

Interpretation: A random variable is a real-valued function from (a sigma-algebra defined on) sample space Ω to $\mathbb{R}$. The sequence of random variables $X_1, \ldots, X_n$ 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.

THEOREM (STRONG LAW OF LARGE NUMBERS)
Suppose that $X_1, \ldots, X_n$ is a sequence of i.i.d. random variables with expectation μ and (finite) variance σ^2. Let Y_n be defined by

$$Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

then, for all $\epsilon > 0$,

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

that is, $Y_n \stackrel{a.s.}{\longrightarrow} \mu$, and thus the mean of $X_1, \ldots, X_n$ converges almost surely to μ.

5.4 Convergence In rth Mean

The sequence of random variables $X_1, \ldots, X_n$ **converges in rth mean** to random variable X, denoted $X_n \stackrel{r}{\longrightarrow} X$ 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

$$X_n \stackrel{r=2}{\longrightarrow} X.$$

In this case, we say that $\{X_n\}$ converges to X *in mean-square* or *in quadratic mean*.

THEOREM
For $r_1 > r_2 \geq 1$,

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

Proof. By Lyapunov's inequality

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

so that

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

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

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

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

THEOREM (RELATING THE MODES OF CONVERGENCE)
For sequence of random variables $X_1, \ldots, X_n$, following relationships hold

$$\left.\begin{array}{c} X_n \stackrel{a.s.}{\longrightarrow} X \\ \text{or} \\ X_n \stackrel{r}{\longrightarrow} X \end{array}\right\} \quad \Longrightarrow \quad X_n \stackrel{p}{\longrightarrow} X \quad \Longrightarrow \quad X_n \stackrel{d}{\longrightarrow} X$$

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.

THEOREM (Partial Converses: NOT EXAMINABLE)

(i) If

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

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

(ii) If, for some positive integer r,

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

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

THEOREM (Slutsky's Theorem)
Suppose that

$$X_n \stackrel{d}{\longrightarrow} X \qquad \text{and} \qquad Y_n \stackrel{p}{\longrightarrow} c$$

Then

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

(ii) $X_n Y_n \stackrel{d}{\longrightarrow} cX$

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

5.5 The Central Limit Theorem

THEOREM (THE LINDEBERG-LÉVY CENTRAL LIMIT THEOREM)
Suppose $X_1, \ldots, X_n$ are i.i.d. random variables with mgf M_X, with expectation μ and variance σ^2, both finite. Let the random variable Z_n be defined by

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

where

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

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

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

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

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 mgf M_Y say, and $\mathbb{E}_{f_Y}[Y_i] = 0$, $\text{Var}_Y[Y_i] = 1$ for each i. Using a Taylor series expansion, we have that for t in a neighbourhood of zero,

$$M_Y(t) = 1 + t\mathbb{E}_Y[Y] + \frac{t^2}{2!}\mathbb{E}_Y[Y^2] + \frac{t^3}{3!}\mathbb{E}_Y[Y^3] + \ldots = 1 + \frac{t^2}{2} + \mathrm{O}(t^3)$$

using the $\mathrm{O}(t^3)$ notation to capture all terms involving t^3 and higher powers. 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 mgf result that

$$M_{Z_n}(t) = \prod_{i=1}^{n}\left\{M_Y\left(\frac{t}{\sqrt{n}}\right)\right\} = \left\{1 + \frac{t^2}{2n} + \mathrm{O}(n^{-3/2})\right\}^n = \left\{1 + \frac{t^2}{2n} + \mathrm{o}(n^{-1})\right\}^n.$$

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

$$M_{Z_n}(t) \longrightarrow \exp\{t^2/2\} \qquad \therefore \qquad Z_n \xrightarrow{d} Z \sim \mathcal{N}(0,1)$$

where no further assumptions on M_X are required. ∎

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

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

so that

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

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

Notes :

(i) The theorem requires the **existence of the mgf** M_X.

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

(iii) 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

$$\overline{X}_n \stackrel{\cdot}{\sim} \mathcal{N}(\mu, \sigma^2/n)$$

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. k-dimensional random variables with mgf $M_{\mathbf{X}}$, with

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

where Σ is a positive definite, symmetric $k \times k$ 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

$$\mathbf{Z}_n \stackrel{d}{\longrightarrow} \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \Sigma)$$

as $n \longrightarrow \infty$.

Appendix (NOT EXAMINABLE)

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,

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

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, also, as $n \longrightarrow \infty$

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

(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 $n \longrightarrow \infty$.

∎

Proof. **(Partial converses)**

(i) Let $\epsilon > 0$. Then for $n \geq 1$,

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

as, by elementary probability theory, $P(A \cup B) \leq P(A) + P(B)$. 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

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

and $X_n \xrightarrow{a.s.} X$.

(ii) Identical to part (i), and using part (b) of the previous theorem that $X_n \xrightarrow{r} X \Longrightarrow X_n \xrightarrow{p} X$.

∎