Probability Theory and Random Processes (MA 225)

Class Notes

September – November, 2020

Instructor
Ayon Ganguly
Department of Mathematics
IIT Guwahati

Contents

Chapter 4

Limit Theorems

This chapter will deal with convergence properties of sequence of RVs. There are several modes of convergence of sequence of RVs. Here, we will discuss four modes of convergence for a sequence of RVs $\{X_n\}$. Then we will see strong law of large numbers and central limit theorem. These are quite useful concepts in probability. They have applications in different other fields including Statistics. In this chapter, we shall not prove most of the theorems. Our main aim will be to understand the theorems and apply them to solve problems. In rest of the chapter, 1_A denotes the indicator function of the set A.

4.1 Modes of Convergence

Definition 4.1 (Almost Sure Convergence)**.** *Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\mathcal{S}, \mathcal{F}, P)$. Let X be a random variable defined on the same probability space $(\mathcal{S}, \mathcal{F}, P)$. We say that X_n converges almost surely or with probability (w.p.) 1 to a random variable X if*

$$P\left(\{\omega \in \mathcal{S} : X_n(\omega) \to X(\omega)\}\right) = 1.$$

Example 4.1. Let $\mathcal{S} = [0, 1], \mathcal{F} = \mathcal{B}([0, 1])$ and P be a uniform probability (for any interval $I \subseteq \mathcal{S}$, $P(I) =$ length of I). Define the sequence of RVs by

$$X_n(\omega) = 1_{[0, \frac{1}{n}]}(\omega) \quad \text{for all } n = 1,\, 2,\, 3,\, \ldots.$$

Then X_n converges almost surely to the zero RV. Here, the zero RV means a RV, say X, defined on the same probability space $(\mathcal{S},\, \mathcal{F},\, P)$ such that $X(\omega) = 0$ for all $\omega \in \mathcal{S}$. To see it, notice that for any fixed $\omega \in (0,\, 1]$, we can find an n_0 such that $\frac{1}{n} < \omega$ for all $n \geq n_0$. Thus, $X_n(\omega) \to 0 = X(\omega)$ as $n \to \infty$. Therefore, $\{\omega \in \mathcal{S} : X_n(\omega) \to X(\omega)\} = (0,\, 1]$ and hence,

$$P\left(\{\omega \in \mathcal{S} : X_n(\omega) \to X(\omega)\}\right) = P\left((0,\, 1]\right) = 1.$$

Thus, $X_n \to 0$ almost surely. ||

Definition 4.2 (Convergence in Probability)**.** *Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\mathcal{S}, \mathcal{F}, P)$. Let X be a random variable defined on the same probability space $(\mathcal{S}, \mathcal{F}, P)$. We say that X_n converges in probability to a random variable X if for any $\epsilon > 0$*,

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

Example 4.2. Let $\mathcal{S} = [0,1], \mathcal{F} = \mathcal{B}([0,1])$ and P be a uniform probability. Define the sequence of RVs using $X_n = 1_{[0,\frac{1}{n}]}$. Then X_n converges in probability to the zero random variable. Let X denote the zero RV defined on the same sample space. To see it, notice that for any fixed $\epsilon > 0$, $|X_n - X| > \epsilon$ only on the interval $[0, \frac{1}{n}]$. Thus,

$$P(|X_n - X| > \epsilon) = \frac{1}{n} \implies \lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0.$$

Therefore, $X_n \to X$ in probability. ||

Example 4.3. Let $\mathcal{S} = [0,1], \mathcal{F} = \mathcal{B}([0,1])$ and P be a uniform probability. Define the sequence of RVs using $X_n = n1_{[0,\frac{1}{n}]}$. Then X_n converges in probability to the zero random variable. Let X denote the zero RV defined on the same sample space. To see it, notice that for any fixed $\epsilon > 0$, $|X_n - X| > \epsilon$ only on the interval $[0, \frac{1}{n}]$. Thus,

$$P(|X_n - X| > \epsilon) = \frac{1}{n} \implies \lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0.$$

Therefore, $X_n \to X$ in probability. ||

It may seem that convergence almost surely and convergence in probability are equivalent. However, this is not true, as the following example shows.

Example 4.4. Let $\mathcal{S} = [0,1], \mathcal{F} = \mathcal{B}([0,1])$ and P be the uniform probability. Define the sequence of RVs by

$$X_{m,n} = 1_{[\frac{m-1}{2^n}, \frac{m}{2^n}]} \quad \text{for } m = 1, 2, \ldots, 2^n;\ n = 1, 2, 3, \ldots.$$

Note that $X_{1,1} = 1_{[0,1/2]}, X_{2,1} = 1_{[1/2,1]}, X_{1,2} = 1_{[0,1/4]}, X_{2,2} = 1_{[1/4,1/2]}, X_{3,2} = 1_{[1/2,3/4]}, X_{4,2} = 1_{[3/4,1]}$ and so on. This sequence of RVs $\{X_{m,n}\}$ can be visualized as follows (see Figure 4.1). We start with the interval $[0, 1]$. First, we divide the interval into two equal parts, $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$. The first RV $X_{1,1}$ is 1 on the first part and 0 on the second part. The second random variable $X_{2,1}$ is 1 on the second part and 0 on the first part. Then, we divide the interval into 2^2 equal parts, *viz.*, $[0, \frac{1}{2^2}]$, $[\frac{1}{2^2}, \frac{2}{2^2}]$, $[\frac{2}{2^2}, \frac{3}{2^2}]$ and $[\frac{3}{2^2}, 1]$. Now, the third RV $X_{1,2}$ is 1 on the first part $[1, \frac{1}{4}]$ and 0 otherwise. The fourth RV $X_{2,2}$ is 1 on the second part $[\frac{1}{4}, \frac{1}{2}]$ and 0 otherwise. The fifth RV $X_{3,2}$ equals 1 on the third part $[\frac{1}{2}, \frac{3}{4}]$ and 0 otherwise. Finally, the sixth RV $X_{4,2}$ is 1 on the fourth part $[\frac{3}{4}, 1]$ and 0 otherwise. Next, we divide the interval $[0, 1]$ into 2^3 equal parts and define the next 8 RVs in the similar manner. This procedure continues.

Let us assume that X be a RV defined on the same probability space and $X = 0$. Then, for any $\epsilon > 0$,

$$P(|X_{m,n} - X| > \epsilon) = \frac{1}{2^n} \implies \lim_{n\to\infty} P(|X_{m,n} - X| > \epsilon) = 0.$$

Therefore, $X_{m,n} \to X$ in probability. However, for any fixed $\omega \in \mathcal{S}$, there exists a subsequence of the sequence of real numbers $\{X_{m,n}(\omega)\}$ that converges to one and another subsequence that converges to zero. Therefore, $\{X_{m,n}(\omega)\}$ does not converge for all $\omega \in \mathcal{S}$. Thus,

$$P(\{\omega \in \mathcal{S} : X_{m,n} \text{ converges}\}) = P(\emptyset) = 0.$$

This shows that $X_{m,n}$ do not converge to any RV almost surely. This example shows that a sequence of RVs, which converges in probability, may not converge almost surely. ||

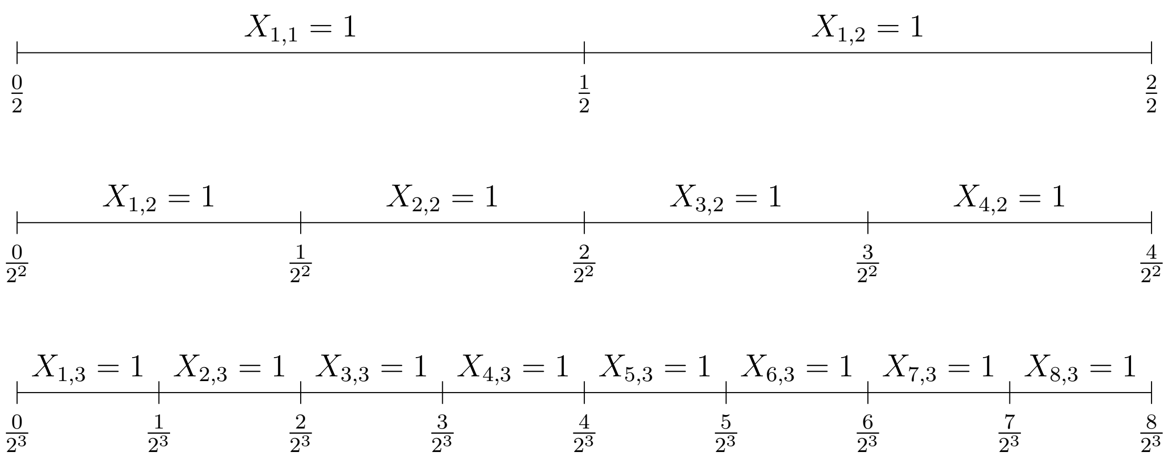

Figure 4.1: Figure for Example 4.4

Definition 4.3 (Convergence in r^{th} Mean)**.** *Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\mathcal{S}, \mathcal{F}, P)$. Let X be a random variable defined on the same probability space $(\mathcal{S}, \mathcal{F}, P)$. For $r = 1, 2, 3, \ldots$, we say that X_n converges in r^{th} mean to a random variable X if*

$$E|X_n - X|^r \to 0 \quad \textit{as } n \to \infty.$$

Example 4.5. Let $\mathcal{S} = [0,1], \mathcal{F} = \mathcal{B}([0,1])$ and P be a uniform measure. Define $X_n = 1_{[0,\frac{1}{n}]}$. Then X_n converges in 1st mean to the zero random variable. To see it, notice that

$$E|X_n - X| = \frac{1}{n} \to 0 \quad \text{as } n \to \infty,$$

where X is a zero RV defined on the same probability space. ||

Definition 4.4 (Convergence in Distribution)**.** *Let $\{X_n\}$ be a sequence of RVs and X be a RV. Let $F_n(\cdot)$ and $F(\cdot)$ denote the CDF of X_n and X, respectively. We say that X_n converges in distribution to a random variable X if*

$$F_n(x) \to F(x) \quad \textit{as } n \to \infty$$

for all x where F is continuous.

Unlike the first three modes of convergence, here X_n's can be defined on different probability spaces. We are only interested if the sequence of CDFs converges to a CDF. This flexibility makes this mode of convergence very useful.

Example 4.6. Suppose X_ns are random variables such that $P(X_n = \frac{1}{n}) = 1$. Then, the CDF of X_n is

$$F_n(x) = \begin{cases} 0 & \text{if } x < \frac{1}{n} \\ 1 & \text{if } x \geq \frac{1}{n}, \end{cases}$$

which converges pointwise to the function

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$$

for all $x \neq 0$, which is the point of discontinuity of the function $F(\cdot)$. Now $F(\cdot)$ is the CDF of the RV X, which takes value 0 with probability one. Therefore, X_n converges in distribution to the zero RV. $||$

The following theorems states the relation between different modes of convergence.

Theorem 4.1. *Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\mathcal{S}, \mathcal{F}, P)$. Let X be a random variable defined on the same probability space $(\mathcal{S}, \mathcal{F}, P)$. Then $X_n \to X$ in probability if $X_n \to X$ almost surely.*

Proof: This prove is skipped here. $\square$

Theorem 4.2. *Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\mathcal{S}, \mathcal{F}, P)$. Let X be a random variable defined on the same probability space $(\mathcal{S}, \mathcal{F}, P)$. Then $X_n \to X$ in probability if $X_n \to X$ in rth mean for any $r = 1, 2, 3, \ldots$.*

Proof: Let $X_n \to X$ in rth mean. Then, using Markov inequality, for any $\epsilon > 0$,

$$P\left(|X_n - X| > \epsilon\right) \leq \frac{E|X_n - X|^r}{\epsilon^r} \to 0 \quad \text{as } n \to \infty.$$

As probability of an event is always non-negative,

$$P\left(|X_n - X| > \epsilon\right) \to 0 \quad \text{as } n \to \infty.$$

Thus $X_n \to X$ in probability. $\square$

Theorem 4.3. *Let $\{X_n\}$ be a sequence of random variables defined on a probability space $(\mathcal{S}, \mathcal{F}, P)$. Let X be a random variable defined on the same probability space $(\mathcal{S}, \mathcal{F}, P)$. Then $X_n \to X$ in distribution if $X_n \to X$ in probability.*

Proof: The proof is skipped here. $\square$

The following figure depicts the relationship between several modes of convergence pictorially. Note that the arrows are one-sided. What about other sides? Moreover, there is no

$$\begin{tikzcd} a.s \arrow[dr] & & mean \arrow[dl] \\ & Prob \arrow[d] & \\ & Dist & \end{tikzcd}$$

arrows between almost sure convergence and rth mean convergence. The following examples show that in general one mode of convergence does not imply other, whenever there is no directed arrows in the above figure. The Example 4.4 shows that probability convergence does not imply almost sure convergence.

Example 4.7. Let $\mathcal{S} = [0,1], \mathcal{F} = \mathcal{B}([0,1])$ and P be a uniform probability. Define the sequence of RVs by

$$X_{m,n} = 1_{[\frac{m-1}{2^n}, \frac{m}{2^n}]} \quad \text{for } m = 1,\, 2,\, \ldots,\, 2^n;\; n = 1,\, 2,\, 3,\, \ldots.$$

Then

$$E|X_{m,n}| = \frac{1}{2^n} \to 0 \quad \text{as } n \to \infty.$$

Thus, $X_{m,n} \to X = 0$ in 1st mean. However, in Example 4.4, we have seen that $X_{m,n}$ does not convergence almost surely. This example shows that rth mean convergence does not imply almost sure convergence. ||

Example 4.8. Let $\mathcal{S} = [0,1], \mathcal{F} = \mathcal{B}([0,1])$ and P be a uniform probability. Define $X_n = n1_{[0,\frac{1}{n}]}$. Now, taking $X = 0$,

$$P\left(|X_n - X| > \epsilon\right) = \frac{1}{n} \to 0 \quad \text{as } n \to \infty,$$

for any $\epsilon > 0$. Thus, $X_n \to X$ in probability. Using the logic used in Example 4.1,

$$P\left(\{\omega \in \mathcal{S} : X_n(\omega) \to X(\omega)\}\right) = P((0,\, 1]) = 1.$$

Thus, $X_n \to X$ almost surely. However, X_n does not converge to X in rth mean. To see it, notice that

$$E|X_n - X|^r = n^{r-1} \to \begin{cases} 1 & \text{if } r = 1 \\ \infty & \text{if } r > 1. \end{cases}$$

This example shows that probability convergence or almost sure convergence do not imply rth mean convergence. ||

Example 4.9. Let X be a $N(0,1)$ RV defined on some probability space $(\mathcal{S}, \mathcal{F}, P)$. Define $X_n = X$ for all n. Notice that the CDFs of X_n are same for all $n = 1,\, 2, \ldots$ and is given by $\Phi(\cdot)$. Moreover, the CDFs of X and $-X$ are also $\Phi(\cdot)$. Thus, X_n converges in distribution to $-X$. However, X_n does not converge to $-X$ in probability. To see it, we can proceed as follows: for $\epsilon > 0$,

$$P\left(|X_n + X| \le \epsilon\right) = P\left(2|X| \le \epsilon\right) = 2\Phi\left(\frac{\epsilon}{2}\right) - 1 \ne 1.$$

This example shows that distribution convergence does not imply probability convergence, even if the random variables are defined on the same probability space. ||

The following theorems provide several properties of different modes of convergence. The proof of the theorems are skipped here.

Theorem 4.4. *Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables defined on a probability space $(\mathcal{S}, \mathcal{F}, P)$. Suppose $X_n \to X$ w. p. 1 and $Y_n \to Y$ w. p. 1. Then*

- *$X_n + Y_n \to X + Y$ w. p. 1.*
- *$X_nY_n \to XY$ w. p. 1.*
- *$f(X_n) \to f(X)$ w. p. 1, for any f continuous.*

Theorem 4.5. *Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables defined on a probability space $(\mathcal{S}, \mathcal{F}, P)$. Suppose $X_n \to X$ in probability and $Y_n \to Y$ in probability. Then*

- $X_n + Y_n \to X + Y$ *in probability.*
- $X_n Y_n \to XY$ *in probability.*
- $f(X_n) \to f(X)$ *in probability, for any f continuous.*

Theorem 4.6. *Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables defined on a probability space $(\mathcal{S}, \mathcal{F}, P)$.*

- *If $X_n \to X$ in r^{th} mean and $Y_n \to Y$ in r^{th} mean, then $X_n + Y_n \to X + Y$ in r^{th} mean.*
- *If $X_n \to X$ in r^{th} mean then $f(X_n) \to f(X)$ in r^{th} mean, for any f bounded continuous.*

Theorem 4.7. *Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of random variables defined on a probability space $(\mathcal{S}, \mathcal{F}, P)$. Suppose $X_n \to X$ in distribution and $Y_n \to c$ in probability for some constant c. Then*

- $X_n + Y_n \to X + c$ *in distribution.*
- $X_n Y_n \to cX$ *in distribution.*
- $f(X_n) \to f(X)$ *in distribution, for any f continuous.*

Example 4.10. Let X, $Y \sim N(0,\, 1)$ and X and Y be independent RVs. Take $X_n = X$ and $Y_n = Y$ for all $n = 1,\, 2,\, 3,\, \ldots$. Then, $X_n \to X$ in distribution and $Y_n \to X$ in distribution. Now, $X_n + Y_n = X + Y \sim N(0,\, 2)$ and $2X \sim N(0,\, 4)$. Thus, $X_n + Y_n$ does not converges to $2X$ in distribution. This example shows that $X_n + Y_n$ may not converge to $X + Y$ in distribution if $X_n \to X$ in distribution and $Y_n \to Y$ in distribution. You can easily check that the same conclusion is also true for product. ||

Theorem 4.8. *Suppose $\{X_n\}$ is a sequence of RVs defined on a probability space and X_n converges in distribution to some constant c, then X_n also converges in probability to c.*

Proof: As X_n converges to a constant c,

$$F_n(x) \to F(x) = \begin{cases} 0 & \text{if } x < c \\ 1 & \text{if } x \geq c \end{cases}$$

as $n \to \infty$. Now, fix $\varepsilon > 0$. Then,

$$\begin{aligned} 0 \leq P\left(|X_n - c| > \varepsilon\right) &= P\left(X_n > c + \varepsilon\right) + P\left(X_n < c - \varepsilon\right) \\ &\leq 1 - F_n(c + \varepsilon) + F_n(c - \varepsilon) \to 1 - 1 + 0 = 0 \end{aligned}$$

as $n \to \infty$. Note that as $c + \varepsilon > c$ and $c - \varepsilon < c$, $F_n(c + \varepsilon) \to 1$ and $F_n(c - \varepsilon) \to 0$. Thus, $X_n \to c$ in probability. □

Corollary 4.1. *Suppose $\{X_n\}$ is a sequence of RVs defined on a probability space. Then, $X_n \to c$ in distribution if and only if $X_n \to c$ in probability, where c is a constant.*

Proof: The proof of the corollary is straight forward by combining the previous theorem and Theorem 4.3. □

Theorem 4.9. *Let X_n be a RV with MGF $M_n(t)$ for $n = 1, 2, 3, \ldots$. Let X be a RV with MGF $M(t)$. If $M_n(t) \to M(t)$ for all t in an open interval containing zero, as $n \to \infty$, then $X_n \to X$ in distribution.*

Theorem 4.10. *Let X_n be a DRV with PMF $f_n(\cdot)$ for $n = 1, 2, 3, \ldots$. Let X be a DRV with PMF $f(\cdot)$. If, for all $x \in \mathbb{R}$, $f_n(x) \to f(x)$ as $n \to \infty$, then $X_n \to X$ in distribution.*

Theorem 4.11. *Let X_n be a CRV with PDF $f_n(\cdot)$ for $n = 1, 2, 3, \ldots$. Let X be a CRV with PDF $f(\cdot)$. If, for all $x \in \mathbb{R}$, $f_n(x) \to f(x)$ as $n \to \infty$, then $X_n \to X$ in distribution.*

Example 4.11. Let $X_n \sim Bin(n, p_n)$, where $p_n \to 0$ and $np_n = \lambda\,(> 0)$. Then, for $n = 1, 2, 3, \ldots$, the MGF of X_n is

$$M_n(t) = \left(1 - p_n + p_n e^t\right)^n = \left(1 + \frac{\lambda}{n}\left(e^t - 1\right)\right)^n \to e^{\lambda\left(e^t - 1\right)}$$

for all $t \in \mathbb{R}$. Note that if $X \sim Poi(\lambda)$, then the MGF of X is

$$M(t) = e^{\lambda\left(e^t - 1\right)} \quad \text{for } t \in \mathbb{R}.$$

Thus, $X_n \to X$ in distribution.

Recall that the motivation of the Poission distribution was not discussed when it was introduced. This example tells us the motivation behind the Poission distribution. We can use Poisson distribution to approximate the probability of a Binomial distribution when probability of success is very small and number of trials is very large. ||

Example 4.12. Under the conditions of the previous example, we can prove that $X_n \to X$ using Theorem 4.10. To see it, we can proceed as follows.

$$\begin{aligned}
P\left(X_n = k\right) &= \binom{n}{k} p_n^k \left(1 - p_n\right)^{n-k} \\
&= \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \times \frac{n(n-1)(n-2)\ldots(n-k+1)}{n^k}\left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&\to \frac{e^{-\lambda}\lambda^k}{k!}.
\end{aligned}$$

Notice that the support of X_n is the set $\{0, 1, 2, \ldots, n\}$. When $n \to \infty$, the support becomes $\{0, 1, 2, \ldots\}$. ||

Example 4.13. Let $X_n \sim U(0, 1 + 1/n)$ for $n = 1, 2, 3, \ldots$. Then the PDF of X_n is

$$f_n(x) = \begin{cases} \frac{1}{1+\frac{1}{n}} & \text{if } 0 < x < 1 + \frac{1}{n} \\ 0 & \text{otherwise} \end{cases} \longrightarrow f(x) = \begin{cases} 1 & \text{if } 0 < x \leq 1 \\ 0 & \text{otherwise,} \end{cases}$$

which is the PDF of a RV X such that $X \sim U(0, 1)$. Thus, $X_n \to X$ in distribution. ||

4.2 Limit Theorems

In this section, we will discuss two very famous and useful theorems. Again, we will skip the proofs, but we will see some applications.

Theorem 4.12 (Strong Law of Large Numbers)**.** *Let $\{X_n\}$ be a sequence of i.i.d. RVs with finite mean μ. Define $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then $\{\overline{X}_n\}$ converges to μ almost surely.*

Proof: The proof is skipped. $\square$

Let us loosely discuss the intuitive idea of the previous theorem. Suppose that we want to find the average height of all Indians. Ideally, we need to go to each and every Indian and record their height. Finally, the average should be calculated based on the observations on height. This average is called population average or population mean. It is a very costly (in terms of money and time) process. Alternatively, we can take a representative sample of the Indian population. Here, sample represents a subset of original population. Then, we can collect the height data for each and every person in the sample and then calculate the mean of those sample observations. This mean is called sample mean. If the number of persons in the sample is very small (say, 5 or 10), the calculated sample mean may not be close to the original population mean. However, if we keep on increasing the sample size (the number of persons in the sample), the sample mean should get closer to population mean. The above theorem provided theoretical justification of this intuitive idea. Note that μ and $\overline{X}$ are population and sample means, respectively. Thus, loosely speaking, the strong law of large numbers (SLLN) states that sample mean converges to population mean almost surely as we increase the sample size.

Example 4.14 (Bernoulli proportion converges to success probability)**.** Suppose that a sequence of independent trials is performed. Let E be a fixed event. Letting

$$X_i = \begin{cases} 1 & \text{if } E \text{ occurs on the } i\text{th trial} \\ 0 & \text{if } E \text{ does not occur on the } i\text{th trial,} \end{cases}$$

we have by the SLLN that, with probability one,

$$\overline{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} \to \mu = E(X_1) = P(E).$$

Since, $X_1 + X_2 + \ldots + X_n$ represents the number of times that the event E occurs in the first n trials, we may interpret it as stating that, with probability one, the limiting proportion of time that the event E occurs is $P(E)$. ||

Example 4.15 (Monte Carlo Integration)**.** Suppose that we want to integrate

$$I = \int_a^b h(x)dx.$$

If we cannot do it explicitly, we can use numerical technique like Simpson's 1/3rd rule. Here, we will see another technique based on the SLLN. Suppose that a and b are finite real numbers. Note that the above integration can be rewritten as

$$I = (b-a)\int_a^b h(x)\frac{1}{b-a}dx = (b-a)E\left(Y\right),$$

where $Y = h(X)$ and $X \sim U(a,\, b)$. Let $\{X_n\}$ be a sequence of *i.i.d.* RVs with common distribution $U(a,\, b)$ and assume that $Y_n = h(X_n)$ for $n = 1,\, 2,\, 3,\, \ldots$. Now, SLLN says that, with probability one,

$$\overline{Y}_n = \frac{Y_1 + Y_2 + \ldots + Y_n}{n} = \frac{1}{n}\sum_{i=1}^{n} h(X_i) \to E(Y) = \frac{I}{b-a} \implies \frac{b-a}{n}\sum_{i=1}^{n} h(X_i) \to I.$$

Thus, we can generate N random numbers from $U(a,\, b)$. The generation from $U(a,\, b)$ can be done using any standard software like R, MATLAB, etc. Here, N is a large integer (the popular choices are 5000 or 10000). Then, the integration I can be approximated using $\frac{b-a}{N}\sum_{i=1}^{N} h(X_i)$. ||

Theorem 4.13 (Central Limit Theorem)**.** *Let $\{X_n\}$ be a sequence of i.i.d. RVs with mean μ and variance $\sigma^2 < \infty$. Then, as $n \to \infty$,*

$$P\left(\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \le a\right) \to \Phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Proof: The proof is skipped. □

The central limit theorem (CLT) says that

$$\frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \to Z \sim N(0,\, 1) \quad \text{in distribution.}$$

Thus, the CDF of standardized sample mean can be approximated (for large sample size) using the CDF of a standard normal distribution, whenever X_n's are i.i.d. RVs with finite mean μ and finite variance σ^2. In other words, the CDF of sample mean can be approximated using the CDF of a $N(\mu,\, \frac{\sigma^2}{n})$ distribution. Note that CLT holds true for any distribution of X_n as long as the variance is finite.

Example 4.16 (Normal Approximation to the Binomial)**.** Let $X_n \sim Bin(n,\, p)$. Then

$$P\left(\frac{X_n - np}{\sqrt{np(1-p)}} \le a\right) \to \Phi(a) \quad \text{as } n \to \infty.$$

We will use CLT to prove this statement. Let $\{Y_n\}$ be a sequence of i.i.d. RVs where $X_1 \sim Bernoulli(p)$. Then

$$\sum_{i=1}^{n} Y_i \stackrel{d}{=} X_n \implies \overline{Y}_n \stackrel{d}{=} \frac{X_n}{n}.$$

Now, $E(Y_n) = p$ and $Var(Y_n) = p(1-p)$. Thus,

$$P\left(\frac{X_n - np}{\sqrt{np(1-p)}} \le a\right) = P\left(\sqrt{n}\frac{\overline{Y}_n - p}{\sqrt{p(1-p)}} \le a\right) \to \Phi(a) \quad \text{as } n \to \infty.$$

The equality in the above line is due to the fact that $\overline{Y}_n$ and $\frac{X_n}{n}$ have same distribution. The convergence is due to the CLT. ||

Example 4.17. The lifetimes of a special type of battery is a RV with mean 40 hours and standard deviation 20 hours. A battery is used until it fails, at which point it is replaced by a new one. Assume a stockpile of 25 such batteries, the lifetimes of which are independent, we want to approximate the probability that over 1100 hours of use can be obtained. Let X_i denote the lifetime of the ith battery to be put in use. Then, we are interested in

$$p = P\left(X_1 + X_2 + \ldots + X_{25} > 1100\right),$$

which can be approximated as follows:

$$\begin{aligned}
p &= P\left(X_1 + X_2 + \ldots + X_{25} > 1100\right) \\
&= P\left(\overline{X}_{25} > 44\right) \\
&= P\left(\sqrt{25}\,\frac{\overline{X}_{25} - 40}{20} > \sqrt{25}\,\frac{44 - 40}{20}\right) \\
&\approx P\left(Z > 1\right), \text{ where } Z \sim N(0,\, 1). \text{ This is due to CLT} \\
&= 1 - \Phi(1) \approx 0.1587,
\end{aligned}$$

as $\Phi(1) \approx 0.8413$. This values can be found from the normal table. ||