1 Probability

Definition. *The **sample space** corresponding to an experiment is the set of all possible outcomes.*

Definition. *A **probabililty measure** on Ω is a function P from subsets of Ω to real numbers satisfying*

1. $P(\Omega) = 1$
2. *If $A \subset \Omega$, then $P(A) \geq 0$*
3. *If $A_i \cap A_j = \emptyset$, i.e. mutally disjoint, then*

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Definition. ***Counting method** Suppose $\Omega = \{\omega_1, \omega_2, \cdots, \omega_N\}$ and $P(\omega_i) = p_i$. To find probability of event A, we simply add the probabilities of ω_i that constitute A. If $P(\omega_i) = \dfrac{1}{N}$. If A can occur in any of mutually exclusive ways, then $P(A) = \dfrac{n}{N}$ or*

$$P(A) = \frac{\textit{number of ways } A \textit{ can occur}}{\textit{total number of outcomes}}$$

Definition. ***Multiplication principle** If there are p experiments and the first has n_1 possible outcomes, the second $n_2, \cdots$, then there are a total of $n_1 \times n_2 \times \cdots \times n_p$ possible outcomes for p experiments*

Proof. By induction. When $p = 2$, The outcome for the two experiments with m and n outcomes can be expressed as an ordered pair (a_i, b_j). These pairs can be exhibited as entries of $m \times n$ rectangular array, in which the pair is in ith row and jth column. There are $m \times n$ entries in this array. Then Assume it is true for $p = k$, that is there are $n_1 \times_2 \times \cdots \times n_k$ possible outcomes for the first k experiments. Now we just apply the multiplication principle regarding the first k experiments and a single experiment and conclude that there are $(n_1 \times_2 \times \cdots \times n_k) \times n_{k+1}$ outcomes for the $k + 1$ experiment. $\square$

Definition. *A **permutation** is an ordered arrangement of objects.*

Proposition. *For a set of size n and a sample size r, there are n^r different ordered samples with replacement and $n(n-1)(n-2)\cdots(n-r+1)$ different ordered samples without replacement. Therefore, the number of ordering of n elements is $n(n-1)\cdots 1 = n!$*

Proposition. ***Combination*** *The number of unordered samples of r objects selected from n objects without replacement is $\binom{n}{r}$.* ***Binomial coefficient*** *occurs in expansion*

$$(a+b)^n = \sum_{k=0}^{\infty} \binom{n}{k} a^k b^{n-k}$$

In particular $2^n = \sum_{k=0}^{\infty} \binom{n}{k}$, which can be interpreted as the number of subsets of a set of n objects.

Example. Suppose n items in a lot and a sample size of r is taken. There are $\binom{n}{r}$ such samples. Suppose the lot contains k defectives. The probability that the sample contains exactly m defectives is modeled by

$$P(A) = \frac{\binom{k}{m}\binom{n-k}{r-m}s}{\binom{n}{r}}$$

Example. Capture/Recapture Method Estimate size of wildlife distribution. **Maximum likelihood** choose what value of n that makes the observed outcome most probable. The probability of the observed outcome as a function of n is called the **likelihood**. Assume there are n animals in the population, t animals are tagged. Then of the second sample of sizse m, r tagged animals are recaptured. We estimate n by the maximizer of the likelihood.

$$L_n = \frac{\binom{t}{r}\binom{n-t}{m-r}}{\binom{n}{m}}$$

Definition. *let A and B be two events with $P(B) \neq 0$. The* ***conditional probability*** *of A given B is defined to be*

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Theorem. ***Multiplication Law*** *Let A and B be events and assume $P(B) \neq 0$. Then*

$$P(A \cap B) = P(A|B)P(B)$$

Theorem. ***Law of Total Probability*** *Let $B_1, B_2, \cdots, B_n$ be such that $\cup_{i=1}^{n} B_i = \Omega$ and $B_i \cap B_j = \emptyset$ for $i \neq j$, with $P(B_i) > 0$ for all i. Then for any event A,*

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$$

Proof. B_i are mutually disjoint partition of sample space. To find the probability of event A, we sum the conditional probability of A given B_i, weighted by $P(B_i)$. Note,

$$\begin{aligned}
P(A) &= P(A \cap \Omega) \\
&= P(A \cap (\bigcup_{i=1}^{n} B_i)\,) \\
&= P(\bigcup_{i=1}^{n} (A \cap B_i)\,) \\
&= \sum_{i=1}^{n} P(A \cap B_i) && \text{(since } A \cap B_i \text{ are disjoint)} \\
&= \sum_{i=1}^{n} P(A|B_i)P(B_i)
\end{aligned}$$

□

Theorem. ***Bayes' Rule*** *Let A and $B_1, \cdots, B_n$ be events where B_i are disjoint, $\cup_{i=1}^{n} B_i = \Omega$, and $P(B_i) > 0$ for all i. Then*

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}$$

Definition. *A and B are said to be* ***independent*** *events if $P(A \cap B) = P(A)P(B)$. In general, $A_1, A_2, \cdots, A_n$ are* ***mutually independent*** *if for any subcollection $A_{i_1} \cap \cdots \cap A_{i_m}$,*

$$P(A_{i_1} \cap \cdots \cap A_{i_m}) = P(A_{i_1}) \cdots P(A_{i_m})$$

2 Random Variables

Definition. *Suppose $X : S \to A$ is a discrete random variable defined on a sample space S. Then the* ***probability mass function*** *$f_X : A \to [0, 1]$ is defined as*

$$f_X(x) = P(X = x)$$

such that $\sum_{x \in A} f_X(x) = 1$
In addition, the ***cumulative distribution function*** *is defined to be*

$$F(x) = P(X \leq x)$$

Definition. *The **density function** $f(x)$ of a continuous random variable is a piecewise continuous function such that $\int_{-\infty}^{\infty} f(x)dx = 1$. Then for any $a < b$, the probability that X faills in the interval (a, b) is the area under the density function*

$$P(a < X < b) = \int_a^b f(x)dx = F(b) - F(as)$$

*The **cumulative distribution function** is defined to be*

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$$

3 Joint Distribution

Definition. *If $X_1, \cdots, X_n$ are **jointly distributd random variables**, their joint cdf is*

$$F(x_1, x_2, \cdots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n)$$

Extrema and Order Statistics

Definition. ***Order statistic*** *In statistics, the k-th order statistic of a statistical sample is equal to its k-th-smallest value. Let $X_1, \cdots, X_n$ be independent samples with a common cdf F_X and density f_X. Then the first order and nth order statistics are*

$$X_{(1)} = V = min(X_1, \cdots, X_n) \qquad X_{(n)} = U = max(X_1, \cdots, X_n)$$

Note that $U \leq u$ iff $X_i \leq u$ for all i, thus

$$F_U(u) = P(U \leq u) = P(X_1 \leq u) \cdots P(X_n \leq u) = [F_X(u)]^n$$

$$f_U(u) = nf_X(u)[F_X(u)]^{n-1}$$

And note that $V \geq v$ iff $X_i \geq v$ for all i, thus

$$(1 - F_V(v)) = [1 - F_U(v)]^n \quad \Rightarrow \quad F_V(v) = 1 - [1 - F_U(v)]^n$$

$$f_V(v) = nf_X(v)[1 - F_X(v)]^{n-1}$$

Definition. *The density of $X_{(k)}$, the kth-order statistic, is*

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} f(x) F^{n-1}(x)[1 - F(x)]^{n-k}$$

4 Expected Value

Definition. *If X is a discrete random variable with frequency function $p(x)$, the expected value of X denoted by $E(X)$, is*

$$E(X) = \sum_i x_i p(x_i)$$

provided that $\sum_i |x_i| p(x_i) < \infty$. If the sum diverges, the expectation is undefined.

Remark. $E(X)$ is also referred to as the **mean** of X and is often denoted by μ or μ_x

Definition. *If X is a continuous random variable with density $f(x)$, then*

$$E(x) = \int_{-\infty}^{\infty} x f(x) dx$$

provided that $\int |x| f(x) dx < \infty$. If the integral diverges, the expectation is undefined.

Theorem. ***Markov's Inequality*** *If X is a random variable with $P(X \geq 0) = 1$ for which $E(X)$ exists, then $P(X \geq t) \leq \dfrac{E(X)}{t}$*

Proof. For discrete case,

$$\begin{aligned} E(X) &= \sum_x xp(x) \\ &= \sum_{x<t} xp(x) + \sum_{x \geq t} xp(x) \\ &\geq \sum_{x \geq t} xp(x) && \text{(because all the sums are nonnegative)} \\ &\geq \sum_{x \geq t} tp(x) = tP(X \geq t) \end{aligned}$$

□

1 4 Expected Value

Theorem. ***Expectation of functions of RV*** *Suppose that $Y = g(X)$*

1. *If X is discrete with frequency function $p(x)$, then*

$$E(Y) = \sum_x g(x) p(x)$$

provided that $\sum |g(x)| p(x) < \infty$

2. *If X is continuous with density function $f(x)$, then*

$$E(Y) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

provided that $\int |g(x)|f(x)dx < \infty$

Proof. Prove for the discrete case. By definition

$$E(Y) = \sum_i y_i p_Y(y_i)$$

Let A_i denote the set of x mapped to y_i by g; that is, $x \in A_i$ if $g(x) = y_i$. Then

$$p_Y(y_i) = \sum_{x \in A_i} p(x)$$

and

$$\begin{aligned} E(X) &= \sum_i \sum_{x \in A_i} p(x) \\ &= \sum_i \sum_{x \in A_i} y_i p(x) && \text{(Note that } \forall x \in A_i,\ g(x) = y_i) \\ &= \sum_i \sum_{x \in A_i} g(x)p(x) \\ &= \sum_x g(x)p(x) \end{aligned}$$

The last step follows because A_i are disjoint and every x belongs to some A_i □

Theorem. *Suppose that $X_1, \cdots, X_n$ are jointly distributed random variable and $Y = g(X_1, \cdots, X_n)$*

1. *If the X_i are discrete with frequency function $p(x_1, \cdots, x_n)$, then*

$$E(Y) = \sum_{x_1, \cdots, x_n} g(x_1, \cdots, g_n)p(x_1, \cdots, x_n)$$

provided that $\sum_{x_1, \cdots, x_n} |g(x_1, \cdots, g_n)|p(x_1, \cdots, x_n) < \infty$

2. *If X_i are continuous with joint density function $f(x_1, \cdots, x_n)$, then*

$$E(Y) = \int \int \cdots \int g(x_1, \cdots, x_n)f(x_1, \cdots, x_n)dx_1 \cdots dx_n$$

provided that the integral with $|g|$ in place of g converges.

Corollary. *If X and Y are independent, $E(XY) = E(X)E(Y)$*

Theorem. ***Expectation of linear combinations of RV*** *If $X_1, \cdots, X_n$ are jointly distributed random variables with expectation $E(X_i)$ and Y is a linear function of X_i, $Y = a + \sum_{i=1}^{n} b_i X_i$, then*

$$E(Y) = a + \sum_{i=1}^{n} b_i E(X_i)$$

Proof. Prove for continuous case in note p125 □

Definition. *If X is a random variable with expected value $E(X)$, the **variance** σ^2 of X is*

$$Var(X) = E\{[X - E(X)]^2\} = \begin{cases} \sum_i (x_i - \mu)^2 p(x_i) & \textbf{\textit{If } } X \textbf{\textit{ discrete}} \\ \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx & \textbf{\textit{If } } X \textbf{\textit{ continuous}} \end{cases}$$

*provided that the expectation exists. The **standard deviation** σ of X is the square root of the variance*

Theorem. *If $Var(X)$ exists and $Y = a + bX$, then $Var(Y) = b^2 Var(X)$*

Proof. Note $E(Y) = a + bE(X)$, hence

$$\begin{aligned} E[(Y - E(Y))^2] &= E\{[a + bX - a - bE(X)]^2\} \\ &= E\{b^2[X - E(X)]^2\} \\ &= b^2 E\{[X - E(X)]^2\} \\ &= b^2 Var(X) \end{aligned}$$

□

Theorem. *The variance of X, if exists, may be calculated with*

$$Var(X) = E(X^2) - [E(X)]^2$$

Theorem. ***Chebyshev's Inequality*** *Let X be a random variable with mean μ and variance σ^2. Then for any $t > 0$,*

$$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$$

Proof. Let $Y = (X - E(X))^2$ such that $E(Y) = E[(X - E(X))^2] = Var(X)$. Now we apply Markov's inequality to Y. Let $t > 0$ be arbitrary,

$$\begin{aligned} P(Y \geq t^2) &\leq \frac{E(Y)}{t^2} \\ P((X - E(X))^2 \geq t^2) &\leq \frac{Var(X)}{t^2} \\ P(|X - \mu| \geq t) &\leq \frac{\sigma^2}{t^2} \end{aligned}$$

□

Remark. The interpretation is that if σ^2 is very small, there is high probability that X will not deviate much from μ. For another interpretation we set $t = k\sigma$ so that

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

This holds for any random variable with any distribution provided the variance exists. However, the real bounds are often much narrower

Definition. *If X and Y are jointly distributed random variables with expectations μ_X and μ_Y, respectively, the* ***covariance*** *of X and Y is*

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y)$$

Remark. Covariance is a measure of joint variability. If X, Y both tends positive, covariance is positive. If they are of different signs, covariance is negative.

Theorem. *Suppose that $U = a + \sum_{i=1}^{n} b_i X_i$ and $V = c + \sum_{j=1}^{m} d_j Y_j$. Then*

$$Cov(U, V) = \sum_{i=1}^{n} \sum_{j=1}^{m} b_i d_j Cov(X_i, Y_j)$$

Remark. One application is since $Var(X) = Cov(X, X)$

$$Var(X + Y) = Cov(X + Y, X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

Corollary.

$$Var(a + \sum_{i=1}^{n} b_i X_i) = \sum_{i=1}^{n} \sum_{i=1}^{n} b_i b_j Cov(X_i, X_j)$$

Corollary. *If X_i are independent, then $Cov(X_i, X_j) = 0$ for $i \neq j$, we have*

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$$

Remark. Note $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$ is true whether or not X_i are independent.

Definition. *If X and Y are jointly distributed random variable and the variances and covariances of both X and Y exist and the variances are nonzero, then the correlation of X and Y, denoted by ρ, is*

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Remark. Correlation is a dimensionless quantity, and as a result, does not change even if X and Y are subject to linear transformation

Theorem. $1 \leq \rho \leq 1$. *Furthermore, $\rho = \pm 1$ if and only if $P(Y = a + bX) = 1$ for some constants a and b*

4.5 Moment-generating function

Definition. *The* ***moment-generating function*** *of a random variable X is $M(t) = E(e^{tX})$ if the expectation is defined,*

$$M(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} p(x) & X \text{ discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x) dx & X \text{ continuous} \end{cases}$$

Remark. The generating function using all moments $m(t)$ can also be defined by a series

$$M(t) = \sum_{k=0}^{\infty} E(X^k) \frac{t^k}{k!} = 1 + E(X)\frac{t}{1!} + E(X^2)\frac{t^2}{2!} + \cdots$$

The idea is that if all moments exist and $E(e^{tX})$ is defined then $M(t)$ completely characterizes the distribution of X

Proposition. *If the moment-generating function exists for t in an open interval containing zero, it uniquely determines the probability distribution*

$$M_X(t) = M_Y(t) \rightarrow F_X(t) = F_Y(t)$$

Remark. If two random variable have same mggf in an open interval containing zero, they have the same distribution

Proposition. *The* ***kth moment*** *of a random variable is $E(X^k)$ if the expectation exists. Specifically, if the moment-generating function exists in* ***an open interval containing zero****, then*

$$E(X^k) = M^{(k)}(0)$$

Remark. Moments can be extracted from the moment generating function using derivatives with respect to t

$$\frac{d^k}{dt^k}M(t) = \frac{d^k}{dt^k}E(e^{tX}) = E(X^k e^{tX})|_{t=0} = E(X^k)$$

Proposition. *If $X_1, \cdots, X_n$ are independent and $S = \sum X_i$ then*

$$M_S(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

Proof. For example, if X and Y are independent ($E(XY) = E(X) + E(Y)$) then

$$M_{X+Y}(t) = E(e^{X+Y}) = E(e^{tX}e^{tY}) = E(e^{tX})E(e^{tY}) = M_X(t)M_Y(t)$$

□

Remark. For example, if $X_1, \cdots, X_n$ are i.i.d.

1. $Bernoulli(p)$ then $\sum X_i \sim Binomial(n, p)$
2. $Geometric(p)$ then $\sum X_i \sim NegBin(n, p)$
3. $Exp(\lambda)$ then $\sum X_i \sim Gamma(n, \lambda)$
4. $Poisson(p)$ then $\sum X_i \sim Poisson(n\lambda)$
5. $Binomial(n_i, p)$ then $\sum X_i \sim Binomial(\sum n_i, p)$
6. $Normal(\mu_i, \sigma_i^2)$ then $\sum X_i \sim Normal(\sum n_i, \sum \sigma_i^2)$
7. $Gamma(\alpha_i, \lambda)$ then $\sum X_i \sim Gamma(\sum \alpha_i, \lambda)$

They are all easy to prove by using this property and the mgf of respective distribution

Proposition. *If X has mgf $M_X(t)$ and $Y = a + bX$, then Y has mgf $M_Y(t) = e^{at}M_X(bt)$*

Definition. *A list of moment generating functions*

1. $Poisson(\lambda)$

$$e^{\lambda(e^t-1)}$$

2. $Normal(\mu, \sigma^2)$

$$e^{t\mu+\frac{1}{2}\sigma^2t^2}$$

3. $Gamma(k, \theta)$

$$(1 - t\theta)^k$$

5 Limit Theorem

Theorem. ***Law of Large Numbers*** *Let $X_1, X_2, \cdots, X_i, \cdots$ be a sequence of independent and identically distributed variables with $E(X_i) = \mu$ and $Var(X_i) = \mu^2$. Let $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, for any $\epsilon > 0$,*

$$P(|\bar{X_n} - \mu| > \epsilon) \xrightarrow{n\to\infty} 0$$

Proof. Notice that X_n is a linear combination of X_i. We can find

$$E(X) = \frac{1}{n}\sum_{i=1}^{\infty} E(X_i) = \mu$$

And since X_i are independent

$$Var(\bar{X_n}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n}$$

By Chebyshev's inequality, which states that for some $\epsilon > 0$

$$P(|\bar{X_n} - \mu| > \epsilon) \leq \frac{Var(\bar{X_n})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \text{ as } n \to \infty$$

□

Remark. The theorem states that the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed. i.e.

$$\bar{X_n} \to \mu \text{ as } n \to \infty$$

Definition. *A sequence $\{X_i\}$ of random variables* ***converges in probability (weakly)*** *towards the random variable X if for all $\epsilon > 0$,*

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

Remark. Law of large number is a special case of convergence in probability

Definition. *The sequence X_n* ***almost surely (strongly)*** *towards X means that*

$$P(\lim_{n\to\infty} X_n = X) = 1$$

Remark. In other words, X_n converges to X strongly if for every $\epsilon > 0$, $|X_n - X| > \epsilon$ only a finitely many times with probability 1; that is, beyond some point in the sequence, the difference is always less than ϵ

Definition. ***Converges in distribution*** *Let $X_1, X_2, \cdots$ be a sequence of random variables with cumulative distribution functions $F_1, F_2, \cdots$ and let X be a random variable with distribution function F. We say X_n converges in distribution to X if*

$$\lim_{n\to\infty} F_n(x) = F(x)$$

for all $x \in \mathbb{R}$ at which F is continuous

Theorem. ***Continuity Theorem*** *Let F_n be a sequence of cumulative distribution functions with corresponding moment-generating function M_n. Let F be a cumulative distribution function with moment generating function M. If $M_n(t) \to M(t)$, i.e. $\lim_{n\to\infty} M_n(t) = M(t)$ for all t in an open interval containing zero, then $F_n(x) \to F(x)$ at all continuity points of F*

Remark. So to prove that X_n converges in distribution to X we prove $M_n(t) \to M(t)$ in open interval containing zero, which implies that $F_n(x) \to F(x)$. One example is that standardized Poisson Random Variable

$$Z_n = \frac{X_n - E(X_n)}{\sqrt{Var(X_n)}} \to N$$

where $\{X_n\}$ is a sequence of Poisson RV with increasing λ and N is RV with standard normal distribution.

Proof. The mgf of X_n is

$$M_{X_n}(t) = e^{\lambda_n(e^t-1)}$$

Then by property of mgf, we have

$$M_{Z_n}(t) = e^{-t\sqrt{\lambda_n}} M_{X_n}(\frac{1}{\sqrt{\lambda_n}}t)$$

By using power series expansion $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, we have

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty} exp(-t\sqrt{\lambda_n} - \lambda_n + \lambda + t\sqrt{\lambda} + \frac{t^2}{2!} + \frac{t^3}{\sqrt{\lambda_n}3!} + \cdots) = e^{\frac{t^2}{2}} = M_N(t)$$

where $M_N(t)$ is the mgf for standard normal. Hence by continuity theorem, we see that $F_{Z_n}(x) \to F_N(x)$, therefore Z_n converges in distribution to N $\square$

Definition. *A random variable is* ***standardized*** *by subtracting its expected value $E[X]$ and dividing the difference by standard deviation*

$$Z = \frac{X - E[X]}{\sqrt{Var(X)}}$$

The effect of standardization on expected value and variance

$$E(Z) = \frac{E(X) - \mu}{\sigma} = 0 \text{ and } Var(Z) = \frac{1}{\sigma^2} Var(X_n) = 1$$

Theorem. ***Central Limit Theorem*** *Let $X_1, X_2, \cdots$ be a sequence of independent random variables where $E(X_i) = \mu$ and variance $Var(X_i) = \sigma^2$ and the common distribution function F and moment generating function M defined in neighborhood of zero. Let*

$$S_n = \sum_{i=1}^{n} X_i$$

Then

$$\lim_{n \to \infty} P(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x) = \Phi(x) \text{ where } -\infty < x < \infty$$

where Φ is the standard normal cdf

Remark. Note the cdf $F(x)$ is defined to be $P(X \leq x)$; and that convergence in distribution is defined to be $\lim_{n \to \infty} F_n(x) \to F(x)$, then

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} \Phi(x)$$

One usage of CLT is to think of binomial random variable as the sum of independent Bernoulli random variable, whose distribution ccan be approximated by a normal distribution.

Theorem. ***Lindeberg-Levy Central Limit Theorem*** *Let $\{X_n\}$ be an independent and identically distributed sequence of random variables such that $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Let sample average be $\bar{X_n} = \frac{1}{n}\sum_{i=0}^{n} X_i$ Then as n approaches infinity, the random variable $\sqrt{n}(\bar{X_n} - \mu)$ converges in distribution to a normal $N(0, \sigma^2)$*

$$\sqrt{n}(\bar{X_n} - \mu) \xrightarrow{d} N(0, \sigma^2) \iff \frac{\bar{X_n} - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0, 1)$$

Remark. Note in this case the average is taken into consideration. Note the denominator is different in this case.