Law of Large Numbers, Central Limit Theorem and Normal Approximation

Definition 1. (1) *Let $(X_n)_{n\geq 1}$ be a sequence of random variables (not necessarily independent), and let a be a real number. We say that the sequence $(X_n)_{n\geq 1}$ converges to a in probability (written as $X_n \xrightarrow{p} a$) if for every $\epsilon > 0$, we have*

$$\lim_{n\to\infty} P(\{|X_n - a| \geq \epsilon\}) = 0.$$

(2) *A sequence $(X_n)_{n\geq 1}$ of random variables is said to be bounded in probability if there exists a $M > 0$ such that*

$$P(\cap_{n=1}^{\infty}\{|X_n| \leq M\}) = 1.$$

(3) *Let $(X_n)_{n\geq 1}$ be a sequence of random variables and let F_n be the d.f. of $X_n, n = 1, 2, \ldots$. Let X be a random variable with d.f. F. We say that the sequence $(X_n)_{n\geq 1}$ converges to X in distribution (written as $X_n \xrightarrow{d} X$) if*

$$\lim_{n\to\infty} F_n(x) = F(x),\ \forall\ x \in C_F,$$

where C_F is the set of continuity point of F.

Example 2. *Consider a sequence $(X_n)_{n\geq 1}$ of independent random variables that are uniformly distributed over $(0, 1)$ and let $Y_n = \min\{X_1, \ldots, X_n\}$. The sequence of values of Y_n cannot increase as n increases. Thus, we intuitively expect that Y_n converges to zero.*
Now, for $\epsilon \geq 1$, $P(X_i \geq \epsilon) = 1 - P(X_i \leq \epsilon) = 0$ and for $0 < \epsilon < 1$, $P(X_i \geq \epsilon) = 1 - P(X_i \leq \epsilon) = (1 - \epsilon)$, $1 \leq i \leq n$.

Hence,

$$\begin{aligned} P(\{|Y_n - 0| \geq \epsilon\}) &= P(X_1 \geq \epsilon, \ldots, X_n \geq \epsilon) \\ &= P(X_1 \geq \epsilon) \cdots P(X_n \geq \epsilon) \\ &= \begin{cases} (1-\epsilon)^n, & \text{if } 0 < \epsilon < 1 \\ 0, & \text{if } \epsilon \geq 1 \end{cases} \end{aligned}$$

Hence, for $\epsilon > 0$, we have

$$\lim_{n\to\infty} P(\{|Y_n - 0| \geq \epsilon\}) = 0.$$

Therefore, $Y_n \xrightarrow{p} 0$.

Theorem 3. *Let $(X_n)_{n\geq 1}$ be a sequence of random variables with $E(X_n) = \mu_n$ and $Var(X_n) = \sigma_n^2, n = 1, 2, \ldots$ Suppose $\lim\limits_{n\to\infty} \mu_n = \mu$ and $\lim\limits_{n\to\infty} \sigma_n^2 = 0$. Then $X_n \xrightarrow{p} \mu$.*

Theorem 4. *Let $(X_n)_{n\geq 1}$ be a sequence of random variables and X be another random variable. Suppose that there exists a $h > 0$ such that m.g.f. $\phi, \phi_1, \phi_2, \ldots$ of $X, X_1, X_2, \ldots$, respectively, are finite on $(-h, h)$.*

(1) *If $\lim\limits_{n\to\infty} \phi_n(t) = \phi(t),\ \forall\ t \in (-h, h)$, then $X_n \xrightarrow{d} X$, where $F, F_1, F_2, \ldots$ are c.d.f. of $X, X_1, X_2, \ldots$, respectively.*
(2) *If $X_1, X_2, \ldots$ are bounded in probability and $X_n \xrightarrow{d} X$, then $\lim\limits_{n\to\infty} \phi_n(t) = \phi(t),\ \forall\ t \in (-h, h)$.*

Continuity Correction: Continuity correction is an adjustment that is made when a discrete distribution is approximated by a continuous distribution.

Table of continuity correction:

Discrete	Continuous
$P(X = a)$	$P(a - 0.5 < X < a + 0.5)$
$P(X > a)$	$P(X > a + 0.5)$
$P(X \leq a)$	$P(X < a + 0.5)$
$P(X < a)$	$P(X < a - 0.5)$
$P(X \geq a)$	$P(X > a - 0.5)$

1. **Law of Large Numbers**

Theorem 5. ***The weak law of large numbers (WLLN):*** *Let $(X_n)_{n\geq 1}$ be a sequence of independent and identically distributed random variables, each having finite mean $E(X_i) = \mu$, $i = 1, 2, \cdots,$. Then, for any $\epsilon > 0$,*

$$P\left(\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\}\right) \to 0 \text{ as } n \to \infty$$

i.e.,

$$\lim_{n\to\infty} P\left(\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\}\right) = 0.$$

equivalently

$$\lim_{n\to\infty} P\left(\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| < \epsilon\right\}\right) = 1.$$

The weak law of large numbers asserts that the sample mean of a large number of independent identically distributed random variables is very close to the true mean with high probability.

Proof. We assume that the random variables have a finite variance σ^2. Now, $E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = \frac{1}{n}E(X_1+X_2+\cdots+X_n) = \frac{E(X_1)+E(X_2)+\cdots+E(X_n)}{n} = \mu$, and $Var\left(\frac{X_1+X_2+\cdots+X_n}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$.

By Chebyshev's Inequality,

$$P\left(\left\{\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \epsilon\right\}\right) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Since $\lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0$, $\lim_{n\to\infty} P\left(\left\{\left|\frac{X_1+X_2+\cdots+X_n}{n} - \mu\right| \geq \epsilon\right\}\right) = 0.$ □

Theorem 6. ***The strong law of large numbers (SLLN):*** *Let $(X_n)_{n\geq 1}$ be a sequence of independent and identically distributed random variables, each having finite mean $E(X_i) = \mu$, $i = 1, 2, \cdots,$. Then,*

$$P\left(\left\{\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} = \mu\right\}\right) = 1$$

i.e.,

$$P\left(\left\{w \in \mathcal{S} \mid \lim_{n\to\infty} \frac{X_1(w) + X_2(w) + \cdots + X_n(w)}{n} = \mu\right\}\right) = 1$$

There is a minor difference between the weak and the strong law. The weak law states that the probability $P\left(\left\{\left|\frac{X_1+X_2+\cdots+X_n}{n} - \mu\right| < \epsilon\right\}\right)$ of a significant deviation of sample mean $\frac{X_1+X_2+\cdots+X_n}{n}$ from μ goes to 1 as $n \to \infty$. Still, for any finite n, this probability can be positive. The weak law provides no conclusive information on the number of such deviations but the strong law does. According to the strong law, $\frac{X_1+X_2+\cdots+X_n}{n}$ converges to μ with probability 1. This implies that for any given $\epsilon > 0$, the probability that the difference $\left|\frac{X_1+X_2+\cdots+X_n}{n} - \mu\right| < \epsilon$ an infinite number of times is equal to 1.

Example 7. *Consider the tossing a coin n-times with S_n the number of heads that turn up. Then the random variable $\frac{S_n}{n}$ represents the fractions of times heads turn up and will have values between 0 and 1. The law of large numbers predicts that the outcomes for this random variable, for large n, will be near $\frac{1}{2}$.*

2. Central Limit Theorem

Theorem 8. ***The Central Limit Theorem (CLT):*** *Let $(X_n)_{n\geq 1}$ be a sequence of independent and identically distributed random variables, each having finite mean μ and variance σ^2. Then*

$$Z_n = \frac{\sqrt{n}(\overline{X_n} - \mu)}{\sigma} \xrightarrow{d} Z = N(0,1), \textit{ where } \overline{X_n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

i.e.,

$$Z_n = \frac{(X_1 + X_2 + \cdots + X_n) - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} Z = N(0,1),$$

i.e.,

$$X_1 + X_2 + \cdots + X_n \approx N(n\mu, n\sigma^2), \textit{ for large } n.$$

The Central Limit Theorem states that irrespective of the nature of the parent distribution, the probability distribution of a normalized version of the sample mean, based on a random sample of large size, is approximately standard normal.

Example 9. *Civil engineers believe that W, the amount of weight (in units of 1000 pounds) that a certain span of a bridge can with stand without structural damage resulting, is normally distributed with mean 400 and standard deviation 40. Suppose that the weight (again, in units of 1000 pounds) of a car is a random variable with mean 3 and standard deviation 0.3. How many cars would have to be on the bridge span for the probability of structural damage to exceed 0.1?*

Solution: Let P_n denote the probability of structural damage when there are n cars on the bridge. That is

$$P_n = P(\{X_1 + X_2 + \cdots + X_n \geq W\}) = P(\{X_1 + X_2 + \cdots + X_n - W \geq 0\})$$

where X_i is the weight of the i-th car, $i = 1, 2, \ldots, n$. Now it follows from central limit theorem that $\sum_{i=1}^{n} X_i$ is approximately normal with mean $3n$ and variance $0.09n$. Hence, since W is independent of the X_i, $i = 1, \ldots, n$, and is also normal, it follows that $\sum_{i=1}^{n} X_i - W$ is approximately normal with mean and variance given by

$$E\left(\sum_{i=1}^{n} X_i - W\right) = 3n - 400$$

$$Var\left(\sum_{i=1}^{n} X_i - W\right) = Var\left(\sum_{i=1}^{n} X_i\right) + Var(W) = 0.09n + 1600$$

Therefore, if we let $Z = \frac{\sum_{i=1}^{n} X_i - W - (3n-400)}{\sqrt{0.09n+1600}}$, then

$$P_n = P(\{X_1 + X_2 + \cdots + X_n - W \geq 0\} = P\left(Z \geq \frac{-(3n - 400)}{\sqrt{0.09n + 1600}}\right)$$

where Y is approximately a standard normal random variable. Now $P(Z \geq 1.28) \approx 0.1$.

$$\{Z \geq 1.28\} \subseteq \{Z \geq \frac{-(3n - 400)}{\sqrt{0.09n + 1600}}\} \Leftrightarrow 0.1 \leq P_n$$

and

$$\{Z \geq 1.28\} \subseteq \{Z \geq \frac{-(3n-400)}{\sqrt{0.09n+1600}}\} \Leftrightarrow \frac{-(3n-400)}{\sqrt{0.09n+1600}} \leq 1.28$$

or

$$n \geq 117$$

Then there is at least 1 chance in 10 that structural damage will occur.

3. Normal approximation to Binomial

Suppose $X \sim \text{Bin}(n, p)$. Then X can be written as $X = X_1 + X_2 + \cdots + X_n$, where

$$X_i = \begin{cases} 1, & \text{if the } i\text{-th trial is success} \\ 0, & \text{otherwise} \end{cases}$$

Also $E(X_i) = p$ and $Var(X_i) = p(1-p), i = 1, 2, \ldots, n$. Therefore, form central limit theorem, the distribution of $\frac{(X_1+X_2+\cdots+X_n)-np}{\sqrt{np(1-p)}}$ approaches the standard normal distribution as $n \to \infty$, i.e., $\frac{X-np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$. Hence, X can be approximated with $N(np, np(1-p))$. In general, the normal approximation will be quite good for values of n satisfying $np(1-p) \geq 10$.

Continuity Correction: Since the Normal distribution (it can take all real numbers) is continuous while Binomial distribution is discrete (it can take positive integer values), we should use the integral for Normal distribution with introducing continuity correction so that the discrete integer x in Binomial becomes the interval $(x - 0.5, x + 0.5)$ in Normal.

Example 10. *A manufacturer makes computer chips of which* 10% *are defective. For a random sample of* 200 *chips, find the approximate probability that more than* 15 *are defective.*

Solution: Let X be the number of defective chips in the sample. Then $X \sim \text{Bin}(200, 0.1)$. therefore, $E(X) = np = 20$ and $Var(X) = np(1-p) = 18$. Then, $\frac{X-20}{\sqrt{18}}$ can be approximated with $Z = N(0,1)$. To allow the continuity correction, we need to calculate $P(X > 15.5)$. So,

$$P(X > 15.5) = P(\frac{X-20}{\sqrt{18}} > \frac{15.5-20}{\sqrt{18}}) = P(Z > -1.06) = P(Z < 1.06) = 0.86.$$

Note: This approximation can also view by using Stirling approximation formula, $n! \approx n^n e^{-n}\sqrt{2\pi n}$, for large n.

4. Normal approximation to Poisson

Suppose $X \sim \text{P}(\lambda)$. Then the m.g.f. of X is $M_X(t) = e^{-\lambda(1-e^t)}, \ \forall\, t \in \mathbb{R}$.

Now, let $Y = \frac{X-\lambda}{\sqrt{\lambda}}$. Then the m.g.f. of Y is $M_Y(t) = e^{-t\sqrt{\lambda}} M_X(\frac{t}{\sqrt{\lambda}})$. Therefore,

$$\lim_{\lambda \to \infty} M_Y(t) = e^{\frac{t^2}{2}}$$

This is the m.g.f. of $N(0,1)$. Hence, $\frac{X-\lambda}{\sqrt{\lambda}} \approx N(0,1)$ for large value of λ. In other words, $X \approx N(\lambda, \lambda)$ for large value of λ. If $\lambda \geq 10$, then the normal approximation will be quite good.

Example 11. *Suppose cars arrive at a parking lot at a rate of* 50 *per hour. Assume that the process is a Poisson with* $\lambda = 50$. *Compute the probability that in the next hour number of cars that arrive at this parking lot will be between* 54 *and* 62.

Solution: Let X be the number of cars that arrive at this parking lot. Then

$$P(54 \le X \le 62) = \sum_{x=54}^{62} \frac{e^{-50} 50^x}{x!}$$

Also, $\frac{X-50}{\sqrt{50}}$ can be approximated with $Z = N(0,1)$. To allow the continuity correction, we need to calculate $P(53.5 \le X \le 62.5)$. Now,

$$P(53.5 \le X \le 62.5) = P\left(\frac{53.5-50}{\sqrt{50}} \le \frac{X-50}{\sqrt{50}} \le \frac{62.5-50}{\sqrt{50}}\right) = \Phi(\frac{12.5}{\sqrt{50}}) - \Phi(\frac{3.5}{\sqrt{50}}) = 0.2717.$$