PROBABILITY DISTRIBUTION AND SAMPLING THEORY

III SEMESTER

2019 Admission Onwards

Complementary Course (STA3 C03)

B Sc MATHEMATICS

UNIVERSITY OF CALICUT
School of Distance Education
Calicut University- P.O,
Malappuram - 673635, Kerala.

19559

UNIVERSITY OF CALICUT

School of Distance Education

Study Material

III SEMESTER

(2019 Admission Onwards)

Complementary Course (STA3 C03)

B Sc MATHEMATICS

PROBABILITY DISTRIBUTION AND SAMPLING THEORY

Prepared by:

Smt. Aparna Aravindakshan.M
Assistant Professor,
Department of Statistics,
St. Joseph's College,
Devagiri, Kozhikode.

Scrutinized by:

Dr. Rajasekharan. K.E,
Assistant Professor,
EMEA College of Arts & Science,
Kondotty.

CONTENTS

3 Sampling Methods 155

4 Sampling Distributions 187

MODULE

ONE

Standard Probability Distributions

A theoretical probability distribution gives an idea about how probability is distributed among the possible values of a random variable (r.v.). It gives us a mathematical expression according to which different values of the random variable are distributed with specified probabilities. Here, we discuss some standard probability distributions that we may often come across. They are of both discrete and continuous type.

1.1 Discrete Probability Distributions

1.1.1 Bernoulli Distribution

Random experiments having exactly two mutually exclusive outcomes are called dichotomous experiments or Bernoulli trials.

Example 1.1.1. *When a coin is tossed, the two possible outcomes are head (H) and tail (T).*

Even if an experiment have more than two mutually exclusive outcomes, we can consider it as a dichotomous experiment. See the following example.

Example 1.1.2. *When a die is thrown, the set of all mutually exclusive outcomes is* $\Omega = \{1, 2, 3, 4, 5, 6\}$*. If we consider getting faces 3 or 5 as an event, say* A*, and its complement in* Ω *as the other event, the experiment reduces to a dichotomous experiment. Here,* $A = \{3, 5\}$ *and* $A' = \{1, 2, 4, 6\}$*.*

In fact all non-trivial experiments can be viewed as Bernoulli trials or dichotomous experiments. The event in which we are interested is labeled as success (S) having probability $P(S) = p$, $0 < p < 1$ and its complementary event as failure (F) with probability $P(F) = 1 - p = q$. If for such an experiment, a random variable X is defined such that it takes value 1 when success occurs and 0 when failure occurs, then X follows a Bernoulli distribution. Hence, Bernoulli distribution, is the discrete probability distribution of a random variable which takes only two values 1 and 0 with respective probabilities p and $1-p$. This distribution is named after Swiss mathematician James Bernoulli (1654-1705).

Definition 1.1.1. *A random variable X is said to follow Bernoulli distribution with parameter p if its probability mass function (p.m.f.) is given by,*

$$f_X(x) = \begin{cases} p^x q^{1-x} & ; \quad x = 0, 1 \\ 0 & ; \quad \textit{otherwise}, \end{cases}$$

where $0 < p < 1$ *and* $p + q = 1$.

In Example 1.1.2, if the die is fair and we are interested in the occurrence of event A, then $p = \frac{1}{3}$.

1.1.2 Binomial Distribution

Binomial distribution is a discrete probability distribution of the number of successes in a sequence of n independent trials of a random experiment having two mutually exclusive outcomes

(Bernoulli trials).

Suppose that a Bernoulli trial is repeated n times keeping the probability 'p' of success constant through out the trials and the trials being independent. We are interested in finding the probability distribution of X, the number of successes in n trials. Hence, the possible values of X depends on the number of independent Bernoulli trials and the random variable X can take the values $x = 0, 1, 2, \ldots, n$. What we require is the probability distribution of X, the probability of getting x successes in n trials for $x = 0, 1, 2, \ldots, n$.

In n trials if we are getting x successes, then there will be $n - x$ failures. Since the trials are independent and p is same in all trials, probability of getting x successes is $p \times p \times \ldots \times p$ (x times) $= p^x$ and probability of getting $n - x$ failures is $q \times q \times \ldots \times q$ ($n - x$ times) $= q^{n-x}$. Hence, the probability of getting x successes and $n - x$ failures is $p^x q^{n-x}$. The number of ways in which x successes can occur in n trials is $\frac{n!}{x!(n-x)!} = \binom{n}{x}$. For example,

- A success in two trials can happen in the following $\binom{2}{1} = 2$ ways:
  SF → a success in the first trial and a failure in the second trial

or
FS → a failure in the first trial and a success in the second trial.

- Two successes in three trials can happen in the following $\binom{3}{2} = 3$ ways:
  SSF → a success in the first trial, a success in the second trial and a failure in the third trial
  or
  SFS → a success in the first trial, a failure in the second trial and a success in the third trial
  or
  FSS → a failure in the first trial, a success in the second trial and a success in the third trial.

Hence, the probability of getting x successes in n trials in any order is given by, $\binom{n}{x}p^x q^{n-x}$. This probability distribution of the random variable X is called Binomial distribution denoted by $X \sim B(n, p)$. This distribution has been discovered by James Bernoulli and the name arises from the fact that the probabilities of $x = 0, 1, \ldots, n$ are successive terms of the binomial expansion $(q + p)^n$.

Definition 1.1.2. *A random variable X is said to follow binomial distribution with parameters n and p if its p.m.f. is given* *by,*

$$f_X(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & ; \quad x = 0, 1, 2, \ldots, n \\ 0 & ; \quad \textit{otherwise}, \end{cases}$$

where $0 < p < 1$ *and* $p + q = 1$.

Remark 1.1.1. *If* $n = 1$*, the binomial random variable reduces to Bernoulli random variable, denoted by* $B(1, p)$.

Remark 1.1.2. *If* $X \sim B(n, p)$*, then* $\sum_{x=0}^{n} f_X(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (q + p)^n = 1$.

Remark 1.1.3. *If* n *Bernoulli trials constitute an experiment and if this experiment is repeated* N *times, then the frequency function or the expected frequency of the Binomial distribution is given by* $N \times P(X = x)$.

Remark 1.1.4. *Let* $X \sim B(n, p)$*. Then* X *gives number of success in* n *independent trials with probability* p *for success in each trial. Note that* $n - X$ *gives number of failures in* n *independent trials with probability* $1 - p = q$ *for failure in each trial. Therefore,* $n - X \sim B(n, q)$.

Moments

Mean

$$\begin{aligned}
\mu_1' = E(X) &= \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x q^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} \\
&= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1} q^{n-x} \\
&= np(q+p)^{n-1} \\
&= np.
\end{aligned}$$

Variance

$$V(X) = \mu_2' - (\mu_1')^2 = E(X^2) - [E(X)]^2$$

$$
\begin{aligned}
\mu_2' = E(X^2) &= \sum_{x=0}^{n} x^2 \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} [x(x-1)+x] \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} x(x-1) \binom{n}{x} p^x q^{n-x} \\
&\qquad + \sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=2}^{n} x(x-1) \frac{n!}{x!(n-x)!} p^x q^{n-x} + E(X) \\
&= n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{(x-2)!(n-x)!} p^{x-2} q^{n-x} \\
&\qquad + np \\
&= n(n-1)p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} q^{n-x} + np \\
&= n(n-1)p^2 + np.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
V(X) &= n(n-1)p^2 + np - (np)^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np - np^2 \\
&= np(1-p) \\
&= npq.
\end{aligned}
$$

The third raw moment,

$$\begin{aligned}
\mu'_3 = E(X^3) &= \sum_{x=0}^{n} x^3 \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} [x(x-1)(x-2) + 3x(x-1) + x] \binom{n}{x} p^x q^{n-x} \\
&= n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mu_3 &= \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3 \\
&= npq(1-2p) \\
&= npq(q-p).
\end{aligned}$$

The fourth raw moment,

$$\begin{aligned}
\mu'_4 = E(X^4) &= \sum_{x=0}^{n} x^3 \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} [x(x-1)(x-2)(x-3) + 6x(x-1)(x-2) \\
&\quad +7x(x-1) + x] \binom{n}{x} p^x q^{n-x} \\
&= n(n-1)(n-2)(n-3)p^4 + 6n(n-1)(n-2)p^3 \\
&\quad +7n(n-1)p^2 + np.
\end{aligned}$$

Therefore,

$$\begin{aligned}\mu_4 &= \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 \\ &= 3n^2p^2q^2 + npq(1-6pq).\end{aligned}$$

Beta and Gamma coefficients

Skewness

$$\begin{aligned}\beta_1 &= \frac{\mu_3^2}{\mu_2^3} \\ &= \frac{[npq(q-p)]^2}{(npq)^3} \\ &= \frac{(q-p)^2}{npq}.\end{aligned}$$

Therefore,

$$\gamma_1 = \sqrt{\beta}_1 = \frac{q-p}{\sqrt{npq}}.$$

Hence, binomial distribution is:

1. Positively skewed if $\gamma_1 > 0$; i.e., $q > p$.
2. Symmetric if $\gamma_1 = 0$, i.e.; $q = p$.

3. Negatively skewed if $\gamma_1 < 0$; i.e., $q < p$.

Kurtosis

$$\begin{aligned}\beta_2 &= \frac{\mu_4}{\mu_2^2} \\ &= \frac{3n^2p^2q^2 + npq(1-6pq)}{(npq)^2} \\ &= 3 + \frac{1-6pq}{npq}.\end{aligned}$$

Therefore,

$$\gamma_2 = \beta_2 - 3 = \frac{1-6pq}{npq}.$$

Hence, binomial distribution is:

1. Leptokurtic if $\beta_2 > 3$; i.e., $\gamma_2 > 0$; i.e., $pq < \frac{1}{6}$.

2. Mesokurtic if $\beta_2 = 3$; i.e., $\gamma_2 = 0$; i.e., $pq = \frac{1}{6}$.

3. Platykurtic if $\beta_2 < 3$; i.e., $\gamma_2 < 0$; i.e., $pq > \frac{1}{6}$.

Moment Generating Function

$$\begin{aligned} M_X(t) &= E\left(e^{tX}\right) \\ &= \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x q^{n-x} \\ &= \sum_{x=0}^{n}\binom{n}{x}(pe^t)^x q^{n-x} \\ &= (q+pe^t)^n. \end{aligned}$$

Additive/Reproductive Property

Theorem 1.1.1. *If $X \sim B(n,p)$, $Y \sim B(m,p)$ and X and Y are independent, then $X+Y \sim B(n+m,p)$.*

Proof. $X \sim B(n,p)$ and $Y \sim B(m,p)$ implies $M_X(t) = (q+pe^t)^n$ and $M_Y(t) = (q+pe^t)^m$ respectively. Since X and Y are independent,

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \times M_Y(t) \\ &= (q+pe^t)^n \times (q+pe^t)^m \\ &= (q+pe^t)^{n+m}, \end{aligned}$$

which is the m.g.f. of $B(n+m,p)$. □

Remark 1.1.5. *If the second parameter is not the same $X+Y$ will not be binomial.*

Remark 1.1.6. *The result can be generalised to k independent binomial random variables having common probability p.*

Remark 1.1.7. *The sum of n independent Bernoulli random variable with parameter p follows $B(n,p)$.*

Remark 1.1.8. *If $X \sim B(n,p)$, $Y \sim B(m,p)$ and X and Y are independent, then the conditional distribution of $X|X+Y=k$ is hypergeometric and $P(X|X+Y=k) = \frac{\binom{n}{x}\binom{m}{k-x}}{\binom{n+m}{k}}$.*

Recurrence Relation for Binomial Central Moments

Theorem 1.1.2. *When $X \sim B(n,p)$, $\mu_{r+1} = pq\left[nr\mu_{r-1} + \frac{d\mu_r}{dp}\right]$.*

Proof.

$$\begin{aligned}
\mu_r &= E[X-E(X)]^r \\
&= E[X-np]^r \\
&= \sum_{x=0}^{n}(x-np)^r\binom{n}{x}p^x q^{n-x}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{d\mu_r}{dp} &= \frac{d}{dp}\left[\sum_{x=0}^{n}(x-np)^r\binom{n}{x}p^x q^{n-x}\right] \\
&= \sum_{x=0}^{n}\frac{d}{dp}\left[(x-np)^r\binom{n}{x}p^x(1-p)^{n-x}\right] \\
&= \sum_{x=0}^{n}\left[r(x-np)^{r-1}(-n)\binom{n}{x}p^x(1-p)^{n-x}\right. \\
&\qquad +(x-np)^r\binom{n}{x}(1-p)^{n-x}xp^{x-1} \\
&\qquad \left. +\,(x-np)^r\binom{n}{x}p^x(n-x)(1-p)^{n-x-1}(-1)\right] \\
&= -nr\sum_{x=0}^{n}\left[(x-np)^{r-1}\binom{n}{x}p^x q^{n-x}\right] \\
&\qquad +\sum_{x=0}^{n}\left[(x-np)^r\binom{n}{x}p^x q^{n-x}\left(\frac{x}{p}\right)\right] \\
&\qquad +\sum_{x=0}^{n}\left[(x-np)^r\binom{n}{x}p^x q^{n-x}\left(-\frac{n-x}{1-p}\right)\right] \\
&= -nr\mu_{r-1}+\sum_{x=0}^{n}\left[(x-np)^r\binom{n}{x}p^x q^{n-x}\left(\frac{x}{p}-\frac{n-x}{1-p}\right)\right] \\
&= -nr\mu_{r-1}+\sum_{x=0}^{n}\left[(x-np)^r\binom{n}{x}p^x q^{n-x}\left(\frac{x-np}{pq}\right)\right] \\
&= -nr\mu_{r-1}+\frac{1}{pq}\sum_{x=0}^{n}\left[(x-np)^{r+1}\binom{n}{x}p^x q^{n-x}\right] \\
&= -nr\mu_{r-1}+\frac{1}{pq}\mu_{r+1}
\end{aligned}$$

Therefore,

$$\mu_{r+1} = pq\left[nr\mu_{r-1} + \frac{d\mu_r}{dp}\right].$$

□

Using the information $\mu_0 = 1$ and $\mu_1 = 0$, we can determine μ_2, μ_3 and μ_4 successively.

Recurrence Relation for Binomial Probabilities

Theorem 1.1.3. *When* $X \sim B(n,p)$, $f_X(x+1) = \frac{n-x}{x+1}\,\frac{p}{q}\,f_X(x)$.

Proof. We have,

$$\begin{aligned}
f_X(x) &= \binom{n}{x}p^x q^{n-x} \\
f_X(x+1) &= \binom{n}{x+1}p^{x+1}q^{n-x-1} \\
\frac{f_X(x+1)}{f_X(x)} &= \frac{\binom{n}{x+1}p^{x+1}q^{n-x-1}}{\binom{n}{x}p^x q^{n-x}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\frac{n!}{(x+1)!(n-x-1)!}p^{x+1}q^{n-x-1}}{\frac{n!}{x!(n-x)!}p^{x}q^{n-x}} \\
&= \frac{n-x}{x+1}\,\frac{p}{q} \\
f_X(x+1) &= \frac{n-x}{x+1}\,\frac{p}{q}\,f_X(x)
\end{aligned}
$$

□

Mode of Binomial Distribution

Mode is the value of the random variable for which p.m.f. is maximum. The maximum of $f_X(x)$ is attained if

1. $f_X(x) \geq f_X(x+1) \Rightarrow np + p - 1 \leq x$
2. $f_X(x) \geq f_X(x-1) \Rightarrow x \leq np + p$.

i.e., $np + p - 1 \leq x \leq np + p$. Therefore, mode is $[(n+1)p]$, the integer part of $(n+1)p$, when $(n+1)p$ is not an integer. But when $(n+1)p$ is an integer, there will be two modes at $(n+1)p - 1$ and $(n+1)p$.

Fitting of Binomial Distribution

Fitting of binomial distribution means to obtain the expected or theoretical binomial frequencies against the given observed data. It is obtained by multiplying the corresponding probabilities with the total frequency. If E_x denote the expected frequency that the random variable X takes the value x, then

$$\begin{aligned} E_x &= N \times f_X(x) \\ &= N \times \binom{n}{x} p^x q^{n-x}; \quad x = 0, 1, 2, \ldots, n;\ 0 < p < 1 \text{ and } p + q = 1. \end{aligned}$$

n can be determined by the values of x in the data and if p is not given it can be obtained by equating mean of the data with np.

Solved Problems

1. Ten coins are tossed simultaneously. Find the probability of getting at least seven heads.
   Solution:
   $p =$ Probability of getting a head $= \frac{1}{2}$
   $q =$ Probability of getting a tail $= \frac{1}{2}$
   The probability of getting x heads in 10 throws of the coin

is

$$\begin{aligned} P(X=x) &= \binom{10}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{10-x} \\ &= \binom{10}{x}\left(\frac{1}{2}\right)^{10}; \quad x=0,1,2,\ldots,10 \end{aligned}$$

Probability of getting at least 7 heads is given by

$$\begin{aligned} P(X\geq 7) &= P(X=7)+P(X=8) \\ &\qquad +P(X=9)+P(X=10) \\ &= \binom{10}{7}\left(\frac{1}{2}\right)^{10}+\binom{10}{8}\left(\frac{1}{2}\right)^{10} \\ &\qquad +\binom{10}{9}\left(\frac{1}{2}\right)^{10}+\binom{10}{10}\left(\frac{1}{2}\right)^{10} \\ &= \left(\frac{1}{2}\right)^{10}\left[\binom{10}{7}+\binom{10}{8}+\binom{10}{9}+\binom{10}{10}\right] \\ &= \frac{120+45+10+1}{1024} \\ &= \frac{176}{1024} \end{aligned}$$

2. The probability that a batsman scores a century in a cricket match is $\frac{1}{4}$. What is the probability that in 6 matches he will score century in exactly 3 matches?
   Solution:
   Here, $X \sim B(6, \frac{1}{4})$. The required probability is $P(X = 3)$.

$$\begin{aligned} P(X=3) &= \binom{6}{3}\left(\frac{1}{4}\right)^3\left(\frac{3}{4}\right)^3 \\ &= 20\left(\frac{3}{16}\right)^3 \\ &= 0.1318 \end{aligned}$$

Exercises

1. Establish the relation concerning Bernoulli and binomial random variables.

2. Ten coins are tossed simultaneously. Find the probability of getting at most three heads.

3. A and B plays a game in which their chances of winning are in the ratio 3:2. Find A's chance of winning at least 3 games out of 5 games played.

4. In a binomial distribution consisting of 5 independent trials, probabilities of 1 and 2 successes are 0.4096 and 0.2048 respectively. Find the parameter 'p' of the distribution.

5. An irregular six faced die is thrown and the expectation that in 10 throws it will give 5 even numbers is twice that of giving 4 even numbers. How many times in 10,000 sets of 10 throws each would you expect it to give no even number.

6. With the usual notations, find 'p' for the binomial variate X, if $n = 6$ and $9P(X = 4) = P(X = 2)$.

7. Mean and variance of binomial distribution are 2.5 and 1.875 respectively. Obtain the binomial probability distribution.

8. If the mean and variance of a binomial distribution are 4 and 2 respectively. Find the probability of

   (a) exactly two successes

   (b) less than two successes

   (c) More than 6 successes

   (d) at least two successes

9. Given the m.g.f of binomial variable $M_X(t) = \left(\frac{1}{3}\right)^5 (2 + e^t)^5$, obtain the mean and variance.

10. Comment on the statement " The mean of a binomial distribution is 3 and variance is 4".

11. Derive the mean and variance of binomial distribution from its m.g.f.

12. In litter of 4 mice the number of litters which contained 0,1,2,3,4 females were noted. The figures are given in the table below.

No. of female mice	0	1	2	3	4	Total
No. of litters	8	29	34	27	7	105

Fit a Binomial distribution to the above data assuming an equal chance for a male or female birth.

13. Four coins are tossed 80 times. The distribution of number of heads is given below.

No. of heads	0	1	2	3	4
Frequency	4	18	32	20	6

Estimate the probability of getting a head and obtain the expected frequencies.

1.1.3 Poisson Distribution

Poisson Distribution is a discrete probability distribution introduced by Simon D. Poisson in 1837. He approached the distribution by considering the limit of a binomial distribution in which n tends to infinity, p tends to zero and np remains finite and equal to λ. There is no need for λ $(= np)$ to be small. It is the largeness of n and the smallness of p that are important.

Another approach to Poisson distribution is as follows: for events occurring independently and randomly, the number of occurrences of an event in a given interval of time or unit of space is given by Poisson distribution. That is, if the occurrence of any one event does not give any information about the occurrence of another event and the probability of occurrence of an event in a given interval of time (or unit of space) does not change through time (or unit of space). ie, the rate of occurrence of the event does not change, then the random variable X, the number of occurrences of events in a fixed time interval, is said to follow Poisson distribution. For example

1. Number of major road accidents on a day in a given city.
2. Number of defective items coming out of a production line in an hour.

3. Number of teak trees per square kilometer.

Hence, the Poisson distribution may be derived in two ways,

1. Poisson distribution as a limiting form of binomial distribution in which n tends to infinity, p tends to zero and np remains finite and equal to λ.
2. For events distributed randomly and independently of one another in time (or space), the distribution of the number of events occurring in fixed time interval (or space) is Poisson.

Definition 1.1.3. *A discrete random variable X is defined to have Poisson distribution if the p.m.f. of X is given by*

$$f_X(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & ; \quad x = 0, 1, 2, \ldots \\ 0 & ; \quad \textit{otherwise}, \end{cases}$$

where $\lambda > 0$ is the parameter of the Poisson distribution.

In this case we can write $X \sim P(\lambda)$.

In real life situations, events having a very small probability of occurrence (i.e., rare events) are modeled using Poisson distribution.

Poisson distribution as a limiting form of Binomial Distribution

The Poisson distribution is obtained as an approximation to the binomial distribution under the conditions:

i) n is very large $(n \to \infty)$

ii) p is very small $(p \to 0)$

iii) $np = \lambda$, a finite quantity.

Proof: For binomial distribution,

$$f(x) = \tbinom{n}{x} p^x q^{n-x} \quad ; \quad x = 0, 1, 2, \ldots, n$$

where $0 < p < 1$ and $p + q = 1$. Now,

$$\begin{aligned} f(x) &= \binom{n}{x} p^x q^{n-x} \\ &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= \frac{n(n-1)(n-2)\ldots(n-\overline{x-1})}{x!} p^x (1-p)^{n-x} \\ &= \frac{n^x(1-\frac{1}{n})(1-\frac{2}{n})\ldots(1-\frac{x-1}{n}) p^x (1-p)^n}{x!(1-p)^x} \qquad (1.1.1) \end{aligned}$$

Now,

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{x-1}{n}\right)=1.$$

Also,

$$np=\lambda\Rightarrow p=\frac{\lambda}{n}.$$

Therefore,

$$\begin{aligned}\lim_{n\to\infty}(1-p)^x &= \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^x=1\\ \lim_{n\to\infty}(1-p)^n &= \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n=e^{-\lambda}.\end{aligned}$$

Applying these limits in equation (1.1.1), we get

$$f(x)=\frac{e^{-\lambda}\lambda^x}{x!};\quad x=0,1,2,\dots,$$

which is the p.m.f. of a Poisson distribution. Hence, binomial distribution tends to Poisson distribution when the conditions stated are satisfied.

Moments

Mean

$$\begin{aligned} \mu'_1 = E(X) &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^x}{x!} \\ &= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda \end{aligned}$$

Variance

$$V(X) = \mu'_2 - (\mu'_1)^2 = E(X^2) - [E(X)]^2$$

$$\begin{aligned} \mu'_2 = E(X^2) &= \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!} \\ &= \sum_{x=0}^{\infty} [x(x-1) + x] \frac{e^{-\lambda}\lambda^x}{x!} \end{aligned}$$

$$\begin{aligned}
&= \sum_{x=2}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{x!} + \sum_{x=1}^{\infty} x\frac{e^{-\lambda}\lambda^x}{x!} \\
&= \lambda^2 e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!} + E(X) \\
&= \lambda e^{-\lambda}e^{\lambda} + \lambda \\
&= \lambda^2 + \lambda
\end{aligned}$$

Therefore,

$$\begin{aligned}
V(X) &= \lambda^2 + \lambda - (\lambda)^2 \\
&= \lambda
\end{aligned}$$

and,

$$S.D. = \sqrt{\lambda}$$

In a similarly way, we can obtain μ_3 and μ_4 as

$$\begin{aligned}
\mu_3 &= \lambda \\
\mu_4 &= 3\lambda^2 + \lambda
\end{aligned}$$

Beta and Gamma coefficients

Skewness

$$\begin{aligned}
\beta_1 &= \frac{\mu_3^2}{\mu_2^3} \\
&= \frac{\lambda^2}{\lambda^3} \\
&= \frac{1}{\lambda}.
\end{aligned}$$

Therefore,

$$\gamma_1 = \sqrt{\beta_1} = \frac{1}{\sqrt{\lambda}}.$$

Since $\lambda > 0$, Poisson distribution is positively skewed.

Note: Poisson distribution is positively skewed. But as $n \to \infty$ the distribution becomes symmetric.

Kurtosis

$$\begin{aligned}\beta_2 &= \frac{\mu_4}{\mu_2^2}\\ &= \frac{3\lambda^2+\lambda}{\lambda^2}\\ &= 3+\frac{1}{\lambda}.\end{aligned}$$

Therefore,

$$\gamma_2 = \beta_2 - 3 = \frac{1}{\lambda}.$$

Since $\lambda > 0$, Poisson distribution is leptokurtic.

Note: Poisson distribution is leptokurtic. ut as $n \to \infty$ the distribution becomes mesokurtic.

Moment Generating Function

$$\begin{aligned} M_X(t) &= E\left(e^{tX}\right) \\ &= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} \\ &= e^{-\lambda} e^{\lambda e^t} \\ &= e^{\lambda(e^t-1)}. \end{aligned}$$

Additive/Reproductive Property

Theorem 1.1.4. *If* $X \sim P(\lambda_1)$, $Y \sim P(\lambda_2)$ *and* X *and* Y *are independent, then* $X + Y \sim P(\lambda_1 + \lambda_2)$.

Proof. $X \sim P(\lambda_1)$ and $Y \sim P(\lambda_2)$ implies $M_X(t) = e^{\lambda_1(e^t-1)}$ and $M_Y(t) = e^{\lambda_2(e^t-1)}$ respectively. Since X and Y are independent,

$$\begin{aligned} M_{X+Y}(t) &= M_X(t) \times M_Y(t) \\ &= e^{\lambda_1(e^t-1)} \times e^{\lambda_2(e^t-1)} \\ &= e^{(\lambda_1+\lambda_2)(e^t-1)}, \end{aligned}$$

which is the m.g.f. of $P(\lambda_1 + \lambda_2)$. Therefore, $X + Y \sim P(\lambda_1 + \lambda_2)$. ☐

Remark 1.1.9. *The result can be generalised to k independent Poisson random variables.*

Recurrence Relation for Poisson Central Moments

Theorem 1.1.5. *When* $X \sim P(\lambda)$, $\mu_{r+1} = \lambda \left[r\mu_{r-1} + \frac{d\mu_r}{d\lambda} \right]$.

Proof.

$$\begin{aligned} \mu_r &= E[X - E(X)]^r \\ &= E[X - \lambda]^r \\ &= \sum_{x=0}^{\infty} (x-\lambda)^r \frac{e^{-\lambda}\lambda^x}{x!} \end{aligned}$$

Therefore,

$$\begin{aligned}
\frac{d\mu_r}{d\lambda} &= \frac{d}{d\lambda}\left[\sum_{x=0}^{\infty}(x-\lambda)^r\frac{e^{-\lambda}\lambda^x}{x!}\right] \\
&= \sum_{x=0}^{\infty}\frac{d}{d\lambda}\left[(x-\lambda)^r\frac{e^{-\lambda}\lambda^x}{x!}\right] \\
&= \sum_{x=0}^{\infty}\frac{1}{x!}\left[r(x-\lambda)^{r-1}(-1)e^{-\lambda}\lambda^x\right. \\
&\qquad +(x-\lambda)^r(-e^{-\lambda})\lambda^x \\
&\qquad \left.+\,(x-\lambda)^r e^{-\lambda}x\lambda^{x-1}\right] \\
&= -r\sum_{x=0}^{\infty}(x-\lambda)^{r-1}\frac{e^{-\lambda}\lambda^x}{x!} \\
&\qquad -\sum_{x=0}^{\infty}(x-\lambda)^r\frac{e^{-\lambda}\lambda^x}{x!} \\
&\qquad +\sum_{x=0}^{\infty}\frac{x}{\lambda}(x-\lambda)^r\frac{e^{-\lambda}\lambda^x}{x!} \\
&= -r\mu_{r-1}+\sum_{x=0}^{\infty}(x-\lambda)^r\frac{e^{-\lambda}\lambda^x}{x!}\left(-1+\frac{x}{\lambda}\right) \\
&= -r\mu_{r-1}+\sum_{x=0}^{\infty}(x-\lambda)^r\frac{e^{-\lambda}\lambda^x}{x!}\left(\frac{x-\lambda}{\lambda}\right) \\
&= -r\mu_{r-1}+\frac{1}{\lambda}\sum_{x=0}^{\infty}(x-\lambda)^{r+1}\frac{e^{-\lambda}\lambda^x}{x!} \\
&= -r\mu_{r-1}+\frac{1}{\lambda}\mu_{r+1}
\end{aligned}$$

Hence,

$$\mu_{r+1} \quad = \quad \lambda \left[r\mu_{r-1} + \frac{d\mu_r}{d\lambda} \right].$$

□

Using the information $\mu_0 = 1$ and $\mu_1 = 0$, we can determine μ_2, μ_3 and μ_4 successively.

Mode of Poisson Distribution

Mode is the value of the random variable for which p.m.f. is maximum. The maximum of $f_X(x)$ is attained if

1. $f_X(x) \geq f_X(x+1) \Rightarrow \lambda - 1 \leq x$
2. $f_X(x) \geq f_X(x-1) \Rightarrow x \leq \lambda$.

i.e., $\lambda - 1 \leq x \leq \lambda$. Therefore, mode is $[\lambda]$, the integer part of λ, when λ is not an integer. But when λ is an integer, there will be two modes at $\lambda - 1$ and λ.

Fitting of Poisson Distribution

Fitting of Poisson distribution means to obtain the expected or theoretical Poisson frequencies against the given observed data. It is obtained by multiplying the corresponding probabilities with the total frequency. If E_x denote the expected frequency that the random variable X takes the value x, then

$$\begin{aligned} E_x &= N \times f_X(x) \\ &= N \times \frac{e^{-\lambda}\lambda^x}{x!}; \quad x = 0, 1, 2, \ldots; \ \lambda > 0. \end{aligned}$$

λ can be estimated as $\bar{x}$, the sample mean.

Solved Problems

1. If $X \sim P(3)$, obtain $P(X \geq 2)$.
   Solution:
   We have, $X \sim P(3)$. Therefore, $\lambda = 3$ and $P(X = x) = \frac{e^{-3}3^x}{x!}$.
   Now,

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - [P(X = 0) + P(X = 1)] \end{aligned}$$

$$
\begin{aligned}
&= 1 - \left[\frac{e^{-3}3^0}{0!} + \frac{e^{-3}3^1}{1!}\right] \\
&= 1 - e^{-3}\left[1+3\right] \\
&= 1 - 4e^{-3} \\
&= 1 - 0.1991 \\
&= 0.8009
\end{aligned}
$$

2. If X is a poisson variate such that $P(X=1) = P(X=2)$, obtain the probability distribution of X.
   Solution:
   Let $X \sim P(\lambda)$. We have,

$$
\begin{aligned}
P(X=1) &= P(X=2) \\
\text{i.e.,} \quad \frac{e^{-\lambda}\lambda^1}{1!} &= \frac{e^{-\lambda}\lambda^2}{2!} \\
\text{i.e.,} \quad 1 &= \frac{\lambda}{2!} \\
\text{i.e.,} \quad \lambda &= 2.
\end{aligned}
$$

Therefore, $X \sim P(2)$ and

$$P(X=x)=\frac{e^{-2}2^x}{x!};\quad x=0,1,2,\ldots.$$

3. Following mistakes per page were observed in a book.

No. of mistakes	0	1	2	3	4
No. of pages	211	90	19	5	0

Fit a Poisson distribution to the above data.

Solution:

In order to fit Poisson distribution to the data given we first calculate the mean $\hat{\lambda}$ from the given data. Then the Poisson distribution fitted to the given data is given by

$$\frac{e^{-\hat{\lambda}}\hat{\lambda}^x}{x!};\quad x=0,1,2,\ldots.$$

Here, $N=\sum f_i = 325$.

$$\begin{aligned}\hat{\lambda} &= \frac{\sum x_i f_i}{N}\\ &= \frac{0\times 211+1\times 90+2\times 19+3\times 5+4\times 0}{325}\\ &= 0.44.\end{aligned}$$

So the distribution to be fitted is $P(0.44)$. Therefore,

$$P(X = x) = \frac{e^{0.44}(0.44)^x}{x!}; \quad x = 0, 1, 2, \ldots$$

The expected frequencies E_x is given by

$$E_x = N \times P(X = x)$$

x	$P(X = x)$	E_x
0	0.644	209
1	0.283	92
2	0.062	20
3	0.009	3
4	0.001	1

Exercises

1. The number of defectives, given by the random variable X, follows Poisson distribution with mean 2. Calculate the probability that there will be

   (a) no defective

   (b) exactly one defective

   (c) at least two defective

   (d) at most two defective.

2. A manufacturer knows that 5% of his products are defective. If the products are sold in boxes of 100 and guarantees that not more than 10 will be defective, what is the probability that a box will fail to meet the guaranteed quality?

3. Six coins are tossed 6,400 times. Using Poisson distribution, find the probability of getting six heads r times.

4. The record of births, over the last 100 years maintained by the municipal council of a town showed that 200 children were born blind during that period. On the assumption that the number of children born blind in an year follows Poisson distribution, estimate the number of years in which there were
   (a) no blind births
   (b) one blind birth
   (c) two blind births
   (d) at least three blind births.

5. A manufacturer who produce medicine bottles finds that 0.1% of the bottles are defective. The bottles are packed in boxes containing 500 bottles. A drug manufacturer buys

100 boxes from the producer of bottles. Using Poisson distribution find how many boxes will contain:

(a) no defective

(b) at least two defective.

6. If X and Y are independent poisson variates such that $P(X = 1) = P(X = 2)$ and $P(Y = 2) = P(Y = 3)$. Find the variance of $X - 2Y$.

7. If X is a Poisson variate such that $P(X = 2) = 9P(X = 4) + 90P(X = 6)$. Find

   (a) λ

   (b) the mean of X

   (c) β_1.

8. A poisson distribution has two modes at $x = 1$ and $x = 2$. What is the probability that x will have one or the other of these two values?

9. Derive the mean and variance of Poisson distribution from its m.g.f.

10. Fit a Poisson distribution to the following data on number of major road accidents in a city on a particular year.

No. of accidents	0	1	2	3	4	5
No. of days	150	120	74	18	2	1

11. Fit a Poisson distribution to the following data.

x	0	1	2	3	4	5	6
f	100	95	75	44	20	3	1

1.1.4 Negative Binomial Distribution

Consider a random experiment having two mutually exclusive outcomes. The outcome in which we are interested is labeled as success (S) having probability $P(S) = p$, $0 < p < 1$ and its complement as failure (F) with probability $P(F) = 1 - p = q$.

Now suppose that a Bernoulli trial is repeated n times keeping the probability, p, of success same through out the trials and the trials independent. If we are interested in finding the probability distribution of X, the number of successes in n trials, then $X \sim B(n, p)$. Here, the number of trials 'n' is fixed.

Instead, suppose we are interested in finding the probability of number of trials required to get r successes. Then we have the negative binomial distribution. Here, the number of successes is fixed, not the number of trials 'n'. Hence the name negative binomial.

Let the random variable Y be the number of trials required to get r successes. Then $y = r, r+1, \ldots$. Suppose Y takes the value y. i.e., y trials are required to get r success. Then, out of these y trials r are successes including the y^{th} one. Hence, there will be $y - r$ failures preceding the r^{th} success. Let X be the number of failures preceding r^{th} success. Clearly, X takes values $0, 1, 2, \ldots$.

Note that both the random variables X and Y follow negative binomial distribution where X assumes values $0, 1, 2, \ldots$ and Y assumes values $1, 2, 3, \ldots$.

$$\begin{aligned}
P(X = x) &= P(\text{ } x \text{ failures preceding the } r^{th} \text{ success}) \\
&= P(\text{Getting } r-1 \text{ successes in } x+r-1 \text{ trials} \\
&\qquad\qquad \text{and a success in } (x+r)^{th} \text{ trial}) \\
&= \binom{x+r-1}{r-1} p^{r-1} q^{(x+r-1)-(r-1)} \times p \\
&= \binom{x+r-1}{x} p^r q^x; \quad x = 0, 1, 2, \ldots
\end{aligned}$$

Definition 1.1.4. *A random variable X is said to follow neg-*

ative binomial distribution with parameters r *and* p *if its p.m.f. is given by,*

$$f_X(x) = \begin{cases} \binom{x+r-1}{x} p^r q^x & ; \quad x = 0, 1, 2, \ldots \\ 0 & ; \quad \textit{otherwise}, \end{cases}$$

where $0 < p < 1$ *and* $p + q = 1$ *and we write* $X \sim NB(r, p)$.

Note: The mean of a random variable following $NB(r, p)$ is given by

$$E(X) = \frac{rq}{p}$$

and the variance is given by

$$V(X) = \frac{rq}{p^2}.$$

1.1.5 Geometric Distribution

Negative binomial random variable gives number of failures preceding the r^{th} success. When $r = 1$ it reduces to geometric random variable and its distribution is called geometric distribution. Hence, it gives the probability distribution of number of failures preceding the 1^{st} success and takes values $0, 1, 2, \ldots$.

Definition 1.1.5. *A random variable* X *is said to follow geo-*

metric distribution with parameter p *if its p.m.f. is given by*

$$f_X(x) = \begin{cases} pq^x & ; \quad x = 0, 1, 2, \ldots \\ 0 & ; \quad \textit{otherwise}, \end{cases}$$

where $0 < p < 1$ *and* $p + q = 1$. *Here we write* $X \sim Geo_0(p)$.

Since the probabilities for $x = 0, 1, 2, \ldots$ are the terms of geometric progression series, the distribution has the name geometric distribution. Sometimes it is called Furry distribution.

Moments

Mean

$$\begin{aligned} \mu_1' = E(X) &= \sum_{x=0}^{\infty} xpq^x \\ &= p[q + 2q^2 + 3q^3 + \ldots] \\ &= pq[1 + 2q + 3q^2 + \ldots] \\ &= pq(1-q)^{-2} \end{aligned}$$

$$= \frac{pq}{p^2}$$

$$= \frac{q}{p}$$

Variance

$$V(X) = \mu_2' - (\mu_1')^2 = E(X^2) - [E(X)]^2$$

$$\begin{aligned}
\mu_2' = E(X^2) &= \sum_{x=0}^{\infty} x^2 pq^x \\
&= \sum_{x=0}^{\infty} [x(x-1)+x] pq^x \\
&= \sum_{x=2}^{\infty} x(x-1)pq^x + \sum_{x=1}^{\infty} xpq^x \\
&= p[2.1q^2 + 3.2q^3 + 4.3q^4 + \ldots] + E(X) \\
&= 2pq^2[1 + 3q + 6q^2 + \ldots] + \frac{q}{p}
\end{aligned}$$

$$
\begin{aligned}
&= 2pq^2(1-q)^{-3} + \frac{q}{p} \\
&= \frac{2q^2}{p^2} + \frac{q}{p}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
V(X) &= \frac{2q^2}{p^2} + \left(\frac{q}{p}\right)^2 \\
&= \frac{q^2}{p^2} + \frac{q}{p} \\
&= \frac{q^2 + pq}{p^2} \\
&= \frac{q(q+p)}{p^2} \\
&= \frac{q}{p^2}
\end{aligned}
$$

Moment Generating Function

$$
\begin{aligned}
M_X(t) &= E\left(e^{tX}\right) \\
&= \sum_{x=0}^{\infty} e^{tx} pq^x \\
&= p\sum_{x=0}^{\infty} (qe^t)^x \\
&= p[1 + qe^t + (qe^t)^2 + \ldots] \\
&= p(1 - qe^t)^{-1} \\
&= \frac{p}{1 - qe^t}
\end{aligned}
$$

Note

1. A random variable X that has geometric distribution is often referred to as a discrete waiting (occurrence) time random variable It represents how long one has to wait for a success (in terms of number of failures).

2. Geometric distribution is a discrete analogue of exponen-

tial distribution (continuous waiting time distribution).

3. Some authors define geometric distribution as the number of trials required to obtain the first success. In this case the random variable X takes values $x = 1, 2, 3, \ldots$ and the p.m.f.is given by

$$f_X(x) = \begin{cases} pq^{x-1} & ; \quad x = 1, 2, 3, \ldots \\ 0 & ; \quad \text{otherwise,} \end{cases}$$

where $0 < p < 1$ and $p + q = 1$. Here we write $X \sim \text{Geo}_1(p)$.

Lack of Memory Property

Theorem 1.1.6. *If X is a geometric random variable with parameter p, then*

$$P[X \geq s + t | X \geq s] = P[X \geq t]; \quad \textit{for } s, t = 0, 1, 2 \ldots$$

Proof.

$$\begin{aligned}
P[X \geq s+t | X \geq s] &= \frac{P[X \geq s+t]}{P[X \geq s]} \\
&= \frac{\sum_{x=s+t}^{\infty} pq^x}{\sum_{x=s}^{\infty} pq^x} \\
&= \frac{q^{s+t} + q^{s+t+1} + \ldots}{q^s + q^{s+1} + \ldots} \\
&= \frac{q^{s+t}[1 + q + q^2 + \ldots]}{q^s[1 + q + q^2 + \ldots]} \\
&= \frac{q^{s+t}}{q^s} \\
&= q^t \\
&= P[X \geq t]
\end{aligned}$$

□

Theorem 1.1.7. *If X is a non-negative integral valued r.v, show that it is geometric if it lacks memory.*

Proof. Given,

$$\begin{aligned} P[X \geq s+t | X \geq s] &= P[X \geq t]; \quad \text{for } s, t = 0, 1, 2 \ldots \\ &\Rightarrow \frac{P[X \geq s+t, X \geq s]}{P[X \geq s]} = P[X \geq t] \\ &\Rightarrow \frac{P[X \geq s+t]}{P[X \geq s]} = P[X \geq t] \quad (1.1.2) \end{aligned}$$

Let $R_k = P[X \geq k] = p_k + p_{k+1} + p_{k+2} + \ldots$

Therefore, from equation (1.1.2) we get,

$$\frac{R_{s+t}}{R_s} = R_t$$

In particular, taking $s = 1$, we get

$$\begin{aligned} R_{1+t} &= R_1 R_t \\ &= (p_1 + p_2 + p_3 + \ldots) R_t \\ &= (1 - p_0) R_t \end{aligned}$$

Therefore,

$$\begin{aligned} R_t &= (1-p_0)R_{t-1} \\ &= (1-p_0)(1-p_0)R_{t-2} \\ &= (1-p_0)^2 R_{t-2} \\ &\vdots \\ &= (1-p_0)^t R_{t-t} \\ &= (1-p_0)^t R_0 \\ &= (1-p_0)^t \\ &\Rightarrow X \sim \mathrm{Geo}_0(p_0) \end{aligned}$$

□

1.1.6 Discrete Uniform Distribution

Discrete uniform distribution is a symmetric probability distribution. It can be used to model situations in which a finite number of values are equally likely to be observed. If the random variable X can assume n possible values, then all these values have the same probability of occurrence $\frac{1}{n}$.

Definition 1.1.6. *The random variable X is said to have uniform distribution on n points $\{x_1, x_2, \ldots, x_n\}$ if its p.m.f. is of the form*

$$f_X(x_i) = \begin{cases} \frac{1}{n} & ; \quad i = 1, 2, \ldots, n \\ 0 & ; \quad \textit{otherwise.} \end{cases}$$

In particular if $x_i = i$, $i = 1, 2, \ldots, n$ the above definition can be modified as follows:

Definition 1.1.7. *The random variable X is said to have uniform distribution on n points $\{1, 2, \ldots, n\}$ if its p.m.f. is of the form*

$$f_X(x) = \begin{cases} \frac{1}{n} & ; \quad x = 1, 2, \ldots, n \\ 0 & ; \quad \textit{otherwise.} \end{cases}$$

Moments of Discrete Uniform Distribution on $\{1, 2, \ldots, n\}$

Mean

$$\begin{aligned}
\mu_1' = E(X) &= \sum_{x=1}^{n} x \frac{1}{n} \\
&= \frac{1}{n} \sum_{x=1}^{n} x \\
&= \frac{1}{n}[1 + 2 + \ldots + n] \\
&= \frac{n(n+1)}{2n} \\
&= \frac{n+1}{2}
\end{aligned}$$

Variance

$$V(X) = \mu_2' - (\mu_1')^2 = E(X^2) - [E(X)]^2$$

$$\begin{aligned}
\mu_2' = E(X^2) &= \sum_{x=1}^{n} x^2 \frac{1}{n} \\
&= \frac{1}{n}\sum_{x=1}^{n} x^2 \\
&= \frac{1}{n}[1^2 + 2^2 + \ldots + n^2]
\end{aligned}$$

$$\begin{aligned}
&= \frac{n(n+1)(2n+1)}{6n} \\
&= \frac{(n+1)(2n+1)}{6}
\end{aligned}$$

Therefore,

$$\begin{aligned}
V(X) &= \frac{(n+1)(2n+1)}{6} - \left[\frac{n+1}{2}\right]^2 \\
&= \frac{n+1}{2}\left[\frac{2n+1}{3} - \frac{n+1}{2}\right] \\
&= \frac{n+1}{2}\left[\frac{4n+2-3n-3}{6}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{n+1}{2}\right)\left(\frac{n-1}{6}\right) \\
&= \frac{n^2-1}{12}
\end{aligned}
$$

Moment Generating Function

$$
\begin{aligned}
M_X(t) &= E\left(e^{tX}\right) \\
&= \sum_{x=1}^{n} e^{tx}\frac{1}{n} \\
&= \frac{1}{n}\sum_{x=1}^{n} e^{tx} \\
&= \frac{1}{n}\left[e^t + e^{2t} + e^{3t} \ldots + e^{nt}\right] \\
&= \frac{1}{n}\, e^t\left[1 + e^t + e^{2t} + \ldots + e^{(n-1)t}\right] \\
&= \frac{e^t}{n}\,\frac{(e^{nt}-1)}{e^t-1}
\end{aligned}
$$

1.2 Continuous Distributions

1.2.1 Continuous Uniform or Rectangular Distribution

The continuous uniform distribution or rectangular distribution is a family of symmetric probability distributions. The distribution describes an experiment where there is an arbitrary number outcome that lies between certain bounds. The bounds are defined by the parameters, a and b. The interval can be either closed (eg. $[a, b]$) or open (eg. (a, b)). Therefore, the distribution is often abbreviated $U(a, b)$. The difference between the bounds defines the interval length; all intervals of the same length on the distribution's support are equally probable.

Definition 1.2.1. *A random variable X is said to have a continuous uniform distribution over an interval (a, b) if its probability density function (p.d.f.) is*

$$f_X(x) = \begin{cases} \frac{1}{b-a} & ; \quad a < x < b \\ 0 & ; \quad otherwise \end{cases},$$

Remark 1.2.1. *a and b, $(a < b)$ are the two parameters of the uniform distribution on (a, b).*

Remark 1.2.2. *The distribution is known as rectangular distribution, since the curve* $y = f(x)$ *describes a rectangle over the* x*-axis and between the ordinates at* $x = a$ *and* $x = b$*.*

Remark 1.2.3. *The cumulative distribution function (c.d.f.) or simply distribution function (d.f.)* $F(x)$ *of* $U(a, b)$ *is given by*

$$F_X(x) = \begin{cases} 0 & ; \quad x < a \\ \frac{x-a}{b-a} & ; \quad a < x < b \\ 1 & ; \quad x > b \end{cases},$$

Since $F(x)$ *is not continuous at* $x = a$ *and* $x = b$*, it is not differentiable at these points. Thus* $\frac{d}{dx}F(x) = f(x) = \frac{1}{b-a} \neq 0$*, exists everywhere except at the points* $x = a$ *and* $x = b$ *and consequently we get the p.d.f.* $f(x)$*.*

Remark 1.2.4. *The graphs of uniform p.d.f. and d.f. are given below.*

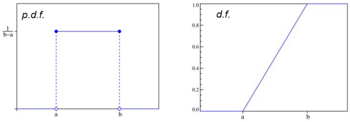


Figure 1.1: p.d.f. and d.f. of Uniform Distribution

Moments

Mean

$$\begin{aligned} E(X) &= \int_a^b x \frac{1}{b-a} dx \\ &= \frac{a+b}{2} \end{aligned}$$

Variance

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

$$\begin{aligned} E(X^2) &= \int_a^b x^2 \frac{1}{b-a} dx \\ &= \frac{a^2+ab+b^2}{3} \end{aligned}$$

Therefore,

$$\begin{aligned} V(X) &= \frac{a^2+ab+b^2}{3} - \left[\frac{a+b}{2}\right]^2 \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

Moment Generating Function

$$\begin{aligned} M_X(t) &= E\left(e^{tX}\right) \\ &= \frac{e^{bt}-e^{at}}{t(b-a)}; \quad t \neq 0 \end{aligned}$$

Result 1.2.1.

If X is a continuous random variable with d.f. $F_X(x)$, then $F_X(x) \sim U[0,1]$.

Exercises

1. Obtain the mean and variance of $U(3,8)$.
2. If X is uniformly distributed on $(1,2)$, find z such that $P(X > z + \mu_X) = \frac{1}{4}$, where $\mu_X = E(X)$.

1.2.2 Gamma Distribution

Gamma distribution is a continuous probability distribution having two parameters. The exponential distribution, Erlang distribution and Chi-square distribution are particular cases of gamma distribution.

Definition 1.2.2. *A random variable X is said to have a gamma distribution with parameters m and p, if its probability density function is given by*

$$f_X(x) = \begin{cases} \frac{m^p}{\Gamma p} e^{-mx} x^{p-1} & ; \quad x > 0 \\ 0 & ; \quad \textit{otherwise,} \end{cases}$$

where $m > 0$ and $p > 0$

Note: Being a p.d.f., we know that

$$\begin{aligned} \int_0^\infty f_X(x) dx &= 1 \\ \text{i.e.,} \int_0^\infty \frac{m^p}{\Gamma p} e^{-mx} x^{p-1} dx &= 1 \\ \text{i.e.,} \int_0^\infty e^{-mx} x^{p-1} dx &= \frac{\Gamma p}{m^p} \end{aligned}$$

Putting $m = 1$ we get,

$$\Gamma p = \int_0^\infty e^{-x} x^{p-1} dx$$

When $p = n$, a positive integer, using integration by parts we get,

$$\begin{aligned}
\Gamma n &= \int_0^\infty e^{-x} x^{n-1} dx \\
&= (n-1)\Gamma(n-1) \\
&= (n-1)(n-2)\Gamma(n-2) \\
&\vdots \\
&= (n-1)(n-2)\ldots 1 \quad \Gamma 1 \\
&= (n-1)!
\end{aligned}$$

Putting $m = 1$ and $p = \frac{1}{2}$ we get,

$$\Gamma \frac{1}{2} = \int_0^\infty e^{-x} x^{\frac{1}{2}-1} dx = \sqrt{\pi}.$$

Moments

Mean

$$\begin{aligned} E(X) &= \int_0^\infty x \frac{m^p}{\Gamma p} e^{-mx} x^{p-1} dx \\ &= \frac{p}{m} \end{aligned}$$

Variance

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

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2 \frac{m^p}{\Gamma p} e^{-mx} x^{p-1} dx \\ &= \frac{p^2}{m^2} + \frac{p}{m^2} \end{aligned}$$

Therefore,

$$\begin{aligned} V(X) &= \frac{p^2}{m^2} + \frac{p}{m^2} - \left[\frac{p}{m}\right]^2 \\ &= \frac{p}{m^2} \end{aligned}$$

Moment Generating Function

$$\begin{aligned} M_X(t) &= E\left(e^{tX}\right) \\ &= \int_0^\infty e^{tx} \frac{m^p}{\Gamma p} e^{-mx} x^{p-1} dx \\ &= \left(1 - \frac{t}{m}\right)^{-p} \end{aligned}$$

1.2.3 Exponential Distribution

Exponential distribution is a continuous probability distribution which has wide utilities. It is the probability distribution of the time between events in a Poisson point process, i.e., a process in which events occur continuously and independently at a constant average rate. It is a particular case of the gamma distribution, continuous analogue of the geometric distribution and is the only continuous distribution having lack of memory property.

Definition 1.2.3. *A random variable X is said to have an exponential distribution with parameter λ, if its probability density function is given by*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & ; \quad x > 0 \\ 0 & ; \quad \textit{otherwise}, \end{cases}$$

where $\lambda > 0$ and is denoted by $\exp(\lambda)$.

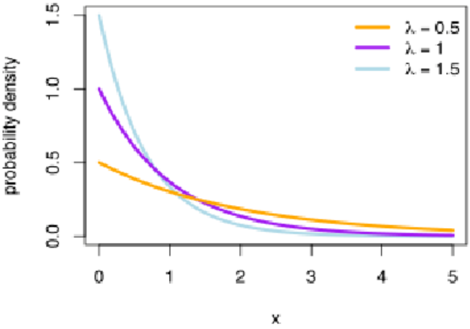


Figure 1.2: The p.d.f. of exponential distribution with parameter λ

The p.d.f. $f_X(x)$ assumes the value 0 for negative values of x, and then for positive values, it starts off at a value equal to λ. This is because if you put $x = 0$ in the p.d.f. expression, you get λ times e^0, which leaves you just with λ. So it starts off with λ, and then it decays at the rate of λ. Notice that when λ is small, the initial value of the p.d.f. is small. But then the decay rate is also small, so that the p.d.f. extends over a large range of x's.

Note: Exponential Distribution is a particular case of Gamma distribution. In a Gamma distribution with parameters m and p, if we put $p = 1$, we get the exponential distribution with parameter m.

Moments

Mean

$$\begin{aligned} E(X) &= \int_0^\infty x\ \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^\infty x\ e^{-\lambda x} dx \end{aligned}$$

$$\begin{aligned} &= \lambda\frac{\Gamma(2)}{\lambda^2} \\ &= \frac{1}{\lambda} \end{aligned}$$

Variance

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

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2\ \lambda e^{-\lambda x} dx \\ &= \lambda\frac{\Gamma(3)}{\lambda^3} \\ &= \frac{2}{\lambda^2} \end{aligned}$$

Therefore,

$$\begin{aligned} V(X) &= \frac{2}{\lambda^2} - \left[\frac{1}{\lambda}\right]^2 \\ &= \frac{1}{\lambda^2} \end{aligned}$$

Moment Generating Function

$$\begin{aligned} M_X(t) &= E\left(e^{tX}\right) \\ &= \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx \\ &= \left(1-\frac{t}{\lambda}\right)^{-1} \\ &= \left(\frac{\lambda}{\lambda-t}\right) \end{aligned}$$

Remark 1.2.5. *If X is an exponential random variable with parameter λ, then its d.f. is given by*

$$F_X(x) = 1 - e^{-\lambda x}, \quad \lambda > 0.$$

This distribution, sometimes called negative exponential distribution occurs in applications such as reliability theory and queueing theory. Reasons for its use include memoryless property and the relation to the poisson distribution. Exponential distribution has the memoryless property like geometric distribution, which is its discrete analogue.

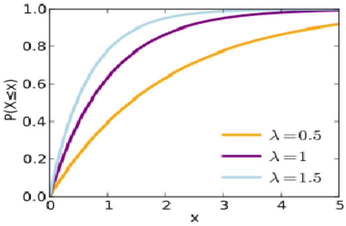


Figure 1.3: The d.f. of exponential distribution with parameter λ

Properties

Lack of Memory Property

Theorem 1.2.1. *If X is an exponential random variable with parameter λ, then*

$$P[X \geq s+t|X \geq s] = P[X \geq t]; \quad \textit{for } s,t > 0.$$

Closure under Minima

Theorem 1.2.2. *If X_1 is an exponential random variable with parameter λ_1 and X_2 is an exponential random variable with parameter λ_2 and X_1 and X_2 are independent, then* $\min(X_1, X_2)$

is an exponential random variable with parameter $\lambda_1 + \lambda_2$.

The above result can be extended to n random variables.

Exercises

1. Write down the p.d.f., d.f., mean variance and m.g.f. of an exponential random variable with parameter $\frac{1}{2}$.
2. Mean of an exponential random variable is $\frac{2}{3}$. Give its p.d.f. and d.f.
3. The m.g.f. of a continuous random variable is given by $\left(1 - \frac{t}{5}\right)^{-1}$. Identify the distribution.
4. If $X_1, X_2, \ldots, X_n$ are n independent exponential random variables with parameter λ, then obtain the distribution of $Y = \sum X_i$.
5. Write down the p.d.f of a random variable having m.g.f. $\frac{5}{5-t}$.

1.2.4 Normal Distribution

The Normal distribution plays a pivotal role in most of the statistical techniques used in applied statistics. The main reason for this is the central limit theorem, according to which normal distribution is found to be the approximation of most of the random variables. We may discuss it in detail later.

Normal distribution was first introduced by a French mathematician, Abraham De-Moivre (1667-1754). He is obtained it while working on certain problems in the games of chance. Later, two mathematical astronomers Pierre Laplace (1749-1827) and Karl Gauss (1777-1855) developed this distribution independently. They found that, it can be used to model errors (the deviation of the observed value from the true value). Hence, this distribution is also known as Gaussian distribution and Laplace's distribution. But, it is most commonly known as the normal distribution.

Definition 1.2.4. *A random variable X is said to follow normal distribution with parameters μ and σ^2 if its p.d.f. is given by*

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad ; \quad -\infty < x < \infty \ ,$$

where $-\infty < \mu < \infty$ and $\sigma > 0$.

In this case we can write $X \sim N(\mu, \sigma^2)$ or $X \sim N(\mu, \sigma)$.

This most important continuous distribution is symmetric about the mean, showing that data near the mean are more frequent in occurrence than data far from the mean. In graphical form, normal distribution will appear as a bell curve.

Properties of Normal Curve

1. The normal curve is symmetrical about the ordinate at $x = \mu$, i.e., $f(\mu + c) = f(\mu - c)$ for any c.
2. The mean, median and mode are identical.
3. The mode of normal curve is at $x = \mu$, and is equal to $\frac{1}{\sigma\sqrt{2\pi}}$.
4. The normal curve extends from $-\infty$ to $+\infty$.
5. For a normal distribution $\beta_1 = 0$ (i.e., symmetric) and $\beta_2 = 3$ (i.e., mesokurtic).
6. x-axis is an asymptote to the curve. That is, the curve touches the x-axis only at $\pm\infty$.
7. In a normal distribution $Q.D. : M.D. : S.D. = 10 : 12 : 15$. Thus, $Q.D. = \frac{2}{3}\ S.D.$ and $M.D. = \frac{4}{5} S.D.$
8. All odd order central moments are zero. i.e.,

$$\mu_{2r+1} = 0, \ r = 0, 1, 2, \ldots$$

9. Even order central moments are given by

$$\mu_{2r} = 1.3.5\ldots(2r-1)\sigma^{2r},\ r = 0, 1, 2, \ldots$$

10. The points of inflection of the curve are $x = \mu \pm \sigma$

11. The lower and upper quartile are equidistant from median.

12. The area under the normal curve is distributed as:

    (a) 68.27% of the items lies between $\mu - \sigma$ and $\mu + \sigma$.

    $$\text{i.e.,} P(\mu - \sigma \leq X \leq \mu + \sigma) = 0.6827.$$

    (b) 95.45% of the items lies between $\mu - 2\sigma$ and $\mu + 2\sigma$.

    $$\text{i.e.,} P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = 0.9545.$$

    (c) 99.73% of the items lies between $\mu - 3\sigma$ and $\mu + 3\sigma$.

    $$\text{i.e.,} P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9973.$$

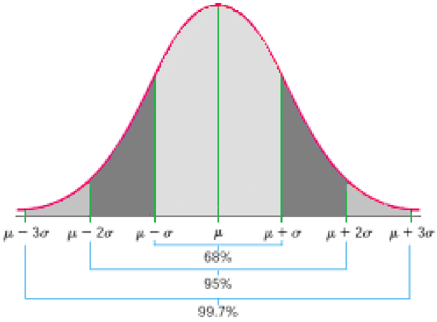


Figure 1.4: Area under Normal curve

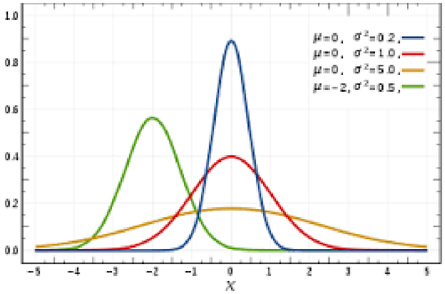


Figure 1.5: Normal curve for different values of μ and σ^2

Moments

Mean

$$
\begin{aligned}
\mu_1' = E(X) &= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx \\
&= \int_{-\infty}^{\infty} (x-\mu+\mu) \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx \\
&= \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx \\
&\qquad +\mu \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx \\
&= \int_{-\infty}^{\infty} \sigma z \frac{1}{\sqrt{2\pi}}\ e^{-\frac{z^2}{2}}\ dz \\
&\qquad +\mu \times 1, \qquad \text{where } \ z = \frac{x-\mu}{\sigma} \\
&= \frac{\sigma}{\sqrt{2\pi}} \times 0 + \mu, \\
&\qquad \text{(being the integral of an odd function)} \\
&= \mu
\end{aligned}
$$

Variance

$$\begin{aligned}
V(X) &= E[X - E(X)]^2 \\
&= E[X - \mu]^2 \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}}\; e^{-\frac{(x-\mu)^2}{2\sigma^2}}\; dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^2 e^{-\frac{z^2}{2}} \sigma dz, \qquad \text{where} \;\; z = \frac{x-\mu}{\sigma} \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}} dz \\
&= \frac{2\sigma^2}{\sqrt{2\pi}} \int_{0}^{\infty} z^2 e^{-\frac{z^2}{2}} dz, \\
&\qquad\qquad \text{(being the integral of an even function)} \\
&= \frac{2\sigma^2}{\sqrt{2\pi}} \int_{0}^{\infty} 2ue^{-u} \frac{du}{\sqrt{2u}}, \qquad \text{where} \;\; u = \frac{z^2}{2} \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \int_{0}^{\infty} u^{\frac{1}{2}} e^{-u} du, \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \int_{0}^{\infty} u^{\frac{3}{2}-1} e^{-u} du, \\
&= \frac{2\sigma^2}{\sqrt{\pi}}\; \frac{\Gamma\left(\frac{3}{2}\right)}{1^{\frac{3}{2}}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{2\sigma^2}{\sqrt{\pi}}\ \frac{1}{2}\ \Gamma\left(\tfrac{1}{2}\right) \\
&= \frac{2\sigma^2}{\sqrt{\pi}}\ \frac{1}{2}\ \sqrt{\pi} \\
&= \sigma^2
\end{aligned}
$$

Odd Order Central Moments

$$
\begin{aligned}
\mu_{2r+1} &= E[X - E(X)]^{2r+1} \\
&= E[X - \mu]^{2r+1} \\
&= \int_{-\infty}^{\infty} (x-\mu)^{2r+1} \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2r+1} e^{-\frac{z^2}{2}}\, \sigma dz,
\end{aligned}
$$

$$\text{where } \ z = \frac{x-\mu}{\sigma}$$

$$= \frac{\sigma^{2r+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2r+1} e^{-\frac{z^2}{2}} dz$$

$$= \frac{\sigma^{2r+1}}{\sqrt{2\pi}} \times 0, \text{(being the integral of an odd function)}$$

Hence,

$$\mu_{2r+1} = 0; \qquad \text{for } r = 0, 1, 2, \ldots.$$

That is, all odd order central moments are zero.

Even Order Central Moments

$$\begin{aligned}
\mu_{2r} &= E[X - E(X)]^{2r} \\
&= E[X - \mu]^{2r} \\
&= \int_{-\infty}^{\infty} (x-\mu)^{2r} \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2r} e^{-\frac{z^2}{2}} \sigma dz, \qquad \text{where } z = \frac{x-\mu}{\sigma} \\
&= \frac{\sigma^{2r}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2r} e^{-\frac{z^2}{2}} dz
\end{aligned}$$

$$= \quad \frac{2\sigma^{2r}}{\sqrt{2\pi}} \int_0^\infty z^{2r} e^{-\frac{z^2}{2}} dz,$$

being the integral of an even function

$$= \quad \frac{2\sigma^{2r}}{\sqrt{2\pi}} \int_0^\infty (2u)^r e^{-u} \frac{du}{\sqrt{2u}}, \qquad \text{where } \ u = \frac{z^2}{2}$$

$$= \quad \frac{2^r\sigma^{2r}}{\sqrt{\pi}} \int_0^\infty u^{r-\frac{1}{2}} e^{-u} du,$$

$$= \quad \frac{2^r\sigma^{2r}}{\sqrt{\pi}} \int_0^\infty u^{r+\frac{1}{2}-1} e^{-u} du,$$

$$= \quad \frac{2^r\sigma^{2r}}{\sqrt{\pi}} \ \frac{\Gamma(r+\frac{1}{2})}{1^{r+\frac{1}{2}}}$$

$$= \quad \frac{2^r\sigma^{2r}}{\sqrt{\pi}} \left(r - \frac{1}{2}\right)\left(r - \frac{3}{2}\right) \ldots \frac{3}{2}.\frac{1}{2} \ \Gamma\left(\tfrac{1}{2}\right)$$

$$= \quad \frac{2^r\sigma^{2r}}{\sqrt{\pi}} \ \frac{(2r-1)(2r-3)\ldots 3.1.\sqrt{\pi}}{2^r}$$

$$= \quad 1.3.5\ldots(2r-1)\sigma^{2r}$$

Recurrence Relation for Even Order Central Moments

We have,

$$\mu_{2r} = 1.3.5\ldots(2r-1)\sigma^{2r}$$

$$\mu_{2r+2} = 1.3.5\ldots(2r-1)\sigma^{2r+2}$$

Therefore,

$$\frac{\mu_{2r+2}}{\mu_{2r}} = (2r+1)\sigma^2$$

$$\mu_{2r+2} = (2r+1)\sigma^2\ \mu_{2r}$$

With this recurrence formula and the information $\mu_0 = 1$ we can calculate μ_2 and μ_4 successively. Putting $r = 0$ we get $\mu_2 = \sigma^2$ and then substituting $r = 1$ we obtain $\mu_4 = 3\sigma^4$.

Beta and Gamma coefficients

Skewness

Since all the odd order central moments are zero,

$$\begin{aligned} \beta_1 &= \frac{\mu_3^2}{\mu_2^3} \\ &= 0. \end{aligned}$$

Hence, $\gamma_1 = \sqrt{\beta}_1 = 0$. That is, normal distribution is symmetric.

Kurtosis

$$\begin{aligned} \beta_2 &= \frac{\mu_4}{\mu_2^2} \\ &= \frac{3\sigma^4}{\sigma^4} \\ &= 3 \end{aligned}$$

Hence, $\gamma_2 = \beta_2 - 3 = 0$. That is, the distribution is mesokurtic.

Moment Generating Function

$$
\begin{aligned}
M_X(t) &= E\left(e^{tX}\right) \\
&= \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sigma\sqrt{2\pi}}\; e^{-\frac{(x-\mu)^2}{2\sigma^2}}\; dx \\
&= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{t(\mu+\sigma z)}e^{-\frac{z^2}{2}}\sigma dz, \qquad \text{where} \;\; z = \frac{x-\mu}{\sigma} \\
&= \frac{e^{\mu t}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{t\sigma z-\frac{z^2}{2}}dz \\
&= \frac{e^{\mu t}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2-2t\sigma z)}dz \\
&= \frac{e^{\mu t}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2-2t\sigma z+t^2\sigma^2)+\frac{1}{2}t^2\sigma^2}dz \\
&= \frac{e^{\mu t+\frac{1}{2}t^2\sigma^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t\sigma)^2}dz \\
&= \frac{e^{\mu t+\frac{1}{2}t^2\sigma^2}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{u^2}{2}}du, \qquad \text{where} \;\; u = z - t\sigma \\
&= \frac{e^{\mu t+\frac{1}{2}t^2\sigma^2}}{\sqrt{2\pi}}2\int_{0}^{\infty} e^{-\frac{u^2}{2}}du,
\end{aligned}
$$

(being the integral of an even function)

$$
\begin{aligned}
&= \frac{e^{\mu t+\frac{1}{2}t^2\sigma^2}}{\sqrt{2\pi}} 2\int_0^\infty e^{-v}\frac{dv}{\sqrt{2v}}, \qquad \text{where } \ v=\frac{u^2}{2} \\
&= \frac{e^{\mu t+\frac{1}{2}t^2\sigma^2}}{\sqrt{\pi}} \int_0^\infty v^{\frac{1}{2}-1}e^{-v}dv \\
&= \frac{e^{\mu t+\frac{1}{2}t^2\sigma^2}}{\sqrt{\pi}} \frac{\Gamma\left(1/2\right)}{1^{1/2}} \\
&= \frac{e^{\mu t+\frac{1}{2}t^2\sigma^2}}{\sqrt{\pi}} \times \sqrt{\pi} \\
&= e^{\mu t+\frac{1}{2}t^2\sigma^2}
\end{aligned}
$$

Central Moment Generating Function

$$
\begin{aligned}
M_{X-\mu}(t) &= E\left[e^{t(X-\mu)}\right] \\
&= e^{-\mu t}E(e^{tX}) \\
&= e^{-\mu t}\times e^{\mu t+\frac{1}{2}t^2\sigma^2} \\
&= e^{\frac{1}{2}t^2\sigma^2}
\end{aligned}
$$

Mean Deviation about Mean

$$
\begin{aligned}
M.D. &= E|X - E(X)| \\
&= E|X - \mu| \\
&= \int_{-\infty}^{\infty} |x-\mu| \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} |x-\mu|\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx, \\
&= \frac{2}{\sigma\sqrt{2\pi}} \int_{\mu}^{\infty} (x-\mu)\ e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ dx, \\
&\qquad\qquad \text{(since the curve is symmetric about } \mu\text{)} \\
&= \frac{2}{\sigma\sqrt{2\pi}} \int_{0}^{\infty} \sigma z\ e^{-\frac{z^2}{2}} \sigma dz, \qquad \text{where} \ \ z = \frac{x-\mu}{\sigma} \\
&= \frac{2\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} z\ e^{-\frac{z^2}{2}} dz \\
&= \frac{2\sigma}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-u} du, \qquad \text{where} \ \ u = \frac{z^2}{2} \\
&= \sqrt{\frac{2}{\pi}}\ \sigma \left[\frac{e^{-u}}{-1}\right]_{0}^{\infty}
\end{aligned}
$$

$$
\begin{aligned}
&= \sqrt{\frac{2}{\pi}}\ \sigma \times 1 \\
&= \sqrt{\frac{2}{\pi}}\ \sigma \\
&= 0.79788\ \sigma
\end{aligned}
$$

Therefore, $M.D. \approx \frac{4}{5} S.D.$

Additive Property

1. If $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$ and if X_1 and X_2 are independent, then

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

Proof. Given $X_1 \sim N(\mu_1, \sigma_1^2)$ implies $M_{X_1}(t) = e^{\mu_1 t + \frac{t^2 \sigma_1^2}{2}}$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ implies $M_{X_2}(t) = e^{\mu_2 t + \frac{t^2 \sigma_2^2}{2}}$. Since

X_1 and X_2 are independent,

$$\begin{aligned} M_{X_1+X_2}(t) &= M_{X_1}(t).M_{X_2}(t) \\ &= e^{\mu_1 t+\frac{t^2\sigma_1^2}{2}} \times e^{\mu_2 t+\frac{t^2\sigma_2^2}{2}} \\ &= e^{(\mu_1+\mu_2)t+\frac{t^2}{2}(\sigma_1^2+\sigma_2^2)}, \end{aligned}$$

which is the m.g.f. of $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. □

2. If X_i, $i = 1, 2, \ldots, n$ are n independent normal variates with mean μ_i and variance σ_i^2 respectively, then $Y = \sum_{i=1}^n X_i$ is normally distributed with mean $\sum_{i=1}^n \mu_i$ and variance $\sum_{i=1}^n \sigma_i^2$.

Proof. Given $X_i \sim N(\mu_i, \sigma_i^2)$ implies $M_{X_i}(t) = e^{\mu_i t+\frac{t^2\sigma_i^2}{2}}$. Now

$$\begin{aligned} M_Y(t) &= M_{\sum_{i=1}^n X_i}(t) \\ &= \prod_{i=1}^n M_{X_i}(t) \end{aligned}$$

$$= \prod_{i=1}^{n} e^{\mu_i t + \frac{t^2\sigma_i^2}{2}}$$

$$= e^{\sum_{i=1}^{n} \mu_i t + \frac{t^2 \sum_{i=1}^{n} \sigma_i^2}{2}},$$

which is the m.g.f. of normal variate with mean $\sum_{i=1}^{n} \mu_i$ and variance $\sum_{i=1}^{n} \sigma_i^2$. ☐

3. If X_i, $i = 1, 2, \ldots, n$ are n independent normal variates with mean μ_i and variance σ_i^2 respectively, then their linear combination, $Y = \sum_{i=1}^{n} a_i X_i$, is normally distributed with mean $\sum_{i=1}^{n} a_i \mu_i$ and variance $\sum_{i=1}^{n} a_i^2 \sigma_i^2$ where a_i's are constants.

Proof. Given $X_i \sim N(\mu_i, \sigma_i^2)$ implies $M_{X_i}(t) = e^{\mu_i t + \frac{t^2\sigma_i^2}{2}}$. Now

$$M_Y(t) = M_{\sum_{i=1}^{n} a_i X_i}(t)$$

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

$$
\begin{aligned}
&= \prod_{i=1}^{n} M_{X_i}(a_i t) \\
&= \prod_{i=1}^{n} e^{\mu_i a_i t + \frac{a_i^2 t^2 \sigma_i^2}{2}} \\
&= e^{\sum_{i=1}^{n} a_i \mu_i t + \frac{t^2 \sum_{i=1}^{n} a_i^2 \sigma_i^2}{2}},
\end{aligned}
$$

which is the m.g.f. of normal variate with mean $\sum_{i=1}^{n} a_i \mu_i$ and variance $\sum_{i=1}^{n} a_i^2 \sigma_i^2$. □

Standard Normal Distribution

Let $X \sim N(\mu, \sigma^2)$. Then the random variable defined by $Z = g(X) = \frac{X-\mu}{\sigma}$ is a standard normal variate. Since X is continuous and $z' = g'(x) = \frac{1}{\sigma} > 0$ for all x, the p.d.f. of Z is given by

$$
f_Z(z) = f_X(g^{-1}(z)) \left| \frac{d}{dz} g^{-1}(z) \right|
$$

Now,

$$z = g(x) = \frac{x-\mu}{\sigma} \quad \Rightarrow \quad x = g^{-1}(z) = \mu + \sigma z$$

$$\Rightarrow \quad \frac{d}{dz} g^{-1}(z) = \sigma$$

and

$$\begin{aligned} f_X(g^{-1}(z)) &= f_X(\mu + \sigma z) \\ &= \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} \end{aligned}$$

Therefore,

$$\begin{aligned} f_Z(z) &= \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} \times \sigma \\ &= \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty \end{aligned}$$

Definition 1.2.5. *A random variable Z is said to follow standard normal distribution if its p.d.f. is given by*

$$f_Z(z) = \tfrac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} \quad ; \quad -\infty < z < \infty \ .$$

We can see that

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = 0$$

and

$$V(Z) = V\left(\frac{X-\mu}{\sigma}\right) = 1.$$

That means, normal distribution with mean zero and variance 1 is called standard normal distribution and we write $Z \sim N(0,1)$. It is for this reason Z is called standard normal variable. In fact, when we say about standardising any random variable, what we mean is, shifting the origin so that the mean of transformed variable is zero and rescaling it so that its variance/S.D. is one.

Moment Generating Function of $N(0,1)$

$$\begin{aligned}
M_Z(t) &= M_{\frac{X-\mu}{\sigma}}(t) \\
&= e^{-\frac{\mu t}{\sigma}} M_X\left(\frac{t}{\sigma}\right) \\
&= e^{-\frac{\mu t}{\sigma}} \times e^{\frac{\mu t}{\sigma} + \frac{\left(\frac{t}{\sigma}\right)^2 \sigma^2}{2}}
\end{aligned}$$

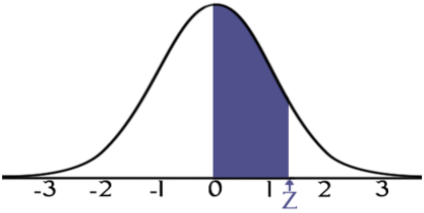
$$= \quad e^{\frac{t^2}{2}}$$

Standard normal distribution satisfies all the properties of normal distribution provided $\mu = 0$ and $\sigma = 1$. Some of them are the following.

1. The curve of $f(z)$ is symmetrical about the ordinate at $z = 0$.
2. The curve of $f(z)$ is maximum at $z = 0$ and the maximum ordinate is $1/\sqrt{2\pi}$.
3. The curve extends from $-\infty$ to $+\infty$.
4. Mean = Median = Mode =0.
5. In a standard normal distribution 68.27% of the items lies between -1 and +1, 95.45% of observations are lying between - 2 and + 2. and 99.73% of observations lies between -3 and + 3.

Area Under the Standard Normal Probability Curve

The table for the areas under the standard normal curve gives the probability of the random variable Z lying between 0 and any positive value of z. The shaded area in the figure represents that probability.



This is the area between the x-axis, the curve of the standard normal p.d.f. and the abscissae 0 and z. This area (probability) can be read from the table of 'Areas under standard normal curve'.

Suppose $X \sim N(\mu, \sigma^2)$ and we are interested in finding the probability of the variate X lying between two values, say, a and b. To determine this, we first make the transformation

$Z = \frac{X-\mu}{\sigma}$. Hence,

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right) \\ &= P\left(z_1 < Z < z_2\right), \end{aligned}$$

where $z_1 = \frac{a-\mu}{\sigma}$ and $z_2 = \frac{b-\mu}{\sigma}$. Therefore, $P(a < X < b)$ is the area under the standard normal curve between the abscissae z_1 and z_2. Hence, to find any probability regarding X, the standard normal table can be made use of.

Solved Problems

1. $X \sim N(20, 4)$. Find the probability that the value taken by X is
   (a) less than 24
   (b) greater than 24
   (c) less than 19
   (d) greater than 19
   (e) between 18 and 22
   (f) between 23 and 29

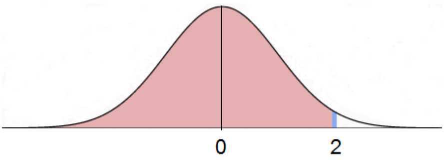
(g) between 17 and 18

Solution:
Given $X \sim N(20, 4)$. Therefore, $\mu = 20$ and $\sigma = 2$.

(a)

$$\begin{aligned} P(X < 24) &= P\left(\frac{X-\mu}{\sigma} < \frac{24-\mu}{\sigma}\right) \\ &= P\left(\frac{X-20}{2} < \frac{24-20}{2}\right) \\ &= P(Z < 2) \end{aligned}$$

which is the area under the standard normal curve from $-\infty$ to 2 as shown in the figure below.

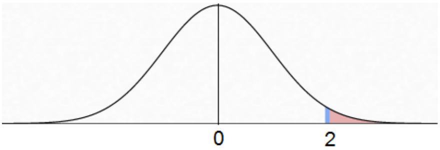


$$\begin{aligned} \therefore \quad P(X < 24) &= P(Z < 0) + P(0 < Z < 2) \\ &= 0.5 + 0.4772 \\ &= 0.9772 \end{aligned}$$

(b)

$$\begin{aligned} P(X > 24) &= P\left(\frac{X-20}{2} > \frac{24-20}{2}\right) \\ &= P(Z > 2) \end{aligned}$$

which is the area under the standard normal curve from 2 to $+\infty$ as shown in the figure below.

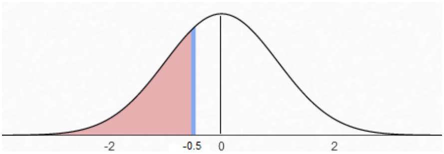


$$\therefore \quad P(X > 24) = P(Z > 0) - P(0 < Z < 2)$$
$$= 0.5 - 0.4772$$
$$= 0.0228$$

(c)

$$P(X < 19) = P\left(\frac{X-20}{2} < \frac{19-20}{2}\right)$$
$$= P(Z < -0.5)$$

which is the area under the standard normal curve from $-\infty$ to -0.5 as shown in the figure below.

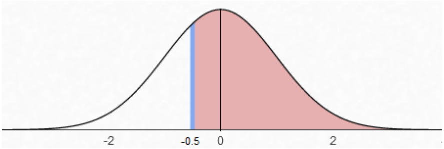


$$\begin{aligned}\therefore \quad P(X < 19) &= P(Z > 0.5) \\ &\qquad \text{(since, symmetric)} \\ &= P(Z > 0) - P(0 < Z < 0.5) \\ &= 0.5 - 0.1915 \\ &= 0.3085\end{aligned}$$

(d)

$$\begin{aligned}P(X > 19) &= P\left(\frac{X - 20}{2} > \frac{19 - 20}{2}\right) \\ &= P(Z > -0.5)\end{aligned}$$

which is the area under the standard normal curve from -0.5 to $+\infty$ as shown in the figure below.

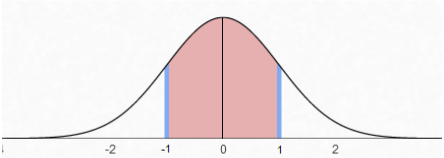


$$\begin{aligned}
\therefore \quad P(X > 19) &= P(-0.5 < Z < 0) + P(Z > 0) \\
&= P(0 < Z < 0.5) + P(Z > 0) \\
&\qquad\qquad \text{(since, symmetric)} \\
&= 0.1915 + 0.5 \\
&= 0.6915
\end{aligned}$$

(e)

$$\begin{aligned}
P(18 < X < 22) &= P\left(\frac{18-20}{2} < \frac{X-20}{2} < \frac{22-20}{2}\right) \\
&= P(-1 < Z < 1)
\end{aligned}$$

which is the area under the standard normal curve from -1 to $+1$ as shown in the figure below.

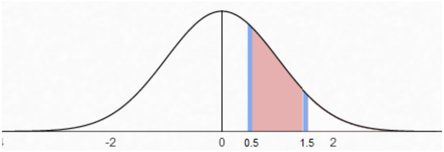


$$
\begin{aligned}
\therefore \quad P(18 < X < 22) &= P(-1 < Z < 0) + P(0 < Z < 1) \\
&= 2P(0 < Z < 1) \\
&\qquad\qquad \text{(since, symmetric)} \\
&= 2 \times 0.3413 \\
&= 0.6826
\end{aligned}
$$

(f)

$$
\begin{aligned}
P(21 < X < 23) &= P\left(\frac{21-20}{2} < \frac{X-20}{2} < \frac{23-20}{2}\right) \\
&= P(0.5 < Z < 1.5)
\end{aligned}
$$

which is the area under the standard normal curve from 0.5 to 1.5 as shown in the figure below.

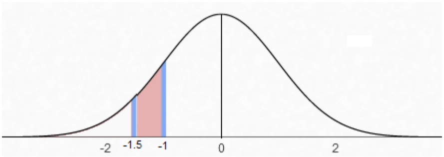


$$\begin{aligned}\therefore \quad P(21 < X < 23) &= P(0 < Z < 1.5) - P(0 < Z < 0.5)\\ &= 0.4332 - 0.1915\\ &= 0.2417\end{aligned}$$

(g)

$$\begin{aligned}P(17 < X < 18) &= P\left(\frac{17-20}{2} < \frac{X-20}{2} < \frac{18-20}{2}\right)\\ &= P(-1.5 < Z < -1)\end{aligned}$$

which is the area under the standard normal curve from -1.5 to -1 as shown in the figure below.



$$
\begin{aligned}
\therefore \quad P(17 < X < 18) &= P(1 < Z < 1.5) \\
& \text{(since, symmetric)} \\
&= P(0 < Z < 1.5) - P(0 < Z < 1) \\
&= 0.4332 - 0.3413 \\
&= 0.0919
\end{aligned}
$$

Exercises

1. X is a normal random variable with mean 40 and standard deviation 4. Find the probability that the value taken by X is
   - (a) less than 50
   - (b) greater than 50
   - (c) less than 30
   - (d) greater than 30
   - (e) between 41 and 46
   - (f) between 38 and 44

   (g) between 35 and 38
   (h) between 35 and 45
   (i) less than 34 and greater than 44

2. Heights of students is normally distributed with mean 164cms and standard deviation 4cms. Find the probability that the height of students is
   (a) more than 176 cms
   (b) less than 162 cms
   (c) between 160 and 175 cms

Median

Since normal distribution is symmetric, mean and median are equal which is equal to μ.

OR

Median M is that value of the random variable such that

$\int_{-\infty}^{M} f(x)dx = \frac{1}{2}$.

$$\text{i.e.,} \quad \int_{-\infty}^{M} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{2}$$

$$\text{i.e.,} \quad \int_{-\infty}^{\frac{M-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} dz = \frac{1}{2}; \qquad \text{where } z = \frac{x-\mu}{\sigma}$$

$$\text{i.e.,} \quad \int_{-\infty}^{\frac{M-\mu}{\sigma}} f(z)dz = \frac{1}{2}$$

But, for a standard normal curve,

$$\int_{-\infty}^{0} f(z)dz = \frac{1}{2}.$$

Hence, we have,

$$\frac{M-\mu}{\sigma} = 0 \;\Rightarrow\; M = \mu.$$

Mode

Mode is the value of the random variable which maximises the p.d.f. So for a continuous distribution it is the solution 'x_0' of

$f'(x) = 0$ if $f''(x_0) < 0$.
Here,

$$\begin{aligned} f(x) &= \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= c\,.\,e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \text{where } c = \frac{1}{\sigma\sqrt{2\pi}} \\ \ln f(x) &= \ln c - \frac{(x-\mu)^2}{2\sigma^2} \end{aligned}$$

Differentiating both sides we get,

$$\begin{aligned} \frac{f'(x)}{f(x)} &= -\frac{2(x-\mu)}{2\sigma^2} \\ &= -\frac{(x-\mu)}{\sigma^2} \qquad (1.2.1) \\ f'(x) = 0 \quad &\Rightarrow \quad x - \mu = 0 \\ &\Rightarrow \quad x = \mu \end{aligned}$$

From equation (1.2.1), we have

$$f'(x) = -\frac{(x-\mu)}{\sigma^2} f(x)$$

$$f''(x) = -\frac{1}{\sigma^2}\left[f(x).1 + (x-\mu)f'(x)\right]$$

When $x = \mu$,

$$\begin{aligned} f''(x) &= -\frac{1}{\sigma^2}[f(x)]_{x=\mu} \\ &= -\frac{1}{\sigma^2} \times \frac{1}{\sigma\sqrt{2\pi}} \\ &< 0 \end{aligned}$$

Therefore, Mode $= \mu$.

Quartile Deviation

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

where Q_1 and Q_3 can be obtained by solving

$$\int_{-\infty}^{Q_1} f(x)dx = 0.25$$

and

$$\int_{-\infty}^{Q_3} f(x)dx = 0.75$$

respectively.

Let $z = \frac{Q_1 - \mu}{\sigma}$. Then,

$$\begin{aligned}
\int_{-\infty}^{Q_1} f(x)dx &= 0.25 \Rightarrow \int_{-\infty}^{z_1} f(z)dz = 0.25, \\
&\qquad\qquad \text{where, } z_1 = \frac{Q_1 - \mu}{\sigma} \\
&\Rightarrow \int_{-z_1}^{\infty} f(z)dz = 0.25 \\
&\Rightarrow \int_{0}^{\infty} f(z)dz - \int_{0}^{-z_1} f(z)dz = 0.25 \\
&\Rightarrow 0.5 - \int_{0}^{-z_1} f(z)dz = 0.25 \\
&\Rightarrow \int_{0}^{-z_1} f(z)dz = 0.5 - 0.25 = 0.25 \\
&\Rightarrow -z_1 = 0.6745 \\
&\Rightarrow z_1 = -0.6745 \\
&\Rightarrow Q_1 = \mu - 0.6745\sigma \qquad (1.2.2)
\end{aligned}$$

Similarly,

$$\int_{-\infty}^{Q_3} f(x)dx = 0.75 \Rightarrow \int_{-\infty}^{z_2} f(z)dz = 0.75,$$

$$\text{where, } z_2 = \frac{Q_3 - \mu}{\sigma}$$

$$\Rightarrow \int_{-\infty}^{0} f(z)dz + \int_{0}^{z_2} f(z)dz = 0.75$$

$$\Rightarrow 0.5 + \int_{0}^{z_2} f(z)dz = 0.75$$

$$\Rightarrow \int_{0}^{z_2} f(z)dz = 0.75 - 0.5 = 0.25$$

$$\Rightarrow z_2 = 0.6745$$

$$\Rightarrow Q_3 = \mu + 0.6745\sigma \qquad (1.2.3)$$

Hence, from equations (1.2.2) and (1.2.3) we get,

$$\begin{aligned} Q.D. &= \frac{Q_3 - Q_1}{2} \\ &= \frac{(\mu + 0.6745\sigma) - (\mu - 0.6745\sigma)}{2} \end{aligned}$$

$$
\begin{aligned}
&= 0.6745\sigma \\
&\approx \frac{2}{3}\sigma
\end{aligned}
$$

Normal Distribution as a Limiting Form of Binomial Distribution

Theorem 1.2.3. *Binomial distribution tends to normal distribution under the following conditions*

1. *n is large ($n \to \infty$)*
2. *neither p nor q is very small*

Proof. Let $X \sim B(n, p)$. Then,

$$f_X(x) = \binom{n}{x} p^x q^{n-x} \quad ; \quad x = 0, 1, 2, \ldots, n$$

where $0 < p < 1$ and $p + q = 1$. Therefore,

$$E(X) = np$$

$$V(X) = npq$$

and

$$M_X(t) = (q + pe^t)^n$$

Define,

$$\begin{aligned} Z &= \frac{X - E(X)}{\sqrt{V(X)}} \\ &= \frac{X - np}{\sqrt{npq}} \\ &= \frac{X - \mu}{\sigma}, \qquad \text{say.} \end{aligned}$$

Now,

$$\begin{aligned} M_Z(t) &= M_{\frac{X-\mu}{\sigma}}(t) \\ &= e^{-\frac{\mu t}{\sigma}} M_X\left(\frac{t}{\sigma}\right) \\ &= e^{-\frac{\mu t}{\sigma}} \left(q + pe^{t/\sigma}\right)^n \end{aligned}$$

Therefore,

$$\ln M_Z(t) = -\frac{\mu t}{\sigma} + n \ln\left(q + pe^{t/\sigma}\right)$$

$$
\begin{aligned}
&= -\frac{\mu t}{\sigma} + n \ln\left[q + p\left(1 + \frac{t/\sigma}{1!} + \frac{(t/\sigma)^2}{2!} + \dots\right)\right] \\
&= -\frac{\mu t}{\sigma} + n \ln\left[q + p + p\left(\frac{t}{\sigma} + \frac{t^2}{2\sigma^2} + \dots\right)\right] \\
&= -\frac{\mu t}{\sigma} + n \ln\left[1 + p\left(\frac{t}{\sigma} + \frac{t^2}{2\sigma^2} + \dots\right)\right] \\
&= -\frac{\mu t}{\sigma} + n \left[p\left(\frac{t}{\sigma} + \frac{t^2}{2\sigma^2} + \dots\right)\right. \\
&\qquad\qquad \left. -\frac{p^2}{2}\left(\frac{t}{\sigma} + \frac{t^2}{2\sigma^2} + \dots\right)^2 + \dots\right] \\
&= -\frac{\mu t}{\sigma} + n \left[\frac{pt}{\sigma} + \frac{pt^2}{2\sigma^2} - \frac{p^2 t^2}{2\sigma^2} + O\left(\frac{1}{n^{3/2}}\right)\right] \\
&= -\frac{\mu t}{\sigma} + \frac{npt}{\sigma} + \frac{npt^2}{2\sigma^2}(1-p) + O\left(\frac{1}{n^{1/2}}\right) \\
&= -\frac{\mu t}{\sigma} + \frac{npt}{\sigma} + \frac{npq}{2\sigma^2}t^2 + O\left(\frac{1}{n^{1/2}}\right) \\
&= -\frac{\mu t}{\sigma} + \frac{\mu t}{\sigma} + \frac{\sigma^2}{2\sigma^2}t^2 + O\left(\frac{1}{n^{1/2}}\right), \\
&\qquad\qquad \text{since } np = \mu \text{ and } npq = \sigma^2 \\
&= \frac{t^2}{2} + O\left(\frac{1}{n^{1/2}}\right) \longrightarrow \frac{t^2}{2} \quad \text{as} \quad n \to \infty
\end{aligned}
$$

$$\therefore \quad M_Z(t) \quad = \quad e^{\frac{t^2}{2}}$$

This is the m.g.f. of a standard normal variate. So $Z \longrightarrow N(0,1)$ as $n \to \infty$.

i.e., $\frac{X-np}{\sqrt{npq}} \longrightarrow N(0,1)$ as $n \to \infty$.

i.e., $X \longrightarrow N(np, npq)$ as $n \to \infty$.

i.e., B.D. tends to N.D. as $n \to \infty$. □

Remark 1.2.6. *The above result can be applied for calculating binomial probabilities when* n *is large.*

Remark 1.2.7. *The normal distribution can also be obtained as a limiting form of Poisson distribution with* $\lambda \to \infty$*.*

Continuity Correction

Continuity correction is an adjustment done while approximating a discrete random variable with a continuous random variable, like approximating Binomial or Poisson random variable with normal random variable. Hence, while calculating the probability of a discrete random variable using normal approximation, correction factor should be applied. This can be done by subtracting -0.5 from the lower limit and adding 0.5 to the upper limit. For example: if $X \sim B(n, p)$, then

$$P(X \leq x) = P(X < x + 1) = P(Y < x + \frac{1}{2}),$$

Where $Y \sim N(np, npq)$.

Exercises

1. A fair coin is tossed 100 times. Calculate the probability of getting 55 to 70 heads using normal approximation.

Fitting of Normal Distribution

When a given frequency distribution is approximated by a normal distribution, we say that normal distribution is fitted to the given data.

In order to fit normal distribution to the given data we calculate the estimates of population mean, $\hat{\mu}$ and population standard deviation, $\hat{\sigma}$ from the given data. Then, the normal curve fitted to the given data is given by

$$f(x) = \frac{1}{\hat{\sigma}\sqrt{2\pi}}\, e^{-\frac{(x-\hat{\mu})^2}{2\hat{\sigma}^2}}; \quad -\infty < x < \infty.$$

To calculate the expected normal frequencies we first find the standard normal variates corresponding to the 'lower limits' of each of the class intervals. i.e., we compute $z_i = \frac{x_i - \hat{\mu}}{\hat{\sigma}}$ where x_i is the lower limit of the, i^{th} class interval. Then, the areas under the normal curve, to the left of $Z = z_i$ $(F(z_i))$, are computed from the tables. Finally, the area for the successive class intervals are obtained by subtraction. Hence, the area for the i^{th} class is given by $F(z_{i+1}) - F(z_i)$; $i = 1, 2, \ldots, n$. Multiplying these areas by N we get the expected normal frequencies.

Exercises

1. Obtain the normal probability curve that may be fitted to the following distribution and hence obtain the corresponding theoretical frequencies.

Class	up to 59	60-69	70-79	80-89	90-99
f	0	6	30	70	230
Class	100-109	110-119	120-129	130-139	140-149
f	350	220	65	25	4

Solution

Class	mid x	f	$d = \frac{x-104.5}{10}$	fd	fd^2
up to 59	-	0	-	0	0
60-69	64.5	6	-4	-24	96
70-79	74.5	30	-3	-90	270
80-89	84.5	70	-2	-140	280
90-99	94.5	230	-1	-230	230
100-109	104.5	350	0	0	0
110-119	114.5	220	1	220	220
120-129	124.5	65	2	130	260
130-139	134.5	25	3	75	225
140-149	144.5	4	4	16	64
Total		1000		-43	1645

$$\begin{aligned} \hat{\mu} = \bar{x} &= A + \frac{\sum fd}{N}.c \\ &= 104.5 + \frac{-43}{1000} \times 10 \\ &= 104.07 \end{aligned}$$

$$\begin{aligned}\hat{\sigma} = s &= c \quad \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2} \\ &= 10 \times \sqrt{\frac{1645}{1000} - \left(\frac{-43}{1000}\right)^2} \\ &= 12.82\end{aligned}$$

Lower limit of actual class (x)	z	$F(z)$	Area	Area × N	Theoretical frequency
-	-	-	0.0003	0.3	0
59.5	-3.48	0.0003	0.0032	3.2	3
69.5	-2.70	0.0035	0.0239	23.9	24
79.5	-1.92	0.0274	0.0997	99.7	100
89.5	-1.14	0.1271	0.2323	232.3	232
99.5	-0.36	0.3594	0.3034	303.4	303+1
109.5	0.42	0.6628	0.2221	222.1	222
119.5	1.20	0.8849	0.0912	91.2	91
129.5	1.98	0.9761	0.0210	21.0	21
139.5	2.76	0.9971	0.0027	2.7	3
149.5	3.54	0.9998	0.0002	0.2	0
Total					1000

1.2.5 Log-normal Distribution

A log-normal (or lognormal) distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. Thus, if the random variable X is log-normally distributed, then $Y = \ln(X)$ has a normal distribution. Equivalently, if Y has a normal distribution, then the exponential function of Y, $X = \exp(Y)$, has a log-normal distribution. A random variable which is log-normally distributed takes only positive real values.

Log-normal distribution has several applications in the field of science as well as social science. It is useful in studying the inequality in income distribution. The hourly median power of received radio signals transmitted between two places follows log-normal distribution.

Derivation

Let $Y \sim N(\mu, \sigma^2)$ and $X = g(Y) = e^Y$. Since X is continuous and $x' = g'(y) = e^y > 0$ for all y. The p.d.f. of X is given by

$$f_X(x) \quad = \quad f_Y(g^{-1}(x))\left|\frac{d}{dx}g^{-1}(x)\right|$$

$$\begin{aligned} x = g(y) = e^y \quad &\Rightarrow \quad y = g^{-1}(x) = \ln x \\ &\Rightarrow \quad \frac{d}{dx} g^{-1}(x) = \frac{1}{x} \end{aligned}$$

and

$$\begin{aligned} f_Y(g^{-1}(x)) &= f_Y(\ln x) \\ &= \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \end{aligned}$$

Therefore,

$$\begin{aligned} f_X(x) &= \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \times \frac{1}{x} \\ &= \frac{1}{x\sigma\sqrt{2\pi}}\, e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, \quad x > 0 \end{aligned}$$

Definition 1.2.6. *A random variable X is said to follow log-normal distribution with parameters μ and σ^2 if its p.d.f. is given by*

$$f_X(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}}\, e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} & ; \quad 0 < x < \infty \\ 0 & ; \quad \textit{otherwise,} \end{cases},$$

where $-\infty < \mu < \infty$ *and* $\sigma > 0$.

When X is distributed log-normally with parameters μ and σ^2, we write $X \sim \Lambda(\mu, \sigma^2)$.

Moments

The r^{th} raw moment is given by

$$\begin{aligned} E(X^r) &= E(e^{Yr}) \\ &= e^{\mu r + \frac{r^2\sigma^2}{2}}, \quad \text{since } \; Y \sim N(\mu, \sigma^2) \end{aligned}$$

Mean

$$E(X) = e^{\mu + \frac{\sigma^2}{2}}$$

Variance

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

$$E(X^2) = e^{2\mu+2\sigma^2}$$

Therefore,

$$\begin{aligned} V(X) &= e^{2\mu+2\sigma^2} - \left[e^{\mu+\frac{\sigma^2}{2}}\right]^2 \\ &= e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} \\ &= e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) \end{aligned}$$

Properties

1. If $X \sim \Lambda(\mu, \sigma^2)$, then $E(\ln X) = \mu$ and $V(\ln X) = \sigma^2$.

2. If $X \sim \Lambda(\mu, \sigma^2)$, then $\frac{1}{X} \sim \Lambda(-\mu, \sigma^2)$.

3. If $X_1 \sim \Lambda(\mu_1, \sigma_1^2)$ and $X_2 \sim \Lambda(\mu_2, \sigma_2^2)$ and if X_1 and X_2 are independent, then $X_1 X_2 \sim \Lambda(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

1.2.6 Beta Distribution

Beta Distribution of First Kind

Beta distribution of first kind is a continuous probability distribution defined on $[0, 1]$.

Definition 1.2.7. *A random variable X is said to have a beta distribution of 1^{st} kind if its p.d.f. is given by*

$$f_X(x) = \begin{cases} \frac{1}{\beta(m,n)} x^{m-1}(1-x)^{n-1} & ; \quad 0 < x < 1 \\ 0 & ; \quad \textit{otherwise,} \end{cases}$$

where $m > 0$ and $n > 0$ are the parameters.

In this case we write $X \sim \beta_1(m, n)$.

Note: Being a p.d.f., we know that

$$\begin{aligned} \int_0^1 f_X(x)dx &= 1 \\ \text{i.e.,} \int_0^1 \frac{1}{\beta(m,n)} x^{m-1}(1-x)^{n-1}dx &= 1 \\ \text{i.e.,} \int_0^1 x^{m-1}(1-x)^{n-1}dx &= \beta(m,n), \end{aligned}$$

which is called the beta function. Note that, $\beta(m,n) = \frac{\Gamma m \Gamma n}{\Gamma(m+n)}$.

Remark 1.2.8. *Beta distribution reduces to* $U(0,1)$ *distribution if* $m = n = 1$.

Moments

Mean

$$\begin{aligned} E(X) &= \int_0^1 x \frac{1}{\beta(m,n)} x^{m-1}(1-x)^{n-1} dx \\ &= \frac{m}{m+n} \end{aligned}$$

Variance

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

$$\begin{aligned} E(X^2) &= \int_0^1 x^2 \frac{1}{\beta(m,n)} x^{m-1}(1-x)^{n-1} dx \\ &= \frac{m(m+1)}{(m+n)(m+n+1)} \end{aligned}$$

Therefore,

$$\begin{aligned} V(X) &= \frac{m(m+1)}{(m+n)(m+n+1)} - \left[\frac{m}{m+n}\right]^2 \\ &= \frac{mn}{(m+n)^2(m+n+1)} \end{aligned}$$

Beta Distribution of Second Kind

Beta distribution of the second kind is a continuous probability distribution defined on $(0, \infty)$. This distribution is also known as beta prime distribution or inverted beta distribution.

Definition 1.2.8. *A random variable* X *is said to have a beta distribution of* 2^{nd} *kind if its p.d.f. is given by*

$$f_X(x) = \begin{cases} \frac{1}{\beta(m,n)} \frac{x^{m-1}}{(1+x)^{m+n}} & ; \quad 0 < x < \infty \\ 0 & ; \quad \textit{otherwise,} \end{cases}$$

where $m > 0$ *and* $n > 0$ *are the parameters.*

In this case we write $X \sim \beta_2(m,n)$ or $X \sim \beta'(m,n)$.

Remark 1.2.9. *If* $X \sim \beta_2(m,n)$ *then* $Y = \frac{1}{1+X} \sim \beta_1(m,n)$.

Moments

Mean

$$E(X) = \frac{m}{n-1}$$

Variance

$$V(X) = \frac{m(m+n-1)}{(n-1)^2(n-2)}$$

1.2.7 Pareto Distribution

Pareto distribution, named after Vilfredo Pareto, is a continuous probability distribution which is skewed and heavy-tailed. Originally used to model the distribution of wealth in a society. The basis of the distribution is that only a few people have very high wealth while a high proportion of a population have low wealth.

Definition 1.2.9. *A random variable X is said to follow pareto* *distribution with parameters* α *and* β *if its p.d.f is given by*

$$f_X(x) = \tfrac{\alpha\ \beta^\alpha}{x^{\alpha+1}} \quad ; \quad x > \beta \ ,$$

where $\alpha > 0$ *and* $\beta > 0$*.*

We write $X \sim Pareto(\alpha, \beta)$ if X follows a pareto distribution with parameters α and β.

The distribution function is given by $F_X(x) = 1 - (\frac{\beta}{x})^\alpha$, $x > \beta$, $\beta > 0$, $\alpha > 0$

Moments

Mean

$$E(X) \quad = \quad \frac{\alpha\beta}{\alpha - 1}; \quad \text{if } \alpha > 1.$$

Variance

$$V(X) \quad = \quad \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)}; \quad \text{if } \alpha > 2.$$

1.2.8 Cauchy Distribution

The Cauchy distribution, named after Augustin Cauchy, is a continuous probability distribution. It is also known, especially among physicists, as the Lorentz distribution, after Hendrik Lorentz. The curve resemble the normal distribution family of curves. While the resemblance is there, it has a taller peak than a normal. And unlike the normal distribution, its fat tails decay much more slowly.

Definition 1.2.10. *A random variable X is said to follow standard Cauchy distribution if its p.d.f is given by*

$$f_X(x) = \tfrac{1}{\pi(1+x^2)} \quad ; \quad -\infty < x < \infty \ .$$

More generally, Cauchy distribution with parameter θ and λ have the p.d.f.

$$f_X(x) = \tfrac{\lambda}{\pi[\lambda^2+(x-\theta)^2)]} \quad ; \quad -\infty < x < \infty.$$

where $-\infty < \theta < \infty$ and $\lambda > 0$.

We represent a Cauchy distribution with parameters θ and λ as $C(\lambda, \theta)$ and standard Cauchy as $C(1, 0)$. If $X \sim C(\lambda, \theta)$, then $\frac{X-\theta}{\lambda} \sim C(1, 0)$.

The Cauchy distribution does not have finite moments of order greater than or equal to one. That is, its expectation and other moments does not exist. The median and mode do exist and are equal. However the central limit theorem (CLT) does not work for the limiting distribution of mean.

MODULE

TWO

Limit Theorems

Limit theorems are about convergence properties of sequence of random variables. It includes, the weak law of large numbers (WLLN), the strong law of large numbers (SLLN) and the central limit theorem (CLT). The modes of convergence for the three limit theorems are convergence in probability, convergence almost surely and convergence in distribution respectively. Before going through the several modes of convergence and limit theorems, let us go through Chebychev's inequality which is used to prove the weak law of large numbers.

2.1 Chebyshev's Inequality

Chebyshev's inequality (also called the Bienaym - Chebyshev inequality) guarantees that, for a wide class of probability distributions, no more than $1/k^2$ of the distribution's values can be more than k standard deviations away from the mean (or equivalently, at least $1 - 1/k^2$ of the distribution's values are within k standard deviations from the mean). The inequality is about the range of standard deviations around the mean and is a particular case of Markov's inequality. The theorem was first stated without proof by Bienaym in 1853 and later proved by Chebyshev in 1867. The inequality has great utility because it can be applied to any probability distribution in which the mean and variance are defined.

Theorem 2.1.1. *If X is a random variable with $E(X) = \mu$ and $V(X) = \sigma^2$ exists, then for any $k > 0$,*

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

or

$$P\{|X-\mu| \le k\sigma\} \ge 1 - \frac{1}{k^2}.$$

Proof. We have,

$$\begin{aligned}
\sigma^2 &= E[X-E(X)]^2 \\
&= E[X-\mu]^2 \\
&= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\ dx \\
&= \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\ dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)\ dx \\
&\qquad + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)\ dx \\
&\ge \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\ dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)\ dx, \\
&\qquad \text{since } (x-\mu)^2 f(x) \text{ is non-negative}
\end{aligned}$$

Now, since $(x-\mu)^2 \ge k^2\sigma^2$ for $x-\mu \le -k\sigma$ or $x-\mu \ge k\sigma \Rightarrow$

$x \leq \mu - k\sigma$ or $x \geq \mu + k\sigma$, it follows that

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)\, dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)\, dx,$$

and hence,

$$\begin{aligned} \frac{1}{k^2} &\geq \int_{-\infty}^{\mu-k\sigma} f(x)\, dx + \int_{\mu+k\sigma}^{\infty} f(x)\, dx, \\ &\geq P(X \leq \mu - k\sigma) + P(X \geq \mu + k\sigma) \\ &\geq P(X - \mu \leq -k\sigma) + P(X - \mu \geq k\sigma) \\ &\geq P(|X - \mu| \geq k\sigma) \end{aligned}$$

That is,

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

which is the same as

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

□

Note: Replacing $k\sigma$ with ϵ, where $k = \epsilon/\sigma$, we have another form of the Chebyshev's inequality:

$$P\{|X-\mu| \geq \epsilon\} \leq \frac{\sigma^2}{\epsilon^2}$$

or

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

where $\epsilon > 0$

Importance of Chebyshev's Inequality

The importance of Chebyshev's inequality lies in its generality. No assumption on the nature of the random variable and the distribution or density of X is made other than that it has a finite variance. This inequality formalises the intuitive meaning of variance: If σ is small there is a high probability for getting a value close to the mean, and if σ is large there is a high probability for getting values farther away from the mean. (And for some σ, the probability of the random variable X taking values outside k times the standard deviation from mean goes down like one over k^2.) This form of the Chebyshev's inequality is useful in establishing the weak law of large numbers.

One limitation with the Chebyshev's inequality is that when

$k < 1$, the inequality does not give any significant result. Also we can see that, generally, the bounds given by the Chebyshev's inequality cannot be improved.

Solved Problems

1. Find the lower bound for $P(-1 \leq X \leq 7)$, where X is a random variable with $E(X) = 3$ and $V(X) = 4$.

Solution:
Given $E(X) = \mu = 3$ and $V(X) = \sigma^2 = 4$. We have to find the lower bound for $P(-1 \leq X \leq 7)$.

$$\begin{aligned} P(-1 \leq X \leq 7) &= P(-1-3 \leq X-3 \leq 7-3) \\ &= P(-4 \leq X-3 \leq 4) \\ &= P(|X-3| \leq 4) \end{aligned}$$

Since the variance is finite, the lower bound of the given probability can be obtained by the Chebychev's inequality

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

Substituting $\mu = 3$, $\sigma^2 = 4$ and $\epsilon = 4$ we get

$$P(-1 \leq X \leq 7) \quad = \quad P(|X-3| \leq 4) \geq 1 - \frac{4}{16} = \frac{3}{4}.$$

OR

If using the Chebychev's inequality of the form

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

on substituting $\mu = 3$ and $\sigma = 2$ we get,

$$P(|X-3| \leq 2k) \geq 1 - \frac{1}{k^2}$$

Taking $2k = 4$, $k = 2$ and $\frac{1}{k} = \frac{1}{2}$. Therefore,

$$P(-1 \leq X \leq 7) \quad = \quad P(|X-3| \leq 4) \geq 1 - \frac{1}{4} = \frac{3}{4}.$$

Note: If probability distribution is given, first obtain mean and variance and then apply Chebychev's inequality.

Exercises

1. Find the least value of probability $P(1 \leq X \leq 7)$, where X is a random variable with $E(X) = 4$ and $V(X) = 4$.

2. A random variable X has mean 50 and variance 81. Use Chebyshev's inequality to obtain appropriate bounds for

   (a) $P\{|X - 50| \geq 15\}$ and

   (b) $P\{|X - 50| < 18\}$.

3. Let X have p.d.f. $f(x) = \begin{cases} \frac{1}{2\sqrt{3}}, & -\sqrt{3} < X < \sqrt{3} \\ 0, & \text{elsewhere} \end{cases}$
   Show by Chebychev's inequality $P\{|X| \geq \frac{3}{2}\}$ has an upper bound $\frac{4}{9}$ where as the true value is $1 - \frac{\sqrt{3}}{2}$.

4. Suppose that the length of an electronic device has p.d.f. $f(x) = e^{-x},\ x > 0$. Determine $P\{|X - 1| \geq 2\}$

   (a) exactly

   (b) approximately using chebychev's inequality.

5. Let X be a random variable taking values $-1,\ 0,\ +1$ with probabilities $\frac{1}{8},\ \frac{6}{8},\ \frac{1}{8}$ respectively. Find using Chebychev's inequality, the upper bound of $P\{|X| \geq 1\}$.

6. For the geometric distribution $f(x) = 2^{-x},\ x = 1, 2, 3, \ldots$ prove that Chebychev's inequality gives $P\{|X-2| \leq 2\} > \frac{1}{2}$ while the actual probability is $\frac{15}{16}$.

7. Two dice are thrown together. Let X denote the sum of the numbers shown up by the two dice. Show by using Chebychev's inequality, that $P\{|X - 7| < 4\} \geq \frac{61}{96}$. Also compute the actual probability.

8. The probability of survival in case of cancer is found to be 0.8. One hundred people are attacked by the disease in a particular area. If X denote the number of survivals, assuming X follows binomial distribution with $n = 100$ and $p = 0.8$, find an upper bound for the probability that the number of survivals will be either less than 68 or greater than 92.

9. How many times a fair coin must be tossed in order to ensure that in 90 percentage of the cases the observed ratio of the number of heads to the number of tosses will lie between 0.4 to 0.6.

10. An unbiased coin is tossed 100 times. Show by using Chebychev's inequality that the probability that the number of heads will be between 30 and 70 is greater than 0.93.

2.2 Modes of Convergence

2.2.1 Convergence in Distribution

Definition 2.2.1. *Let* $\{F_n\}$ *be a sequence of distribution functions. If there exists a d.f.* F *such that as* $n \to \infty$, $F_n(x) \to F(x)$ *at every point* x *at which* F *is continuous, we say that* F_n *converges in law (or, weakly), to* F*, and we write* $F_n \xrightarrow{w} F$.

If $\{X_n\}$ is a sequence of random variables and $\{F_n\}$ is the corresponding sequence of d.f.s, we say that X_n converges in distribution (or law) to X if there exists an random variable X with d.f. F such that $F_n \xrightarrow{w} F$. We write $X_n \xrightarrow{L} X$.

Note: It is quite possible for a given sequence of d.f.s to converge to a function that is not a d.f.

2.2.2 Convergence in Probability

Definition 2.2.2. *Let* $\{X_n\}$ *be a sequence of random variables defined on some probability space* (Q, S, P)*. We say that the sequence* $\{X_n\}$ *converges in probability to a constant 'a' if for every* $\epsilon > 0$, $P\{|X_n - a| > \epsilon\} \to 0$ *as* $n \to \infty$*. We write* $X_n \xrightarrow{P} a$.

Definition 2.2.3. *Let* $\{X_n\}$ *be a sequence of random variables* *defined on some probability space* (Q, S, P). *We say that the sequence* $\{X_n\}$ *converges in probability to the random variable* X *if for every* $\epsilon > 0$, $P\{|X_n - X| > \epsilon\} \to 0$ *as* $n \to \infty$. *We write* $X_n \xrightarrow{P} X$.

This is also called stochastic convergence.

Remark 2.2.1. *The definition (2.2.3) says nothing about the convergence of the sequence of random variables* $\{X_n\}$ *to the random variable* X *in the sense in which it is understood in real analysis. Thus* $X_n \xrightarrow{P} X$ *does not imply that given* $\epsilon > 0$, *we can find an* N *such that* $|X_n - X| < \epsilon$ *for* $n > N$. *Definition (2.2.3) speaks only of the convergence of the sequence of probabilities* $P\{|X_n - X| > \epsilon\}$ *to 0.*

2.3 Laws of Large Numbers

According to the laws of large numbers, 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 to the expected value as more trials are performed. The weak law of large numbers and the strong law of large numbers differ in the mode of convergence. Bernoulli's law of large numbers, Khintchine's law of large numbers, Kolmogorov's law of large numbers etc.

are some of the important weak laws of large numbers. One weak law differs from the other only in the assumptions used about the random variables.

2.3.1 Weak Law of Large Numbers

The weak laws of large numbers discuss the convergence in probability of the sequence of partial sums or partial averages. That is, as the sample size n grows to infinity, the probability that the sample mean $\bar{X}$ differs from the population mean μ by some small amount epsilon is equal to 0. Suppose an experiment can be performed as frequently as we want. Assume these performances are independent and that some probability measure describes this performance. Let A be any desired event that may or may not occur on each performance of the experiment. The probability of A occurring is $P(A)$, the same for each repetition. For any fixed number n of repetitions of the experiment, let $n(A)$ be the number of times that event A occurs. Thus $n(A)/n$ is the proportion of repetitions of the experiment for which event A occurred. The intuitive result refer to the value of proportion $n(A)/n$, as n increases without limit. It seems reasonable that this proportion should in fact converge to $P(A)$, the probability that A occurs for each performance of the experiment. This phenomenon was earlier referred to as ‘statistical regularity’; one way of trying to show that this convergence should occur is given by the weak law of large numbers (WLLN), which can be stated as the following theorem and in particular, is given by Bernouli’s weak law of large numbers.

Theorem 2.3.1. *Let* $X_1, X_2, \ldots, X_n$ *be a sequence of independent and identically distributed random variables with* $E(X_i) = \mu$ *and* $V(X_i) = \sigma^2$, $i = 1, 2, \ldots, n$. *Define* $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. *Then for any* $\epsilon > 0$, *however small,*

$$P\{|\bar{X}_n - \mu| \geq \epsilon\} \to 0 \quad as \quad n \to \infty$$

Proof. Given $E(X_i) = \mu$ and $V(X_i) = \sigma^2$, $i = 1, 2, \ldots, n$. Therefore,

$$E(\bar{X}_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \frac{n\mu}{n} = \mu.$$

$$V(\bar{X}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Now, from Chebychev’s inequality

$$P\{|\bar{X}_n - \mu| \geq k\sigma_{\bar{X}_n}\} \leq \frac{1}{k^2}.$$

That is,

$$P\{|\bar{X}_n - \mu| \geq k\frac{\sigma}{\sqrt{n}}\} \leq \frac{1}{k^2}.$$

Choose $\epsilon = \frac{k\sigma}{\sqrt{n}}$. Then, $\frac{1}{k^2} = \frac{\sigma^2}{n\epsilon^2}$. Therefore,

$$P\{|\bar{X}_n - \mu| \geq \epsilon\} \leq \frac{\sigma^2}{n\epsilon^2} \to 0 \quad \text{as} \quad n \to \infty.$$

That is,

$$P\{|\bar{X}_n - \mu| \geq \epsilon\} \to 0 \quad \text{as} \quad n \to \infty.$$

□

Note: This result describes the limiting value for a sequence of probabilities as distinguished from the limiting values of $\bar{X}_n$ (if it exists). That is, if we were to compute the difference $|\bar{X}_n - \mu|$ for successively larger values of n, the probability that this value exceeds for any fixed ϵ gets smaller and smaller and converges to zero. This does not necessarily implied that for any given realised sequence $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots$ the actual observed difference $|\bar{X}_n - \mu|$ must converge to 0. A stronger notion of convergence is needed for this to be true, which is beyond our scope of study.

A more general form of the theorem are as follows: (for further reading)

Theorem 2.3.2 (Chebychev's Weak Law of Large Numbers). *Let* $X_1, X_2, \ldots, X_n$ *be a sequence of pairwise uncorrelated random variables with* $E(X_i) = \mu_i$ *and* $V(X_i) = \sigma_i^2$, $i = 1, 2, \ldots, n$. *If* $n^{-2}\sum_{i=1}^{n}\sigma_i^2 \to 0$, *as* $n \to \infty$, *then*

$$\bar{X} \xrightarrow{P} n^{-1}\sum_{i=1}^{n}\mu_i.$$

And if the random variables are identically distributed,

$$\bar{X}_n \xrightarrow{P} \mu.$$

Theorem 2.3.3 (Khintchine's Weak Law of Large Numbers). *Let* $X_1, X_2, \ldots, X_n$ *be a sequence of i.i.d. random variables with finite mean* μ. *Then, as* $n \to \infty$

$$\bar{X}_n \xrightarrow{P} \mu.$$

Note: Let $X_1, X_2, \ldots, X_n$ be a sequence of pairwise uncorrelated random variables with $E(X_i) = \mu_i$ and $V(X_i) = \sigma_i^2$, $i = 1, 2, \ldots, n$. We say that $\{X_n\}$ obeys WLLN if $\frac{B_n}{n^2} \to 0$ as $n \to \infty$, where $B_n = V(X_1 + X_2 + \ldots + X_n)$.

2.3.2 Bernoulli's Law of Large Numbers

Bernoulli's law of large numbers is a particular case of the weak law of large numbers obtained simply as an extension of Chebychev's inequality to the binomial distribution.

Theorem 2.3.4 (Bernoulli's Law of Large Numbers). *Consider n trials of a random experiment, each trial resulting in a success or failure. Let X_n be the number of success in n trials with constant probability p of success for each trial, then for any $\epsilon > 0$*

$$P\left\{\left|\frac{X_n}{n} - p\right| < \epsilon\right\} \to 1 \quad as \quad n \to \infty.$$

Proof. Here, $X_n \sim B(n,p)$. Therefore, $E(X_n) = np$ and $V(X) = npq$.

$$E\left(\frac{X_n}{n}\right) = \frac{1}{n}E(X_n) = \frac{1}{n}np = p$$

$$V\left(\frac{X_n}{n}\right) = \frac{1}{n^2}V(X_n) = \frac{1}{n^2}npq = \frac{pq}{n}$$

By Chebychev's inequality,

$$P\left\{\left|\frac{X_n}{n} - p\right| < k\sigma_{\frac{X}{n}}\right\} \geq 1 - \frac{1}{k^2}$$

That is,

$$P\left\{\left|\frac{X_n}{n}-p\right|<k\sqrt{\frac{pq}{n}}\right\}\geq 1-\frac{1}{k^2}.$$

Put, $k\sqrt{\frac{pq}{n}}=\epsilon$. Then, $\frac{1}{k^2}=\frac{pq}{n\epsilon^2}$ and

$$P\left\{\left|\frac{X_n}{n}-p\right|<\epsilon\right\}\geq 1-\frac{pq}{n\epsilon^2}.$$

But $pq\leq\frac{1}{4}$ for all $0\leq p\leq 1$. Therefore,

$$P\left\{\left|\frac{X_n}{n}-p\right|<\epsilon\right\}\geq 1-\frac{1}{4n\epsilon^2}\to 1\quad\text{as}\quad n\to\infty$$

or

$$P\left\{\left|\frac{X_n}{n}-p\right|\geq\epsilon\right\}\to 0\quad\text{as}\quad n\to\infty.$$

□

Note: Euler Maclaurin's formula

1. If m and n are natural numbers and $f(x)$ is a real or complex valued continuous function for real numbers x in the interval $[m,n]$, then the integral

$$I=\int_m^n f(x)dx$$

can be approximated by the sum (or vice versa)

$$S = f(m+1) + f(m+2) + \ldots + f(n).$$

2. In particular, when $m = 0$,

$$S = f(1) + f(2) + \ldots + f(n)$$

can be approximated by the sum (or vice versa)

$$I = \int_0^n f(x)dx.$$

That is,

$$f(1) + f(2) + \ldots + f(n) \approx \int_0^n f(x)dx.$$

Solved Problems

1. Examine whether WLLN hold for the sequence of independent random variables $\{X_k\}$, $k = 1, 2, 3, \ldots$ where $P(X_k = \pm\sqrt[3]{k}) = \frac{1}{2}$.

Solution:
$P(X_k = \sqrt[3]{k}) = \frac{1}{2}$ and $P(X_k = -\sqrt[3]{k}) = \frac{1}{2}$.

$$\begin{aligned} E(X_k) &= \sqrt[3]{k} \times \frac{1}{2} - \sqrt[3]{k} \times \frac{1}{2} \\ &= 0. \\ E(X_k^2) &= \sqrt[3]{k}^2 \times \frac{1}{2} - \sqrt[3]{k}^2 \times \frac{1}{2} \\ &= k^{2/3}. \end{aligned}$$

Therefore,

$$\begin{aligned} V(X_k) &= E(X_k^2) - [E(X_k)]^2 \\ &= k^{2/3}. \\ B_k &= V(X_1 + X_2 + \ldots + X_k) \\ &= V(X_1) + V(X_2) + \ldots + V(X_k) \\ &= 1^{2/3} + 2^{2/3} + \ldots + k^{2/3} \\ &\approx \int_0^k x^{2/3} dx \quad \text{(By Euler Maclaurin's formula)} \end{aligned}$$

$$= \left[\frac{3}{5}x^{5/3}\right]_0^k$$

$$= \frac{3}{5}k^{5/3}$$

Now,

$$\lim_{k\to\infty}\frac{B_k}{k^2} = \lim_{k\to\infty}\frac{\frac{3}{5}k^{5/3}}{k^2} = \frac{3}{5}\lim_{n\to\infty}\frac{1}{k^{1/3}} \to 0$$

as $k \to \infty$. Therefore, the sequence obeys WLLN.

Exercises

1. Let X_i assumes the values i^α and $-i^\alpha$ with equal probability. Check whether the law of large numbers hold to the independent random variables $X_1, X_2, \ldots$ for $\alpha < \frac{1}{2}$.

2. Let $\{X_n\}$ be a sequence of mutually independent and identically distributed random variables with mean μ and finite variance σ^2. If $S_n = \sum_{i=1}^n X_i$, prove that WLLN does not hold for the sequence $\{S_n\}$.

3. Examine whether the WLLN hold good for the sequence $\{X_n\}$ of the independent random variables where $P\{X_n = \frac{1}{\sqrt{n}}\} = \frac{2}{3}$ and $P\{X_n = \frac{-1}{\sqrt{n}}\} = \frac{1}{3}$.

2.4 Central Limit Theorem

'Central Limit Theorem' is a name used to describe a series of theorems saying that the sum of a large number of random variables is approximately normally distributed with mean equal to the sum of the means and variance equal to the sum of the variances, under specified conditions. In other words, the standardised sum of a large number of random variables is approximately standard normally distributed, under specified conditions.

The first attempt in this direction is on the convergence in distribution of a sequence of i.i.d. Bernoulli random variables known as the Bernoulli's central limit theorem, by DeMoivre and Laplace. Later Lévy identified that this result holds not only for i.i.d. Bernoulli random variables but also for any sequence of i.i.d. random variables. The Lévy central limit theorem is the most useful version of the celebrated central limit theorem in the i.i.d. case. By relaxing the identically distributed condition, Lindeberg obtained a set of sufficient conditions for the convergence of suitably centered and normalized S_n to the normal random variable These conditions were later proved to be necessary by Feller and is known by the name Lindeberg-Feller central limit theorem.

2.4.1 Lévy Central Limit Theorem

Theorem 2.4.1. *Let* $X_1, X_2, \ldots, X_n$ *be a sequence of i.i.d. random variables with* $E(X_i) = \mu$ *and* $V(X_i) = \sigma^2 < \infty$, $i = 1, 2, \ldots, n$. *Also let* $S_n = X_1 + X_2 + \ldots + X_n$. *Then*

$$Z = \frac{S_n - n\mu}{\sqrt{n}\sigma} \rightarrow \ N(0,1)$$

as $n \rightarrow \infty$.

Proof. Given, $X_1, X_2, \ldots, X_n$ is a sequence of i.i.d. random variables, $E(X_i) = \mu$, $V(X_i) = \sigma^2 < \infty$, $i = 1, 2, \ldots, n$ and

$$S_n = X_1 + X_2 + \ldots + X_n = \sum_{i=1}^{n} X_i.$$

Assume that $M_{X_i}(t)$ exists for $i = 1, 2, \ldots, n$. Now,

$$\begin{aligned} M_Z(t) &= M_{\frac{S_n - n\mu}{\sqrt{n}\sigma}}(t) \\ &= M_{S_n - n\mu}\left(\frac{t}{\sqrt{n}\sigma}\right) \\ &= e^{\frac{-n\mu t}{\sqrt{n}\sigma}} M_{S_n}\left(\frac{t}{\sqrt{n}\sigma}\right) \end{aligned}$$

$$= \quad e^{\frac{-\sqrt{n}\mu t}{\sigma}} M_{\sum_{i=1}^{n} X_i}\left(\frac{t}{\sqrt{n}\sigma}\right)$$

$$= \quad e^{\frac{-\sqrt{n}\mu t}{\sigma}} \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{\sqrt{n}\sigma}\right)$$

(Since X_i's are independent)

$$= \quad e^{\frac{-\sqrt{n}\mu t}{\sigma}} \left[M_X\left(\frac{t}{\sqrt{n}\sigma}\right)\right]^n$$

(Since X_i's are identically distributed)

Therefore,

$$M_Z(t) \quad = \quad e^{\frac{-\sqrt{n}\mu t}{\sigma}} \left[1 + \frac{t}{\sqrt{n}\sigma}\mu_1' + \frac{1}{2}\left(\frac{t}{\sqrt{n}\sigma}\right)^2 \mu_2' + o\left(\frac{1}{n^{3/2}}\right)\right]^n$$

where $o\left(\frac{1}{n^{3/2}}\right)$ denotes terms with $\frac{1}{n^{3/2}}$ and its higher powers.

$$\begin{aligned} \ln M_Z(t) \quad = \quad & \frac{-\sqrt{n}\mu t}{\sigma} + n \ln \left[1 + \frac{t}{\sqrt{n}\sigma}\mu_1' \right. \\ & \left. + \frac{1}{2}\left(\frac{t}{\sqrt{n}\sigma}\right)^2 \mu_2' + o\left(\frac{1}{n^{3/2}}\right)\right] \end{aligned}$$

$$
\begin{aligned}
&= \frac{-\sqrt{n}\mu t}{\sigma} \\
&\quad +n\left[\left\{\frac{t}{\sqrt{n}\sigma}\mu'_{1} + \frac{1}{2}\left(\frac{t}{\sqrt{n}\sigma}\right)^2 \mu'_{2} + o\left(\frac{1}{n^{3/2}}\right)\right\}\right. \\
&\quad \left. -\frac{1}{2}\left\{\frac{t}{\sqrt{n}\sigma}\mu'_{1} + \frac{1}{2}\left(\frac{t}{\sqrt{n}\sigma}\right)^2 \mu'_{2} + o\left(\frac{1}{n^{3/2}}\right)\right\}^2 + \dots\right] \\
&= \frac{-\sqrt{n}\mu t}{\sigma} + \frac{\sqrt{n}\mu t}{\sigma} + \frac{t^2}{2\sigma^2}\left[\mu'_{2} - (\mu'_{1})^2\right] + o\left(\frac{1}{n^{1/2}}\right); \\
&\qquad\qquad (\text{Since, } \mu'_{1} = \mu) \\
&= \frac{t^2}{2} + o\left(\frac{1}{n^{1/2}}\right); \quad (\text{Since, } \mu'_{2} - (\mu'_{1})^2 = \sigma^2) \\
&\to \frac{t^2}{2} \quad \text{as } n \to \infty
\end{aligned}
$$

Therefore,

$$
M_Z(t) = e^{\frac{t^2}{2}} \quad \text{as } n \to \infty,
$$

which is the m.g.f. of a standard normal random variable. Hence, $Z \to N(0,1)$ as $n \to \infty$. □

Assumptions of Lévy Central Limit Theorem

1. random variables are i.i.d.

2. Mean and variance both exist and finite.

Corollary 2.4.1. *Asymptotically,* $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

Corollary 2.4.2. *Asymptotically,* $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

Corollary 2.4.3. *Asymptotically,* $S_n \sim N(n\mu, n\sigma^2)$.

Corollary 2.4.4. *Asymptotically,* $\frac{S_n - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$.

Note: The above corollaries can be used when $n \geq 30$.

2.4.2 De-Moivre's-Laplace Central Limit Theorem

Theorem 2.4.2. *Let* $X_1, X_2, \ldots, X_n$ *be a sequence of i.i.d. Bernoulli random variables for* $i = 1, 2, \ldots, n$. *Also let* $S_n = X_1 + X_2 + \ldots + X_n$. *Then*

$$Z = \frac{S_n - np}{\sqrt{npq}} \rightarrow \ N(o,1) \quad as \quad n \rightarrow \infty.$$

Proof. Since, X_i; $i = 1, 2, \ldots, n$ are i.i.d. Bernoulli random variables, $M_{X_i}(t) = (q + pe^t)$. Therefore,

$$M_{S_n}(t) = (q + pe^t)^n.$$

Also, $E(S_n) = np$ and $V(S_n) = npq$. Let

$$Z = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - np}{\sqrt{npq}}.$$

$$\begin{aligned}
M_Z(t) &= M_{\frac{S_n - np}{\sqrt{npq}}}(t) \\
&= e^{\frac{-npt}{\sqrt{npq}}} M_{S_n}\left(\frac{t}{\sqrt{npq}}\right) \\
&= e^{\frac{-npt}{\sqrt{npq}}}\left(q + pe^{\left(\frac{t}{\sqrt{npq}}\right)}\right)^n \\
&= \left[e^{\frac{-pt}{\sqrt{npq}}}\left(q + pe^{\left(\frac{t}{\sqrt{npq}}\right)}\right)\right]^n \\
&= \left(qe^{\frac{-pt}{\sqrt{npq}}} + pe^{\frac{(1-p)t}{\sqrt{npq}}}\right)^n \\
&= \left(qe^{\frac{-pt}{\sqrt{npq}}} + pe^{\frac{qt}{\sqrt{npq}}}\right)^n
\end{aligned}$$

$$= \left[q\left\{1-\frac{pt}{\sqrt{npq}}+\frac{p^2t^2}{2npq}+o'\left(\frac{1}{n^{3/2}}\right)\right\}+\right.$$

$$\left.+p\left\{1+\frac{qt}{\sqrt{npq}}+\frac{q^2t^2}{2npq}+o''\left(\frac{1}{n^{3/2}}\right)\right\}\right]^n$$

where $o'\left(\frac{1}{n^{3/2}}\right)$ and $o''\left(\frac{1}{n^{3/2}}\right)$ denotes terms with $\frac{1}{n^{3/2}}$ and their higher powers. Therefore,

$$\begin{aligned}\lim_{n\to\infty} M_Z(t) &= \lim_{n\to\infty}\left[1+\frac{t^2}{2n}+o\left(\frac{1}{n^{3/2}}\right)\right]^n \\ &= \lim_{n\to\infty}\left[1+\frac{t^2}{2n}\right]^n \\ &= e^{t^2/2}\end{aligned}$$

which is the m.g.f of a standard normal variate. Hence, by uniqueness theorem

$$Z=\frac{S_n-np}{\sqrt{npq}} \to N(o,1) \quad \text{as} \quad n\to\infty.$$

Hence, $S_n \to N(np, npq)$ as $n\to\infty$. □

Remark 2.4.1. *The theorem says that, standardised sequence* *of the partial sums of Bernoulli variates tends to standard normal variate as $n \to \infty$. In other words binomial distribution tends to normal distribution as $n \to \infty$.*

Solved Problems

1. 60 real numbers are randomly selected from the interval (0,1). Find the probability of sum of these numbers greater than 35 approximately.

   Solution:
   Given, $n = 60$ and $X_i \sim U(0,1)$, for $i = 1, 2, \ldots, 60$

   Therefore,

   $$E(X_i) = \frac{1}{2} \text{ and } V(X_i) = \frac{1}{12}.$$

   Let,

   $$S_{60} = X_1 + X_2 + \ldots + X_{60}$$

   Now,

   $$E(S_{60}) = 60\ E(X_i) = 60 \times \frac{1}{2} = 30$$

   $$V(S_{60}) = 60\ V(X_i) = 60 \times \frac{1}{12} = 5$$

Since n is large, $S_{60} \sim N(30, 5)$. Therefore,

$$\begin{aligned} P(S_{60} > 35) &= P\left(\frac{S_{60} - 30}{\sqrt{5}} - \frac{35 - 30}{\sqrt{5}}\right) \\ &= P(Z > 2.236) \\ &= 0.5 - P(0 < Z < 2.236) \\ &= 0.5 - 0.4873 \\ &= 0.0127 \end{aligned}$$

Exercises

1. Let $\{X_n, n \geq 1\}$ be a sequence of i.i.d. Poisson random variables with mean λ. Obtain the limiting distribution of $\sum_{i=1}^{n} X_i$ with the help of CLT.

2. Let $\bar{X}$ be the mean of 16 independent observations from a Poisson distribution with mean 4. Apply CLT to find approximation to $P(3 < \bar{X} < 5)$. Also approximate it using Chebychev's inequality.

3. A distribution with unknown mean μ has variance equal to 1.5. By using CLT, how large a sample should be taken

in order that the probability will be at least 0.95 that the sample mean will not differ from the population men by more than 0.5.

4. If $f(x) = \frac{1}{x^2}$, $1 < x < \infty$ and zero elsewhere be the p.d.f. of a random variable X. A random sample of size 72 is chosen from this population. determine the probability that more than 50 of them are less than 3.

5. One thousand rounds are fired from a gun at a target, the probability of a hit on each round being 0.7. Use the CLT to determine the probability that the number of hits will be

   (a) between 675 and 725.

   (b) more than 680.

6. What is the probability of obtaining more than 520 heads in 1000 tosses of a fair coin?

MODULE

THREE

Sampling Methods

Statistics is a branch of science which deals with collection, organisation, analysis and interpretation of data, which are the raw materials for any statistical study. The intension is to study some characteristics of the population with the help of data. This can be collected either by census survey or sample survey.

Population

Population means the aggregate of all elements having one or more common characteristics. If we consider only one characteristic, then by a statistical population we mean a collection of numerical values enumerated with respect to that characteristic from each and every element of the group under the topic of investigation. Hence, a statistical population can be considered as the set of admissible values of a random variable and the distribution of this random variable is called the distribution of the population.

Example 3.0.1. *Suppose that the government is planning to introduce a new teaching method in the primary schools of the state. But before introducing it there is a need to know the I.Q. levels of children in the state. This is to get an idea whether the new method would be fruitful or not.*

In this example, the collection of all primary school children of the state is the population under consideration.

Population can be of any size. The number of units of the population is called population size, usually denoted N. The population to be considered should be made clear before any statistical study.

Census

In a statistical investigation, one will be interested in studying the characteristic of a population called parameter. This can be done either by studying the entire population or a representative part of it. If data is collected from each and every unit of a population, then it is known as complete enumeration or census survey or census method. In many situations it is not possible to enumerate the entire population due to tediousness, excessive cost, perishable nature etc. For example, in the case of the problem in example 3.0.1, it is not feasible to check the IQ level of each an every primary school child of the state. In such situations, data is collected from a representative part of a population called sample.

Sample

A part of the population chosen at random is called a sample. Sample selected should be such that it is capable of exhibiting the characteristic of the population. The number of units in a sample is called sample size (n).

Sampling

The process of selecting sample is known as sampling. The method of drawing inferences about the characteristics of the population by observing only a part of it is called sample survey or sample method or method of sampling. And often they provide the same results if done systematically.

Sampling Theory

Sampling theory is a prime field of study in statistics and deals with the following problems:

1. Choosing proper sampling design or various methods of selecting a sample from a population.
2. Various methods for constructing estimate for the unknown population parameters.
3. Errors in the estimates cannot be measured without knowing the true values of the parameters. True values of the parameters are never available in a sample survey. Against this theory developed various methods of assessing the errors in the estimate with the help of sampling theory and

sampling distribution.

4. In a survey, cost is an important factor. Sampling theory enables us to determine the sample without violating the cost restriction.

5. Non-sampling errors is a serious threat in both census and sampling methods. Sampling theory suggest various remedial methods to avoid the presence of non-sampling errors.

Various Assumptions in Sampling Theory

The basic assumptions are as follows:

1. The population under consideration is assumed to be finite and all the units of the population are distinguishable from one another.

2. The sample selected may be either non-probability sample (judgement) or a probability sample (random sample). Usually the word 'sample' is considered to be a 'random sample' in modern theory. By random sampling, we mean a sample selected by some random mechanism. This random mechanism is used in such a way that specify probability is assigned to each unit in the population to come

under the sample.

3. It is assumed that the measurements collected from the sample survey are free from all types of errors. This assumption, though unrealistic is essential for the development of the theory.
4. The assumption of normality in sampling theory gives the advantage of setting up of confidence interval for the unknown population parameters for large values of n. For small values of n, in practical problems we use t tables.

Advantages of Sampling over Census

1. Sampling is less time consuming.
2. Sampling is more economical.
3. Sampling ensures completeness and the high degree of accuracy due to the small area of operation.
4. It is possible to obtain more detailed information in sample survey then complete enumeration.
5. Sampling is also advocated where census is neither necessary nor desirable.

6. In some cases sampling is the only feasible method. For example, we have to test the sharpness of blades- if you test each blade perhaps the whole of the product will be wasted in such circumstances the sampling techniques will be more useful.
7. Sampling is much more scientific than census because in it the extent of the reliability of the result can be known where as this is not always possible in census.

Limitations

1. Improper planning of sample survey will lead to incorrect and misleading results.
2. It requires service of experts.
3. Sampling procedure cannot be used if you want to obtain information about each and every units of the population.
4. If the sample is not a true representative of the population, then the sample will fail to give the true characteristics of the population.
5. Chance of sampling error is more.

Principles of Sampling Theory

The three important basic principles are:

1. Principle of validity - The sample should be so chosen that the estimates could be interpreted objectively and in terms of probability. This principle ensures that there is some definite and preassigned probability for each individual in the sampling design.
2. Principle of statistical regularity - This principle stress upon the desirability and importance of selecting sample design where inclusion of sampling units in the sample is based on probability theory.
3. Principle of optimisation. - This principle stress upon obtaining optimum results with minimisation of loss in terms of cost and mean square error (M.S.E.).

Principal Steps in a Sample Survey

1. Statement of objectives
2. Definition of population to be studied.
3. Determination of sampling frame and sampling units

4. Selection of proper sampling design.

5. Organisation of field work.

6. Summary and analysis of data.

Sampling Units

These units may be natural units of the population such as individuals of a locality or natural aggregates of such units such as family or they may be artificial units such as a farm etc. Before selecting the sample the population must be divided into parts which are distinct, unambiguous and non-overlapping such that every element (smallest component part in which a population can be divided) of the population belongs to one and only one sampling unit.

Sampling Frame

A complete list of sampling units which represent the population to be covered is called the sampling frame popularly known as frame.

Sampling and Non-sampling Errors

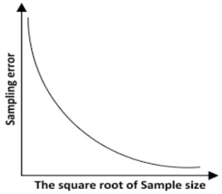
The errors involved in collection, processing and analysis of the data may be classified as:

1. Sampling error

2. Non-sampling error

Sampling error

The error which arises due to the fact that, only a sample being used to estimate the population parameter is termed sampling error or sampling fluctuations. Whatever may be the degree of consciousness in selecting a sample, there will always be a difference between the population value and its corresponding estimate. This error is inherent and unavoidable in any and every sampling scheme. A sample with the smallest sampling error will always be considered a good representative of the population. This error can be reduced by increasing the size of the sample. In fact, the decrease in sampling error is inversely proportional to the square of the sample size.



Non-sampling error

Besides sampling error, the sample estimate may be subject to other error which are grouped together and termed non-sampling errors. The main source of non-sampling errors are:

1. Failure to measure some of the units in the selected sample.
2. Observational errors due to defective measurement techniques.
3. Errors introduced in editing, coding and tabulating the results.

In practice, the census survey results may suffer from non-sampling error although this may be free from sampling error. The non-sampling error is likely to increase with increase in sample size but sampling error decreases with increasing sample size.

Desiderata in Planning of Sample Surveys

The main stages of a survey are planning, data collection and data processing. Some of the important aspects in the planning and execution of a sample survey are:

1. Specification of data requirements - When specifying the data requirements, the following points should be considered
    (a) Statistical statement of desired information
    (b) Clear specification of the domain of study
    (c) Form of data to be collected and limitations of budget
    (d) Degree of precision required
2. Survey references and reporting periods - From the operational point of view, it is desirable to decide these periods well in advance.

   (a) Survey period - The time period during which data required are collected.
   (b) Reference period - The time period to which the data information should refer. Many surveys intend to measure frequencies of events or instances within a given period of time.
   (c) Reporting period - The time period for which the information is collected for a unit.

3. Preparation of sample frame - Sampling frame is the structure of the survey. Adequate attention must be paid for the preparation of an up-to-date and accurate sampling frame.

4. Choice of sampling design - Principle of optimisation should be always kept in mind when the choice of sampling design is made
   (a) Either a degree of precision with a minimum cost, or
   (b) The maximum possible precision with a fixed cost

5. Method of data collection - After a very careful examination of the frame, design, budget and objectives, a decision should be made regarding the method of data collection. That is, to collect primary data or to use secondary data.

6. Field work and training of personal - It is essential that the personnel should be trained well in locating sampling units, method of collection of data and eliciting correct information from different sources before the field work.

7. Processing of survey data - processing the data collected include

    (a) Scrutiny and editing the data

    (b) Tabulation of data

    (c) Statistical analysis

8. Preparation of reports - The report may have sections such as objectives; scope; subject coverage; method of data collection; survey reference and reporting periods; sampling design and estimation procedure; tabulation procedure; presentation of results; accuracy; cost structure; responsibility; and references.

Methods of Sampling

3.1 Non-probability Sampling

This is the method of selecting samples in which the choice of selection of sampling units depends entirely on the judgement of the sampler. This method is also sometimes called purposive or judgement sampling. In this procedure, the sampler inspects the entire population and select a sample of typical units which he considers as close to the average of the population. This sampling method is mainly used for opinion surveys, but cannot be recommended for general use as it is subject to the drawbacks of prejudice and bias of the sampler. Some methods of probability sampling are:

1. Convenience Sampling
2. Purposive Sampling
3. Judgement Sampling
4. Quota Sampling
5. Snowball Sampling.

3.2 Probability Sampling

This is the method of selecting samples according to certain law of probability in which each unit of the population has some definite non-zero probability of being selected in the sample. A clear specification of all possible samples of a given type along with their corresponding probabilities of selection is said to constitutes a sampling design. Some methods of probability sampling are:

1. Simple Random Sampling
2. Stratified Random Sampling
3. Systematic Random Sampling
4. Cluster Sampling

3.2.1 Simple Random Sampling

The simplest and common most method of sampling is simple random sampling in which the sample is drawn unit by unit, with equal probability of selection for each unit at each draw. Therefore, simple random sampling is a method of selecting n units out of a population of size N by giving equal probability to all units, or a sampling procedure in which all possible combinations of n units that may be formed from the population of N units have the same probability of selection. It is also sometimes referred to as random sampling. Since the word random is used in the literature in many different senses, an extra qualifying adjective is advisable. Some writers prefer the phrase unrestricted random sampling.

If a unit is selected and noted and then returned to the population before the next drawing is made and this procedure is repeated n times, it gives rise to a simple random sampling of n units. This procedure is generally known as simple random sampling with replacement (SRSWR). If this procedure is repeated till n distinct units are selected and all repetitions are ignored, it is called a simple random sampling without replacement (SRSWOR). In SRSWR the number of distinct samples of size n that can be drawn from a population of size N is N^n and in SRSWOR the number of distinct samples of size n that can be drawn from a population of size N is $\binom{N}{n}$.

Remark 3.2.1. *The probability that a specified unit of the population being selected at any given draw is equal to the probability of its being selected at the first draw equal to* $1/N$.

Remark 3.2.2. *The probability of a specified unit being included in the sample ignoring order is equal to* n/N.

Remark 3.2.3. *The probability of a specified sample of n units ignoring order is* $1/\binom{N}{n}$.

Procedures of selecting a random sample

Some of the procedures for selecting random samples are as follows:

1. Lottery method
2. Use of random number tables

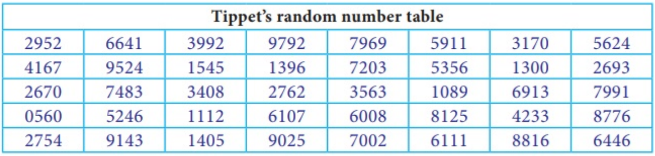
Lottery Method

In this method, a ticket or chit may be associated with each unit of the population. Thus, each sampling unit has its identification mark from 1 to N. The procedure of selecting an individual is simple. All the tickets or chits are placed in a container, drum or metallic spherical devices in which a thorough mixing or reshuffling is possible, before each draw. Draws of tickets or chits may be continued until a sample of the required size is obtained. When the number of unit is large, this method is not suitable as thorough shuffling will not be possible.

Use of Random Number Table

Random number table is an arrangement of digits 0 to 9 in either a linear or rectangular pattern, where each position is filled with one of these digits. A table of random numbers is so constructed that all the numbers $0, 1, \ldots, 9$ appear independent of each other. Some random number tables in common use are:

1. Tippet's random number table
2. Fisher and Yate's table
3. Kendall and Smith table
4. A million random digits.



Tippet's random number table							
2952	6641	3992	9792	7969	5911	3170	5624
4167	9524	1545	1396	7203	5356	1300	2693
2670	7483	3408	2762	3563	1089	6913	7991
0560	5246	1112	6107	6008	8125	4233	8776
2754	9143	1405	9025	7002	6111	8816	6446

To ascertain whether the series of random numbers are really random, the following tests may be applied:

1. Frequency test

2. Serial test
3. Gap test
4. Poker test

A practical method of selecting a random sample is to choose units one by one with the help of a table of random numbers. By considering two-digit numbers, we can obtain numbers from 00 to 99, all having the same frequency. Similarly, three or more digit numbers may be obtained by combining three or more rows or columns of these tables.

The simplest way of selecting a sample of the required size is by selecting a random number from 1 to N and then taking the unit bearing that number. This procedure involves a number of rejections, since all numbers greater than N appearing in the table are not considered for selection. The use of random numbers, is therefore modified and some of this modified procedures are:

1. Remainder approach
2. Quotient approach
3. Independent choice of digits

Remainder approach

Let N be an r-digit number and let its r digit highest multiple be N'. A random number k is chosen from 1 to N' and the unit with the serial number equal to the remainder obtained on dividing k by N is selected. If the remainder is 0 the last unit is selected. As an illustration, let $N = 123$, the highest 3 digit multiple of 123 is 984. For selecting a unit, one random number from 001 to 984 has to be selected. Let the random number selected be 287. Dividing 287 by 123 the reminder is 41. Hence, the unit with serial number 41 is selected in the sample.

Quotient approach

let N be an r-digit number and let its r digit highest multiple be N' such that $\frac{N'}{N} = q$. A random number k is chosen from 0 to $(N' - 1)$. Dividing k by q the quotient r is obtained and the unit bearing the serial number $r - 1$ is selected in the sample. As an illustration, let $N = 16$ and hence $N' = 96$ and $q = \frac{96}{16} = 6$. Let the two digit random number chosen be 65 which lies between 0 and 95. Dividing 65 by 6, the quotient is 10 and has the unit bearing the number $(10 - 1) = 9$ is selected in the sample.

Note:

1. Inflation factor $= \frac{N}{n}$, also called expansion or raising.
2. Sampling fraction $f = \frac{n}{N}$.

Advantages and Disadvantages

Advantages

1. Main advantage of simple random sampling is its ease of use.
2. All units have same probability of selection.

Disadvantages

1. Time consuming
2. Costly.

3.2.2 Stratified Random Sampling

Of all the methods of sampling, the procedure most commonly used in surveys is stratified sampling. In stratified sampling, the population of N units is subdivided into k sub-populations called strata, the i^{th} sub-population having N_i units $(i = 1, 2, \ldots, k)$. These sub-populations are non-overlapping so that they comprise the whole population such that

$$N_1 + N_2 + \ldots + N_k = N.$$

A sample is drawn from each stratum independently, the sample size within the i^{th} stratum being n_i $(i = 1, 2, \ldots, k)$ such that

$$n_1 + n_2 + \ldots + n_k = n.$$

The procedure of taking samples in this way is known as stratified sampling. If the sample is taken randomly from each stratum, the procedure is known as stratified random sampling.

The main objective of stratification is to give a better cross section of the population size to give a higher degree of relative precision. To achieve this the following points are to be examined carefully.

1. Formation of strata
2. Number of strata to be made
3. Allocation of sample size within each stratum

4. Analysis of data from a stratified design

There are four methods of allocation of sample sizes to different strata in stratified sampling procedure. These are:

1. Equal allocation
2. Proportional allocation
3. Neyman allocation
4. Optimum allocation.

Principles of stratification

1. The strata should be non overlapping and should together comprise the whole population.
2. Stratification of population should be done in such a way that strata are homogeneous within themselves, with respect to the characteristic under study.
3. In many practical situations when it is difficult to stratify with respect to the characteristic under the study, administrative convenience may be considered as the base of stratification.

4. If the limit of precision for certain sub-population is given, it will be better to treat each sub-population as a stratum.

Advantages of stratification

1. Stratification may be desired for administrative convenience.
2. Stratification by natural characteristics help in improving the sampling design.
3. Stratification is particularly more effective when there are extreme values in the population which can be segregated into separate stata, thereby reducing the variability within strata.
4. Stratification makes it possible to use different sampling design in different strata.
5. Stratification ensures adequate representation to various groups of the population which may be of some interest or importance.
6. Stratification also ensure selection of a better cross section of the population.

7. Stratification brings a gain in the precision in estimation of a characteristic of a population.

3.2.3 Systematic Random Sampling

Systematic random sampling is a sampling technique in which only the first unit is selected with the help of random numbers and the rest get selected automatically according to some predesigned pattern. It is also referred to as systematic sampling. Systematic sampling is simple and planned so well that nothing can go wrong. This procedure in many situations provides estimates more efficient than simple random sampling and is widely used in various types of surveys. Systematic sampling has the drawback of not building an unbiased estimate of sampling variance with single sample.

Linear systematic sampling

Suppose N units of the population are numbered from 1 to N in some order. Let $N = nk$, where n is the sample size and k is an integer, and a random number less than or equal to k be selected and every k^{th} unit thereafter. The resultant sample is called k^{th} systematic sample and such a procedure termed linear systematic sampling. Let i $(\leq k)$ be the randomly selected first unit, k being called the sampling interval. Then, the sample comprises the units $i, i+k, \ldots, i+(n-1)k$. The technique will generate k systematic samples with equal probability, which may be shown in the schematic diagram below.

Sample Number					
1	2	$\ldots$	i	$\ldots$	k
y_1	y_2	$\cdots$	y_i	$\cdots$	y_k
y_{1+k}	y_{2+k}	$\cdots$	y_{i+k}	$\cdots$	y_{2k}
y_{1+2k}	y_{2+2k}	$\cdots$	y_{i+2k}	$\cdots$	y_{3k}
$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\ldots$	$\vdots$
$y_{1+(j-1)k}$	$y_{2+(j-1)k}$	$\cdots$	$y_{i+(j-1)k}$	$\cdots$	y_{jk}
$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\ldots$	$\vdots$
$y_{1+(n-1)k}$	$y_{2+(n-1)k}$	$\cdots$	$y_{i+(n-1)k}$	$\cdots$	y_{nk}

Another practical situation is that N is not expressible in the form $N = nk$. Let k be taken as the integer nearest to N/n. Then a random number is chosen from 1 to k and every k^{th} unit is drawn in the sample. Under this condition, the sample size is not necessarily n and in some cases it may be $(n-1)$. Hence, in this case, the present sampling scheme will give rise to samples of unequal size. For example, if $N = 11$, $n = 4$, then the value of $k = 3$ and possible samples are $(1, 4, 7, 10)$; $(2, 5, 8, 11)$; $(3, 6, 9)$, which are not of the same size. An improvement of this method is circular systematic sampling.

Circular systematic sampling

If $N \neq nk$, and every k^{th} unit be included in a circular manner till the whole list is exhausted, it will be called circular systematic sampling. To overcome the difficulty of varying sample size under the situation $N \neq nk$, the procedure is modified slightly by which a sample of constant size is always obtained. The procedure consists in selecting a unit, by a random start, from 1 to N and then there after selecting every k^{th} unit, k being an integer nearest to N/n, in a circular manner, until a sample of n units is obtained. This method is generally known as circular systematic sampling. Suppose that a unit with random number i is selected. The sample will then consists of the units corresponding to the serial numbers

$$i + jk; \quad \text{if } i + jk \leq N$$

$$i + jk - N; \quad \text{if } i + jk > N$$

for $j = 0, 1, 2, \ldots, (n-1)$.

Every unit has got an equal probability of selection $(1/N)$ in this method. As an illustration, let $N = 11$, $n = 4$. Then $k = 3$. The possible samples are, therefore, $(1, 4, 7, 10)$; $(2, 5, 8, 11)$; $(3, 6, 9, 1)$; $(4, 7, 10, 2)$; $(5, 8, 11, 3)$; $(6, 9, 1, 4)$; $(7, 10, 2, 5)$; $(8, 11, 3, 6)$; $(9, 1, 4, 7)$; $(10, 2, 5, 8)$ and $(11, 3, 6, 9)$.

Advantages and Disadvantages

Advantages

1. The main advantage of systematic sampling is its simplicity of selection, operational convenience and even spread of the sample over the population.

2. Another advantage is that except for population with periodicities, systematic sampling provides an efficient estimate as compared to alternative designs.

3. Sometimes systematic sampling variances are much smaller than the variance for random selection of units within strata.

Disadvantages

1. In case of periodicity in the population, systematic sampling has to be used with considerable care.
2. A serious disadvantage of systematic sampling lies in its use with population having an unforeseen periodicity which may substantially contribute bias to the estimate of the population mean value.
3. Another disadvantage concerns the drawback of estimating the sampling variance of estimators with a single sample.

3.2.4 Cluster Sampling

The smallest unit into which a population can be divided is called an element of the population. A group of such elements is known as a cluster. When the sampling unit is a cluster, the procedure is called cluster sampling. If the entire area containing the population under study is divided into smaller segments and each element in the population belongs to one and only one segment the procedure is sometimes called area sampling.

Clusters are generally made up of neighboring elements and, therefore, the elements within a cluster tend to have similar characteristics. As a simple rule, number of elements in a cluster should be small and number of clusters should be large. After dividing the population into specified clusters, the required number of clusters can be selected either by equal or unequal probabilities of selection. Either all the elements in the selected clusters are enumerated or only a handful of elements are chosen from each cluster by implementing systematic or simple random sampling.

The technique is widely used in statistics where the researcher can't collect data from the entire population as a whole. It is the most economical and practical solution for statisticians doing research. Various sampling procedures, viz. simple random sampling, stratified random sampling and systematic random sampling can be applied to clusters sampling by treating the clusters as sampling units.

Advantages and Disadvantages

Advantages

1. Easier, cheaper and faster
2. Efficiency per unit cost is more.

Disadvantages

1. Less efficient than simple random sampling.
2. Efficiency decrease with increase in cluster size.

MODULE

FOUR

Sampling Distributions

A statistical population means, a collection of numerical values enumerated with respect to some characteristics from each and every element of the group under the topic of investigation. Characteristic of the individuals of the population is called a variable. Variables are of two types:

1. Categorical or Qualitative variables usually referred to as

attributes.
eg. gender, eye colour, nationality etc.

2. Quantitative variables.
eg. height, income, I.Q. level etc.

Parameter

The summary value of the variable for the population (or the summary value of the statistical population) is called a parameter. It gives a very good idea about the population. For example, population mean μ, population variance σ^2, population standard deviation σ, population correlation coefficient ρ etc. In example (3.0.1), our parameter could be average I.Q. level of all primary school children in the state. Hence we can say that, variable is a characteristic of an individual and parameter is a characteristic of the population.

Our goal is to get the summary value of the variable for the population under consideration. Usually, population is too large to cover due to constraints like cost, time etc. Therefore, all we can do is to study a part of the population called sample.

Statistic

A statistic is the characteristic of the sample. It is a function of sample observations. In fact, it is a measurable function of a random sample and hence itself a random variable The common use of statistic is to estimate a particular population parameter. For example, sample mean $\bar{X}$, sample variance S^2, sample standard deviation S, sample correlation coefficient r etc. In example (3.0.1), our statistic could be average I.Q. level of all primary school children in the sample. Statistic used for estimating a population parameter is called an estimator and the value taken by a statistic when a particular sample is realised is called an estimate. Another use of statistic is in testing of hypothesis.

Sampling Distributions

We can see that, the population elements included in different samples from the same population may be different. So the value of the statistic is liable to vary from one sample to another. Thus, if a number of samples, each of the same size, is taken from the same population and for each sample the value of the statistic is calculated, a series of values of statistic will be obtained. If the number of samples is large, this may be arranged into a frequency table. The frequency distribution of the statistic that will be obtained, if the number of samples, each of the same size, were large is called the sampling distribution of a statistic.

Mathematically, a probability distribution is a function that describes the likelihood of obtaining the possible values that a random variable can assume. A statistic, being a Borel-measurable function of a random variables (random sample) is also a random variable The probability distribution of this random variable is known as sampling distribution.

Standard Error

Standard deviation of the sampling distribution of a statistic is called the standard error of the statistic. If $T = T(X_1, X_2, \ldots, X_n)$ is a statistic, then

$$S.E.\ (T) = \sqrt{V(T)}.$$

The standard deviation measures the dispersion or amount of variability of individual data values from its mean. While standard error measures how far the value of the statistic is likely to be from the true parameter value. For example, standard error of the sample mean measures how far the sample mean of the data is likely to be from the true population mean. Note that, the standard error of mean is always less than the standard deviation.

Uses of Standard Error

S.E. plays an important role in large sample theory. It forms the basis of testing of hypothesis.

1. S.E. is inversely proportional to the sample size, ($S.E. \propto \frac{1}{n}$), it helps to determine the size of the sample.
2. Used for testing a given hypothesis.
3. Reliability of a sample is the reciprocal of the S.E.
4. Used for determining the confidence interval of population parameters.

Sampling Distribution of Small Samples Drawn from Normal Population

4.1 Sampling Distribution of Sample Mean

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$. Then the statistic 'sample mean' is given by

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

Then, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

Proof. Since $X_1, X_2, \ldots, X_n$ is a random sample of size n from $N(\mu, \sigma^2)$, we can consider them as i.i.d. random variables having the same distribution $N(\mu, \sigma^2)$. Therefore,

$$M_{X_i}(t) = e^{\mu t + \frac{1}{2}t^2\sigma^2}; \quad i = 1, 2, \ldots, n$$

and

$$M_{\bar{X}}(t) \;=\; M_{\frac{\sum_{i=1}^{n} X_i}{n}}(t)$$

$$
\begin{aligned}
&= M_{\sum_{i=1}^{n} X_i}\left(\frac{t}{n}\right) \\
&= \prod_{i=1}^{n} M_{X_i}\left(\frac{t}{n}\right) \\
&= \prod_{i=1}^{n}\left(e^{\mu\frac{t}{n}+\frac{\left(\frac{t}{n}\right)^2\sigma^2}{2}}\right) \\
&= \left(e^{\mu\frac{t}{n}+\frac{t^2\sigma^2}{2n^2}}\right)^n \\
&= e^{\mu t+\frac{t^2\sigma^2}{2n}},
\end{aligned}
$$

which is the m.g.f. of $N(\mu, \frac{\sigma^2}{n})$. Therefore,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

and its p.d.f. is given by

$$f_{\bar{X}}(x) = \frac{\sqrt{n}}{\sigma\sqrt{2\pi}}\ e^{-\frac{n(x-\mu)^2}{2\sigma^2}}; \quad -\infty < x < \infty.$$

□

Note:

1. When the population is $N(\mu, \sigma^2)$,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

   for any sample of size n.

2. Let $\bar{X}$ be the sample mean of a sample of size n from a population which is **not normal** whose mean is μ and variance σ^2. Then by CLT,

$$\bar{X} \to N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{as} \quad n \to \infty.$$

3. Let $\bar{X}$ be the sample mean of a sample of size n from any population. Then, if σ^2 is unknown, when n is large ($n \geq 30$) we can replace σ^2 by $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, the sample variance, and

$$\bar{X} \to N\left(\mu, \frac{S^2}{n}\right) \quad \text{as} \quad n \to \infty.$$

4. $E(\bar{X}) = \mu$, $V(\bar{X}) = \frac{\sigma^2}{n}$ and $S.E.(\bar{X}) = \frac{\sigma}{\sqrt{n}}$.

4.2 Chi-square Distribution

Karl Pearson in about 1900 described a well known probability distribution called “Chi-square distribution” or “distribution of Chi-square”. Chi-square is a random variable used as a test statistic. The square of a standard normal variate is known as the Chi-square (χ^2) variate with 1 degree of freedom.

Thus, if $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

and

$$Z^2 = \left(\frac{X - \mu}{\sigma}\right)^2 \sim \chi^2(1)$$

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$. Then,

$$X_i \sim N(\mu, \sigma^2) \Rightarrow Z_i = \frac{X_i - \mu}{\sigma} \sim N(0, 1).$$

We define χ^2 statistic with n degrees of freedom as the sum of the squares of n **independent** standard normal variates. That is,

$$\chi^2 = \sum_{i=1}^{n} Z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2(n).$$

We shall use m.g.f. to obtain the distribution of χ^2.

$$\begin{aligned} M_{\chi^2}(t) &= M_{\sum_{i=1}^n Z_i^2}(t) \\ &= \prod_{i=1}^n M_{Z_i^2}(t) \\ &= \left[M_{Z^2}(t)\right]^n \qquad (4.2.1) \end{aligned}$$

Now, since $M_{g(X)}(t) = \int_x e^{tg(x)} f_X(x)dx$,

$$\begin{aligned} M_{Z^2}(t) &= E\left[e^{tZ^2}\right] \\ &= \int_{-\infty}^{\infty} e^{tz^2} f_Z(z)dz \\ &= \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}(1-2t)} dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \frac{du}{\sqrt{1-2t}}, \\ &\qquad \text{where } z\sqrt{1-2t} = u \end{aligned}$$

$$= \frac{1}{\sqrt{1-2t}}, \quad \text{since } \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = 1$$

$$= (1-2t)^{-1/2}. \tag{4.2.2}$$

Therefore, from (4.2.1) and (4.2.2) we have,

$$M_{\chi^2}(t) \quad = \quad (1-2t)^{-n/2},$$

which is the m.g.f. of a gamma distribution with $m = \frac{1}{2}$ and $p = \frac{n}{2}$. Hence, the p.d.f. of χ^2 is

$$f_{\chi^2}(x) = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}} x^{\left(\frac{n}{2}-1\right)} \quad ; \quad 0 < x < \infty \ ,$$

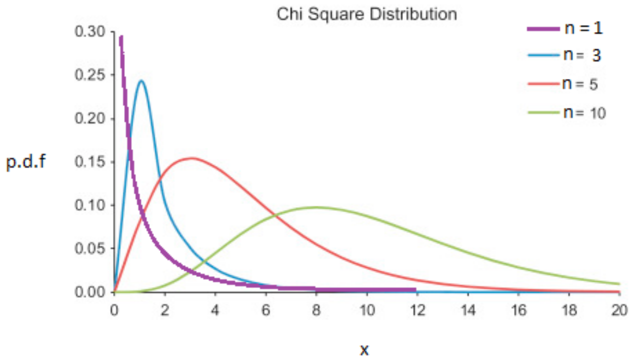
Definition 4.2.1. *A random variable X is said to follow chi-square distribution with n degrees of freedom if its p.d.f. is given by*

$$f_X(x) = \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}} x^{\left(\frac{n}{2}-1\right)} \quad ; \quad 0 < x < \infty \ ,$$

and we write $X \sim \chi^2(n)$.

Remark 4.2.1. *Chi-square distribution is a particular case of Gamma distribution with parameters $m = \frac{1}{2}$ and $p = \frac{n}{2}$.*

Properties of Chi-square Curve



1. The shape of the χ^2 curve depends on the value of n.
2. For small n the curve is positively skewed.
3. As n increases, the curve approaches symmetry.
4. For large n the curve is approximately normally distributed.
5. The distribution is unimodal.

Degrees of Freedom

Degrees of freedom is the number of values in the final calculation of a statistic, that are free to vary. It is equal to the sample size minus the number of restrictions (eg. number of parameters estimated in intermediate steps, number of cells pooled). It is usually denoted by ν.

For example, originally there are 'n' degrees of freedom in a sample of n observations. But, 1 degree of freedom is used up in calculating $\bar{X}$. Therefore, there is a restriction of the constraint $\sum_{i=1}^{n}(X_{i} - \bar{X}) = 0$ in the calculation of $S^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(X_{i} - \bar{X})^2$, leaving $(n-1)$ degrees of freedom for the residuals $(X_{i} - \bar{X})$ to calculate S^2. Where as, $\sum_{i=1}^{n}(X_{i} - \mu)^2$ has n degrees of freedom, since $\sum_{i=1}^{n}(X_{i} - \mu)$ need not be zero.

Moments

Mean

$$\mu'_{1} = E(X) \quad = \quad \int_{0}^{\infty} x f_{X}(x) dx$$

$$
\begin{aligned}
&= \int_0^\infty x \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}} x^{\left(\frac{n}{2}-1\right)} dx \\
&= \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \int_0^\infty e^{-\frac{x}{2}} x^{\left(\frac{n}{2}+1-1\right)} dx \\
&= \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \; \frac{\Gamma\left(\frac{n}{2}+1\right)}{\left(\frac{1}{2}\right)^{\frac{n}{2}+1}} \\
&= \frac{n/2}{1/2} \\
&= n
\end{aligned}
$$

Variance

$$V(X) = \mu_2' - (\mu_1')^2 = E(X^2) - [E(X)]^2$$

$$
\begin{aligned}
\mu_2' = E(X^2) &= \int_0^\infty x^2 \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}} x^{\left(\frac{n}{2}-1\right)} dx \\
&= \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \int_0^\infty e^{-\frac{x}{2}} x^{\left(\frac{n}{2}+2-1\right)} dx
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \; \frac{\Gamma\left(\frac{n}{2}+2\right)}{\left(\frac{1}{2}\right)^{\frac{n}{2}+2}} \\
&= \frac{\left(\frac{n}{2}+1\right)\frac{n}{2}}{\left(\frac{1}{2}\right)^2} \\
&= n(n+2)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
V(X) &= E(X^2) - [E(X)]^2 \\
&= n(n+2) - n^2 \\
&= 2n
\end{aligned}
$$

Moment Generating Function

$$
\begin{aligned}
M_X(t) &= E\left(e^{tX}\right) \\
&= \int_0^\infty e^{tx} \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} e^{-\frac{x}{2}} x^{\left(\frac{n}{2}-1\right)} dx
\end{aligned}
$$

$$\begin{aligned} &= \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\int_0^\infty e^{-(1-2t)\frac{x}{2}}x^{\left(\frac{n}{2}-1\right)}dx \\ &= \frac{\left(\frac{1}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\;\frac{\Gamma\left(\frac{n}{2}\right)}{\left(\frac{1-2t}{2}\right)^{\frac{n}{2}}} \\ &= (1-2t)^{n/2} \end{aligned}$$

Additive/Reproductive Property

Theorem 4.2.1. *If* $X_1 \sim \chi^2(n_1)$, $X_2 \sim \chi^2(n_2)$ *and* X_1 *is independent of* X_2, *then* $X_1 + X_2 \sim \chi^2(n_1 + n_2)$.

Proof. $X_1 \sim \chi^2(n_1)$, $X_2 \sim \chi^2(n_2)$ implies $M_{X_1}(t) = (1-2t)^{n_1}$ and $M_{X_2}(t) = (1-2t)^{n_2}$ respectively. Since X_1 and X_2 are independent,

$$\begin{aligned} M_{X_1+X_2}(t) &= M_{X_1}(t) \times M_{X_2}(t) \\ &= (1-2t)^{\frac{n_1}{2}} \times (1-2t)^{\frac{n_2}{2}} \\ &= (1-2t)^{\frac{n_1+n_2}{2}}, \end{aligned}$$

which is the m.g.f. of $\chi^2(n_1 + n_2)$. Therefore, $X_1 + X_2 \sim \chi^2(n_1 + n_2)$. $\square$

Note:

1. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$. Then,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

   and

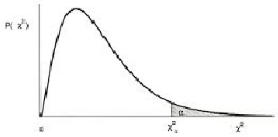
$$\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}\right)^2 \sim \chi^2(1).$$

2. If $X \sim \chi^2(n)$, then $E(X) = n$ and $V(X) = 2n$.

3. If $X \sim \chi^2(n)$, then as $n \to \infty$ by CLT

$$\frac{X - n}{\sqrt{2n}} \to N(0, 1).$$

4. $P(\chi^2 > \chi^2_\alpha) = \alpha$.



4.3 Sampling Distribution of Sample Variance

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$. Let $\bar{X}$ be the sample mean. Then,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

is called the sample variance.

Theorem 4.3.1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ random variables. Then $\bar{X}$ and $((X_1 - \bar{X}), (X_2 - \bar{X}), \ldots, (X_n - \bar{X}))$ are independent.*

Proof. Let $M(t, t_1, t_2, \ldots, t_n)$ be the m.g.f. of $(\bar{X}, (X_1 -$

$\bar{X}), (X_2 - \bar{X}), \ldots, (X_n - \bar{X}))$.

$$\begin{aligned}
M(t, t_1, t_2, \quad \ldots \quad , t_n) = E[\exp\{t\bar{X} + t_1(X_1 - \bar{X}) + t_2(X_2 - \bar{X})& \\
+ \ldots + t_n(X_n - \bar{X})\}]& \\
&= E\left[\exp\left\{\sum_{i=1}^{n} t_i X_i - \left(\sum_{i=1}^{n} t_i - t\right)\bar{X}\right\}\right] \\
&= E\left[\exp\left\{\sum_{i=1}^{n} t_i X_i - \left(\sum_{i=1}^{n} t_i - t\right)\frac{\sum_{i=1}^{n} X_i}{n}\right\}\right] \\
&= E\left[\exp\left\{\sum_{i=1}^{n} t_i X_i - \left(\frac{\sum_{i=1}^{n} t_i - t}{n}\sum_{i=1}^{n} X_i\right)\right\}\right] \\
&= E\left[\exp\left\{\sum_{i=1}^{n} t_i X_i - \left(\frac{n\bar{t} - t}{n}\sum_{i=1}^{n} X_i\right)\right\}\right] \\
&= E\left[\exp\left\{\sum_{i=1}^{n} X_i \left(t_i - \frac{n\bar{t} - t}{n}\right)\right\}\right] \\
&= E\left[\exp\left\{\sum_{i=1}^{n} X_i \left(\frac{nt_i - n\bar{t} + t}{n}\right)\right\}\right] \\
&= \prod_{i=1}^{n} E\left[\exp\left\{X_i \left(\frac{nt_i - n\bar{t} + t}{n}\right)\right\}\right]
\end{aligned}$$

$$
\begin{aligned}
&= \prod_{i=1}^{n} \exp\left[\mu\left(\frac{nt_i - n\bar{t} + t}{n}\right) + \left(\frac{nt_i - n\bar{t} + t}{n}\right)^2 \frac{\sigma^2}{2}\right] \\
&= \prod_{i=1}^{n} \exp\left[\frac{\mu}{n}\left\{n\left(t_i - \bar{t}\right) + t\right\} + \frac{\sigma^2}{2n^2}\left\{n\left(t_i - \bar{t}\right) + t\right\}^2\right] \\
&= \exp\left[\frac{\mu}{n}\left\{n\sum_{i=1}^{n}\left(t_i - \bar{t}\right) + nt\right\}\right. \\
&\qquad\qquad \left. + \frac{\sigma^2}{2n^2}\sum_{i=1}^{n}\left\{n\left(t_i - \bar{t}\right) + t\right\}^2\right] \\
&= \exp\left[\mu t + \frac{\sigma^2}{2n^2}\sum_{i=1}^{n}\left\{t^2 + 2nt\left(t_i - \bar{t}\right) + n^2\left(t_i - \bar{t}\right)^2\right\}\right] \\
&= \exp(\mu t)\ \exp\left[\frac{\sigma^2}{2n^2}\left\{nt^2 + n^2\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^2\right\}\right] \\
&= \exp(\mu t)\ \exp\left(\frac{t^2\sigma^2}{2n}\right)\ \exp\left[\frac{\sigma^2}{2}\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^2\right] \\
&= \exp\left(\mu t + \frac{t^2\sigma^2}{2n}\right)\ \exp\left[\frac{\sigma^2}{2}\sum_{i=1}^{n}\left(t_i - \bar{t}\right)^2\right] \\
&= M_{\bar{X}}(t)\ M_{(X_1 - \bar{X}),(X_2 - \bar{X}),\ldots,(X_n - \bar{X})}(t_1, t_2, \ldots, t_n) \\
&= M(t, 0, 0, \ldots, 0)\ M(0, t_1, t_2, \ldots, t_n)
\end{aligned}
$$

Therefore, $\bar{X}$ and $(X_1 - \bar{X}), (X_2 - \bar{X}), \ldots, (X_n - \bar{X})$ are independent. $\square$

Corollary 4.3.1. *Let $X_1, X_2, \ldots, X_n$ be i.i.d. $N(\mu, \sigma^2)$ random variables. Then $\bar{X}$ and S^2 are independent.*

Corollary 4.3.2. *Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$. Then,*

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

Proof. We have, $X_i \sim N(\mu, \sigma^2),\ i = i, 2, \ldots, n$. Then,

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n).$$

Also, as $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$,

$$\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \right)^2 \sim \chi^2(1).$$

Consider,

$$\begin{aligned}
\sum_{i=1}^{n}(X_i-\mu)^2 &= \sum_{i=1}^{n}(X_i-\bar{X}+\bar{X}-\mu)^2\\
&= \sum_{i=1}^{n}(X_i-\bar{X})^2+\sum_{i=1}^{n}(\bar{X}-\mu)^2\\
&\qquad +2(\bar{X}-\mu)\sum_{i=1}^{n}(X_i-\bar{X})\\
&= (n-1)S^2+n(\bar{X}-\mu)^2
\end{aligned}$$

Dividing each term by σ^2, we have

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2}+\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

Since $\bar{X}$ and S^2 are independent, we have

$$\begin{aligned}
M_{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}(t) &= M_{\frac{(n-1)S^2}{\sigma^2}} \times M_{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2}\\
\text{i.e.,}(1-2t)^{-n/2} &= M_{\frac{(n-1)S^2}{\sigma^2}} \times (1-2t)^{-1/2}.
\end{aligned}$$

Therefore,

$$M_{\frac{(n-1)S^2}{\sigma^2}} = (1-2t)^{-(n-1)/2}.$$

Hence,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

□

The p.d.f. of $\frac{(n-1)S^2}{\sigma^2}$ is given by

$$f_{\frac{(n-1)S^2}{\sigma^2}}(x) = \frac{\left(\frac{1}{2}\right)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{x}{2}} x^{\left(\frac{n-1}{2}-1\right)} \quad ; \quad 0 < x < \infty \ \cdot$$

Now, let $X = \frac{(n-1)S^2}{\sigma^2}$. Then, $S^2 = \frac{\sigma^2 X}{n-1}$ and

$$\begin{aligned} M_{S^2}(t) &= M_{\frac{\sigma^2 X}{n-1}}(t) \\ &= M_X\left(\frac{\sigma^2 t}{n-1}\right) \\ &= \left(1 - \frac{2\sigma^2 t}{n-1}\right)^{-(n-1)/2}. \end{aligned}$$

Therefore, $S^2 \sim Gamma(\frac{n-1}{2\sigma^2}, \frac{n-1}{2})$ with p.d.f.

$$f_{S^2}(x) = \frac{\left(\frac{n-1}{2\sigma^2}\right)^{\frac{n-1}{2}}}{\Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{(n-1)x}{2\sigma^2}} x^{\left(\frac{n-1}{2}-1\right)} \quad ; \quad 0 < x < \infty \;\cdot$$

Notes:

1. $E(S^2) = \frac{\frac{n-1}{2}}{\frac{n-1}{2\sigma^2}} = \sigma^2.$

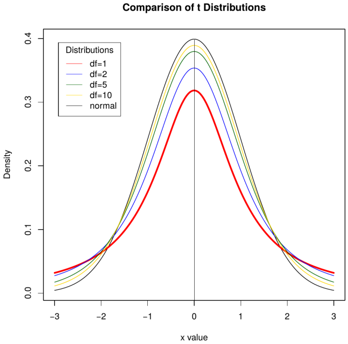
2. $V(S^2) = \frac{\frac{n-1}{2}}{\left(\frac{n-1}{2\sigma^2}\right)^2} = \frac{2\sigma^4}{(n-1)}.$

4.4 Student's *t*-distribution

According to CLT, if a simple random sample of size n is taken from a population whose mean and variance are μ and σ^2 respectively, then the sample mean $\bar{X}$ will be distributed normally with mean μ and variance $\frac{\sigma^2}{n}$, for large n. In other words, for a population which is not normal

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1) \quad \text{as} \quad n \to \infty.$$

When the population standard deviation σ is not known and S is the sample standard deviation, then also $\frac{\bar{X}-\mu}{S/\sqrt{n}} \to N(0,1)$ provided n, the sample size, is sufficiently large (i.e., $n \geq 30$). But, in the case when the population is normal, σ is unknown and sample size is small (i.e., $n < 30$) the distribution of $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ will not be normal. The distribution of the statistic in such cases is known as Student's t-distribution. It is the ratio of **independent** standard normal random variable and the square root of χ^2 random variable divided by its degrees of freedom. The distribution was derived by William Zealy Gosset, who wrote under the pen name 'Student', as his employer did not give him permission to publish his papers.



Definition 4.4.1. *Let $X \sim N(0, 1)$ and $Y \sim \chi^2(n)$, and let X*

and Y *be independent. Then, the statistic*

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

is said to have a t*-distribution with* n *degrees of freedom and we write* $T \sim t(n)$.

The p.d.f. of the random variable T with n degrees of freedom is given by

$$\begin{aligned} f_T(t) &= \frac{1}{\sqrt{n}\ \beta\left(\frac{1}{2},\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} \quad ; \quad -\infty < t < \infty \\ &= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\ \Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2} \quad ; \quad -\infty < t < \infty \end{aligned}.$$

Notes:

1. When $n = 1$, the above p.d.f. reduces to

   $$f(t) = \frac{1}{\pi(1+t^2)}; \quad -\infty < t < \infty,$$

   which is the standard Cauchy distribution $C(1, 0)$.

2. Let $X \sim t(n)$, $n > 1$. Then. $E(X^r)$ exists for $r < n$.

(a) If $r < n$ is odd, then

$$E(X^r) = 0.$$

(b) If $r < n$ is even, then

$$E(X^r) = n^{r/2} \frac{\Gamma\left(\frac{r+1}{2}\right)\ \Gamma\left(\frac{n-r}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\ \Gamma\left(\frac{n}{2}\right)}.$$

3. If $X \sim t(1)$, then $E(X)$ does not exist. (Cauchy)

4. If $X \sim t(n)$, $n > 1$, then $E(X) = 0$.

5. If $X \sim t(n)$, $n > 2$, then $E(X) = 0$ and $V(X) = E(X^2) = \frac{n}{n-2}$.

6. The curve representing t- distribution is symmetric about zero.

7. t- curve is unimodal.

8. Mean = Median = Mode.

9. For large n the curve is approximately normally distributed.

10. For t distribution m.g.f. does not exist.

Statistic following t-distribution

1. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$, where σ^2 is unknown. Let $\bar{X}$ be the sample mean and S^2 be the sample variance. Then,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1).$$

Proof. Here, $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$. Also $\bar{X}$ and S^2 are independent. Then, by the definition of T- statistic.

$$\begin{aligned} T &= \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/n-1}} \\ &= \frac{(\bar{X}-\mu)}{S/\sqrt{n}} \\ &= \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1). \end{aligned}$$

$\square$

Note: Since,

$$(n-1)S^2 = ns^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2,$$

$$\frac{(n-1)S^2}{\sigma^2} = \frac{ns^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

and

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n-1}} \sim t(n-1).$$

2. Let $X_1, X_2, \ldots, X_m$ be a random sample of size m from $N(\mu_1, \sigma^2)$ and $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from $N(\mu_2, \sigma^2)$ where σ^2 is unknown. Then,

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}\left(\frac{1}{m} + \frac{1}{n}\right)}} \sim t(m+n-2).$$

Proof. Here, $\bar{X} \sim N\left(\mu_1, \frac{\sigma^2}{m}\right)$ and $\bar{Y} \sim N\left(\mu_2, \frac{\sigma^2}{n}\right)$. Therefore,

$$(\bar{X} - \bar{Y}) \sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{m} + \frac{\sigma^2}{n}\right),$$

$$\frac{(m-1)S_1^2}{\sigma^2} \sim \chi^2(m-1)$$

and

$$\frac{(n-1)S_2^2}{\sigma^2} \sim \chi^2(n-1).$$

Hence, by the definition of T- statistic

$$\begin{aligned}
T &= \frac{\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma^2}{m}+\frac{\sigma^2}{n}}}}{\sqrt{\frac{\frac{(m-1)S_1^2}{\sigma^2}+\frac{(n-1)S_2^2}{\sigma^2}}{m+n-2}}} \\
&= \frac{\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}}{\frac{1}{\sigma}\sqrt{\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}}} \\
&= \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}} \sim t(m+n-2)
\end{aligned}$$

□

Now if $\mu_1 = \mu_2$, then

$$T = \frac{(\bar{X}-\bar{Y})}{\sqrt{\frac{(m-1)S_1^2+(n-1)S_2^2}{m+n-2}\left(\frac{1}{m}+\frac{1}{n}\right)}} \sim t(m+n-2).$$

4.5 F-distribution

F-distribution is named in honor of Prof. Ronald A. Fisher. It is the ratio of two **independent** χ^2 random variables divided by their respective degrees of freedom.

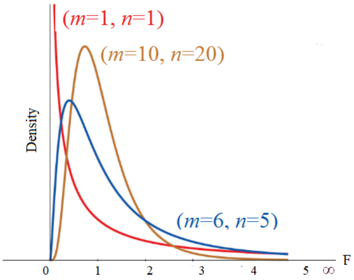
Definition 4.5.1. *Let X and Y be independent χ^2 random variables with m and n degrees of freedom respectively. Then the random variable*

$$F = \frac{X/m}{Y/n}$$

is said to have an F-distribution with (m, n) degrees of freedom, and we write $F \sim F(m, n)$.

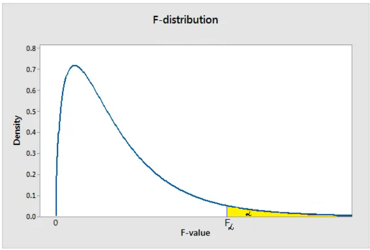
The p.d.f. of F with (m, n) degrees of freedom is given by,

$$f_F(x) = \frac{\left(\frac{m}{n}\right)^{m/2}}{\beta\left(\frac{m}{2}, \frac{n}{2}\right)} \frac{x^{\frac{m}{2}-1}}{\left(1+\frac{m}{n}x\right)^{\frac{m+n}{2}}} \quad ; \quad 0 < x < \infty \ .$$



Notes:

1. If $X \sim F(m,n)$, then $\frac{1}{X} \sim F(n,m)$.
2. If $m = 1$, i.e., $X \sim F(1,n)$, then $F = [t(n)]^2$.
3. $X \sim C(1,0) \ \Rightarrow \ X \sim t(1) \ \Rightarrow \ X^2 \sim F(1,1)$.
4. $P(F > F_{\alpha}) = \alpha$.



5. Let $X \sim F(m,n)$. Then for $k > 0$, integer,

$$E(X^k) = \frac{\Gamma\left(k+\frac{m}{2}\right)\ \Gamma\left(\frac{n}{2}-k\right)}{\Gamma\left(\frac{m}{2}\right)\ \Gamma\left(\frac{n}{2}\right)} \left(\frac{n}{m}\right)^k, \quad n > 2k.$$

6.

$$E(X) = \frac{n}{n-2}, \quad n > 2$$

$$V(X) = \frac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}, \quad n > 4.$$

7. No mean exists for $n \leq 2$.

8. F- distribution is also known ar variance-ratio distribution.

9. Let $X \sim F(m,n)$. Then the random variable $Z = \frac{1}{2}\log X$ is known as Fisher's Z- statistic.

Statistic Following F-distribution

1. Let $X_1, X_2, \ldots, X_m$ be a random sample of size m from $N(\mu_1, \sigma_1^2)$ and $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from $N(\mu_2, \sigma_2^2)$. Then,

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F(m-1, n-1).$$

If $\sigma_1 = \sigma_2$, then

$$F = \frac{S_1^2}{S_2^2} \sim F(m-1, n-1).$$

Uses of F Statistic

1. To test the equality of 2 normal population variances.
2. To test the equality of 3 or more normal population means.

Exercises

1. If $X_1, X_2, \ldots, X_n$ are independent $N(\mu, \sigma^2)$ random variables, find the distribution of $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$ and $Y = \frac{\bar{X}-\mu}{\sqrt{\sum_{i=1}^{n}(X_i-\mu)^2}}$.
2. If X_1, X_2, X_3 and X_4 are independent observations from a univariate normal population with mean zero and unit variance. Find the distribution of
   (a) $U = \frac{\sqrt{2}X_3}{\sqrt{X_1^2+X_2^2}}$
   (b) $V = \frac{3X_4^2}{X_1^2+X_2^2+X_3^2}$.

3. If X_1 and X_2 are independent χ^2 random variables each with 1 degree of freedom, find λ such that $P(X_1 + X_2 > \lambda) = \frac{1}{2}$.

4. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ and $\bar{X}$ and S^2, respectively be the sample mean and sample variance. Let $X_{n+1} \sim N(\mu, \sigma^2)$ and assume that $X_1, X_2, \ldots, X_n, X_{n+1}$ are independent. Find the sampling distribution of $\frac{X_{n+1} - \bar{X}}{S}\sqrt{\frac{n}{n+1}}$.

5. If $X_1, X_2, \ldots, X_n$ are independent $N(0, \sigma^2)$ random variables, what is the distribution of $Y = \frac{\sum_{i=1}^{n} X_i^2}{\sigma^2}$?

6. Let X and Y be independent standard normal variates. What is the distribution of $\frac{X^2}{Y^2}$ and write down p.d.f.

7. Let the random variable X has p.d.f. $f(x) = 1$, $0 \leq x \leq 1$. Show that $-2 \ln X$ has χ^2 distribution with 2 degrees of freedom.

8. Show that the m.g.f. of $\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ is $\left(1 - \frac{2t\sigma^2}{n}\right)^{-\left(\frac{n-1}{2}\right)}$. Obtain the mean as well as variance.

BIBLIOGRAPHY

[1] Rohatgi V. K. and Saleh, A.K. Md. E. (2009): *An Introduction to Probability and Statistics.* 2^{nd} Edn. (Reprint), John Wiley & Sons.

[2] S.C.Gupta and V. K. Kapoor, *Fundamentals of Mathematical Statistics*, Sultan Chand & Sons.

[3] Mood, A.M. Graybill, F.A. and Boes, D.C. (2007): *Introduction to the Theory of Statistics,* 3^{rd} Edn., (Reprint), Tata McGraw-Hill Pub. Co. Ltd.

[4] John E Freund, *Mathematical Statistics*, Pearson Edn, New Delhi.

[5] Cochran W.G. (1984):*Sampling Techniques*, 3^{rd} Edn, Wiley Eastern.

[6] K.X. Joseph: *Subsidiary Statistics.*