Some notes on asymptotic theory in probability

Alen Alexanderian*

Abstract

We provide a precise account of some commonly used results from asymptotic theory in probability.

Contents

1 Introduction and basic definitions

This brief note summarizes some important results in asymptotic theory in probability. The main motivation of this theory is to approximate distribution of large sample statistics with a limiting distribution which is often much simpler to work with.

The structure of this note is as follows. First, we recall some basic definitions from probability theory in Section 2. Then, we proceed by discussing the notion of convergence in Probability and $o_p(n)$ and $O_p(n)$ notations in Section 3. Section 4 presents the probabilistic Taylor formula. In Section 5 we discuss convergence in distributions along with some related results, and in Section 6, we follow up by presenting the Central Limit Theorem along with some accompanying results.

2 Basic definitions from probability theory

Throughout this note, we let $(\Omega, \mathcal{F}, P)$, be a probability space, where Ω denotes a sample space, $\mathcal{F}$ is a suitable σ-algebra on Ω, and P is a probability measure. We call a measurable function $X : \Omega \mapsto \mathbf{R}$ a random variable; sometimes, we work with

*The University of Texas at Austin, USA. E-mail: alen@ices.utexas.edu
Last revised: April 16, 2015

random k-vectors (vectors valued random variables), which are measurable functions $\boldsymbol{X} : \Omega \mapsto \mathbf{R}^k$. The expectation (mean) of a random variable X is

$$\mathrm{E}\left[X\right] = \int_\Omega X(\omega)\, dP(\omega).$$

Variance of a random variable measures the deviation of X from its mean,

$$\mathrm{Var}\left[X\right] = \mathrm{E}\left[(X - \mathrm{E}\left[X\right])^2\right].$$

It is trivial to show,

$$\mathrm{Var}\left[X\right] = \mathrm{E}\left[X^2\right] - (\mathrm{E}\left[X\right])^2.$$

Next, we recall the notion of the covariance of two random variables. Let X and Y be two random variables,

$$\mathrm{Cov}\left[X, Y\right] = \mathrm{E}\left[(X - \mathrm{E}\left[X\right])(Y - \mathrm{E}\left[Y\right])\right].$$

Note that $\mathrm{Cov}\left[X, X\right] = \mathrm{Var}\left[X\right]$.

The variance and covariance for random k-vectors are defined by,

$$\mathrm{Var}\left[\boldsymbol{X}\right] = \mathrm{E}\left[(\boldsymbol{X} - \mathrm{E}\left[\boldsymbol{X}\right]) \cdot (\boldsymbol{X} - \mathrm{E}\left[\boldsymbol{X}\right])\right],$$
$$\mathrm{Cov}\left[\boldsymbol{X}, \boldsymbol{Y}\right] = \mathrm{E}\left[(\boldsymbol{X} - \mathrm{E}\left[\boldsymbol{X}\right]) \cdot (\boldsymbol{Y} - \mathrm{E}\left[\boldsymbol{Y}\right])\right].$$

Here $\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{k} x_i y_i$.

Definition 2.1 (Distribution Function)**.** *The distribution function of a random variable X is the function $F_X : \mathbf{R} \mapsto [0, 1]$ defined by*

$$F_X(z) := P(X \leq z)$$

For a random k-vector $\boldsymbol{X}$ we similarly define $F_{\boldsymbol{X}} : \mathbf{R}^k \mapsto [0, 1]$ by

$$F_{\boldsymbol{X}}(\boldsymbol{z}) := P(\boldsymbol{X} \leq \boldsymbol{z})$$

Note that for two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbf{R}^k$, we write $\boldsymbol{x} \leq \boldsymbol{y}$ to mean $x_i \leq y_i$, for $i = 1, \ldots, k$.

Another equivalent definition of expectation is given by

$$\mathrm{E}\left[\boldsymbol{X}\right] = \int_\Omega X(\omega)\, dP(\omega) = \int_{\mathbf{R}^k} \boldsymbol{z}\, dF_{\boldsymbol{X}}(\boldsymbol{z}).$$

We also have, for a measurable function g

$$\mathrm{E}\left[g(\boldsymbol{X})\right] = \int_\Omega g(X(\omega))\, dP(\omega) = \int_{\mathbf{R}^k} g(\boldsymbol{z})\, dF_{\boldsymbol{X}}(\boldsymbol{z}). \tag{2.1}$$

The next definition concerns the very important notion of the characteristic function of a random variable.

Definition 2.2. *Let $\boldsymbol{X}$ be an random k-vector. Then*

$$\phi_{\boldsymbol{X}}(\boldsymbol{\xi}) := \mathrm{E}\left[e^{i\boldsymbol{\xi}\cdot\boldsymbol{X}}\right], \quad \boldsymbol{\xi} \in \mathbf{R}^k$$

is the characteristic function of X.

Using (2.1) we have,

$$\phi_{\boldsymbol{X}}(\boldsymbol{\xi}) = \mathrm{E}\left[e^{i\boldsymbol{\xi}\cdot\boldsymbol{X}}\right] = \int_{\mathbf{R}^k} e^{i\boldsymbol{\xi}\cdot\boldsymbol{z}}\, dF_{\boldsymbol{X}}(\boldsymbol{z}).$$

3 Convergence in probability and $o_p(n)$, $O_p(n)$ notations

Let us begin by defining the notion of convergence in probability to zero [2].

Definition 3.1. *Let $\{X_n\}$ be a sequence of random variables defined on $(\Omega, \mathcal{F}, P)$. We say X_n converges in probability to zero, written $X_n = o_p(1)$ or $X_n \xrightarrow{P} 0$, if for every $\epsilon > 0$,*

$$P(|X_n| > \epsilon) \to 0, \quad \text{as } n \to \infty.$$

A closely related notion is that of a sequence being bounded in probability [2].

Definition 3.2. *Let $\{X_n\}$ be a sequence of random variables defined on $(\Omega, \mathcal{F}, P)$. We say X_n is bounded in probability (or tight), if for every $\epsilon > 0$, there exists $M_\epsilon > 0$ such that,*

$$P(|X_n| > M_\epsilon) < \epsilon, \quad \text{for all } n.$$

We can look at the idea of boundedness in probability in the following alternative way. We say X_n is bounded in probability if for every $\epsilon > 0$, there is a set $F \in \mathcal{F}$ and a number M_ϵ such that for all $\omega \in F$,

$$|X_n(\omega)| \leq M_\epsilon, \quad \text{for all } n,$$

and

$$P(F^c) < \epsilon.$$

To get used to this ideas, let us prove the following basic fact: if $X_n \xrightarrow{P} 0$, then X_n is bounded in probability. That is if $X_n = o_p(1)$, then $X_n = O_p(1)$. This corresponds to the basic idea from real analysis that says a convergent sequence is bounded. The argument is as follows; suppose $X_n \xrightarrow{P} 0$. Let $\epsilon > 0$ be given; we know $P(|X_n| > 1) \to 0$, therefore, there is some n_0 such that

$$P(|X_n| > 1) < \epsilon, \quad \text{for } n \geq n_0.$$

Now, pick M_0 sufficiently large so that $P(|X_i| > M_0) < \epsilon$ for $i = 1, \ldots, n_0 - 1$. Then, we have for $M = \max(1, M_0)$ that

$$P(|X_n| > M) < \epsilon, \quad \text{for all } n,$$

which completes the proof that $X_n = O_p(1)$.

Now, we use the above developments to define convergence in probability and order in probability [2].

Definition 3.3. *Let $\{X_n\}$ be a sequence of random variables on $(\Omega, \mathcal{F}, P)$ and let $\{a_n\}$ be a sequence of strictly positive reals.*

1. *X_n converges is probability to X if and only if $X_n - X = o_p(1)$.*
2. *$X_n = o_p(a_n)$ if and only if $a_n^{-1} X_n = o_p(1)$.*
3. *$X_n = O_p(a_n)$ if and only if $a_n^{-1} X_n = O_p(1)$.*

The following proposition[2] shows the similarities between the notions of o_p and O_p with their counterparts from real analysis (the little-oh and big-oh notations).

Proposition 3.4. *Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables on $(\Omega, \mathcal{F}, P)$, and let $\{a_n\}$ and $\{b_n\}$ be sequences of positive real numbers.*

1. *Suppose $X_n = o_p(a_n)$ and $Y_n = o_p(b_n)$; then,*

   (a) $X_nY_n = o_p(a_nb_n)$.
   (b) $X_n + Y_n = o_p(\max(a_n, b_n))$.
   (c) $|X_n|^r = o_p(a_n^r)$ for $r > 0$.
2. If $X_n = o_p(a_n)$ and $Y_n = O_p(b_n)$, then $X_nY_n = o_p(a_nb_n)$.

We can extend the above developments to random k-vectors. For a random k-vector, $\boldsymbol{X}$, denote its j^{th} component by X^j.

Definition 3.5. *Let $\{\boldsymbol{X}_n\}$ be a sequence of random k-vectors on $(\Omega, \mathcal{F}, P)$, and let $\{a_n\}$ be a sequence of strictly positive reals.*

1. $\boldsymbol{X}_n = o_p(a_n)$ *if and only if* $X_n^j = o_p(a_n)$ *for* $j = 1, \ldots, k$.
2. $\boldsymbol{X}_n = O_p(a_n)$ *if and only if* $X_n^j = O_p(a_n)$ *for* $j = 1, \ldots, k$.
3. $\boldsymbol{X}_n$ *converges in probability to* $\boldsymbol{X}$*, written* $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}$*, if and only if* $\boldsymbol{X}_n - \boldsymbol{X} = o_p(1)$.

Recall that in analysis, convergence in $\mathbf{R}^k$ is characterized by componentwise convergence. Same idea applies when we talk about convergence in probability of random k-vectors.

Let $\|X\|$ denote the Euclidean norm, $\|X\|^2 = \sum_{j=1}^{k} |X^j|^2$. The following result characterizes convergence in probability in terms of the Euclidean distance [2].

Proposition 3.6. *Let $\{\boldsymbol{X}_n\}$ be a sequence of random k-vectors on $(\Omega, \mathcal{F}, P)$ and $\boldsymbol{X}$ a random k-vector on the same probability space. Then, $\boldsymbol{X}_n - \boldsymbol{X} = o_p(1)$ if and only if $\|\boldsymbol{X}_n - \boldsymbol{X}\| = o_p(1)$.*

Proof. Suppose $\boldsymbol{X}_n - \boldsymbol{X} = o_p(1)$. Then, for every $\epsilon > 0$,

$$\lim_{n\to\infty} P(|X_n^j - X^j|^2 > \epsilon/k) = 0, \quad \text{for } j = 1, \ldots, k.$$

Claim: $P(\sum_{j=1}^{k} |X_n^j - X^j|^2 > \epsilon) \leq \sum_{j=1}^{k} P(|X_n^j - X^j|^2 > \epsilon/k)$.

This follows from the fact that $\sum_{j=1}^{k} |X_n^j - X^j|^2 > \epsilon$ implies at least one of the summands is larger than ϵ/k. To make this clearer, let

$$E_n = \{\omega : \sum_{j=1}^{k} |X_n^j(\omega) - X^j(\omega)|^2 > \epsilon\}, \quad \text{and} \quad E_n^j = \{\omega : |X_n^j - X^j|^2 > \epsilon/k\},$$

and note that $E_n \subseteq \cup_{j=1}^{k} E_n^j$. Therefore, $P(E_n) \leq \sum_{j=1}^{k} P(E_n^j)$, which establishes the claim. Then, we have,

$$P(\|\boldsymbol{X}_n - \boldsymbol{X}\|^2 > \epsilon) \leq \sum_{j=1}^{k} P(|X_n^j - X^j|^2 > \epsilon/k) \to 0, \quad \text{as } n \to \infty.$$

Therefore, $\|\boldsymbol{X}_n - \boldsymbol{X}\|^2 = o_p(1)$, and by proposition 3.4, $\|\boldsymbol{X}_n - \boldsymbol{X}\| = o_p(1)$. Conversely, if $\|\boldsymbol{X}_n - \boldsymbol{X}\| = o_p(1)$, we use $|X_n^j - X^j| \leq \|\boldsymbol{X}_n - \boldsymbol{X}\|$ to get

$$P(|X_n^j - X^j| > \epsilon) \leq P(\|\boldsymbol{X}_n - \boldsymbol{X}\| > \epsilon) \to 0 \quad \text{as } n \to \infty.$$ □

Proposition 3.7. *Let $\{\boldsymbol{X}_n\}$ and $\{\boldsymbol{Y}_n\}$ be sequences of random k-vectors on $(\Omega, \mathcal{F}, P)$. If $\boldsymbol{X}_n - \boldsymbol{Y}_n \xrightarrow{P} 0$ and $\boldsymbol{Y}_n \xrightarrow{P} \boldsymbol{Y}$, where Y is a random k-vector on $(\Omega, \mathcal{F}, P)$ then, $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{Y}$ also.*

Proof. Note that by Proposition 3.6, $\|\boldsymbol{X}_n - \boldsymbol{Y}_n\| = o_p(1)$ and $\|\boldsymbol{Y}_n - \boldsymbol{Y}\| = o_p(1)$. Then, we have

$$\|\boldsymbol{X}_n - \boldsymbol{Y}\| \leq \|\boldsymbol{X}_n - \boldsymbol{Y}_n\| + \|\boldsymbol{Y}_n - \boldsymbol{Y}\| = o_p(1),$$

where the last equality follows from Proposition 3.4. Therefore, again applying Proposition 3.6, we have $\boldsymbol{X}_n - \boldsymbol{Y} = o_p(1)$. □

We also have the following continuity result [2].

Proposition 3.8. *Let $\{X_n\}$ be a sequence of random variables on $(\Omega, \mathcal{F}, P)$, such that $X_n \xrightarrow{P} X$. If $g : \mathbf{R} \mapsto \mathbf{R}$ is a continuous mapping, then $g(X_n) \xrightarrow{P} g(X)$.*

Proof. Let $\epsilon > 0$ be given. For any $K \in \mathbf{R}_+$,

$$\begin{aligned} P(|g(X_n) - g(X)| > \epsilon) \leq P(|g(X_n) - g(X)| > \epsilon, |X| \leq K, |X_n| \leq K) \\ + P(\{|X| > K\} \cup \{|X_n| > K\}). \end{aligned}$$

Now we note that g is uniformly continuous on $\{x : |x| \leq K\}$, and thus there exists $\delta = \delta(\epsilon)$ such that $|g(X_n) - g(X)| \leq \epsilon$ for $|X_n - X| < \delta$ (where we consider the set where $|X| \leq K$ and $|X_n| \leq K$). Therefore,

$$\{|g(X_n) - g(X)| > \epsilon, |X| \leq K, |X_n| \leq K\} \subseteq \{|X_n - X| > \delta\}.$$

Therefore, we have

$$\begin{aligned} P(|g(X_n) - g(X)| > \epsilon) &\leq P(|X_n - X| > \delta) + P(|X| > K) + P(|X_n| > K) \\ &\leq P(|X_n - X| > \delta) + P(|X| > K) + P(|X| > \frac{K}{2}) + P(|X_n - X| > \frac{K}{2}) \end{aligned}$$

Now, let $\gamma > 0$ be arbitrary; we can choose K sufficiently large so that the second and third terms are both less than $\gamma/2$. Thus, we have for sufficiently large K),

$$P(|g(X_n) - g(X)| > \epsilon) \leq P(|X_n - X| > \delta) + P(|X_n - X| > \frac{K}{2}) + \gamma.$$

Taking the limit of both sides of above equation and using the fact that $|X_n - X| \xrightarrow{P} 0$ gives,

$$\lim_{n \to \infty} P(|g(X_n) - g(X)| > \epsilon) \leq \gamma,$$

and since $\gamma > 0$ was arbitrary the proof is complete. □

Suppose $X_n = a + o_p(1)$, then using the above result we have for any continuous function g,

$$g(X_n) = g(a) + o_p(1).$$

Now, what if $g(x)$ is also differentiable at $x = a$? In the next section, we discuss a probabilistic analogue of Taylor's Theorem.

Remark 3.9. *The following technical note is relevant in what follows. Let Let $\{X_n\}$ be a sequence of of random variables such that $X_n = O_p(r_n)$, where $0 < r_n \to 0$. Then, $X_n = o_p(1)$; i.e., $X_n \xrightarrow{P} 0$. To show this, we let $\delta > 0$ be fixed but arbitrary, and let $\epsilon > 0$ be given. By $X_n = O_p(r_n)$, there exists $M_\epsilon \in (0, \infty)$ such that $P(r_n^{-1}|X_n| > M_\epsilon) < \epsilon$, for all n. Now, take $n_0(\epsilon)$, sufficiently large, such that for $n \geq n_0(\epsilon)$, we have $r_n M_\epsilon < \delta$. Then, we have for $n \geq n(\epsilon)$,*

$$P(|X_n| > \delta) \leq P(|X_n| > M_\epsilon r_n) < \epsilon,$$

and hence the claim.

4 Probabilistic Taylor expansion

Proposition 4.1 (Proposition 6.1.5 in [2]). *Let $\{X_n\}$ be a sequence of of random variables such that $X_n = a + O_p(r_n)$ where $a \in \mathbf{R}$ and $0 < r_n \to 0$ as $n \to \infty$. If g is a function with s continuous derivatives at a then,*

$$g(X_n) = \sum_{j=0}^{s} \frac{g^{(j)}(a)}{j!}(X_n - a)^j + o_p(r_n^s).$$

Proof. Define the function h as below,

$$h(x) = \begin{cases} \frac{g(x) - \sum_{j=0}^{s} \frac{g^{(j)}(a)}{j!}(x-a)^j}{\frac{(x-a)^s}{s!}}, & x \neq a \\ 0, & x = a. \end{cases}$$

The function h defined above is continuous at $x = a$ (by repeated application of L'Hôpital's rule we can check $\lim_{x \to a} h(a) = 0$). Therefore, $h(X_n) = h(a) + o_p(1)$, and since $h(a) = 0$, we have $h(X_n) = o_p(1)$. But then, using Proposition 3.4(2), we have

$$(X_n - a)^s h(X_n) = o_p(r_n^s),$$

that is

$$g(X_n) - \sum_{j=0}^{s} \frac{g^{(j)}(a)}{j!}(X_n - a)^j = o_p(r_n^s),$$

which completes the proof. □

We can also state an analogue of the above result for random k-vectors, the proof of which is similar [2].

Proposition 4.2. *Let $\{\boldsymbol{X}_n\}$ be a sequence of of random k-vectors such that $\boldsymbol{X}_n = \boldsymbol{a} + O_p(r_n)$ where $a \in \mathbf{R}^k$ and $0 < r_n \to 0$ as $n \to \infty$. If $g : \mathbf{R}^k \mapsto \mathbf{R}$ is a function such that the partial derivatives $\frac{\partial g}{\partial x_i}$ are continuous in a neighborhood of a then,*

$$g(\boldsymbol{X}_n) = g(\boldsymbol{a}) + \nabla g(\boldsymbol{a}) \cdot (\boldsymbol{X}_n - \boldsymbol{a}) + o_p(r_n).$$

5 Convergence in distribution

When we talk about convergence in probability, we always talk about a sequence of random variables, $X_1, X_2, \ldots$, defined on the same probability space. In this section, we discuss convergence in distribution, which makes sense even if X_i are not defined on the same probability space.

Definition 5.1. *Let $\{\boldsymbol{X}_n\}$ be a sequence of random k-vectors, with distribution functions $\{F_{\boldsymbol{X}_n}(\cdot)\}$. We say $\boldsymbol{X}_n$ converges in distribution to $\boldsymbol{X}$ if,*

$$\lim_{n\to\infty} F_{\boldsymbol{X}_n}(\boldsymbol{z}) = F_{\boldsymbol{X}}(\boldsymbol{z})$$

for all $\boldsymbol{z} \in C$ where C is the set of continuity points of $F_{\boldsymbol{X}_n}$. We use the notation, $\boldsymbol{X}_n \overset{D}{\to} \boldsymbol{X}$ or $F_{\boldsymbol{X}_n} \overset{D}{\to} F_{\boldsymbol{X}}$ to denote convergence in distribution.

Speaking in practical terms, $\boldsymbol{X}_n \overset{D}{\to} \boldsymbol{X}$ tells us that the distribution of $\boldsymbol{X}_n$ is well approximated by the distribution of $\boldsymbol{X}$ for large n. The following Theorem [2] is an important result linking convergence in distribution to convergence of characteristic functions.

Theorem 5.2. *Let $\{F_n\}$, $n = 0, 1, 2, \ldots$, be distribution functions on $\mathbf{R}^k$ corresponding to characteristic functions, $\phi_n(\boldsymbol{\xi}) = \int_{\mathbf{R}^k} e^{i\boldsymbol{\xi}\cdot\boldsymbol{x}}\, dF_n(\boldsymbol{x})$, $n = 0, 1, 2, \ldots$, then the following are equivalent.*

1. $F_n \xrightarrow{D} F_0$,
2. $\int_{\mathbf{R}^k} g(x)\, dF_n(\boldsymbol{x}) \to \int_{\mathbf{R}^k} g(\boldsymbol{x})\, dF_0(\boldsymbol{x})$, *for every bounded continuous function g on $\mathbf{R}^k$.*
3. $\lim_{n\to\infty} \phi_n(\boldsymbol{\xi}) = \phi_0(\boldsymbol{\xi})$ *for every $\boldsymbol{\xi} \in \mathbf{R}^k$.*

Let us get some insight into the notion of convergence in distribution by proving the following simple result, which says that a sequence that converges in distribution is bounded in probability.

Lemma 5.3. *Suppose $X_n \xrightarrow{D} X$. Then, $X_n = O_p(1)$.*

Proof. Let $\epsilon > 0$ be given. Choose M_0 sufficiently large so that $P(|X| > M_0) < \epsilon$. Now, we know by convergence in distribution that,

$$P(|X_n| > M_0) \to P(|X| > M_0).$$

So there exists some n_0 such that for $n \geq n_0$,

$$P(|X_n| > M_0) < \epsilon$$

Next, choose M_1 sufficiently large so that $P(|X_i| > M_1) < \epsilon$, for $i = 1, \ldots, n_0 - 1$. Then, for $M = \max(M_0, M_1)$, we have

$$P(|X_n| > M) < \epsilon, \quad \text{for all } n. \qquad \square$$

Proposition 5.4. *Let $\{X_n\}$ be a sequence of random variables. Suppose $X_n \xrightarrow{P} X$. Then,*

1. $\mathrm{E}\left[|e^{itX_n} - e^{itX}|\right] \to 0$ *as $n \to \infty$, for all $t \in \mathbf{R}$, and*
2. $X_n \xrightarrow{D} X$.

Proof. Let $t \in \mathbf{R}$ and $\epsilon > 0$ be fixed but arbitrary. Pick $\delta = \delta(\epsilon) > 0$ such that,

$$|x - y| < \delta \Longrightarrow |e^{itx} - e^{ity}| = |1 - e^{it(y-x)}| < \epsilon.$$

Then, we have

$$\begin{aligned}
\mathrm{E}\left[|e^{itX_n} - e^{itX}|\right] &= \mathrm{E}\left[|1 - e^{it(X-X_n)}|\right] \\
&= \mathrm{E}\left[|1 - e^{it(X-X_n)}|\chi_{\{|X_n-X|<\delta\}}\right] + \mathrm{E}\left[|1 - e^{it(X-X_n)}|\chi_{\{|X_n-X|\geq\delta\}}\right] \\
&\leq \epsilon + \mathrm{E}\left[|1 - e^{it(X-X_n)}|\chi_{\{|X_n-X|\geq\delta\}}\right] \\
&\leq \epsilon + 2P(|X_n - X| \geq \delta).
\end{aligned}$$

Therefore, taking limit as $n \to \infty$ gives,

$$\lim_{n\to\infty} \mathrm{E}\left[|e^{itX_n} - e^{itX}|\right] \leq \epsilon,$$

and since $\epsilon > 0$ was arbitrary, the proof of the first statement of the Proposition is complete.

To prove convergence in distribution, we show the convergence of characteristic functions:

$$\begin{aligned}|\phi_{X_n}(t) - \phi_X(t)| &= |\mathrm{E}\left[e^{itX_n}\right] - \mathrm{E}\left[e^{itX}\right]| \\ &= |\mathrm{E}\left[e^{itX_n} - e^{itX}\right]| \\ &\leq \mathrm{E}\left[|e^{itX_n} - e^{itX}|\right] \to 0, \quad \text{as } n \to \infty,\end{aligned}$$

where the last conclusion follows from the first statement of the Proposition; this completes the proof. □

Remark 5.5. *By Proposition 5.4 we have that convergence in probability implies convergence in distribution. While the converse is not true in general, we mention the following important special case. Let $\{X_n\}$ be a sequence of random variables such that $X_n \overset{D}{\to} c$ where c is a constant. Then, we have that $X_n \overset{P}{\to} c$ also. To see this, let $\delta > 0$ be fixed but arbitrary, and note that*

$$P(|X_n - c| > \delta) \leq P(X_n - c > \delta) + P(X_n - c \leq -\delta) = 1 - F_{X_n}(c + \delta) + F_{X_n}(c - \delta) \to 0,$$

as $n \to \infty$.

Proposition 5.6. *Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random variables such that $X_n - Y_n = o_p(1)$, and $X_n \overset{D}{\to} X$; then, $Y_n \overset{D}{\to} X$ also.*

Proof. Again, we will show the convergence of the characteristic functions. First, we claim,

$$|\phi_{Y_n}(t) - \phi_{X_n}(t)| \to 0, \quad \text{as } n \to \infty, \quad \forall t \in \mathbf{R}.$$

This follows from,

$$\begin{aligned}|\phi_{Y_n}(t) - \phi_{X_n}(t)| &= |\mathrm{E}\left[e^{itY_n}\right] - \mathrm{E}\left[e^{itX_n}\right]| \\ &= |\mathrm{E}\left[e^{itY_n} - e^{itX_n}\right]| \\ &\leq \mathrm{E}\left[|e^{itY_n} - e^{itX_n}|\right] \\ &= \mathrm{E}\left[|1 - e^{it(X_n - Y_n)}|\right] \to 0, \quad \text{as } n \to \infty,\end{aligned}$$

where the last conclusion follows from Proposition 5.4. The rest of the proof is simple; for any $t \in \mathbf{R}$,

$$|\phi_{Y_n}(t) - \phi_Y(t)| \leq |\phi_{Y_n}(t) - \phi_{X_n}(t)| + |\phi_{X_n}(t) - \phi_X(t)| \to 0, \quad \text{as } n \to \infty.$$ □

The following continuity result is very useful.

Proposition 5.7. *Let $\{X_n\}$ be a sequence of random variables such that $X_n \overset{D}{\to} X$. If $h : \mathbf{R} \mapsto \mathbf{R}$ is a continuous function, then $h(X_n) \overset{D}{\to} h(X)$.*

Proof. We will show $\phi_{h(X_n)}(t) \to \phi_{h(X)}(t)$ for all t. Note that

$$\phi_{X_n}(t) = \mathrm{E}\left[e^{ith(X_n)}\right] = \int_{\mathbf{R}} e^{ith(x)}\, dF_{X_n}.$$

Now, for every $t \in \mathbf{R}$, the function $g(x) = e^{ith(x)}$ is a bounded and continuous function of x. Therefore, by Theorem 5.2(2), we have

$$\int_{\mathbf{R}} e^{ith(x)}\, dF_{X_n} \to \int_{\mathbf{R}} e^{ith(x)}\, dF_X,$$

but the right hand side of the above equation is none but, $\phi_{h(X)}(t)$. Hence, we have shown that $\phi_{X_n}(t) \to \phi_X(t)$ for every t and consequently, by Theorem 5.2, $h(X_n) \overset{D}{\to} h(X)$. □

6 Central Limit Theorem and related results

Definition 6.1. *A sequence $\{X_n\}$ of random variables is said to be asymptotically normal with* mean μ_n *and* standard deviation σ_n *if $\sigma_n > 0$ for n sufficiently large and*

$$\frac{X_n - \mu_n}{\sigma_n} \xrightarrow{D} Z, \quad Z \sim N(0,1).$$

In this case we use the notation,

$$X_n \text{ is } AN(\mu_n, \sigma_n^2).$$

Remark 6.2. *[2] If X_n is $AN(\mu_n, \sigma_n^2)$ it is not necessarily the case that $\mu(X_n) = \mu_n$ and* $\mathrm{Var}[X_n] = \sigma_n^2$.

The following Central Limit Theorem, which is stated without proof is well known (I will at some point put a carefully written proof here). Elementary proofs of this Theorem can be found in [3] or [4], for more rigorous, measure theoretic, proofs see [1, 2, 5].

Theorem 6.3 (Central Limit Theorem)**.** *Let $\{X_n\}$ be a sequence of independent and identically distributed random variables with mean μ and variance σ^2 (notation: $\{X_n\} \sim IID(\mu, \sigma^2)$). Define $\bar{X}_n = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$. Then,*

$$\bar{X}_n \text{ is } AN(\mu, \frac{\sigma^2}{n})$$

The following result [2], known as the delta-method in statistics literature, is useful in applications.

Proposition 6.4. *If X_n is $AN(\mu, \sigma_n^2)$, where $\sigma_n \to 0$ as $n \to \infty$, and if g is a function continuously differentiable at μ, and $g'(\mu) \neq 0$, then*

$$g(X_n) \text{ is } AN(g(\mu), [g'(\mu)]^2 \, \sigma_n^2).$$

Proof. Let $Z_n = \frac{X_n - \mu}{\sigma_n}$, we know that $Z_n \xrightarrow{D} Z \sim N(0,1)$. Thus, by Lemma 5.3, $Z_n = O_p(1)$. Therefore, $X_n = \mu + O_p(\sigma_n)$. Then, by Taylor formula given in Proposition 4.1, we have $g(X_n) = g(\mu) + g'(\mu)(X_n - \mu) + o_p(\sigma_n)$. Therefore,

$$\frac{g(X_n) - g(\mu)}{\sigma_n} = g'(\mu)\left(\frac{X_n - \mu}{\sigma_n}\right) + o_p(1).$$

we rewrite the above equation as

$$\frac{g(X_n) - g(\mu)}{g'(\mu)\sigma_n} = \left(\frac{X_n - \mu}{\sigma_n}\right) + o_p(1),$$

and recall that $\dfrac{X_n - \mu}{\sigma_n} \xrightarrow{D} Z \sim N(0,1)$; therefore, by Proposition 5.6, we have

$$\frac{g(X_n) - g(\mu)}{g'(\mu)\sigma_n} \xrightarrow{D} Z \sim N(0,1),$$

That is,

$$g(X_n) \text{ is } AN\left(g(\mu), [g'(\mu)]^2 \, \sigma^2\right).$$

□

References

[1] L. Breiman, *Probability*, Siam, 1992.

[2] P. Brockwell and R. Davis, *Time Series: Theory and Methods*, Springer, 1991.

[3] P. G. Hoel, S. C. Port, and C. J. Stone, *Introduction to Probability Theory*, Houghton Mifflin Company, 1971.

[4] S. Ross, *A First Course in Probability*, Macmillan Publishing Company, 2nd ed., 1984.

[5] D. Williams, *Probability with Martingales*, Cambridge University Press, 1991.