STANDARD DISTRIBUTIONS (DISCRETE) SUMMARY

DISTRIBUTION	Pmf	Mean	Variance	Mgf
Binomial distribution	$\binom{n}{x}p^xq^{n-x}$ $x=0,1,2,...n$	np	npq	$(q+pe^t)^n$
Point binomial/ Bernoulli distribution	p^xq^{1-x} $x=0,1$	p	pq	$q+pe^t$
Poisson distribution	$\frac{e^{-\lambda}\lambda^x}{x!}$ $x=0,1,2.....\infty$	λ	λ	$e^{\lambda(e^t-1)}$
Uniform distribution	$\frac{1}{n}$ $x=1,2,....n$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\frac{1}{n}\frac{e^t(e^{nt}-1)}{e^t-1}$
Geometric distribution	q^xp $x=0,1,2.....\infty$	$\frac{q}{p}$	$\frac{q}{p^2}$	$\frac{p}{1-qe^t}$
Negative binomial distribution	$\binom{x+r-1}{r}p^rq^x$ $x=0,1,2.....\infty$	$\frac{rq}{p}$	$\frac{rq}{p^2}$	--

Notes

- ***The number of successes out of n trials*** **of a Bernoulli trial (outcomes -- success and failure) follows a binomial distribution.**
- **For a binomial distribution** ***mean > variance .***
- ***The number of occurrences of an event in an interval of time*** **follows a poisson distribution.**
- **For a poisson distribution** ***mean = variance .***
- **Poisson distribution is also called** ***the distribution of rare events.***
- **BD tends to PD when n $\rightarrow\infty$ and p$\rightarrow$ 0**
- ***The number of failures preceding the first success*** **follows a geometric distribution.**
- ***The number of failures preceding the rth success*** **follows a negative binomial distribution.**

STANDARD DISTRIBUTIONS (CONTINUOUS) SUMMARY

DISTRIBUTION	Pdf	Mean	Variance	Mgf
Uniform/ Rectangular distribution	$\frac{1}{b-a}$, $a\le x\le b$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{t(b-a)}$
Exponential distribution	$\theta e^{-\theta x}$, $0<x<\infty$	$\frac{1}{\theta}$	$\frac{1}{\theta^2}$	$\left(1-\frac{t}{\theta}\right)^{-1}$
Gamma distribution	$\frac{m^p}{\Gamma p}x^{p-1}e^{-mx}$ $0<x<\infty$	$\frac{p}{m}$	$\frac{p}{m^2}$	$\left(1-\frac{t}{m}\right)^{-p}$
Normal distribution	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty<x<\infty$	μ	σ^2	$e^{\mu t+\frac{1}{2}t^2\sigma^2}$
Beta distribution of first kind	$\frac{1}{\beta(m,n)}x^{m-1}(1-x)^{n-1}$, $0<x<1$	-	-	-
Beta distribution of second kind	$\frac{1}{\beta(m,n)}\frac{x^{m-1}}{(1-x)^{m+n}}$ $0<x<\infty$	-	-	-
Lognormal distribution	$\frac{1}{x\sigma\sqrt{2\pi}}e^{-\frac{(\log x-\mu)^2}{2\sigma^2}}$ $0<x<\infty$	-	-	-
Pareto distribution	$\frac{\theta}{x_0}\left(\frac{x_0}{x}\right)^{\theta-1}$ $0<x<\infty$	-	-	*Mgf does not exist*
Cauchy distribution	$\frac{1}{\pi\beta\{1+\left[\frac{x-\alpha}{\beta}\right]^2\}}$, $-\infty<x<\infty$	*Mean does not exist*	*variance does not exist*	*Mgf does not exist*

Note :

- **Exponential distribution is also called** ***waiting time distribution.***
- **Exponential distribution is the continuous analogue of geometric distribution,**
- **Gamma distribution is the continuous analogue of negative binomial distribution,**
- **Sum of independent exponential random variables follows gamma distribution.**
- **BD tends to ND when n$\rightarrow\infty$ and p is not so small.**
- **If** ***logX*** **follows** ***normal*** **distribution then** ***X*** **follows** ***lognormal*** **distribution.**
- **Lognormal distribution is used for studying** ***the inequalities of income*** **distribution.**
- **Pareto distribution is used to study the distribution of** ***incomes if it exceeds certain limit x_0.***

MODULE 2 LIMIT THEOREMS

1. **Tchebychef's Inequality**
   Let X be a random variable with mean μ and variance σ², then for a positive real number t
   P{ |X-μ| ≥ tσ } ≤ $\frac{1}{t^2}$ or P{ |X-μ| ≤ tσ } ≥ 1 − $\frac{1}{t^2}$

2. **Convergence in probability**
   A sequence of random variables $\{X_n\}$ is said to converge in probability to a constant **a** if
   P{$|X_n-a|<\varepsilon$}$\rightarrow$1 as n $\rightarrow\infty$ Or P{$|X_n-a|>\varepsilon$}$\rightarrow$0 as n$\rightarrow\infty$

3. **Convergence in distribution**
   Let $\{X_n\}$ be a sequence of random variables having cdf $F_n(x)$. Then this sequence $\{X_n\}$ of random variables is said to converge in distribution to a random variable X having cdf F(x) if
   $F_n(x)\rightarrow$ F(x) as n $\rightarrow\infty$

4. **Weak law of large numbers**
   Let $\{X_n\}$ be a sequence of independent random variables with $E(X_i)$ = μ and $V(X_i)$ = σ² , i = 1,2,3.....Define $X_n=\frac{\Sigma X_i}{n}$. Then for any ε >0, however small, P{$|X_n-\mu|>\varepsilon$}$\rightarrow$0 as n$\rightarrow\infty$

5. **Bernoulli's law of large numbers**
   Consider n repetitions of a Bernoulli trial with probability 'p' of success in each trial. Let X_n denote the number of successes out of n trials . Then for any ε > 0, however small ,
   P{$|\frac{X_n}{n}-p|>\varepsilon$}$\rightarrow$0 as n$\rightarrow\infty$

6. **Central limit theorem**
   Central limit theorem states that the sum of a very large number of random variables is approximately normally distributed with mean equal to sum of the means and variance equal to sum of the variances provided the random variables follow some very general conditions .

7. **Lindberg Levy central limit theorem**
   Let $\{X_n\}$ be a sequence of independent and identically distributed (iid) random variables with $E(X_i)$ = μ and $V(X_i)$ = σ² , i = 1,2,3... and let $S_n=X_1+X_2+\cdots+X_n$
   Then S_n will be normally distributed with mean = nμ and variance = nσ². (SD = √nσ) .
   That is $S_n\rightarrow$ N(nμ , √nσ)

MODULE 3 SAMPLING METHODS

Simple random sampling

Simple random sampling is a type of probability sampling in which each unit of the population has equal chance of being selected in to the sample. The sample obtained using simple random sampling is called simple random sample.

Simple random sampling is classified in to two – SRSWR and SRSWOR.

Systematic random sampling

If we can arrange the items of the population in a definite order, say alphabetical, chronological ,geographical etc this method can be used. Once the items of the population are arranged in some order, assign them numbers 1,2,3.... Divide them in to a number of groups which is equal to the required sample size. From the first group choose an item at random using lottery method. If we select the 4th item , choose the 4th item of every group systematically. This will constitute a systematic sample.

Stratified random sampling

When the population is heterogeneous with respect to some characteristic, but can be divided into a number of homogeneous subgroups this method can be applied. Here we divide the population in to different homogeneous sub groups known as 'strata' and from each stratum select the items proportionally using lottery method. This will constitute a stratified sample.

Cluster sampling

This type of sampling can be applied when the population is spread over a large geographical area. In this method of sampling the population is first divided into different subgroups called clusters based on some common characteristics. Then a random selection of clusters is made using simple random sampling method. Data is then collected from every item in the selected clusters.

MODULE 4 SAMPLING DISTRIBUTIONS

1. **Chi square distribution**
   Pdf $f(\chi^2)=\frac{(\frac{1}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}(\chi^2)^{\frac{n}{2}-1}e^{-\frac{\chi^2}{2}}$, $\chi^2>0$

   Mean, $E(\chi^2)$ = n
   Variance, $V(\chi^2)$ = 2n
   Mgf , Mx(t) = $(1-2t)^{-\frac{n}{2}}$

2. **Student's t distribution**
   Pdf , $f(t)=\frac{1}{\sqrt{n}\beta(\frac{1}{2},\frac{n}{2})}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}$, $-\infty<t<\infty$

3. **Snedcor's F distribution**
   Pdf , $f(F)=\frac{n_1^{\frac{n_1}{2}}n_2^{\frac{n_2}{2}}}{\beta(\frac{n_1}{2},\frac{n_2}{2})}\frac{F^{\frac{n_1}{2}-1}}{(n_1F+n_2)^{\frac{n+n_2}{2}}}$, F > 0

4. **Relation between normal, chi square,t and F distributions**
   (i) Let Z $\rightarrow$ N(0,1) ,then $Z^2\rightarrow\chi^2(1)$
   (ii) Let $Z_1^2\rightarrow\chi^2(1)$, $Z_2^2\rightarrow\chi^2(1)$, ... $Z_n^2\rightarrow\chi^2(1)$
   then $Z_1^2+Z_2^2+....+Z_n^2\rightarrow\chi^2(n)$
   (iii) Let Z $\rightarrow$ N(0,1) , Standard normal distribution and $\chi^2\rightarrow\chi^2(n)$, chi square distribution with n df Then $t=\frac{Z}{\sqrt{(\chi^2/n)}}\rightarrow t(n)$
   (iv) Let $\chi_1^2\rightarrow\chi^2(n_1)$ and $\chi_2^2\rightarrow\chi^2(n_2)$ then the random variable
   $F=\frac{\chi_1^2/n_1}{\chi_2^2/n_2}\rightarrow F(n_1, n_2)$.
   (v) Let Z $\rightarrow$ N(0,1) and $\chi^2\rightarrow\chi^2(n)$ Then $t=\frac{Z}{\sqrt{(\chi^2/n)}}\rightarrow t(n)$
   Squaring $t^2=\frac{Z^2}{\chi^2/n}=\frac{Z^2/1}{\chi^2/n}=\frac{\chi^2(1)/1}{\chi^2/n}$ = F(1,n)

5. **Population:** By a statistical population we mean a set of numerical observations which can be treated as admissible values of random variable, where the random variable represents the characteristic under study.

6. **Sample:** Sample is the representative team subset of the population

7. **Parameter :** Any function of population observation is called parameter

8. **Statistic :** Any function of sample observations is called a statistic.

9. **Sampling distribution :** Sampling distribution is the distribution of the statistic calculated from the random sample taken from the population.

Sampling distribution of sample mean

When samples are taken from normal population N(μ, σ) , σ known

$\bar{X}\rightarrow$ N(μ, σ/√n) or $Z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\rightarrow$N(0,1)

When samples are taken from normal population N(μ, σ) , σ unknown , n large

$\bar{X}\rightarrow$ N(μ, s/√n) or $Z=\frac{\bar{x}-\mu}{s/\sqrt{n}}\rightarrow$N(0,1)

When samples are taken from normal population N(μ, σ) , σ unknown , n small

$t=\frac{\bar{x}-\mu}{s/\sqrt{(n-1)}}\rightarrow$t(n-1)

When large samples are taken from any population σ known

$\bar{X}\rightarrow$ N(μ, σ/√n) or $Z=\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\rightarrow$N(0,1)

When large samples are taken from any population σ unknown

$\bar{X}\rightarrow$ N(μ, s/√n) or $Z=\frac{\bar{x}-\mu}{s/\sqrt{n}}\rightarrow$N(0,1)

Sampling distribution of sample variance

When samples are taken from normal population N(μ, σ)

$\chi^2=\frac{ns^2}{\sigma^2}\rightarrow\chi^2(n-1)$