Convergence in Distribution and Central Limit Theorems

7.1 Convergence in Distribution

Definition 7.1. *Let (X_n) be a sequence of random variables. Let $F_n(x), F(x)$ be the CDF of X_n and X, respectively. Then X_n converges to X in distribution, if $F_n(x)$ converges to $F(x)$ at all continuity points x of F, i.e.,*

$$X_n \to X \text{ in distribution} \quad \Leftrightarrow \quad F_n(x) \to F(x)\ \forall x \text{ s.t. } F \text{ is continuous at } x,$$

Convergence in distribution is also denoted by $X_n \overset{(\mathrm{d})}{\to} X$.

Lemma 7.2. *Almost sure convergence implies convergence in probability and convergence in probability implies convergence in distribution.*

$$a.s \implies \text{in } \mathbb{P} \implies \text{in distribution.}$$

Proof. (a.s. $\implies$ in $\mathbb{P}$) Proved in the homework.
(in $\mathbb{P} \implies$ in dist.) Suppose $X_n \xrightarrow{\mathbb{P}} X$. Let $F_n(X)$ denote the CDF of X_n and $F(x)$ denote the CDF of X. Then,

$$\begin{aligned} |F_n(x) - F(x)| = |\mathbb{P}(X_n \leqslant x) - \mathbb{P}(X \leqslant x)| &\leqslant \mathbb{E}\,|\mathbb{1}_{X_n \leqslant x} - \mathbb{1}_{X \leqslant x}| \\ &\leqslant \mathbb{E}(\mathbb{1}_{|X_n - X| > \varepsilon} + \mathbb{1}_{x-\varepsilon < X \leqslant x+\varepsilon}) \\ &= \mathbb{P}(|X_n - X| > \varepsilon) + \mathbb{P}(x - \varepsilon < X \leqslant x + \varepsilon). \end{aligned}$$

As $n \to \infty$,

$$\limsup_{n\to\infty} |F_n(x) - F(x)| \leqslant 0 + \mathbb{P}(x - \varepsilon < X \leqslant x + \varepsilon) = F(x+\varepsilon) - F(x-\varepsilon)$$

where convergence in $\mathbb{P}$ is used. If x is continuous point of $F(x)$, the RHS is zero as $\varepsilon \to 0$. ■

Example 7.3. *Consider $X_n = \delta_{1/n}, n \geqslant 1$ with $\mathbb{P}(X_n = 1/n) = 1$ and $F_n(x) = \mathbb{1}_{\{x \geqslant 1/n\}}$. Then $X_n \xrightarrow{a.s.} X = \delta_0, \mathbb{P}(X = 0) = 1, F(x) = \mathbb{1}_{\{x\geqslant 0\}}$. Indeed, F_n converges to F at all continuity points $x \neq 0$ of F, i.e., $F_n(x) \to F(x),\ \forall\ x \neq 0$. But at $x = 0$, $F_n(0) = 0\ \forall\ n,$ but $F(0) = 1$. In general, it is not necessarily true that F_n wouldn't converge to F at discontinuity points. For example, let $X_n = \delta_{-1/n}$. Then $F_n(x) \to F(x)$ for all x, even $x = 0$.*

Theorem 7.4. *Let (X_n) be a sequence of r.v.s. Then $X_n \xrightarrow{\mathbb{P}} X$ iff,*

$$\forall \text{ sequence } (n_k)_{k\geqslant 1},\ \exists \text{ sub-sequence } (n_{k_i})_{i \geqslant 1} \text{ s.t. } X_{n_{k_i}} \xrightarrow{a.s.} X \text{ as } i \to \infty.$$

Proof. ($\implies$) WLOG assume $X = 0$. By assumption, we have $\lim_{n\to\infty} \mathbb{P}(|X_n| > \varepsilon) = 0$ for all $\varepsilon > 0$. Then for any given sub-sequence (X_{n_k}), choose n_{k_i} s.t,

$$\mathbb{P}(|X_{n_{k_i}}| > i^{-2}) < i^{-2} \text{ for all } i \geqslant 1.$$

Then $\sum_{i=1}^{\infty} \mathbb{P}(|X_{n_{k_i}}| > i^{-2}) < \infty$ and by first Borel-Cantelli Lemma,

$$\mathbb{P}(|X_{n_{k_i}}| > i^{-2} \text{ i.o.}) = 0,$$

which is equivalent to, $X_{n_{k_i}} \xrightarrow{a.s.} 0$ as $i \to \infty$.

($\Leftarrow$) Proof by contradiction. Suppose $X_n \not\to 0$ in Probability. Thus there exists $\varepsilon > 0$, such that $\mathbb{P}(|X_n| > \varepsilon) \not\to 0$. So there exist $\delta > 0$ and a sub-sequence (X_{n_k}) s.t,

$$\mathbb{P}(|X_{n_k}| > \varepsilon) > \delta, \quad \forall k \tag{7.1}$$

However by assumption, there exist a further sub-sequence $(X_{n_{k_i}})$ s.t $X_{n_{k_i}} \xrightarrow{a.s.} 0 \implies X_{n_{k_i}} \xrightarrow{\mathbb{P}} 0$ which contradicts (**??**). ■

Theorem 7.5. *Let (X_n) be a sequence of r.v.s. Then $X_n \overset{(d)}{\to} X$ iff, there exists a sample space $(\Omega, \mathcal{F}, \mathbb{P})$ and r.v.s (Y_n) such that,*

$$X_n \overset{d}{=} Y_n, \quad X \overset{d}{=} Y, \quad Y_n \xrightarrow{a.s.} Y$$

Proof. $\Leftarrow$ is easy. For $\Rightarrow$ consider the probability space $((0,1), \mathcal{B}, \mathbb{P} = \lambda)$ where λ is the Lebesgue measure and the rv, $U(\omega) = \omega,\ \omega \in \Omega$. Let $F_n(x)$ be CDF of X_n and $F(x)$ be CDF of X. Then,

$$F_n^{-1}(U) \xrightarrow{a.s.} F^{-1}(U),$$

since by assumption $F_n(x)$ converges to $F(x)$ for all continuity points of F. Define,

$$Y_n := F_n^{-1}(U), \quad Y := F^{-1}(U)$$

Then $Y_n \xrightarrow{a.s.} Y$. ■

Exercise 7.6. *Show that if $F_n(x)$ converges to $F(x)$ for all continuity points of F, then, $F_n^{-1}(y) \to F^{-1}(y)$ for all $y \in (0,1)$ but a countable number of points.*

Corollary 7.7. *Let $(X_n)_{n\geqslant 1}$ be a sequence of r.v.s, and f be a bounded continuous function. If $X_n \overset{(d)}{\to} X$ then,*

$$\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X)). \tag{7.2}$$

Proof. Use Theorem **??** to construct $Y_n \xrightarrow{a.s.} Y$. Then since f is continuous, $f(Y_n) \xrightarrow{a.s.} f(Y)$. Then since f is bounded, use BCT to show,

$$\mathbb{E}(f(Y_n)) \to \mathbb{E}(f(Y))$$

which is equivalent to (**??**). ■

Definition 7.8. *Let $\mathbb{P}, \mathbb{P}_n, n \geqslant 1$ be a sequence of probability measures. Then $\mathbb{P}_n$ converges weakly to $\mathbb{P}$, i.e., $\mathbb{P}_n \xrightarrow{w} \mathbb{P}$, if*

$$\int f \mathrm{d}\,\mathbb{P}_n \to \int f \mathrm{d}\,\mathbb{P} \text{ for all bounded and continuous function } f.$$

Theorem 7.9. *The following are equivalent,*

(i) $X_n \xrightarrow{d} X$.

(ii) $X_n \xrightarrow{w} X$

(iii) $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$, *for all* $f \in C_c^\infty(\mathbb{R})$.

(iv) $\mathbb{P}(X_n \in C) \to \mathbb{P}(X \in C)$, *for all closed subsets* $C \in \mathbb{R}$ *such that* $\mathbb{P}(X \in \partial C) = 0$.

(v) $\mathbb{P}(X_n \in O) \to \mathbb{P}(X \in O)$, *for all open subsets* $O \in \mathbb{R}$ *such that* $\mathbb{P}(X \in \partial O) = 0$

Proof. ((iii) $\to$ (i)) Let $F(x) = \mathbb{P}(X \leqslant x)$ be CDF of X. Let x be a fixed continuity point of F. Fix $\varepsilon > 0$ small. We can approximate $f_0(y) := \mathbb{1}_{y \leqslant x}$ by a C^∞ function,

$$f_\varepsilon(y) = \begin{cases} 1 & y \leqslant x - \varepsilon \\ g((x-y)/\varepsilon) & y \in (x-\varepsilon, x) \\ 0 & y \geqslant x \end{cases}$$

where

$$g(y) = \frac{\int_0^y e^{-1/t(1-t)} dt}{\int_0^1 e^{-1/t(1-t)} dt} \in [0,1], \quad y \in [0,1].$$

Then, $0 \leqslant f_0(y) - f_\varepsilon(y) \leqslant \mathbb{1}_{y \in [x-\varepsilon, x]}$ and

$$\begin{aligned} F_n(x) - F(x) = \mathbb{E}\left(f_0(X_n) - f_0(X)\right) &\geqslant \mathbb{E}\left(f_\varepsilon(X_n) - f_0(X)\right) \\ &\geqslant \mathbb{E}\left(f_\varepsilon(X_n) - f_\varepsilon(X) + f_\varepsilon(X) - f_0(X)\right) \\ &\geqslant \mathbb{E}\left(f_\varepsilon(X_n) - f_\varepsilon(X)\right) - \mathbb{P}(X \in [x-\varepsilon, x]). \end{aligned}$$

Note that, even though $f_\varepsilon \in C^\infty(\mathbb{R})$, it is not C_c^∞. We choose $K > 0$ large and define

$$\hat{f}_{\varepsilon,K}(y) = \begin{cases} 1 & x - K < y \leqslant x - \varepsilon \\ g((x-y)/\varepsilon) & y \in (x-\varepsilon, x) \\ g(y - x + K + 1) & y \in (x-K-1, x-K) \\ 0 & \text{o.w.} \end{cases}$$

Then $0 \leqslant f_\varepsilon(y) - \hat{f}_{\varepsilon,K}(y) \leqslant \mathbb{1}_{y \leqslant x-K}$ and $\hat{f}_\varepsilon \in C_c^\infty$. Thus

$$\begin{aligned} F_n(x) - F(x) &\geqslant \mathbb{E}\left(f_\varepsilon(X_n) - f_\varepsilon(X)\right) - \mathbb{P}(X \in [x-\varepsilon, x]) \\ &\geqslant \mathbb{E}(\hat{f}_{\varepsilon,K}(X_n) - \hat{f}_{\varepsilon,K}(X)) - \mathbb{P}(X \in [x-\varepsilon, x]) - \mathbb{P}(X \leqslant x-K). \end{aligned}$$

Taking the limit $n \to \infty$,

$$\liminf_{n\to\infty}(F_n(x) - F(x)) \geqslant -\mathbb{P}(X \in [x-\varepsilon, x]) - \mathbb{P}(X \leqslant x-K) = F(x-\varepsilon) - F(x) - F(x-K)$$

where the assumption about the convergence of expectation of C_c^∞ functions is used. Taking the limit $\varepsilon \to 0, K \to \infty$, we get

$$\liminf_{n\to\infty}(F_n(x) - F(x)) \geqslant 0 \tag{7.3}$$

where the fact that F is continuous at x is used. To obtain a upper bound, we use the fact that $F_n(x) - F(x) = \mathbb{E}(\mathbb{1}_{X>x} - \mathbb{1}_{X_n>x})$ and approximate $\mathbb{1}_{y>x}$ by $\hat{f}_{\varepsilon,K}(x-y)$ to get

$$\limsup_{n\to\infty}(F_n(x) - F(x)) \leqslant 0.$$

This combined with (**??**) would yield the desired result. $\lim_{n\to\infty} F_n(x) = F(x)$. ■

7.2 Central Limit Theorems

Definition 7.10 (Normal distribution). *$X \sim N(0,1)$ if density function of X is,*

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, x \in \mathbb{R}$$

and the distribution function is $\Phi(x) = \int_{-\infty}^{x} \phi(t)\mathrm{d}t$ and $X \sim N(\mu, \sigma^2)$ if $\frac{X-\mu}{\sigma} \sim N(0,1)$.

Proposition 7.11. *Let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$. If $X_1 \perp X_2$, then*

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

In particular if $X_1, \ldots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$, then $X_1 + \ldots + X_n \sim N(n\mu, n\sigma^2)$ or,

$$\frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n\sigma^2}} \sim N(0,1).$$

Theorem 7.12 (Basic Central Limit Theorem). *If $X_1, \ldots, X_n$ are i.i.d. random variables with $\mathbb{E}(X_1) = \mu$ and $\mathrm{Var}(X_1) = \sigma^2$. Then,*

$$\frac{X_1 + \ldots + X_n - n\mu}{\sqrt{n\sigma^2}} \overset{d}{\to} Z \sim N(0,1).$$

Theorem 7.13 (Lindeberg's Central Limit Theorem). *Let $X_1, \ldots, X_n$ be independent random variables with $\mathbb{E}(X_i) = \mu_i$ and $\mathrm{Var}(X_i) = \sigma_i^2$. Define $s_n^2 := \sum_{i=1}^n \sigma_i^2$. Then,*

$$\frac{1}{s_n}\sum_{i=1}^n (X_i - \mu_i) \xrightarrow{(d)} N(0,1),$$

if $s_n \to \infty$ and $\forall \varepsilon > 0$,

$$\frac{1}{s_n^2}\sum_{i=1}^n \mathbb{E}\left(|X_i - \mu_i|^2 \mathbb{1}_{|X_i-\mu_i| \geqslant \varepsilon s_n}\right) \to 0.$$