U.P. Rajarshi Tandon Open
University, Prayagraj

MScSTAT – 202N/ MASTAT –202N
Non Parametrics

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics,
D. D. U. Gorakhpur University, Gorakhpur

Prof. Shruti **Member-Secretary**
Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Prof. Shashi Bhushan **Writer**
Department of Statistics (Unit 1, 2, 3, 5, 6, 7, 8 & 9*)*
University of Lucknow, Lucknow

Dr. Pratyasha Tripathi **Writer**
Department of Statistics (Unit 4*)*
Tilka Manjhi Bhagalpur University, Bhagalpur, Bihar

Prof. S. K. Pandey **Editor**
Department of Statistics (Unit 5, 6, 7 & 8*)*
University of Lucknow, Lucknow

Prof. Shruti **Editor**
School of Sciences (Unit 1, 2, 3, 4 & 9*)*
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Shruti **Course Coordinator**
School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 202N/ MASTAT – 202N **NON PARAMETRICS**

First Edition: ***July 2023***
ISBN :

Printed By:

Blocks & Units Introduction

The present SLM on ***Non-Parametrics*** consists of eleven units with three blocks.

The ***Block - 1 – Order Statistics,*** is the first block, which is divided into four units.

The ***Unit - 1 – Basic Distribution Theory,*** is the first unit of present self learning material, which describes Order statistics, Distribution of maximum, minimum and r-th order statistic, Joint distribution of r-th and s-th order statistic.

In ***Unit – 2 – Asymptotic Distribution Theory,*** the main emphasis on the Moments of order statistics, non parametric estimation of distribution function, Glivenko-Cantelli fundamental theorem

In ***Unit – 3 – Distribution Free Intervals,*** we have focussed mainly on Distribution of range function of order statistics, distribution free confidence intervals for quintiles, distribution free tolerance interval

In ***Unit – 4 – Rank Order Statistics*** is discussed with rank order statistics, Dwass technique, Ballot theorem and its generatiuon, extention and application to fluctuation of sums of random variable.

The ***Block - 2 – Sequential Analysis*** is the second block with two units.

In ***Unit – 5 – Sequential Tests*** is discussed with SPRT and its properties, Wald's Fundamental identity, OC and ASN functions, Wald's equation.

In ***Unit – 6 – Sequential Estimation*** has been discussed, Cramer Rao Inequality of sequential estimation, Stein's two stage procedure.

The ***Block - 3 – Nonparametric Tests and Inference*** has three units.

Unit – 7 – One- Sample and Two-Sample Location Tests dealt with One and two sample location tests, Sign test. Wilcoxon test, Median test.

Unit – 8 – Other Non- Parametric Tests dealt with Mann- Whitney U- Test, Application of U-statistic to rank tests. One sample and two sample Kolmagorov-Smirnov tests. Run tests.

Unit – 9 – Non-Parametric Inference, The Kruskal-Wallis one way ANOVA Test, Friedman's two-way analysis of variance by ranks, efficiency criteria and theoretical basis for calculating ARE, Pitman ARE.

At the end of every block/unit the summary, self assessment questions and further readings are given.

References & Further Readings

- Bagdonavicius, V., Kruopis, J., Nikulin, M.S. (2011). "Non-parametric tests for complete data", ISTE & WILEY: London & Hoboken. ISBN 978-1-84821-269-5.
- Corder, G. W.; Foreman, D. I. (2014). Nonparametric Statistics: A Step-by-Step Approach. Wiley. ISBN 978-1118840313.
- Gibbons, Jean Dickinson; Chakraborti, Subhabrata (2003). Nonparametric Statistical Inference, 4th Ed. CRC Press. ISBN 0-8247-4052-1.
- Hettmansperger, T. P.; McKean, J. W. (1998). Robust Nonparametric Statistical Methods. Kendall's Library of Statistics. Vol. 5 (First ed.). London: Edward Arnold. New York: John Wiley & Sons. ISBN 0-340-54937-8. MR 1604954. also ISBN 0-471-19479-4.
- Hollander M., Wolfe D.A., Chicken E. (2014). Nonparametric Statistical Methods, John Wiley & Sons.
- Sheskin, David J. (2003) Handbook of Parametric and Nonparametric Statistical Procedures. CRC Press. ISBN 1-58488-440-1
- Wasserman, Larry (2007). All of Nonparametric Statistics, Springer. ISBN 0-387-25145-6.

U.P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 202N/ MASTAT –202N
Non Parametrics

Block: 1 Order Statistics

Unit – 1 : Basic Distribution Theory

Unit – 2 : Asymptotic Distribution Theory

Unit – 3 : Distribution Free Intervals

Unit – 4 : Rank Order Statistics

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics,
D. D. U. Gorakhpur University, Gorakhpur

Prof. Shruti **Member-Secretary**
Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Prof. Shashi Bhushan **Writer**
Department of Statistics (Unit 1, 2 & 3*)*
University of Lucknow, Lucknow

Dr. Pratyasha Tripathi **Writer**
Department of Statistics (Unit 4*)*
Tilka Manjhi Bhagalpur University, Bhagalpur, Bihar

Prof. Shruti **Editor**
School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Shruti **Course Coordinator**
School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 202N/ MASTAT – 202N **NON PARAMETRICS**
©UPRTOU
First Edition: ***July 2023***
ISBN :

Printed By:

Block & Units Introduction

The ***Block - 1 – Order Statistics,*** is the first block, which is divided into four units.

The ***Unit - 1 – Basic Distribution Theory,*** is the first unit of present self learning material, which describes Order statistics, Distribution of maximum, minimum and r-th order statistic, Joint distribution of r-th and s-th order statistic.

In ***Unit – 2 – Asymptotic Distribution Theory,*** the main emphasis on the Moments of order statistics, non parametric estimation of distribution function, Glivenko-Cantelli fundamental theorem

In ***Unit – 3 – Distribution Free Intervals,*** we have focussed mainly on Distribution of range function of order statistics, distribution free confidence intervals for quintiles, distribution free tolerance interval

In ***Unit – 4 – Rank Order Statistics*** is discussed with rank order statistics, Dwass technique, Ballot theorem and its generatiuon, extention and application to fluctuation of sums of random variable.

At the end of every unit the summary, self assessment questions and further readings are given.

UNIT:1 BASIC DISTRIBUTION THEORY

Structure

1.1 Introduction

The subject of order statistics deals with the properties and applications of these observations and their functions. For example, the median which is the middle most observation, the extreme observations X(1) and X(n), the sample range which is the difference of these extreme observations, the mid range which is the average of the two extreme observations X(1) and X(n)etc.

Applications of order statistics are in the theory of extremes, reliability theory, theory of outliers etc. The theory of extremes deals with floods, draughts, extreme breaking strength and fatigue failure etc. In reliability, suppose there are n components which can be useful only if at most r ($1 \leq r < n$) components are failing – for example consider the illumination bulbs used during festivals which are connected in series, if few of them fails, the string of bulbs can still be used by removing such defective bulbs; but beyond certain number of such removal the string of bulbs are quite useless. In outlier theory, outliers are those observations which deviate from the rest of the observations. Hence the extreme deviates from the estimate of the location parameters are useful tools in outlier detection procedures.

1.2 Objectives

The objective of this unit is to provide a basic understanding of concepts related to Order statistics. The concepts of Distribution of maximum, minimum and r-th order statistic, Joint distribution of r-th and s-th order statistic should be clear after study of this material.

1.3 Order Statistics

***Definition*:** The observation occupying r^{th} place in ascending order of the sample values is known as the r^{th} order statistic. We denote it by Y_r or $X_{(r)}$ so that $Y_1 = X_{(1)}$ represents the minimum of the sample observations while $Y_n = X_{(n)}$ is the maximum of sample observations.

The definition of order statistics does not require that the *X’s* to be identically distributed, nor do we need them to be independent. Also, it was not presumed that the parent distributions be continuous, nor that their densities exist. Although, most of the classical results dealing with order statistics were originally derived in more restrictive settings. Generally, it is assumed that the X’s were independent and identically distributed (i.i.d.) with common continuous (cumulative) distribution function *F(x)*, and having a density function *f(x)* and, henceforth, we will assume the *X’s* to be so.

The following list, though, not exhaustive, but may serve help to convince the reader that this text will not be focusing on some abstract concepts of little *practical utility*:

1. ***Robust Location Estimates:***

Suppose that *n* independent measurements are available, and we wish to estimate their assumed common mean. It has long been recognized that the sample mean, though attractive from many viewpoints, suffers from an extreme sensitivity to outliers and model violations. Estimates based on the median or the average of central order statistics are less sensitive to model assumptions.

2. *Detection of Outliers:*

If one is confronted with a set of measurements and is concerned with determining whether some have been incorrectly made or reported, attention naturally focuses on certain order statistics of the sample. Usually, the largest one or two and/or the smallest one or two are deemed most likely to be outliers.

3. *Censored Sampling:*

Fifty expensive machines are started up in an experiment to determine the expected life of a machine. If, as is to be hoped, they are fairly reliable, it would take an enormously long time to wait for all machines to fail. Instead, great savings in time and machines can be effected if we base our estimates on the first few failure times (i.e., the first few order statistics from the conceptual sample of i.i.d. failure times).

4. *Waiting for the Big One:*

Disastrous floods and destructive earthquakes recur throughout history. Dam construction has long focused on so called 100-year floods. Presumably the dams are built big enough and strong enough to handle any water flow to be encountered except for a level expected to occur only once every 100 years. Whether one agrees or not with the 100-year disaster philosophy, such inferences are concerned with the distribution of large order statistics from a possibly dependent, possibly not identically distributed sequence.

5. *Strength of Materials:*

The adage that a chain is no stronger than its weakest link underlies much of the theory of strength of materials, whether they be threads, sheets, or blocks. By considering failure potential in infinitesimally small sections of the material, one quickly is led to strength distributions associated with limits of distributions of sample minima, which is again an order statistic.

6. *Reliability:*

The example of a cord composed of n threads can be extended to lead us to reliability applications of order statistics. It may be that failure of one thread will cause the cord to break (the weakest link), but more likely the cord will function as long as *k* (a number less than *n)* of the threads remain unbroken.

7. ***Quality Control:***

Each candy bar should weigh 2.1 ounces; just a smidgen over the weight stated on the wrapper. No matter how well the candy pouring machine was adjusted at the beginning of the shift, minor fluctuations will occur, and potentially major aberrations might be encountered (if a peanut gets stuck in the control valve). We must be alert for correctable malfunctions causing unreasonable variation in the candy bar weight. Enter the quality control man with his *X and R charts* or his *median and R charts.* If the median (or perhaps the mean) is far from the target value, we must shut down the line.

8. ***Selecting the Best:***

Field trials of corn varieties involved carefully balanced experiments to determine which of several varieties is most productive. Obviously, we are concerned with the maximum of a set of probably not identically distributed variables in such a setting.

9. ***Inequality Measurement:***

The income distribution in India (where a few individuals earn most of the money) is clearly more unequal than that of United Kingdom (where progressive taxation has a leveling effect). How does one make such statements precise? The usual approach involves order statistics of the corresponding income distributions. The particular device used is called a *Lorenz curve.* It summarizes the percent of total income accruing to the poorest *p* percent of the population for various values of *p.* Mathematically this is just the scaled integral of the empirical quantile function. A high degree of convexity in the Lorenz curve signals a high degree of inequality in the income distribution.

10. ***Olympic Records:***

Bob Beamon's 1968 long jump remains on the *Olympic* record book. Few other records last that long. If the best performances in each Olympic Games were modeled as independent identically distributed random variables, then records would become more and more scarce as time went by. Such is not the case. The simplest explanation involves improving and increasing populations. Thus the 1964 high jumping champion was the best of, say, *Nx* active international-caliber jumpers. In 1968 there were more high-caliber jumpers of probably higher caliber. So, we are looking, most likely, at a sequence of not identically distributed random variables. But in any case, we are focusing on maxima, that is, on certain order statistics.

11. *Allocation of Prize Money:*

At the end of the annual Bob Hope golf tournament the player with the lowest score gets first prize. The second lowest score gets second prize, etc. In 1991 the first five prizes were: $198,000, $118,800, $74,800, $52,800, and $44,000. Obviously, we are dealing with order statistics here. Presumably the player with the highest ability level will most likely post the lowest score.

12. *Characterizations and Goodness of Fit:*

The exponential distribution is famous for its so-called lack of memory. The usual model involves a light bulb or other electronic device. For example, if $X_1, ,..., X_n$ are i.i.d. exponential, then their spacings $(X_{(i)} - X_{(j)})$ are again exponential and, remarkably, are independent. It is only in the case of exponential random variables that such spacings properties are encountered. A vast literature of exponential characterizations and related goodness-of-fit tests has consequently developed. It is interesting to note that most tests of goodness of fit for any parent distribution implicitly involve order statistics, since they often focus on deviations between the empirical quantile function and the hypothesized quantile function.

Let X_1, X_2,…, X_n be a random sample of size n taken from a continuous population whose p. d. f. is $f(x)$ and c .d. f. is $F(x)$ for $a<x<b$. Let Y_1 be the minimum of X_1, X_2,…., X_n is called the first order statistics, Y_2 be the next minimum is called the second order statistics and so on so that Y_n be the maximum of X_1, X_2,…., X_n is called the n^{th} order statistics. Then $Y_1 < Y_2 < .... < Y_n$ is known as order statistics of the random sample X_1, X_2,…, X_n.

1.4 Distribution of Maximum

Let X_1, $X_2,\ldots,$ X_n be a random sample of size n taken from a continuous population whose p. d. f. is $f(x)$ and c .d. f. is $F(x)$ for $a<x<b$. Let $F_{Y_n}(x)$ be the c .d. f. of maximum or the n-th order statistics Y_n at the point x is given by

$$F_{Y_n}(x)=P[Y_n \le x]$$

$$=\mathrm{P}[Max\ of\ X_1, X_2,\ldots, X_n \le x]$$

$$=\mathrm{P}[X_1 \le x, X_2 \le x,\ldots, X_n \le x]$$

$$=\mathrm{P}[X_1 \le x]P[X_2 \le x]\ldots P[X_n \le x] \qquad (X_1, X_2,\ldots, X_n \text{ are mutually independent})$$

$$=F_{X_1}(x)F_{X_2}(x)\ldots F_{X_n}(x) \qquad (1)$$

where $F_{X_i}(x)$ is c. d. f. of X_i for i= 1, 2,…, n and since $X_1, X_2,\ldots, X_n$ are identically distributed with c. d. f. $F(x)$ such that

$$F_{X_1}(x)=F_{X_2}(x)=\ldots=F_{X_n}(x)=F(x) \qquad (2)$$

Using (2) in (1), we get

$$F_{Y_n}(x)=\{F(x)\}^n \qquad (3)$$

Since in case continuous population, the density function is given by

p. d. f. of Y_n = $f_{Y_n}(x)$

$$=\frac{dF_{Y_n}(x)}{dx}$$

$$=n\{F(x)\}^{n-1}\frac{dF(x)}{dx} \qquad \left[\because f(x)=\frac{dF(x)}{dx}\right]$$

$$=\begin{cases} n[F(x)]^{n-1} f(x) & ; a<x<b \\ 0 & ; \text{otherwise} \end{cases} \qquad (4)$$

Example: *Find out the p.d.f. of the maximum of the sample values from a sample of size n drawn from $U(0,\theta)$ parent.*

Solution: Consider the p.d.f. of $U(0,\theta)$ given by

$$f(x)=\begin{cases}\dfrac{1}{\theta} & ;0<x<\theta\\ 0 & ;otherwise\end{cases}$$

so that its d.f. is given by

$$F(x)=\begin{cases}0 & ;x\le 0\\ \dfrac{x}{\theta} & ;0<x\le\theta\\ 1 & ;x\ge\theta\end{cases}$$

Let $X_{(n)}$ be the n^{th} order statistics or maximum of sample values in a sample of size *n*. Then

$$f_{X(n)}(x)=n\{F(x)\}^{n-1}f(x)\quad ;0<x<\theta$$

$$=n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta}\qquad ;0<x<\theta$$

$$=n\,\frac{x^{n-1}}{\theta^n}\qquad ;0<x<\theta$$

Hence, the p.d.f. of $X_{(n)}$ is given by

$$f_{X_{(n)}}(x)=\begin{cases}\dfrac{nx^{n-1}}{\theta^n} & ;0<x<\theta\\ 0 & ;otherwise\end{cases}$$

1.5 Probability Integral Transform

If X is a random variable of a continuous type having p.d.f. $f(x)$ and distribution function $F(x)$, then $Z=F(x)$ has a uniform distribution $U(0,1)$.

Proof: Given X is a continuous random variable with p. d. f. $f(x)$ and c. d. f. $F(x)$ then we wish to prove that $Z = F(x)$ follows $U(0,1)$. Consider the p.d.f. of *z* given by

$f(z) = (\operatorname{mod} J)$ (Put *x* in terms of *z* in $f(x)$)

where J stands for the jacobian of transformation. For particular (or specific) values of *z* and *x*, we may write

$z = F(x)$

$$J = \left|\frac{dx}{dz}\right| = \left|\frac{1}{dz/dx}\right| = \left|\frac{1}{dF(x)/dx}\right| = \frac{1}{f(x)} \qquad (\because f(x) \geq 0)$$

so that

$$f(z) = \begin{cases} 1 & 0 < z < 1 \\ 0 & otherwise \end{cases}$$

$\Rightarrow Z = F(x) \sim U(0,1)$ Q.E.D.

Remark: The importance of probability integral transform is that the order statistics $X_{(1)}, \ldots, X_{(n)}$ in a sample from any continuous distribution with c.d.f. $F(x)$ are transformed by order preserving probability integral transform $u = F(x)$ into $U_{(1)}, \ldots, U_{(n)}$.

Example: *If X is a uniform random variable with distribution function F(x), prove that*

$$E\left[X_{(r)}\right] = \frac{n!}{(r-1)!(n-r)!}\int_0^1 Y^{r-1}(1-Y)^{n-r}\, h(Y)\, dY$$

where $h(Y) = F^{-1}(Y)$

Solution: Let $X \sim U(a,b)$ and consider

$$E\left[X_{(r)}\right] = \int_a^b x\, f_{X_{(r)}}(x)\, dx \qquad ; a<x<b$$

$$= \int_a^b x \frac{n!}{(r-1)!(n-r)!}\{F(x)\}^{r-1}\{1 - F(x)\}^{n-r} f(x)\, dx$$

Let $F(x) = y$

$$\Rightarrow \frac{dy}{dx} = \frac{dF(x)}{dx} = f(x) \Rightarrow dy = f(x)dx$$

so that

$$E\left[X_{(r)}\right] = \int_0^1 F^{-1}(y) \frac{n!}{(r-1)!(n-r)!} y^{r-1}(1-y)^{n-r}\, dy$$

$$= \frac{n!}{(r-1)!(n-r)!} \int_0^1 y^{r-1}(1-y)^{n-r}\, h(y)\, dy \qquad \left[\because F^{-1}(y) = h(y)\right] \qquad \text{Q.E.D.}$$

Example: *Let* $x_{1,} x_2, x_3$ *be independent random variable with p. d. f.*

$$f(x) = \exp\left[-(x-\theta)\right] I_{(\theta,\infty)}(x)$$

Determine the constant $c = c(\theta)$ *for which* $P\left(\theta < x_{(3)} < c\right) = 0.96$

Solution: Since

$$f(x) = \mathrm{e}^{\left[-(x-\theta)\right]} I_{(\theta,\infty)}(x)$$

where $I_{(\theta,\infty)}(x) = \begin{cases} 1 & if\ \theta \le x \le \infty \\ 0 & otherwise \end{cases}$

In other words

$$f(x) = \begin{cases} e^{-(x-\theta)} & ;\theta \le x \le \infty \\ 0 & ;otherwise \end{cases}$$

$$F(x) = \Pr[X \le x] = \int_{-\infty}^{x} f(x)dx$$

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

$$= 0 + \int_{\theta}^{x} e^{-(x-\theta)}\, dx$$

$$= \left[-e^{-(x-\theta)} \right]_{\theta}^{x}$$

$$= \left[1 - e^{-(x-\theta)} \right]$$

Therefore, $F(x) = \begin{cases} 1 - e^{-(x-\theta)} & ; \theta \le x \le \infty \\ 0 & ; otherwise \end{cases}$

We may write $P\left(\theta < x_{(3)} < c\right) = 0.96$ as

$$\int_{\theta}^{c} f_{x_{(3)}}(x)\, dx = 0.96$$

where $f_{x_{(3)}}(x)$ stands for the p. d. f. of third order statistic where n=3, so that

$$\int_{\theta}^{c} 3\left\{F(x)\right\}^2 f(x) dx = 0.96$$

Let $F(x) = t \Rightarrow f(x)dx = dt$ so that

$$\int_{F(\theta)}^{F(c)} 3t^2 dt = 0.96$$

$$\Rightarrow \int_{0}^{F(c)} 3t^2 dt = 0.96$$

where $F(\theta) = 0$ and $F(c) = 1 - e^{-(c-\theta)}$ so we get

$$t^3\Big|_{0}^{F(c)} = 0.96$$

$$\left[F(c)\right]^3 = 0.96$$

$$1 - e^{-(c-\theta)} = (0.96)^{1/3}$$

$$1 - (0.96)^{1/3} = e^{-(c-\theta)}$$

$$c = \theta - \ln\left[1 - (0.96)^{1/3}\right]$$

is the required value of c, such that $\Pr\left[\theta < x_{(3)} < c\right] = 0.96$.

1.6 Distribution of Minimum

Let $X_1,\ X_2,\ldots,\ X_n$ be a random sample of size n taken from a continuous population whose p. d. f. is $f(x)$ and c .d. f. is $F(x)$ for $a<x<b$. Let $F_{Y_1}(x)$ be the c .d. f. of minimum or the first order statistics Y_1 at the point x is given by

$$F_{Y_1}(x) = \mathrm{P}\left[Y_1 \le x\right]$$

$$= 1-\mathrm{P}\left[Y_1 > x\right]$$

$$= 1-\mathrm{P}\left[\min\left(X_1, X_2,..., X_n\right) > x\right]$$

$$= 1-\mathrm{P}\left[X_1 > x, X_2 > x,..., X_n > x\right]$$

$$= 1-P\left[X_1 > x\right]P\left[X_2 > x\right]...P\left[X_n > x\right] \quad (X_1, X_2,...., X_n \text{ are mutually ind.})$$

$$= 1-\left\{1-P\left[X_1 \le x\right]\right\}\left\{1-P\left[X_2 \le x\right]\right\}...\left\{1-P\left[X_n \le x\right]\right\}$$

$$= 1-\left\{1-F_{X_1}(x)\right\}\left\{1-F_{X_2}(x)\right\}...\left\{1-F_{X_n}(x)\right\}$$

where $F_{X_i}(x)$ is c. d. f. of X_i for i= 1, 2, …, n and since $X_1, X_2,..., X_n$ are identically distributed with c. d. f. $F(x)$ such that

$$F_{X_1}(x) = F_{X_2}(x) = ... = F_{X_n}(x) = F(x) \qquad (2)$$

Using (2) in (1), we get

$$F_{Y_1}(x) = 1-\left\{1-F(x)\right\}^n \qquad ; a<x<b$$

Since in case continuous population, the density function is given by

p. d. f. of Y_1 = $f_{Y_1}(x)$

$$= \frac{dF_{Y_1}(x)}{dx}$$

$$= n\{1-F(x)\}^{n-1}\frac{dF(x)}{dx} \qquad \left[\because f(x)=\frac{dF(x)}{dx}\right]$$

$$= \begin{cases} n\{1-F(x)\}^{n-1} f(x) & ; a<x<b \\ 0 & ;\text{otherwise} \end{cases} \tag{4}$$

Example: *Let $X_j\,(j=1,2,...,n)$ be i.i.d. negative exponential random variable with parameter λ then show that the distribution of $X_{(1)}$ is a negative exponential distribution with parameter $n\lambda$. Conversely, show that if $X_j\,(j=1,2,...,n)$ are i.i.d. random variables and $X_{(1)}$ follows a negative exponential distribution with parameter $n\lambda$ then the common distribution of X's is negative exponential with parameter λ.*

Solution:

It is given as $f(x)=\begin{cases}\lambda e^{-\lambda x} & ;0<x<\infty \\ 0 & ;elsewhere\end{cases}$

Therefore $F(x)=\begin{cases}1-e^{-\lambda x} & ;0<x\le\infty \\ 0 & ;otherwise\end{cases}$

Let $F_{X_{(1)}}(x)$ be the c. d. f. of $X_{(1)}$ and since, $X_1, X_2, ..., X_n$ are identically distributed with c. d. f. $F(x)$. Therefore,

$F_{X_{(1)}}(x)=F_{X_{(2)}}(x)=....=F_{X_{(n)}}(x)=F(x)$

Hence,

$$F_{X_{(1)}}(x)=1-\{1-F(x)\}^n$$

$$=1-\{1-1+e^{-\lambda x}\}^n$$

$$=1-e^{-\lambda xn}$$

therefore,

$$f_{X_{(1)}}(x) = \frac{dF_{X_{(1)}}(x)}{dx}$$

$$= \frac{d}{dx}\left(1 - e^{-\lambda xn}\right)$$

$$= n\lambda e^{-\lambda xn}$$

$X_{(1)}$ follows negative exponential with parameter $n\lambda$.

Conversely suppose it is given that $X_{(1)}$ follows negative exponential with parameter $n\lambda$ so that $F_{X_{(1)}}(x) = 1 - e^{-\lambda xn}$

Also, $F_{X_{(1)}}(x) = 1 - \{1 - F(x)\}^n$

Equating both we have,

$$1 - \{1 - F(x)\}^n = 1 - e^{-n\lambda x}$$

$$1 - F(x) = e^{-x\lambda}$$

$$F(x) = 1 - e^{-\lambda x}$$

$$\Rightarrow f(x) = \frac{d}{dx} F(x)$$

$$= \frac{d}{dx}\left(1 - e^{-\lambda x}\right)$$

$$= \lambda e^{-\lambda x}$$

which is negative exponential with parameter λ. Hence, the common distribution of X's is negative exponential with parameter λ.

1.7 Distribution of r-th Order Statistic

Let $X_1,\ X_2,\ldots,\ X_n$ be a random sample of size n taken from a continuous population whose p. d. f. is $f(x)$ and c .d. f. is $F(x)$ for $a<x<b$. Let $F_{Y_r}(x)$ be the c. d. f. and $f_{Y_r}(x)$ be the p. d. f. of the r^{th} order statistic Y_r at the point x is given by

$$f_{Y_r}(x)=\frac{dF_{Y_r}(x)}{dx}$$

$$=\lim_{h\to 0}\frac{F(x+h)-F(x)}{h}$$

$$=\lim_{h\to 0}\frac{\Pr(Y_r\le x+h)-\Pr(Y_r\le x)}{h}$$

$$=\lim_{h\to 0}\frac{1}{h}\mathrm{P}(x\le Y_r\le x+h)$$

$$\overset{\downarrow}{a}\underbrace{\qquad\qquad\qquad\qquad}_{r-1\ X\text{'s}<x}\overset{\downarrow}{x}\underbrace{\qquad\quad}_{\text{one } X<x+h}\overset{\downarrow}{x+h}\underbrace{\qquad\qquad\qquad\qquad}_{n-r\ X\text{'s}>x+h}\overset{\downarrow}{b}$$

$$=\lim_{h\to 0}\frac{1}{h}\mathrm{P}\left[r-1\text{of the } X's\le x,\ \text{one } X \text{ in}(x,x+h],\ \ n-r \text{ of the } X's>x+h\right]$$

Using multinomial law, we have

$$f_{Y_r}(x)=\lim_{h\to 0}\frac{1}{h}\frac{n!}{(r-1)!(n-r)!}\left\{\mathrm{P}(X\le x)\right\}^{r-1}\mathrm{P}(x\le X\le x+h)\left\{\mathrm{P}(X>x+h)\right\}^{n-r}$$

$$=\lim_{h\to 0}\frac{1}{h}\frac{n!}{(r-1)!(n-r)!}\left\{F(x)\right\}^{r-1}\left\{F(x+h)-F(x)\right\}\left\{1-F(x+h)\right\}^{n-r}$$

$$=\lim_{h\to 0}\frac{n!}{(r-1)!(n-r)!}\left\{F(x)\right\}^{r-1}\left\{\frac{F(x+h)-F(x)}{h}\right\}\left\{1-F(x)\right\}^{n-r}$$

$$=\begin{cases}\dfrac{n!}{(r-1)!(n-r)!}\left\{F(x)\right\}^{r-1}\left\{1-F(x)\right\}^{n-r}f(x) & ;a<x<b\\ 0 & ;\text{otherwise}\end{cases}$$

Remark: It is interesting to note that the $F_{Y_1}(x) = 1 - \{1 - F(x)\}^n$ and $F_{Y_n}(x) = \{F(x)\}^n$ are special cases of the general result of $F_{Y_r}(x)$ given by

$$
\begin{aligned}
F_{Y_r}(x) &= P(Y_r \le x) \\
&= P\{\text{at least } r \text{ of the } X_i \text{ are less than or equal to } x\} \\
&= \sum_{i=r}^{n} \binom{n}{i} F^i(x)\left[1 - F(x)\right]^{n-i}
\end{aligned}
$$

since the summand is the binomial probability of getting exactly i of the $X_1, \ldots, X_n$ less than or equal to x. Also, one more useful relationship that exists between the binomial sums and incomplete beta functions is

$$F_{Y_r}(x) = I_{F(x)}(r, n-r+1)$$

where $I_p(a,b)$ is an incomplete beta function defined as

$$I_p(a,b) = \int_0^p y^{a-1}(1-y)^{b-1}\, dy \quad ; a > 0, b > 0$$

Therefore, $F_{Y_r}(x)$ can be calculated from the tables of $I_p(a,b)$ and the percentage points of Y_r can be obtained by inverse interpolation of these tables.

Example: *Obtain the upper 5% point of Y_4 in sample of 5 from standard normal distribution.*

Solution: We need to find x such that

$$
\begin{aligned}
F_{Y_4}(x) &= 0.95 \\
&= I_{F(x)}(4, 5-4+1) = I_{F(x)}(4, 2)
\end{aligned}
$$

so that

$$0.05 = I_{1-F(x)}(2, 4)$$

thereby giving

$$0.0764 = 1 - F(x)$$

Hence, from normal tables, we have x = 1.43.

Example: *Let* $Y_1 < Y_2 < Y_3 < Y_4$ *denote the order statistics of the random sample of size 4 from the population with p. d. f.*

$$f(x) = \begin{cases} 2x & ; 0 \le x \le 1 \\ 0 & ; otherwise \end{cases}$$

Obtain the p. d. f. of Y_3 *and* $P(1/2 < Y_3)$

Solution: It is given that n=4 and

$$f(x) = \begin{cases} 2x & ; 0 \le x \le 1 \\ 0 & ; otherwise \end{cases}$$

Hence, for $x < 0$

$$F(x) = \Pr[X \le x] = \int_{-\infty}^{x} 0\,dx = 0$$

For $x \le 1$

$$F(x) = \Pr[X \le x] = \int_{-\infty}^{0} f(x)dx + \int_{0}^{x} f(x)dx = x^2$$

For $x > 1$

$$F(x) = \Pr[X \le x] = \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{x} f(x)dx = 1$$

Therefore, $F(x) = \begin{cases} 0 & ; x < 0 \\ x^2 & ; 0 \le x \le 1 \\ 1 & ; x \ge 1 \end{cases}$

Putting r=3 and n=4 in (1) we have,

p. d. f. of $Y_3 = f(Y_3)$

$$= \frac{4!}{2!\,1!}\{F(y_3)\}^2\{1 - F(y_3)\}f(y_3) \qquad ; \ 0 \le y_3 \le 1$$

$$=12\left[y_3^2\right]^2\left[1-y_3^2\right]2y_3 \qquad ;0\le y_3\le 1$$

$$=24\left[y_3^5\right]\left[1-y_3^2\right] \qquad ;0\le y_3\le 1$$

so that

$$f(Y_3)=\begin{cases}24y_3^5\left(1-y_3^2\right) & ;0\le y_3\le 1\\ 0 & ;otherwise\end{cases}$$

Now, $\Pr\left(Y_3>1/2\right)=1-\Pr\left(Y_3\le 1/2\right)$

$$=1-\int_0^{1/2} f\left(y_3\right)\,dy_3$$

$$=1-24\int_0^{1/2}\left(1-y_3^2\right)y_3^5\,dy_3$$

Let $y_3^2=t\ \therefore 2y_3\,dy_3=dt$

$$\Pr\left(Y_3>1/2\right)=1-12\int_0^{1/4}t^2\left(1-t^2\right)dt$$

$$=1-\left[4t^3-3t^4\right]_0^{1/4}=1-\frac{1}{64}\left[4-\frac{3}{4}\right]=1-\frac{13}{256}$$

$$=\frac{243}{256}$$

Example: *Let* $f\left(x,\theta\right)=1/\theta\ ,0<x<\theta$ *and* x_1,x_2,x_3 *be a random sample of size 3 from this parent distribution and let* Y_1,Y_2,Y_3 *be order statistic of this sample so that* $Y_1=\min\left(x_1,x_2,x_3\right)$ *and* $Y_3=\max\left(x_1,x_2,x_3\right)$. *Obtain* $\Pr\left(Y_2\ge\theta/2\right)$.

Solution: Given

$$f\left(x\right)=f\left(x,\theta\right)=1/\theta\ ,0<x<\theta$$

Now for $x<\theta$

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(x)dx = 0$$

$$= \int_{-\infty}^{0} f(x)dx + \int_{0}^{x} f(x)dx = 0 + \int_{0}^{x} \frac{1}{\theta}dx$$

$$= \frac{x}{\theta}$$

For $x > \theta$

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{\theta} f(x)dx = \int_{\theta}^{x} f(x)dx = 1$$

Therefore, $F(x) = \begin{cases} 0 & ; x \le 0 \\ x/\theta & ; 0 < x \le \theta \\ 1 & ; x \ge \theta \end{cases}$

Thus, $\Pr(Y_2 \ge \theta/2) = \int_{\theta/2}^{\theta} f(w)\, dw$

where $f(w)$ is the p. d. f. of second order statistic for n=3. So

$$f(w) = \frac{3!}{(2-1)!(3-2)!}\{F(w)\}^{2-1}\{1-F(w)\}^{3-2} f(w)$$

$$= 6\{F(w)\}\{1-F(w)\} f(w) \qquad ; 0 < w < \theta$$

So, $\Pr(Y_2 \ge \theta/2) = \int_{\theta/2}^{\theta} 6F(w)\{1-F(w)\} f(w)\, dw$

Let $F(w) = t \Rightarrow f(w)dw = dt$ and $F(\theta/2) = 1/2, F(\theta) = 1$

so that

$$\Pr(Y_2 \ge \theta/2) = \int_{1/2}^{1} 6t(1-t)dt$$

$$=6\left[\frac{t^2}{2}-\frac{t^3}{3}\right]_{1/2}^{1}=\left[3t^2-2t^3\right]_{1/2}^{1}=3-2-\frac{3}{4}+\frac{1}{4}$$

$$=\frac{1}{2}$$

Example: *Let* $Y_1<Y_2<Y_3<Y_4$ *be the order statistics of the random sample of size 4 from the distribution having probability density function*

$$f(x)=\begin{cases} e^{-x} & ;0<x<\infty \\ 0 & ;otherwise \end{cases}$$

Find $\Pr(3\le Y_4)$.

Solution: Given that

$$f(x)=\begin{cases} e^{-x} & ;0<x<\infty \\ 0 & ;otherwise \end{cases}$$

Now for $x\le 0$

$$F(x)=\Pr[X\le x]=\int_{\infty}^{0} f(x)\,dx+\int_{0}^{x} f(x)\,dx$$

$$=0+\int_{0}^{x} e^{-x}dx$$

$$=1-e^{-x}$$

Therefore,

$$\Pr[Y_4\ge 3]=\Pr[3\le 4]$$

$$=\int_{3}^{\infty} f(w)\,dw$$

where $f(w)$ is the p. d. f. of 4th order statistic Y_4 for n=4

$$\because f(w) = 4\left[F(w)\right]^{4-1} f(w)dw \qquad ; 0 < w \le \infty$$

so that

$$\Pr\left[Y_4 \ge 3\right] = \int_3^{\infty} 4\left[F(w)\right]^3 f(w)dw$$

Putting $F(w) = t \;\therefore f(w)dw = dt$, we have

$$\Pr\left[Y_4 \ge 3\right] = \int_{F(3)}^{F(\infty)} 4[t]^3 dt = \left[t^4\right]_{F(3)}^{F(\infty)} \qquad \text{since } \left[F(3)\right] = 1 - e^{-3} \text{ and } \left[F(\infty)\right] = 1$$

$$= 1 - \left(1 - e^{-3}\right)^4$$

1.8 Joint Distribution of r-th and s-th Order Statistic

Let $X_1,\ X_2, \ldots,\ X_n$ be a random sample of size n taken from a continuous population whose p. d. f. is $f(x)$ and c .d. f. is $F(x)$ for $a < x < b$. Let $F_{rs}(x, y)$ be the joint c. d. f. and $f_{rs}(x, y)$ be the joint p. d. f. of the r^{th} and s^{th} $(r < s)$ order statistic Y_r and Y_s at the point (x, y), $x<y$, is given by

$$f_{rs}(x, y) = \frac{d^2 F_{rs}(x, y)}{dxdy}$$

$$= \lim_{\substack{h \to 0 \\ k \to 0}} \frac{\mathrm{P}\left(x < Y_r \le x + h,\ y < Y_s \le y + k\right)}{hk}$$

$$= \lim_{\substack{h \to 0 \\ k \to 0}} \frac{1}{hk} \mathrm{P}\left[\begin{matrix} r-1 \text{ of the } X\text{'s} \le x,\ \text{one } X \text{ in} \left(x, x+h\right], s-r-1 \text{ of the } X\text{'s in} \left(x+h, y\right], \\ \text{one } X \text{ in} \left(y, y+k\right],\ \ n-s \text{ of the } X\text{'s} > y+k \end{matrix}\right]$$

$$\begin{matrix} \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ a & \underbrace{\hspace{3em}}_{r-1\ X\text{'s} < x} & x & \underbrace{\hspace{3em}}_{\text{one } X < x+h} & x+h & \underbrace{\hspace{3em}}_{s-r-1\ X\text{'s} < y} & y & \underbrace{\hspace{3em}}_{\text{one } X < y+k} & y+k & \underbrace{\hspace{3em}}_{n-s\ X\text{'s} > y+k} & b \end{matrix}$$

Using multinomial law, we have

$$=\lim_{\substack{h\to 0\\ k\to 0}}\frac{1}{hk}\frac{n!}{(r-1)!1!(s-r-1)!1!(n-s)!}\begin{bmatrix}\{P(X\le x)\}^{r-1}P(x\le X\le x+h)\{P(x+h\le X\le y)\}^{s-r-1}\\ P(y\le X\le y+k)\{P(X>y+k)\}^{n-s}\end{bmatrix}$$

$$=\lim_{\substack{h\to 0\\ k\to 0}}\frac{1}{hk}\frac{n!}{(r-1)!(s-r-1)!(n-s)!}\begin{bmatrix}\{F(x)\}^{r-1}\{F(x+h)-F(x)\}\{F(y)-F(x+h)\}^{s-r-1}\\ \{F(y+k)-F(y)\}\{1-F(y+k)\}^{n-s}\end{bmatrix}$$

$$=\lim_{\substack{h\to 0\\ k\to 0}}\frac{1}{hk}\frac{n!}{(r-1)!(s-r-1)!(n-s)!}\{F(x)\}^{r-1}\lim_{h\to 0}\{F(y)-F(x+h)\}^{s-r-1}$$

$$\lim_{k\to 0}\{1-F(y+k)\}^{n-s}\lim_{k\to 0}\frac{1}{k}\{F(y+k)-F(y)\}\lim_{h\to 0}\frac{1}{h}\{F(x+h)-F(x)\}$$

$$=\begin{cases}\dfrac{n!}{(r-1)!(s-r-1)!(n-s)!}\{F(x)\}^{r-1}\{F(y)-F(x)\}^{s-r-1}\{1-F(y)\}^{n-s}\ f(x)f(y); & a<x<y<b\\ 0 & ;\text{otherwise}\end{cases}$$

Example: Let $Y_1<Y_2<Y_3<Y_4$ be the order statistics of a random sample of size 4 from the probability distribution function

$$f(x)=\begin{cases}e^{-x} & ;0<x<\infty\\ 0 & ;otherwise\end{cases}$$

Show that Y_2 and Y_4-Y_2 are stochastically independent.

Solution: For $x\le\infty$

$$F(x)=\Pr(X\le x)=\int_{-\infty}^{x}f(x)dx$$

$$=\int_{-\infty}^{0}f(x)dx+\int_{0}^{x}f(x)dx$$

$$=0+\left[-e^{-x}\right]_0^x$$

$$=1-e^{-x}$$

Hence, $F(x)=\begin{cases}1-e^{-x} & ;0<x\le\infty \\ 0 & ;otherwise\end{cases}$

Let $Z_1=Y_2$ and $Z_2=Y_4-Y_2$

Then, the joint p. d. f. of Y_2 and Y_4 is given by

$$g_{24}(y_2,y_4)=\frac{5!}{(5-9)!(2-1)!(9-2-1)!}\{F(y_2)\}^{2-1}\{F(y_4)-F(y_2)\}^1\{1-F(y_4)\}^1 f(y_2)f(y_4)$$

$$;0<y_2<y_4<\infty$$

$$=120\{1-e^{-y_2}\}\{1-e^{-y_4}-1+e^{-y_2}\}\{1-1+e^{-y_4}\}\{e^{-y_2}\}\{e^{-y_4}\}$$

$$=120\{e^{-y_2}\}\{1-e^{-y_2}\}\{e^{-y_2}-e^{-y_4}\}\{e^{-2y_4}\} \qquad ;0<y_2<y_4<\infty \qquad (1)$$

For specific values we may write

$z_1=y_2$ and $z_2=y_4-y_2$

i.e. $y_2=z_1$ and $y_4=z_2+z_1$

The jacobian of transformation is

$$J=\frac{\partial(y_2,y_4)}{\partial(z_1,z_2)}=\begin{vmatrix}1 & 0\\ 1 & 1\end{vmatrix}=1$$

Hence the joint p. d. f. of z_1 and z_2 is

$f(z_1,z_2)=$ (mod J)(put y_2 and y_4 in terms of z_1 and z_2 in $g_{24}(y_2,y_4)$)

$$=120e^{-z_1}\left(1-e^{-z_1}\right)\left(e^{-z_1}-e^{-z_1-z_2}\right)e^{-2(z_1+z_2)} \qquad 0<z_1<\infty, 0<z_2<\infty$$

$$=120e^{-4z_1}\left(1-e^{-z_1}\right)\left(1-e^{-z_2}\right)e^{-2z_2}$$

Now,

$$f(z_1) = \int_0^\infty f(z_1, z_2)\, dz_2$$

$$= 120\, e^{-4z_1}\left(1 - e^{-z_1}\right)\int_0^\infty \left(1 - e^{-z_2}\right) e^{-2z_2}\, dz_2 \qquad ;\ 0 < z_1 < \infty$$

Let $e^{-z_2} = t$ and $e^{-z_2} dz_2 = dt$ so that

$$f(z_1) = 120\, e^{-4z_1}\left(1 - e^{-z_1}\right)\int_1^0 t(1-t)(-dt) \qquad ;\ 0 < z_1 < \infty$$

$$= 120\, e^{-4z_1}\left(1 - e^{-z_1}\right)\int_0^1 \left(t - t^2\right) dt \qquad ; 0 < z_1 < \infty$$

$$= 120\, e^{-4z_1}\left(1 - e^{-z_1}\right)\left(\frac{t^2}{2} - \frac{t^3}{3}\right)_0^1 \qquad ; 0 < z_1 < \infty$$

$$= 120\, e^{-4z_1}\left(1 - e^{-z_1}\right)\left(\frac{1}{2} - \frac{1}{3}\right) \qquad ;\ 0 < z_1 < \infty$$

$$= 20\, e^{-4z_1}\left(1 - e^{-z_1}\right) \qquad ; 0 < z_1 < \infty \qquad (3)$$

Similarly,

$$f(z_2) = \int_0^\infty f(z_1, z_2)\, dz_1$$

$$= 120\, e^{-2z_2}\left(1 - e^{-z_2}\right)\int_0^\infty \left(1 - e^{-z_1}\right) e^{-4z_1}\, dz_1 \qquad ; 0 < z_2 < \infty$$

Let $e^{-z_1} = t \Rightarrow -e^{-z_1} dz_1 = dt$

$$f(z_2) = 120\, e^{-z_2}\left(1 - e^{-z_2}\right)\int_0^1 t^3 (1-t)\, dt \qquad ; 0 < z_2 < \infty$$

$$= 120\, e^{-z_2}\left(1 - e^{-z_2}\right)\left(\frac{t^4}{4} - \frac{t^5}{5}\right)_0^1 \qquad ; 0 < z_2 < \infty$$

$$= 6e^{-2z_2}\left(1-e^{-z_2}\right) \qquad ; \; 0 < z_2 < \infty \qquad (4)$$

From (2), (3) and (4), we have,

$$f\left(z_1, z_2\right) = f\left(z_1\right) f\left(z_2\right)$$

showing $Z_1 = Y_2$ and $Z_2 = Y_2 - Y_4$ are stochastically independent.

1.9 Joint Distribution of Order Statistics

Let $X_1,\ X_2, \ldots,\ X_n$ be a random sample of size n taken from a continuous population whose p. d. f. is $f\left(x\right)$ and c .d. f. is $F\left(x\right)$ for $a < x < b$. Let $F(y_1, \ldots, y_n)$ be the joint c. d. f. and $f(y_1, \ldots, y_n)$ be the joint p. d. f. of all the order statistics $Y_1, \ldots, Y_n$ at the point $\left(y_1, \ldots, y_n\right)$ is given by

$$f\left(y_1, \ldots, y_n\right) = \frac{\partial^n F\left(y_1, \ldots, y_n\right)}{\partial y_1 \ldots \partial y_n}$$

$$= \lim_{\substack{\Delta y_1 \to 0 \\ \ldots \\ \Delta y_n \to 0}} \frac{\mathrm{P}\left(y_1 < Y_1 \le y_1 + \Delta y_1, \ldots, y_n < Y_n \le y_n + \Delta y_n\right)}{\Delta y_1 \ldots \Delta y_n}$$

$$= \lim_{\substack{\Delta y_1 \to 0 \\ \ldots \\ \Delta y_n \to 0}} \frac{\mathrm{P}\left\{\text{one } X \text{ in } (y_1, y_1 + \Delta y_1], \ldots, \text{one } X \text{ in } (y_n, y_n + \Delta y_n]\right\}}{\Delta y_1 \ldots \Delta y_n}$$

$$\overset{\downarrow}{a} \underbrace{\qquad\qquad} \overset{\downarrow}{y_1} \underbrace{\qquad\quad}_{\text{one } X} \overset{\downarrow}{y_1 + \Delta y_1} \underbrace{\qquad\qquad}_{\ldots} \overset{\downarrow}{y_n} \underbrace{\qquad\quad}_{\text{one } X} \overset{\downarrow}{y_n + \Delta y_n} \underbrace{\qquad\qquad} \overset{\downarrow}{b}$$

Using multinomial law, we have

$$= \lim_{\substack{\Delta y_1 \to 0 \\ \ldots \\ \Delta y_n \to 0}} \frac{n!}{1! \ldots 1!} \frac{\mathrm{P}\left(y_1 < Y_1 \le y_1 + \Delta y_1\right)}{\Delta y_1} \ldots \frac{\mathrm{P}\left(y_n < Y_n \le y_n + \Delta y_n\right)}{\Delta y_n}$$

$$= \frac{n!}{1! \ldots 1!} \lim_{\Delta y_1 \to 0} \left\{\frac{F\left(y_1 + \Delta y_1\right) - F\left(y_1\right)}{\Delta y_1}\right\} \ldots \lim_{\Delta y_n \to 0} \left\{\frac{F\left(y_n + \Delta y_n\right) - F\left(y_n\right)}{\Delta y_n}\right\}$$

$$= \begin{cases} n!\, f(y_1) \ldots f(y_n) & ;\ a < y_1 < \ldots < y_n < b \\ 0 & ;\text{otherwise} \end{cases}$$

1.10 Summary

This unit provides a thorough understanding of concepts related to Order statistics. The concepts of Distribution of maximum, minimum and r-th order statistic, Joint distribution of r-th and s-th order statistic are described in detail. The learner should try to solve the self-assessment problems given in the next section.

1.11 Self-Assessment Exercises

Q1. Derive the distribution of maximum and minimum order statistics for a random sample of size n drawn from a population having pdf $f(x|\theta)$ and cdf $F(x|\theta)$.

Q2. What do you understand by order statistics. Explain the use of order statistics by giving suitable examples. Also, derive the distribution of r-th order statistic.

Q3. Derive the joint distribution of r-th and s-th order statistic.

UNIT:2 ASYMPTOTIC DISTRIBUTION THEORY

Structure

2.1 Introduction

The exact probability density function of the r^{th} order statistic of a sample of size n from a continuous population with *cdf* $P(x)$ was given as

$$f_r(x) = \frac{1}{\beta(r, n-r+1)} P(x)^{r-1}\ p(x)\left\{1-P(x)\right\}^{n-r}.$$

But when n is large, this form is rather cumbersome to deal with. Hence, the asymptotic form of the density has to be derived for dealing large sample cases. While dealing with asymptotic distribution for any n, two distinct situations arise. They are

(i) As n approaches infinity, r/n remains fixed.

(ii) As n approaches infinity, r or $(n - r)$ remains fixed.

The first situation arises when dealing with distribution of quantiles, while the second situation will arise when dealing with distribution of extreme values. Here we intend to consider only on the first situation. The distribution of the r^{th} order statistic, when n approaches infinity with r/n remaining fixed can be obtained as follows.

2.2 Objectives

The objective of this unit is to provide a basic understanding of concepts related to Asymptotic Distribution Theory. The concepts of the moments of order statistics, non parametric estimation of distribution function, Glivenko-Cantelli fundamental theorem should be clear after study of this material.

2.3 Moments of Order Statistics

Let $X_1, X_2, \ldots\ldots, X_n$ be a random sample of size n from an absolutely continuously population with pdf $f(x)$ and cdf $F(x)$, and let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the corresponding order statistics. From the pdf of $X_{i:n}$ we have, for $1 \leq i \leq n$ and $m = 1,2, \ldots.,$

$$\mu_{i:n}^{(m)} = E(X_{i:n}^{m}) = \int_{-\infty}^{\infty} x^m\, f_{i:n}(x)\, dx$$

$$= \frac{n!}{(i-1)!\,(n-i)!} \int_{-\infty}^{\infty} x^m \{F(x)\}^{i-1}\{1-F(x)\}^{n-i} f(x) dx. \tag{1}$$

we will denote $\mu_{i:n}^{(1)}$ by $\mu_{i:n}$ for convenience. From the first moments, we can determine the variance of $X_{i:n}$ by

$$\sigma_{i,i:n} = \sigma_{i:n}^{(2)} = var(X_{i:n}) = \mu_{i:n}^{(2)} - \mu_{i:n}^{2} \quad 1 \leq i \leq n \tag{2}$$

Similarly, from the joint density function of $X_{i:n}$ and $X_{j:n}$ we have, for $1 \leq i \leq j \leq n$ and m_i, $m_j = 1,2, \ldots,$

$$\mu_{i,j:n}^{(m_i,m_j)} = E(X_{i;n}^{m_i}\, X_{j;n}^{m_j})$$

$$= \iint_{-\infty<x_i<x_j<\infty} x_i^{m_i} x_j^{m_j} f_{i,j:n}\left(x_i, x_j\right) dx_i dx_j$$

$$= \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!} \times \iint_{-\infty<x_i<x_j<\infty} x_i^{m_i} x_j^{m_j} \{F(x_i)\}^{i-1} \left\{F\left(x_j\right) - F(x_i)\right\}^{j-i-1} \times$$

$$\left\{1 - F\left(x_j\right)\right\}^{n-j} f(x_i) f\left(x_j\right) dx_i dx_j \tag{3}$$

Once again, we use $\mu_{i,j:n}$ instead of $\mu_{i,j:n}^{(1,1)}$. The covariance of $X_{i:n}$ and $X_{j:n}$ may then be determined by

$$\sigma_{i,j:n} = cov(X_{i:n}, X_{j:n}) = \mu_{i,j:n} - \mu_{i:n}\mu_{j:n}\ , \qquad 1 \le i \le j \le n. \quad (4)$$

The formulas in (1) and (3) will enable one to derive exact explicit expression for the single and the product moments of order statistics, respectively, in many cases. Also, in situations where it is not possible to derive such explicitly expressions for the moments, the formulas in (1) and (3) can be used to compute the necessary moments by employing some numerical methods of integration.

The expressions for the single and the product moments of order statistics in Eqs (1) and (3) can be easily modified to the case when the population distribution is discrete and written as

$$\mu_{i:n}^{(m)} = E(X_{i:n}^{m}) = \textstyle\sum_{L_1 \le x \le L_2} x^m f_{i:n}(x), \qquad 1 \le i \le n, \quad m = 1,2,\ldots\ldots, \quad (5)$$

and

$$\mu_{i,j:n}^{(m_i,m_j)} = E\left(X_{i:n}^{m_i}\, X_{j:n}^{m_j}\right) = \sum_{L_1 \le x_i \le x_j \le L_2}\sum x_i^{m_i}\, x_j^{m_j} f_{i,j:n}(x_i, x_j),$$

$$1 \le i < j \le n,\ \ m_i, m_j = 1,2,\ldots, \qquad (6)$$

where $[L_1, L_2]$ is the support of the discrete population distribution; in Eqs (5) and (6), $f_{i:n}(x)$ and $f_{i,j:n}(x_i, x_j)$ are the pmf of $X_{i:n}$ and the joint pmf of $X_{i:n}$ and $X_{j:n}$, respectively.

Now, by defining the inverse cumulative distribution function of the population as

$$F^{-1}(u) = sup\{x: F(x) \le u\}, \qquad 0 < u < 1, \qquad (7)$$

and define the single the product moments of order statistics as

$$\mu_{i:n}^{(m)} = E(X_{i:n}^{m}) = \frac{n!}{(i-1)!\ (n-i)!}\int_0^1 \{F^{-1}(u)\}^m u^{i-1}(1-u)^{n-i} du,$$

$$1 \le i \le n,\ \ m = 1,2,\ldots., \qquad (8)$$

and

$$\mu_{i,j:n}^{(m_i,m_j)} = E(X_{i;n}^{m_i}\, X_{j;n}^{m_j})$$

$$= \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!} \times \iint_{0<u_i<u_j<1} \{F^{-1}(u_i)\}^{m_i} \{F^{-1}(u_i)\}^{m_j} u_i^{i-1} \left(u_j - u_i\right)^{i-j-1} \times$$

$$\left(1 - u_j\right)^{n-j} du_i\, du_j \quad 1 \le i \le j \le n, \quad m_i, m_j = 1,2, \dots. \tag{9}$$

Example: Let $X_{(r)}$ denote the r^{th} order statistic in a random sample of size n from a uniform distribution in the interval (0,1). Compute $\mu_{r,n}$.

Solution: Here $p(x)$ =1, $0 \le x \le 1$ and $P(x) = x$.

$\therefore \quad f_{r,n}(x) = n\binom{n-1}{r-1} x^{r-1}(1-x)^{n-r} = \frac{n!}{(r-1)!(n-r)!}\, x^{r-1}(1-x)^{n-r}$

$\mu_{r,n} = E\{X_{(r)}\} = \int_0^1 x\, f_r(x)dx = \frac{n!}{(r-1)!(n-r)!} \int_0^1 x\,.\,x^{r-1}(1-x)^{n-r}dx$

$= \frac{n!}{(r-1)!(n-r)!} \frac{r!(n-r)!}{(n+1)!} = \frac{r}{n+1}; \; r = 1,2,\cdots,n.$

This shows that the n order statistics $X_1, X_2, \cdots, X_n$ divides the area under the *pdf* into n+1 parts each of which is on the average $\frac{1}{n+1}$.

Example: Let $X_{(r)}, X_{(s)}, X_{(t)}$ and $X_{(u)}$ $(r \le s \le t \le u)$ be four order statistics from a sample of size n from a uniform distribution with *pdf* $p(x) = 1$, $0 \le x \le 1$. Suppose $Y_1 = \frac{X_{(r)}}{X_{(s)}}$, $Y_2 = \frac{X_{(s)}}{x_{(t)}}$, $Y_3 = \frac{X_{(t)}}{X_{(u)}}$ and $Y_4 = X_{(u)}$, then prove that Y_1, Y_2, Y_3 and Y_4 are independently distributed each having a beta distribution.

Solution*:* Consider the joint distribution of the four order statistics $X_{(r)}, X_{(s)}, X_{(t)}$ and $X_{(u)}$ $(r \le s \le t \le u)$ of a sample of size n from a uniform distribution *pdf* $p(x) = 1$, $0 \le x \le 1$ and *cdf* $P(x) = x$, given by

$f_{r,s,t,u;n}(x_1, x_2, x_3, x_4) = \frac{n!}{(r-1)!(s-r-1)!(t-s-1)!(u-t-1)!(n-u)!} P^{r-1}(x_1)p(x_1)[P(x_2) - P(x_1)]^{s-r-1} \times p(x_2)[P(x_3) - P(x_2)]^{t-s-1}p(x_3)[P(x_4) - P(x_3)]^{u-t-1}p(x_4)[1 - P(x_4)]^{n-u}.$

Let $C = \frac{n!}{(r-1)!(s-r-1)!(t-s-1)!(u-t-1)!(n-u)!}$. Then,

$f_{r,s,t,u;n}(x_1, x_2, x_3, x_4) = C.x_1^{r-1}(x_2 - x_1)^{s-r-1}(x_3 - x_2)^{t-s-1}(x_4 - x_3)^{u-t-1}(1 - x_4)^{n-u}$.

On making the transformation $y_1 = \frac{x_1}{x_2}$, $y_2 = \frac{x_2}{x_3}$, $y_3 = \frac{x_3}{x_4}$ and $y_4 = x_4$, the inverse transformation is $x_4 = y_4$, $x_3 = y_3 y_4$, $x_2 = y_2 y_3 y_4$ and $x_1 = y_1 y_2 y_3 y_4$. The jacobian of the transformation is given by

$$|J| = \left|\frac{\partial(x_1,x_2,x_3,x_4)}{\partial(y_1,y_2,y_3,y_4)}\right| = \begin{vmatrix} y_2y_3y_4 & y_1y_3y_4 & y_1y_2y_4 & y_1y_2y_3 \\ 0 & y_3y_4 & y_2y_4 & y_2y_3 \\ 0 & 0 & y_4 & y_3 \\ 0 & 0 & 0 & 0 \end{vmatrix} = y_2 y_3^2 y_4^3;\ 0 \le y_i \le 1, \text{ for}$$

i = 1,2,3,4. Hence the joint density of Y_1, Y_2, Y_3 and Y_4 is given by

$$g(y_1, y_2, y_3, y_4) = f_{r,s,t,u}(x_1, x_2, x_3, x_4)\ |J|$$

$$= C.(y_1y_2y_3y_4)^{r-1}(y_2y_3y_4 - y_1y_2y_3y_4)^{s-r-1}(y_3y_4 - y_2y_3y_4)^{t-s-1}$$

$$\times (y_4 - y_3y_4)^{u-t-1}(1 - y_4)^{n-u} y_2 y_3^2 y_4^3$$

$$= C.y_1^{r-1}(1 - y_1)^{s-r-1} y_2^{r-1+s-r-1+1}(1 - y_2)^{t-s-1} y_3^{r-1+s-r-1+t-s-1+2}(1 - y_3)^{u-t-1}$$

$$\times y_4^{r-1+s-r-1+t-s-1+u-t-1+3}(1 - y_4)^{n-u}$$

$$= C.y_1^{r-1}(1 - y_1)^{s-r-1} y_2^{s-1}(1 - y_2)^{t-s-1} y_3^{t-1}(1 - y_3)^{u-t-1} y_4^{u-1}(1 - y_4)^{n-u}.$$

Now, $C = \frac{n!}{(r-1)!(s-r-1)!(t-s-1)!(u-t-1)!(n-u)!}$

$$= \frac{(s-1)!}{(r-1)!(s-r-1)!}\frac{(t-1)!}{(s-1)!(t-s-1)!}\frac{(u-1)!}{(t-1)!(u-t-1)!}\frac{n!}{(u-1)!(n-u)!}$$

$$= \frac{1}{\beta(r,s-r)}\frac{1}{\beta(s,t-s)}\frac{1}{\beta(t,u-t)}\frac{1}{\beta(u,n-u+1)}.$$

$$\therefore\ g(y_1, y_2, y_3, y_4) = \frac{1}{\beta(r,s-r)} y_1^{r-1}(1 - y_1)^{s-r-1} \frac{1}{\beta(s,t-s)} y_2^{s-1}(1 - y_2)^{t-s-1}$$

$$\times \frac{1}{\beta(t,u-t)} y_3^{t-1}(1 - y_3)^{u-t-1} \frac{1}{\beta(u,n-u+1)} y_4^{u-1}(1 - y_4)^{n-u};$$

$$0 \le y_i \le 1, \text{ for } i = 1,2,3,4.$$

This proves that the variables Y_1, Y_2, Y_3 and Y_4 are independently distributed as beta distribution.

Using this result, the moment $E\left[X_{(r)}^{a} X_{(s)}^{b} X_{(t)}^{c} X_{(u)}^{d}\right]$ can be obtained as follows.

$$E\left[X_{(r)}^{a} X_{(s)}^{b} X_{(t)}^{c} X_{(u)}^{d}\right] = \int_0^1 \int_0^{x_4} \int_0^{x_3} \int_0^{x_2} x_1^a x_2^b x_3^c x_4^d \, f_{r,s,t,u}(x_1, x_2, x_3, x_4) dx_1 dx_2 dx_3 dx_4.$$

On making the transformation $y_1 = \frac{x_1}{x_2}$, $y_2 = \frac{x_2}{x_3}$, $y_3 = \frac{x_3}{x_4}$ and $y_4 = x_4$, the inverse transformation is $x_4 = y_4$, $x_3 = y_3 y_4$, $x_2 = y_2 y_3 y_4$ and $x_1 = y_1 y_2 y_3 y_4$. The jacobian of the transformation is given by

$$|J| = \left|\frac{\partial(x_1,x_2,x_3,x_4)}{\partial(y_1,y_2,y_3,y_4)}\right| = \begin{vmatrix} y_2y_3y_4 & y_1y_3y_4 & y_1y_2y_4 & y_1y_2y_3 \\ 0 & y_3y_4 & y_2y_4 & y_2y_3 \\ 0 & 0 & y_4 & y_3 \\ 0 & 0 & 0 & 0 \end{vmatrix} = y_2 y_3^2 y_4^3;\ 0 \le y_i \le 1, \text{ for}$$

i = 1,2,3,4. Thus,

$$E\left[X_{(r)}^{a} X_{(s)}^{b} X_{(t)}^{c} X_{(u)}^{d}\right]$$

$$= \int_0^1 \int_0^1 \int_0^1 \int_0^1 (y_1y_2y_3y_4)^a \, (y_2y_3y_4)^b (y_3y_4)^c (y_4)^d \, g(y_1, y_2, y_3, y_4) dy_1 dy_2 dy_3 dy_4$$

$$= \int_0^1 \frac{1}{\beta(r,s-r)} y_1^{a+r-1}(1-y_1)^{s-r-1}\, dy_1 \int_0^1 \frac{1}{\beta(s,t-s)} y_2^{s-1}(1-y_2)^{a+b+t-s-1} dy_2 \times$$

$$\int_0^1 \frac{1}{\beta(t,u-t)} y_3^{a+b+c+t-1}(1-y_3)^{u-t-1}\, dy_3 \int_0^1 \frac{1}{\beta(u,n-u+1)} y_4^{a+b+c+d+u-1}(1-y_4)^{n-u}\, dy_4$$

$$= \frac{\beta(a+r,s-r)}{\beta(r,s-r)} \frac{\beta(a+b+s,t-s)}{\beta(s,t-s)} \frac{\beta(a+b+c+t,u-t)}{\beta(t,u-t)} \frac{\beta(a+b+c+d+u,n-u+1)}{\beta(u,n-u+1)}.$$

This result can be generalized for *k* variables. Suppose $X_{(r1)} \le X_{(r2)} \le \cdots \le X_{(rk)}$ be *k* order statistics of a sample of size *n* from a standard uniform distribution, then

$$E\left[X_{r1}^{a1} X_{r2}^{a2} \cdots X_{rk}^{ak}\right] = E\left[\prod_{i=1}^{k} X_{(ri)}^{ai}\right]$$

$$= \frac{\beta(a1+r1,r2-r1)}{\beta(r1,r2-r1)} \frac{\beta(a1+a2+r2,r3-r2)}{\beta(r2,r3-r2)} \cdots \frac{\beta(a1+a2+\cdots+ak,n-rk+1)}{\beta(rk,n-rk+1)}$$

$$= \frac{n!}{(n+\sum_{i=1}^{k} ai)!} \prod_{i=1}^{k} \frac{(ri-1+\sum_{j=1}^{i} aj)!}{(ri-1+\sum_{j=1}^{i-1} aj)!}.$$

This is valid only for uniform distribution. In particular for $1 \le r \le s \le n$, and taking $p_r = \frac{r}{n+1}$ and $q_r = 1 - p_r = 1 - \frac{r}{n+1} = \frac{n-r+1}{n+1}$, we have

$\mu_{r,n} = p_r = \frac{r}{n+1}$ and

$\sigma_{r,s;n}$ = Cov$(X_{(r)}, X_{(s)}) = E[(X_{(r)} - \mu_{r,n})(X_{(s)} - \mu_{s,n})]$

$$= E[X_{(r)}X_{(s)}] - \mu_{r,n}\mu_{s,n} = \frac{n!}{(n+2)!}\prod_{i=1}^{2}\frac{(ri-1+i)!}{(ri-1+i-1)!} - \frac{r}{n+1}.\frac{s}{n+1}$$

$$= \frac{n!}{(n+2)!}\frac{r!}{(r-1)!}\frac{(s+1)!}{s!} - \frac{r}{n+1}.\frac{s}{n+1} = \frac{1}{(n+2)(n+1)}\frac{r}{1}\frac{(s+1)}{1} - \frac{r}{n+1}.\frac{s}{n+1}$$

$$= \frac{r}{n+1}\left[\frac{(s+1)}{(n+2)} - \frac{s}{n+1}\right] = \frac{r}{n+1}\frac{(n+1-s)}{(n+1)(n+2)} = \frac{r}{(n+1)(n+2)}\left(1 - \frac{s}{n+1}\right) = \frac{p_r p_s}{n+2}.$$

When $r = s$, we get the variance of $X_{(r)}$ as

$$\sigma_{r,r;n} = \sigma^2_{r,n} = \frac{p_r^2}{n+2} = \frac{r^2}{(n+1)^2(n+2)}.$$

Note that the computation of the moments of order statistics could be simplified if it is known that the *pdf* $p(x)$ is symmetric about origin, as in this case, $p(-x) = p(x)$ and $P(-x) = 1 - P(x)$;

$$f_{r,n}(-x) = \frac{1}{\beta(r,n-r+1)}P^{r-1}(-x)p(-x)[1 - P(-x)]^{n-r}$$

$$= \frac{1}{\beta(r,n-r+1)}[1 - P(x)]^{r-1}p(x)\ [1 - 1 + P(x)]^{n-r}$$

$$= \frac{1}{\beta(n-r+1,r)}P^{n-r}(x)p(x)[1 - P(x)]^{r-1} = f_{n-r+1,n}(x).$$

Hence, $\mu_{r,n} = E(X_{(r)}) = \int_{-\infty}^{\infty} x f_r(x)dx = \int_{-\infty}^{\infty}(-x)f_{n-r+1}(x)dx = -\mu_{n-r+1,n}$.

Similarly, $\sigma_{r,s;n}$ = Cov$(X_{(r)}, X_{(s)}) = E[(X_{(r)} - \mu_{r,n})(X_{(s)} - \mu_{s,n})]$

$$= E[X_{(r)}X_{(s)}] - \mu_{r,n}\mu_{s,n} = \int_{-\infty}^{\infty}\int_{-y}^{\infty}(-x)\,(-y)f_{n-r+1,n-s+1}(x,y)dxdy - \mu_{n-r+1,n}\mu_{n-s+1,n}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{y} x\,y f_{n-r+1,n-s+1}(x,y)dxdy - \mu_{n-r+1,n}\mu_{n-s+1,n}$$

$$= \sigma_{n-r+1,n-s+1;n}.$$

Using this result, the mean and variance of the range $R = X_{(n)} - X_{(1)}$ of a sample of size n from a symmetric distribution is given by

$E(R) = E[X_{(n)}] - E[X_{(1)}] = \mu_{n,n} - \mu_{1,n} = \mu_{n,n} - (-\mu_{n,n}) = 2\mu_{n,n}$ and

$$V(R) = V[X_{(n)} - X_{(1)}] = V[X_{(n)}] - 2Cov[X_{(n)}, X_{(1)}] + V[X_{(1)}]$$

$= \sigma_{n,n}^2 - 2\,\sigma_{1,n;n} + \sigma_{1,n}^2 = \sigma_{n,n}^2 - 2\,\sigma_{1,n;n} + \sigma_{n,n}^2 = 2\left(\sigma_{n,n}^2 - \sigma_{1,n;n}\right).$

Theorem: If $F(x)$ is the *cdf* of a random variable and if $E(X)$ exists, then the following limits hold: (*i*) $\lim_{x \to -\infty} xF(x) = 0$ and (*ii*) $\lim_{x \to \infty} x[1 - F(x)] = 0$.

Proof: If $E(X)$ exists, then the integral $\int_{-\infty}^{\infty} |x| dF(x) < \infty$. Hence, if we consider the integral $\int_{-\infty}^{x} y dF(y) \leq \int_{-\infty}^{x} x dF(y)$, since $y \leq x$;

$= x \int_{-\infty}^{x} dF(y) = x[F(y)]|_{-\infty}^{x} = xF(x)$ as $F(-\infty) = 0$.

Or $\int_{-\infty}^{x} y dF(y) \leq xF(x)$. Now,

$0 = \lim_{x \to -\infty} \int_{-\infty}^{x} y dF(y) \leq \lim_{x \to -\infty} xF(x) \leq 0$. Hence,

$0 \leq \lim_{x \to -\infty} xF(x) \leq 0 \Rightarrow \lim_{x \to -\infty} xF(x) = 0$. This proves (*i*).

Now, to prove (*ii*) consider the integral

$\int_{x}^{\infty} y dF(y) \geq \int_{x}^{\infty} x dF(y)$, since $y \geq x$;

$= x \int_{x}^{\infty} dF(y) = x[F(y)]|_{x}^{\infty} = x[F(\infty) - F(x)]$

$= x[1 - F(x)]$ as $F(\infty) = 1$.

Or $\int_{x}^{\infty} y dF(y) \geq x[1 - F(x)]$. Now,

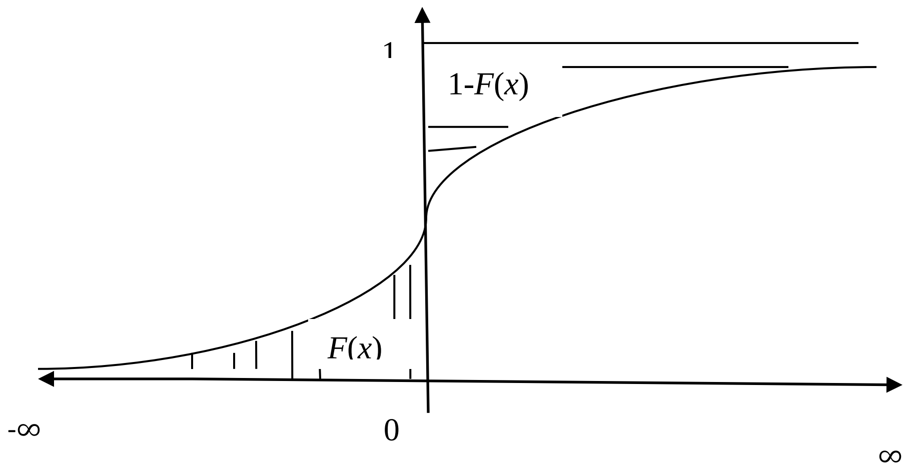
$0 = \lim_{x \to \infty} \int_{x}^{\infty} y dF(y) \geq \lim_{x \to \infty} x[1 - F(x)] \geq 0$. Hence,

$0 \geq \lim_{x \to \infty} x[1 - F(x)] \geq 0 \Rightarrow \lim_{x \to \infty} x[1 - F(x)] = 0$.

Theorem: If $E(X)$ exists, then it can be expressed as

$\int_{-\infty}^{\infty} x dF(x) = E(X) = \int_{0}^{\infty} [1 - F(y)]\, dy - \int_{-\infty}^{0} F(y) dy.$

Proof*:*



Consider the integral $\int_0^x ydF(y)$ and integrating it by parts, we get

$$\int_0^x ydF(y) = y.F(y)|_0^x - \int_0^x F(y)dy = xF(x) - \int_0^x F(y)dy$$

$$= -x[1-F(x)] + x - \int_0^x F(y)dy = -x[1-F(x)] + \int_0^x dy - \int_0^x F(y)dy.$$

Or $\int_0^x ydF(y) = -x[1-F(x)] + \int_0^x [1-F(y)]dy.$ (1)

Also by considering the integral $\int_{-x}^0 ydF(y) = y.F(y)|_{-x}^0 - \int_{-x}^0 F(y)dy$

$$= xF(-x) - \int_{-x}^0 F(y)dy.$$

Hence,

$$\int_0^x ydF(y) + \int_{-x}^0 ydF(y) = -x[1-F(x)] + xF(-x) + \int_0^x [1-F(y)]dy - \int_{-x}^0 F(y)dy.$$

Or $\int_{-x}^x ydF(y) = -x[1-F(x)] + xF(-x) + \int_0^x [1-F(y)]dy - \int_{-x}^0 F(y)dy.$

Now letting $x \to \infty$, we have

$$\int_{-\infty}^{\infty} ydF(y) = -\lim_{x\to\infty} x[1-F(x)] - \lim_{x\to\infty}[-xF(-x)] + \int_0^{\infty}[1-F(y)]dy - \int_{-\infty}^0 F(y)dy.$$

Or $E(X) = \int_0^{\infty}[1-F(y)]dy - \int_{-\infty}^0 F(y)dy$, since $\lim_{x\to\infty} x[1-F(x)] = 0$ and

$\lim_{x\to\infty}[-xF(-x)] = \lim_{x\to-\infty}[xF(x)] = 0$, by Theorem1.

Hence the theorem is proved.

Now consider the expected value of $X_{(r)}$ from a sample of size n given by

$$\mu_{r,n} = \int_0^\infty [1 - F_{r,n}(x)]dx - \int_{-\infty}^0 F_{r,n}(x)dx = \int_0^\infty [1 - F_{r,n}(x)]dx - \int_0^\infty F_{r,n}(-x)dx$$

$$= \int_0^\infty [1 - F_{r,n}(x) - F_{r,n}(-x)]dx.$$

When $p(x)$ is symmetrical about $x = 0$, then $f_{r,n}(-x) = f_{n-r+1,n}(x)$ and hence, $F_{r,n}(-x) = 1 - F_{n-r+1,n}(x)$. Therefore,

$$\mu_{r,n} = \int_0^\infty [1 - F_{r,n}(x) - 1 + F_{n-r+1,n}(x)]dx = \int_0^\infty [F_{n-r+1,n}(x) - F_{r,n}(x)]dx.$$

Further, when $r = 1$, $\mu_{1,n} = \int_0^\infty [F_{n,n}(x) - F_{1,n}(x)]dx$ and when $r = n$,

$$\mu_{n,n} = \int_0^\infty [F_{1,n}(x) - F_{n,n}(x)]dx.$$

$$\therefore \quad E(R) = \mu_{n,n} - \mu_{1,n} = \int_0^\infty [F_{1,n}(x) - F_{n,n}(x)]dx - \int_0^\infty [F_{n,n}(x) - F_{1,n}(x)]dx$$

$$= \int_0^\infty [F_{1,n}(x) - F_{n,n}(x) - F_{n,n}(x) + F_{1,n}(x)]dx = 2\int_0^\infty [F_{1,n}(x) - F_{n,n}(x)]dx.$$

2.4 Some Basic Relations

The following checks can be applied while computing the moments of order statistics, by noting that

$$\left[\sum_{r=1}^n X_{(r)}^k\right]^m = \left[X_{(1)}^k + X_{(2)}^k + \cdots + X_{(n)}^k\right]^m = [X_1^k + X_2^k + \cdots + X_n^k]^m = \left[\sum_{i=1}^n X_i^k\right]^m . \qquad (1)$$

Now for different pairs of values of k and m, *i.e.* (k, m) = (1,1), (2,1) and (1,2) this expression reduces to:

(i) For the pair(1,1),

$\sum_{r=1}^n X_{(r)} = \sum_{i=1}^n X_i$. Hence,

$E\left[\sum_{r=1}^n X_{(r)}\right] = E[\sum_{i=1}^n X_i]$.

Or $\sum_{r=1}^n E\left(X_{(r)}\right) = \sum_{i=1}^n E(X_i)$.

Or $\sum_{r=1}^n \mu_{r,n} = \sum_{i=1}^n \mu = n\mu$, (2) where $E(X) = \mu$.

(ii) For the pair (2,1),

$\left[\sum_{r=1}^{n} X_{(r)}^{2}\right] = \left[\sum_{i=1}^{n} X_{i}^{2}\right]$. Hence,

$$E\left[\sum_{r=1}^{n} X_{(r)}^{2}\right] = E\left[\sum_{i=1}^{n} X_{i}^{2}\right] = \sum_{i=1}^{n} E(X_{i}^{2}) = nE(X^{2}). \tag{3}$$

Or $E\left[\sum_{r=1}^{n} X_{(r)}^{2}\right] = n(\sigma^{2} + \mu^{2})$. (4)

(iii) For the pair (1,2)

$\left[\sum_{r=1}^{n} X_{(r)}\right]^{2} = \left[\sum_{i=1}^{n} X_{i}\right]^{2}$.

Or $\left[\sum_{r=1}^{n} X_{(r)}\right]\left[\sum_{s=1}^{n} X_{(s)}\right] = \left[\sum_{i=1}^{n} X_{i}\right]^{2}$

Or $\sum_{r=1}^{n} X_{(r)}^{2} + 2\sum_{r=1}^{n}\sum_{s=r+1}^{n} X_{(r)} X_{(s)} = \sum_{i=1}^{n} X_{i}^{2} + 2\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i<j}}^{n} X_{i}X_{j}$.

On taking expectation and using (4), we get

$$E\left[\sum_{r=1}^{n} X_{(r)}^{2}\right] + 2\sum_{r=1}^{n}\sum_{s=r+1}^{n} E\left[X_{(r)}X_{(s)}\right] = \sum_{i=1}^{n} E[X_{i}^{2}] + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} E[X_{i}X_{j}].$$

Or $n(\sigma^{2} + \mu^{2}) + 2\sum_{r=1}^{n}\sum_{s=r+1}^{n} E\left[X_{(r)}X_{(s)}\right] = n(\sigma^{2} + \mu^{2}) + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n} E[X_{i}]E[X_{j}]$.

Or $\sum_{r=1}^{n}\sum_{s=r+1}^{n} E\left[X_{(r)}X_{(s)}\right] = \frac{1}{2}n(n-1)\mu^{2}$. (5)

Also, since

$\sum_{r=1}^{n}\left[X_{(r)} - \mu_{r,n}\right] = \sum_{r=1}^{n} X_{(r)} - \sum_{r=1}^{n} \mu_{r,n} = \sum_{i=1}^{n} X_{i} - n\mu$.

Or $\sum_{r=1}^{n}\left[X_{(r)} - \mu_{r,n}\right] = \sum_{i=1}^{n}(X_{i} - \mu)$. (6)

Squaring both sides of (6), we get

$$\sum_{r=1}^{n}\sum_{s=1}^{n}\left(X_{(r)} - \mu_{r,n}\right)\left(X_{(s)} - \mu_{s,n}\right) = \sum_{i=1}^{n}(X_{i} - \mu)^{2} + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n}(X_{i} - \mu)\left(X_{j} - \mu\right).$$

On taking expectation on both sides of the above equation, we get

$\sum_{r=1}^{n}\sum_{s=1}^{n} E\left\{\left(X_{(r)} - \mu_{r,n}\right)\left(X_{(s)} - \mu_{s,n}\right)\right\}$

$= \sum_{i=1}^{n} E\{(X_i-\mu)^2\} + \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} E(X_i-\mu)E\left(X_j-\mu\right).$

Or $\sum_{r=1}^{n}\sum_{s=1}^{n}\sigma_{r,s;n} = \sum_{i=1}^{n}\sigma^2 = n\sigma^2$. [Since, $E(X_i-\mu) = E\left(X_j-\mu\right) = 0$.] (7)

Relations (1) to (7) hold for both continuous as well as discrete variates.

2.5 Recurrence Relations

Recurrence relations between the moments of order statistics are studied for reducing the number of calculations required for the evaluation of the moments. Such relations may also be used as checks on direct calculations. To prove the recurrence relations we need the following lemma.

Lemma: Suppose $I_x(m,n) = \frac{1}{\beta(m,n)}\int_0^x t^{m-1}(1-t)^{n-1}dt$ denotes the incomplete beta function up to the value *x* with parameters *m* and *n* then

$aI_y(a+1,b) + bI_y(a,b+1) = (a+b)I_y(a,b).$

Proof: We have,

$aI_y(a+1,b) + bI_y(a,b+1) = \frac{a}{\beta(a+1,b)}\int_0^y t^a(1-t)^{b-1}dt + \frac{b}{\beta(a,b+1)}\int_0^y t^{a-1}(1-t)^b dt.$

Now, $\frac{a}{\beta(a+1,b)} = \frac{a\Gamma(a+b+1)}{\Gamma(a+1)\Gamma(b)} = \frac{a(a+b)\Gamma(a+b)}{a\Gamma(a)\Gamma(b)} = \frac{a+b}{\beta(a,b)}$. Similarly,

$\frac{b}{\beta(a,b+1)} = \frac{b\Gamma(a+b+1)}{\Gamma(a)\Gamma(b+1)} = \frac{b(a+b)\Gamma(a+b)}{b\Gamma(a)\Gamma(b)} = \frac{a+b}{\beta(a,b)}.$

Hence, the L.H.S. of the above expression is given by

$$\frac{a+b}{\beta(a,b)}\int_0^y t^a(1-t)^{b-1}dt + \frac{a+b}{\beta(a,b)}\int_0^y t^{a-1}(1-t)^b dt$$

$$= \frac{a+b}{\beta(a,b)}\int_0^y t^{a-1}(1-t)^{b-1}.t dt + \frac{a+b}{\beta(a,b)}\int_0^y t^{a-1}(1-t)^{b-1}.(1-t)dt$$

$= \frac{a+b}{\beta(a,b)}\int_0^y t^{a-1}(1-t)^{b-1}.(t+1-t)dt = \frac{a+b}{\beta(a,b)}\int_0^y t^{a-1}(1-t)^{b-1}dt = (a+b)I_y(a,b).$

Hence the lemma is proved.

Relation 1. For an arbitrary distribution with finite k^{th} moment

$$(n-r)\mu_{r;n}^{k} + r\mu_{r+1;n}^{k} = n\mu_{r;n-1}^{k},$$

where $r = 1, 2, \cdots, n-1$ and $k = 1, 2, \cdots$.

Proof*:* By definition,

$$\mu_{r;n}^{k} = E\left[X_{(r)}^{k}\right] = \int_{-\infty}^{\infty} x^{k} f_{r}(x)dx = \int_{-\infty}^{\infty} x^{k}\left[\frac{d}{dx} I_{P(x)}(r, n-r+1)\right]dx,$$

where $I_{P(x)}(r, n-r+1) = \frac{1}{\beta(r,n-r+1)}\int_{0}^{P(x)} t^{r-1}(1-t)^{n-r}dt.$

We shall now make use of the above lemma with $a = r$, $b = n - r$ and $y = P(x)$. Then, we have

$$rI_{P(x)}(r+1, n-r) + (n-r)I_{P(x)}(r, n-r+1) = nI_{P(x)}(r, n-r). \tag{1}$$

Differentiating both sides

$$r\frac{d}{dx}I_{P(x)}(r+1, n-r) + (n-r)\frac{d}{dx}I_{P(x)}(r, n-r+1) = n\frac{d}{dx}I_{P(x)}(r, n-r).$$

Or $\quad r\frac{1}{\beta(r+1,n-r)}x^{r}(1-x)^{n-r-1} + (n-r)\frac{1}{\beta(r,n-r+1)}x^{r-1}(1-x)^{n-r}$

$$= n\frac{1}{\beta(r,n-r)}x^{r-1}(1-x)^{n-r-1}.$$

Or $rf_{r+1,n}(x) + (n-r)f_{r,n}(x) = nf_{r,n-1}(x)$

On multiplying both sides by x^{k} and integrating out, we get

$$r\int_{-\infty}^{\infty} x^{k} f_{r+1,n}(x)dx + (n-r)\int_{-\infty}^{\infty} x^{k} f_{r,n}(x)dx = n\int_{-\infty}^{\infty} x^{k} f_{r,n-1}(x)dx.$$

Or $r\mu_{r+1;n}^{k} + (n-r)\mu_{r;n}^{k} = n\mu_{r;n-1}^{k}$.

This proves recurrence relation 1.

Corollary 1: For n even

$$\frac{1}{2}\left(\mu_{\frac{n}{2}+1;n}^{k} + \mu_{\frac{n}{2};n}^{k}\right) = \mu_{\frac{n}{2};n-1}^{k}.$$

Proof: On taking $r = \frac{n}{2}$ in recurrence relation1, we get

$\left(n-\frac{n}{2}\right)\mu_{r;n}^{k}+\frac{n}{2}\mu_{r+1;n}^{k}=n\mu_{r;n-1}^{k}.$

Or $\frac{n}{2}\left(\mu_{\frac{n}{2}+1;n}^{k}+\mu_{\frac{n}{2};n}^{k}\right)=n\mu_{\frac{n}{2};n-1}^{k}\ \Rightarrow \frac{1}{2}\left(\mu_{\frac{n}{2}+1;n}^{k}+\mu_{\frac{n}{2};n}^{k}\right)=\mu_{\frac{n}{2};n-1}^{k}.$

Hence the corollary is proved.

In view of the above corollary, on taking k = 1, we get the result that the expected value of the median in samples of size n (where n even) and n – 1 are equal. This is because, when n is even, the median is given by $\frac{1}{2}\left(X_{\left(\frac{n}{2}\right)}+X_{\left(\frac{n}{2}+1\right)}\right)$. Then the expected value is given by

$$\frac{1}{2}\left[E\left(X_{\left(\frac{n}{2}\right)}\right)+E\left(X_{\left(\frac{n}{2}+1\right)}\right)\right]=\frac{1}{2}\left(\mu_{\frac{n}{2};n}+\mu_{\frac{n}{2}+1;n}\right). \tag{2}$$

Also the median in a sample of size n -1, when n is even is given by $X_{\frac{n}{2};n-1}$ and its expected value is given by

$$E\left(X_{\frac{n}{2};n-1}\right)=\mu_{\frac{n}{2};n-1}. \tag{3}$$

Thus from above corollary, (2) and (3) are equal. Thus the expected value of the median of a sample of size n is same as the expected value of the median of a sample of size n – 1.

Corollary 2: If the parent population is symmetric about origin and n is even, then

$$\mu_{\frac{n}{2};n-1}^{k}=\begin{cases}\mu_{\frac{n}{2};n}^{k} \text{ if } k \text{ is even}\\ 0 \text{ if } k \text{ is odd}.\end{cases}$$

Proof: When the parent distribution is symmetric about origin, then $f_{r,n}(-x)=f_{n-r+1,n}(x)$ and

$$\mu_{\frac{n}{2}+1;n}^{k}=E\left(X_{\left(\frac{n}{2}+1\right)}^{k}\right)=\int_{-\infty}^{\infty}x^{k}f_{\frac{n}{2}+1;n}(x)dx=\int_{-\infty}^{\infty}(-1)^{k}x^{k}f_{n-\frac{n}{2}-1+1;n}(-x)d(-x)$$

$$=\int_{-\infty}^{\infty}(-1)^{k}x^{k}f_{\frac{n}{2};n}(x)dx=(-1)^{k}\int_{-\infty}^{\infty}x^{k}f_{\frac{n}{2};n}(x)dx=(-1)^{k}E\left(X_{\frac{n}{2}}^{k}\right)=(-1)^{k}\mu_{\frac{n}{2};n}^{k}.$$

Or $$\mu_{\frac{n}{2}+1;n}^{k}=\begin{cases}\mu_{\frac{n}{2};n}^{k} \text{ if } k \text{ is even}\\ -\mu_{\frac{n}{2};n}^{k} \text{ if } k \text{ is odd}.\end{cases} \tag{4}$$

Thus from Corollary1,

$\mu_{\frac{n}{2};n-1}^{k}=\frac{1}{2}\left(\mu_{\frac{n}{2}+1;n}^{k}+\mu_{\frac{n}{2};n}^{k}\right).$

Now from (4), we have $\mu_{\frac{n}{2}+1;n}^{k}=\mu_{\frac{n}{2};n}^{k}$ for k even and hence,

$\mu^k_{\frac{n}{2};n-1} = \frac{1}{2}\left(\mu^k_{\frac{n}{2};n} + \mu^k_{\frac{n}{2};n}\right) = \mu^k_{\frac{n}{2};n}$.

For k odd, $\mu^k_{\frac{n}{2}+1;n} = -\mu^k_{\frac{n}{2};n}$ gives,

$\mu^k_{\frac{n}{2};n-1} = \frac{1}{2}\left(-\mu^k_{\frac{n}{2};n} + \mu^k_{\frac{n}{2};n}\right) = 0$.

This proves Corollary2.

Relation 2. For an arbitrary distribution

$$\mu^k_{r;n} = \sum_{i=r}^{n} \binom{i-1}{r-1}\binom{n}{i}(-1)^{i-r}\mu^k_{i;i}. \tag{1}$$

Thus the moments of $X_{r;n}$ are expressible in terms of simpler moments of the largest in samples of size $r, r+1, \cdots, n$.

Proof: By definition

$$\mu^k_{r,n} = E(X^k_{(r)}) = \int_{-\infty}^{\infty} x^k f_{r,n}(x)dx = \int_{-\infty}^{\infty} x^k \left[\frac{d}{dx} I_{P(x)}(r, n-r+1)\right] dx, \tag{2}$$

where $I_{P(x)}(r, n-r+1) = \frac{1}{\beta(r,n-r+1)}\int_0^{P(x)} t^{r-1}(1-t)^{n-r}dt.$ (3)

Now consider the integrand in (3), which is given by

$$\frac{1}{\beta(r,n-r+1)} t^{r-1}(1-t)^{n-r} = \frac{n!t^{r-1}}{(r-1)!(n-r)!}\sum_{j=0}^{n-r}\binom{n-r}{j}(-1)^j t^j$$

$$= \sum_{j=0}^{n-r} \frac{n!}{(r-1)!(n-r)!}\frac{(n-r)!}{j!(n-r-j)!}(-1)^j t^{r-1+j}.$$

Now let $i = j + r$. Then $j = i - r$, which gives

$$\frac{1}{\beta(r,n-r+1)} t^{r-1}(1-t)^{n-r} = \sum_{i=r}^{n} \frac{n!}{(r-1)!}\frac{1}{(i-r)!(n-i)!}(-1)^{i-r}t^{r-1+i-r}$$

$$= \sum_{i=r}^{n} \frac{n!}{i!(n-i)!}\frac{i!}{(i-r)!(r-1)!}(-1)^{i-r}t^{i-1} = \sum_{i=r}^{n} \frac{n!}{i!(n-i)!}\frac{(i-1)!}{(i-r)!(r-1)!}(-1)^{i-r}it^{i-1}$$

$$= \sum_{i=r}^{n} \binom{n}{i}\binom{i-1}{r-1}(-1)^{i-r}it^{r-1+i-r}\ . \tag{4}$$

Substituting (4) in (3), we get

$$I_{P(x)}(r, n-r+1) = \int_0^{P(x)}\left(\sum_{i=r}^{n}\binom{n}{i}\binom{i-1}{r-1}(-1)^{i-r}it^{i-1}\right)dt.$$

Then,

$$\frac{d}{dx} I_{P(x)}(r, n-r+1) = \sum_{i=r}^{n} \binom{n}{i} \binom{i-1}{r-1} (-1)^{i-r} i P(x)^{i-1} p(x). \qquad (5)$$

Substituting (5) in (2), we get

$$\mu_{r,n}^{k} = \int_{-\infty}^{\infty} x^k \sum_{i=r}^{n} \binom{n}{i} \binom{i-1}{r-1} (-1)^{i-r} i P(x)^{i-1} p(x) dx$$

$$= \sum_{i=r}^{n} \binom{n}{i} \binom{i-1}{r-1} (-1)^{i-r} \int_{-\infty}^{\infty} x^k \, i P(x)^{i-1} p(x) dx$$

$$= \sum_{i=r}^{n} \binom{n}{i} \binom{i-1}{r-1} (-1)^{i-r} \int_{-\infty}^{\infty} x^k \, f_{i;i}(x) dx$$

$= \sum_{i=r}^{n} \binom{n}{i} \binom{i-1}{r-1} (-1)^{i-r} E(X_{i;i}^{k})$, where $X_{i;i}$ denotes the i^{th} order statistic from a sample of size *i*. Hence,

$\mu_{r,n}^{k} = \sum_{i=r}^{n} \binom{n}{i} \binom{i-1}{r-1} (-1)^{i-r} \mu_{i;i}^{k}$, which proves relation2.

Relation 2. For an arbitrary distribution,

$$\mu_{r;n}^{k} = \sum_{i=n-r+1}^{n} \binom{i-1}{n-r} \binom{n}{i} (-1)^{i-n+r-1} \mu_{1;i}^{k}. \qquad (1)$$

i.e. the moments of $X_{(r)}$ are expressible in terms of the moments of the smallest in samples of sizes $n-r+1, n-r+2, \cdots, n$.

Proof*:* By definition

$$\mu_{r,n}^{k} = E(X_{(r)}^{k}) = \int_{-\infty}^{\infty} x^k f_{r,n}(x) dx = \int_{-\infty}^{\infty} x^k \left[\frac{d}{dx} I_{P(x)}(r, n-r+1)\right] dx, \qquad (2)$$

where $I_{P(x)}(r, n-r+1) = \frac{1}{\beta(r,n-r+1)} \int_{0}^{P(x)} t^{r-1}(1-t)^{n-r} dt. \qquad (3)$

Now consider the integrand in (3), which is given by

$$\frac{1}{\beta(r,n-r+1)} t^{r-1}(1-t)^{n-r} = \frac{n!}{(r-1)!(n-r)!} \{1-(1-t)\}^{r-1}(1-t)^{n-r}$$

$$= \frac{n!}{(r-1)!(n-r)!} \sum_{j=0}^{r-1} \binom{r-1}{j} (-1)^{j} (1-t)^{j} \, (1-t)^{n-r}$$

$$= \sum_{j=0}^{n-r} \frac{n!}{(r-1)!(n-r)!} \frac{(r-1)!}{j!(r-1-j)!} (-1)^j (1-t)^{n-r+j}.$$

Now let $i = j + n - r + 1$. Then $j = i - n + r - 1$, which gives

$$\frac{1}{\beta(r,n-r+1)} t^{r-1}(1-t)^{n-r} = \sum_{i=n-r+1}^{n} \frac{n!}{(n-r)!} \frac{1}{(i-n+r-1)!(r-1-i+n-r+1)!} (-1)^{i-n+r-1}(1-t)^{n-r+i-n+r-1}$$

$$= \sum_{i=r}^{n} \frac{n!}{i!(n-i)!} \frac{i!}{(i-r)!(r-1)!} (-1)^{i-r} t^{i-1} = \sum_{i=n-r+1}^{n} \frac{n!}{i!(n-i)!} \frac{(i-1)!}{(n-r)!(i-i-n+r)!} (-1)^{i-n+r-1} i(1-t)^{i-1}$$

$$= \sum_{i=n-r+1}^{n} \binom{n}{i} \binom{i-1}{n-r} (-1)^{i-n+r-1} i(1-t)^{i-1} \ . \tag{4}$$

Substituting (4) in (3), we get

$$I_{P(x)}(r, n-r+1) = \int_0^{P(x)} \left(\sum_{i=r}^{n} \binom{n}{i} \binom{i-1}{n-r} (-1)^{i-n+r-1} i(1-t)^{i-1} \right) dt.$$

Then,

$$\frac{d}{dx} I_{P(x)}(r, n-r+1) = \sum_{i=n-r+1}^{n} \binom{n}{i} \binom{i-1}{n-r} (-1)^{i-n+r-1} i\{1-P(x)\}^{i-1} p(x). \tag{5}$$

Substituting (5) in (2), we get

$$\mu_{r,n}^{k} = \int_{-\infty}^{\infty} x^k \left[\frac{d}{dx} I_{P(x)}(r, n-r+1) \right] dx$$

$$= \int_{-\infty}^{\infty} x^k \sum_{i=n-r+1}^{n} \binom{n}{i} \binom{i-1}{n-r} (-1)^{i-n+r-1} i\{1-P(x)\}^{i-1} p(x) dx$$

$$= \sum_{i=n-r+1}^{n} \binom{n}{i} \binom{i-1}{n-r} (-1)^{i-n+r-1} \int_{-\infty}^{\infty} x^k \, i\{1-P(x)\}^{i-1} p(x) dx$$

$$= \sum_{i=n-r+1}^{n} \binom{n}{i} \binom{i-1}{n-r} (-1)^{i-n+r-1} \int_{-\infty}^{\infty} x^k \, f_{1;i}(x) dx$$

$$= \sum_{i=n-r+1}^{n} \binom{n}{i} \binom{i-1}{n-r} (-1)^{i-n+r-1} E\left[X_{1;i}^{k}\right]$$

$$= \sum_{i=n-r+1}^{n} \binom{n}{i} \binom{i-1}{n-r} (-1)^{i-n+r-1} \mu_{1;i}^{k}.$$

This proves relation 2`.

Relation 3. For an arbitrary distribution and $1 \le r < s \le n$,

$(r-1)\mu_{r,s;n} + (s-r)\mu_{r-1,s;n} + (n-s+1)\mu_{r-1,s-1;n} = n\mu_{r-1,s-1;n-1}$.

Proof: By definition

$$\mu_{r-1,s-1;n-1} = E\left(X_{r-1;n-1}X_{s-1;n-1}\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{r-1,s-1;n-1}(x,y)dxdy \quad (1)$$

$$= \frac{(n-1)!}{(r-2)!(s-r-1)!(n-s)!}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyP(x)^{r-2}p(x)[P(y)-P(x)]^{s-r-1}p(y)[1-P(x)]^{n-s}dxdy.$$

Or $n\mu_{r-1,s-1;n-1}$

$$= \frac{n!}{(r-2)!(s-r-1)!(n-s)!}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyP(x)^{r-2}p(x)[P(y)-P(x)]^{s-r-1}p(y)[1-P(x)]^{n-s}dxdy. \quad (2)$$

Writing 1 as

$1 = P(x) + P(y) - P(x) + 1 - P(y)$, we have

$$n\mu_{r-1,s-1;n-1} = \frac{n!}{(r-2)!(s-r-1)!(n-s)!}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy[P(x) + \{P(y)-P(x)\}\{1-P(y)\}] \times$$

$$P(x)^{r-2}p(x)[P(y)-P(x)]^{s-r-1}p(y)[1-P(x)]^{n-s}dxdy.$$

$$= \frac{n!}{(r-2)!(s-r-1)!(n-s)!}\Big[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyP(x)^{r-1}p(x)[P(y)-P(x)]^{s-r-1}p(y)[1-P(x)]^{n-s}dxdy +$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyP(x)^{r-2}p(x)[P(y)-P(x)]^{s-r}p(y)[1-P(x)]^{n-s}dxdy +$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyP(x)^{r-2}p(x)[P(y)-P(x)]^{s-r-1}p(y)[1-P(x)]^{n-s+1}dxdy\Big].$$

Multiplying and dividing by $(r-1)$ in the first term, $(s-r)$ in the second term and $(n-s+1)$ in the third term of the above expression, we have

$n\mu_{r-1,s-1;n-1}$

$$= \frac{n!(r-1)}{(r-1)!(s-r-1)!(n-s)!}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyP(x)^{r-1}p(x)[P(y)-P(x)]^{s-r-1}p(y)[1-P(x)]^{n-s}dxdy$$

$$+ \frac{n!(s-r)}{(r-2)!(s-r)!(n-s)!}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyP(x)^{r-2}p(x)[P(y)-P(x)]^{s-r}p(y)[1-P(x)]^{n-s}dxdy$$

$$+ \frac{n!(n-s+1)}{(r-2)!(s-r-1)!(n-s+1)!}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyP(x)^{r-2}p(x)[P(y)-P(x)]^{s-r-1}p(y)[1-P(x)]^{n-s+1}dxdy$$

$$= (r-1)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{r,s;n}(x,y)dxdy + (s-r)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{r-1,s;n}(x,y)dxdy$$

$$+ (n-s+1)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{r-1,s-1;n}(x,y)dxdy$$

$$= (r-1)\mu_{r,s;n} + (s-r)\mu_{r-1,s;n} + (n-s+1)\mu_{r-1,s-1;n}.$$

This proves relation 3.

Remark*:* In the above relation if we take r = 1 and s = r + 1, we get

$(1-1)\mu_{1,r+1;n} + (r+1-r)\mu_{1-1,r+1;n} + (n-r-1+1)\mu_{1-1,r+1-1;n} = n\mu_{1-1,r+1-1;n-1}$

Or $\mu_{r+1;n} + (n-r)\mu_{r;n} = n\mu_{r;n-1}$.

Lemma 1*:* The r^{th} order statistic $X_{(r)}$ from an exponential population with *pdf*

$p(x) = e^{-x},\ \ x \geq 0;$

Can be expressed as

$X_{(r)} = \sum_{i=1}^{r}\frac{V_i}{n-i+1}$, where $V_1, V_2, \cdots, V_n$ are mutually independent random variables which are distributed as standard exponential variates with *pdf*

$p(v_i) = e^{-v_i},\ \ v_i \geq 0.$

Proof*:* Since X follows an exponential distribution, the *cdf* is given by

$P(x) = \int_0^x e^{-y}dy = 1 - e^{-x}.$

Now consider the joint distribution of two order statistics $X_{(r)}$ and $X_{(s)}$, $1 \leq r < s \leq n$, which is given by

$$f_{r,s}(x,y) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!}P^{r-1}(x)p(x)[P(y) - P(x)]^{s-r-1}p(y)[1-P(y)]^{n-s},$$

$$0 < x < y < \infty.$$

Let $X_{(0)}$ = 0 and $r = i-1, s = i; i = 1,2,\cdots,n$. Then,

$$f_{i-1,i}(x,y) = \frac{n!}{(i-2)!0!(n-i)!}P^{i-2}(x)p(x)[P(y) - P(x)]^0p(y)[1-P(y)]^{n-i}$$

$$= \frac{n!}{(i-2)!(n-i)!}[1 - e^{-x}]^{i-2}e^{-x}e^{-y}[1 - 1 + e^{-y}]^{n-i}$$

$$= \frac{n!}{(i-2)!(n-i)!}[1 - e^{-x}]^{i-2}e^{-x}e^{-(n-i+1)y}.$$

Now making a transformation $Z_i = X_{(i)} - X_{(i-1)}$ and $X_{(i-1)} = X_{(i-1)}$ for all $i = 1,2 \cdots, n$, *i.e.* z = $y - x$ and $x = x$, the inverse transformation is given by $x = x$ and $y = z + x$. The Jacobean of transformation is $|J|$ =1 and the range of Z_i is given by $Z_i > 0$ *i.e.* $0 < z < \infty$ and $0 < x < \infty$. Thus the joint density of Z_i and $X_{(i-1)}$ is given by

$$g(z,x) = \frac{n!}{(i-2)!(n-i)!}[1- e^{-x}]^{i-2}e^{-x}e^{-(n-i+1)(z+x)}$$

$$= \frac{n!}{(i-2)!(n-i)!}[1- e^{-x}]^{i-2}e^{-(n-i+2)x}e^{-(n-i+1)z}.$$

On integrating out *x*, we get the density of Z_i as

$$h(z) = \frac{n!}{(i-2)!(n-i)!}e^{-(n-i+1)z}\int_0^\infty[1- e^{-x}]^{i-2}e^{-(n-i+2)x}dx.$$

Now using the probability integral transformation, which will transform each of the order statistics $X_{(1)}, X_{(2)} \cdots, X_{(n)}$ to $U_{(1)}, U_{(2)} \cdots, U_{(n)}$, which are the new order statistics from the uniform population in the interval (0,1), we have

$u = P(x) = 1 - e^{-x}$, which gives $x = -\log(1-u)$; $\frac{du}{dx} = p(x) = e^{-x} = 1 - u$. Hence,

$$h(z) = \frac{n!}{(i-2)!(n-i)!}e^{-(n-i+1)z}\int_0^1 u^{i-2}(1-u)^{(n-i+2)}\frac{du}{(1-u)}$$

$$= \frac{n!}{(i-2)!(n-i)!}e^{-(n-i+1)z}\int_0^1 u^{i-2}(1-u)^{(n-i+1)}du$$

$$= \frac{n!}{(i-2)!(n-i)!}e^{-(n-i+1)z}\beta(i-1, n-i+2)$$

$$= (n-i+1)e^{-(n-i+1)z};\ z > 0.$$

Thus if $V_i =,(n-i+1)Z_i$ then $V_i > 0$ for all $i = 1,2,\cdots,n$; and the distribution of V_i will be

$f(v_i) = e^{-v_i},\ v_i > 0.$

Also, for $i = 1$, we have $\frac{V_1}{n-1+1} = Z_1 = X_{(1)} - X_{(0)} = X_{(1)}$.

For $i = 2$, $\frac{V_2}{n-2+1} = Z_2 = X_{(2)} - X_{(1)}$; for $i = 3$, $\frac{V_3}{n-3+1} = Z_3 = X_{(3)} - X_{(2)}$; $\cdots$; for $i = n - 1$, $\frac{V_{n-1}}{n-n+1+1} = Z_{n-1} = X_{(n-1)} - X_{(n-2)}$; and for $i = n$, $\frac{V_n}{n-n+1} = Z_n = X_{(n)} - X_{(n-1)}$. Hence,

$X_{(1)} = \frac{V_1}{n-1+1}$,

$X_{(2)} = Z_2 + X_{(1)} = \frac{V_2}{n-2+1} + \frac{V_1}{n-1+1}$,

$\vdots$

$X_{(r)} = Z_r + Z_{r-1} + \cdots + Z_1 = \sum_{i=1}^{r} \frac{V_i}{n-i+1}$.

Hence the lemma is proved.

Example: For a random sample of size *n* from an exponential distribution with *pdf*

$p(x) = e^{-x}$ for $x \geq 0$

$= 0$ for $x < 0$.

Show that $\mu_{r;n} = \sum_{i=n-r+1}^{n} \frac{1}{i}$ and for $r < n$,

$\sigma_{r,r;n} = \sigma_{r;n}^2 = \sum_{i=n-r+1}^{n} \frac{1}{i^2}$.

Solution: We have $\mu_{r,n} = E\left(X_{(r)}\right)$ and since the variable X is distributed as a standard exponential, by above lemma, we can express $X_{(r)}$ as

$X_{(r)} = \sum_{i=1}^{r} \frac{V_i}{n-i+1}$, where $V_1, V_2, \cdots, V_n$ are mutually independent random variables which are distributed as standard exponential variates with *pdf*

$p(v_i) = e^{-v_i}$, $v_i \geq 0$.

Here, $E(V_i) = \int_0^\infty v e^v \, dv = 1$ and $V(V_i) = E(V_i^2) - 1 = 1$. Hence,

$$\mu_{r,n} = E\left[\sum_{i=1}^{r} \frac{V_i}{n-i+1}\right] = \sum_{i=1}^{r} \frac{E(V_i)}{n-i+1}$$

$$= \sum_{i=1}^{r} \frac{1}{n-i+1} = \frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{n-r+1} = \sum_{i=n-r+1}^{n} \frac{1}{i}.$$

$$\sigma_{r,r;n} = \sigma_{r;n}^2 = V(X_{(r)}) = V\left[\sum_{i=1}^{r} \frac{V_i}{n-i+1}\right] = \sum_{i=1}^{r} \frac{V(V_i)}{(n-i+1)^2}$$

$$= \sum_{i=1}^{r} \frac{1}{(n-i+1)^2} = \frac{1}{n^2} + \frac{1}{(n-1)^2} + \frac{1}{(n-2)^2} + \cdots + \frac{1}{(n-r+1)^2} = \sum_{i=n-r+1}^{n} \frac{1}{i^2}.$$

2.6 Non-Parametric Estimation of Distribution Function

The statistical problem is estimation of the entire distribution function F, or its values $F(x)$ at a specific argument x, or on the other hand, quantiles, i.e., the argument x where $F(x)$ takes specific values.

2.6.1 The Empirical Distribution Function

By repeating the experiment n times independently under the same conditions, one obtains the realization $x_1, x_2, \dots, x_n$ of a random sample $X_1, X_2, \dots, X_n$, from the distribution function under investigation, say $F(x)$. Its values at the real x may be estimated in the following way.

Define the empirical distribution function (e.d.f)

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I[X_i \leq x].$$

Essentially it estimates $F(x) = P[X \leq x]$ by the relative frequency of the event $[X_i \leq x, i = 1,2, \dots, n]$ in the random sample. If the order statistics of the random sample are $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ and its sample realization is $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$, then an equivalent representation of e.d.f. is given by

$$F_n(x) = \begin{cases} 0, & \text{if } x < X_{(1)}, \\ \dfrac{i}{n} & \text{if } X_{(i)} \leq x < X_{(i+1)},\ i = 1, \dots, n-1 \\ 1 & \text{if } X_{(n)} \leq x. \end{cases} \quad (1)$$

It is a jump function, with each jump equal to $1/n$ and located at the n order statistics $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$. Thus $F_n(x)$ will always yield a discrete (right continuous) distribution function giving probability $1/n$ to each of the order statistics. In case of ties, the appropriate adjustment to the jump size at the tied observations will be made, more specifically if k observations are tied, the jump size is taken to be k/n. The distribution function $F(x)$, which is being estimated, may or may not be discrete. However, we shall see later that in all cases $F_n(x)$ tends to be closer and closer to $F(x)$ at all x, with probability 1 as n, the sample size, becomes larger and larger. Hence it is a very attractive estimator of $F(x)$.

2.6.1.1 Properties of the Empirical Distribution Function

The empirical distribution function $F_n(x)$ is an unbiased and a weakly consistent estimator of the unknown distribution function $F(x)$.

For a fixed x

$$
\begin{aligned}
E[F_n(x)] &= \frac{1}{n}\sum_{i=1}^{n} E\big[I[X_i \leq x]\big] \\
&= P[X \leq x] \\
&= F(x)\,. \qquad (2)
\end{aligned}
$$

Hence $F_n(x)$ is unbiased for $F(x)$. It is also weakly consistent, since

$$
\begin{aligned}
\mathrm{Var}\big(F_n(x)\big) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} I[X_i \leq x]\right) \\
&= \frac{1}{n^2} n \mathrm{Var} I[X \leq x] \\
&= \frac{F(x)\big(1-F(x)\big)}{n} \\
&\to 0 \text{ as } n \to \infty. \qquad (3)
\end{aligned}
$$

The Borel strong law of large numbers also applies, giving $F_n(x) \to F(x)$ as $n \to \infty$ with probability 1 at fixed x.

The following theorem shows that the empirical distribution function $F_n(x)$ is a uniformly strongly consistent estimator of the unknown distribution function $F(x)$.

2.7 Glivenko-Cantelli Fundamental Theorem

Statement: Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables from distribution $F(x)$. Let $F_n(x)$ be the corresponding empirical distribution function. Then

$$P\left[\sup_{-\infty<x<\infty} |F_n(x) - F(x)| \to 0 \text{ as } n \to \infty\right] = 1.$$

Proof: Let $j = 1,2,\dots,k$ and $k = 1,2,\dots$ Let x_{kk} be ∞. Define x_{jk} to be the largest value of x such that

$$F(x-0) \leq \frac{j}{k} \leq F(x).$$

Thus, for every k, the points $x_{1k}, x_{2k}, \dots, x_{k-1k}$ provides a partition of the real line given by $(-\infty, x_{1k}], (x_{1k}, x_{2k}], \dots, (x_{k-1k}, \infty)$. The convergence with probability 1 of $F_n(x)$ to $F(x)$ at each end point of the above interval follows from the Borel strong law of large numbers. This is so because $F(x)$ is the probability of the event $[X \leq x]$ and $F_n(x)$ is the relative frequency of this event in independent trials. Then by elementary rules of intersections and unions of events (of probability 1) we get the uniform convergence of $F_n(x)$ to $F(x)$ at all the endpoints of the above intervals with probability 1. Using the nondecreasing nature of the functions $F_n(x)$ and $F(x)$ and the definition of x_{jk} it is assured that the absolute difference $|F_n(x) - F(x)|$ for x within any of the above intervals is not more than the absolute difference at one of the endpoints of one of the intervals plus $1/k$. Since k can be arbitrarily large, we can chose it so as to make $1/k$ as small as we please ensuring the uniform convergence over $(-\infty, \infty)$ of $F_n(x)$ to $F(x)$ with probability 1.

This theorem shows that $F_n(x), -\infty < x < \infty$ is a good estimator of the true distribution function $F(x)$, especially if the number of observations is not too small. The functionals (parameters) of the true distribution function may then be estimated by the corresponding functionals of $F_n(x)$. For example, the mean of F may be estimated by the mean of $F_n(x)$ which turns out to be $\bar{X}$, the sample mean. This is a point estimator.

We have seen that $nF_n(x) = \sum_{i=1}^{n} I[X_i \leq x]$. This can be interpreted as the number of successes in n independent trials with probability of success equal to $F(x)$ at each trial. Hence, for a fixed $x, nF_n(x)$ has a Binomial $(n, F(x))$ distribution.

Therefore, from De-Moivre Laplace Theorem it follows that

$$\frac{nF_n(x) - nF(x)}{\sqrt{nF_n(x)\big(1 - F_n(x)\big)}} \to N(0,1) \text{ as } n \to \infty\,. \tag{4}$$

2.8 Summary

This unit provides a thorough understanding of concepts related to Asymptotic Distribution Theory. The concepts of moments of order statistics, non-parametric estimation of distribution function and Glivenko-Cantelli fundamental theorem are described in detail. The learner should try to solve the self-assessment problems given in the next section.

2.9 Self-Assessment Exercises

Q1. Describe the moments of order statistics and their utility.

Q2. Explain estimation of distribution function under non-parametric theory.

Q3. State and prove Glivenko-Cantelli fundamental theorem.

Q4. Show that the mean and variance of the empirical distribution function F_n, are

$$E[F_n(x)] = F(x), \qquad var[F_n(x)] = \frac{F(x)[1 - F(x)]}{n}.$$

Hence show that $F_n(x)$ is a consistent estimator of $F(x)$.

UNIT:3 DISTRIBUTION FREE INTERVALS

Structure

3.1 Introduction

A statistical procedure or a method is called *distribution free* if the statistic used has a distribution which does not depend on the distribution function (or the density or mass function) of the population from which the sample is drawn.

For example, suppose we draw a sample of size n from a continuous population with median M and call the total number of observations in the sample greater than the median value M as r, then the distribution of r will be a binomial with parameter n and 1/2, whatever be the parent distribution be. Hence r is a distribution free statistic.

3.2 Objectives

The objective of this unit is to provide a basic understanding of concepts related to Distribution Free Intervals. The concepts of the distribution of range function of order statistics, distribution free confidence intervals for quintiles, distribution free tolerance interval should be clear after study of this material.

3.3 Distribution of Range Function of Order Statistics

Let $Y_i\left(i=1,2,...,n\right)$ be an ith order statistic of the random sample $X_1, X_2,..., X_n$ drawn from a continuous population whose c. d. f. $F\left(x\right)$ and p. d. f. is $f\left(x\right)$ for $a<x<b$. We define the sample range as

$R=Y_n-Y_1$.

In order to find the p. d. f. of R we first need to find the joint p. d. f. of Y_1 and Y_n is given by

$$g\left(y_1,y_n\right)=\frac{n!}{(n-2)!}\left\{F\left(y_n\right)-F\left(y_1\right)\right\}^{n-2} f\left(y_n\right) f\left(y_1\right) \qquad ;a<y_1<y_n<b$$

$$=n\left(n-1\right)\left\{F\left(y_n\right)-F\left(y_1\right)\right\}^{n-2} f\left(y_n\right) f\left(y_1\right) \qquad ;a<y_1<y_n<b$$

Let us now consider the transformation

$R=Y_n-Y_1$

$U=Y_n$

For specific values, we write

$r=y_n-y_1$

$u=y_n$

Then,

$y_1=u-r$ and $y_n=u$

Thus, the transformation $y_1=u-r$ and $y_n=u$ maps $\left\{\left(y_1,y_n\right);a<y_1<y_n<b\right\}$ onto $\left\{\left(r,u\right);a<r<u<b\right\}$ so that the joint p. d. f. of R and U is given by

$f_{RU}\left(r,u\right)=$ (mod J){putting y_1 and y_n in terms of u and r in $g\left(y_1,y_n\right)$ }

where J stands for the jacobian of transformation given by

$$J=\frac{\partial\left(y_1,y_n\right)}{\partial\left(r,u\right)}=\begin{vmatrix}\frac{\partial y_1}{\partial r} & \frac{\partial y_1}{\partial u}\\ \frac{\partial y_2}{\partial r} & \frac{\partial y_2}{\partial u}\end{vmatrix}=\begin{vmatrix}-1 & 1\\ 0 & 1\end{vmatrix}=1$$

so the joint p. d. f. of R and U takes form

$$f_{RU}\left(r,u\right)=n\left(n-1\right)\left\{F\left(u\right)-F\left(u-r\right)\right\}^{n-2} f\left(u-r\right) f\left(u\right) \qquad ;a<r<u<b$$

In order to obtain the p. d. f. of R we integrate out U from the joint p. d. f. of R and U and get

$$f_R(r) = \int_a^b f_{RU}(r,u)\,du$$

$$= \begin{cases} n(n-1)\int_r^b \{F(u)-F(u-r)\}^{n-2} f(u-r)\,f(u)\,du & ; a<r<b \\ 0 & ; \text{otherwise} \end{cases}$$

Example: *If* $f(x) = \lambda e^{-\lambda x} \qquad ; 0 < x < \infty$

and $F(x) = 1 - e^{-\lambda x} \qquad ; 0 < x < \infty$

Find the distribution of sample range.

Solution: We know distribution of sample range

$$f(r) = n(n-1)\int_r^\infty \left[F(u)-F(u-r)\right]^{n-2} f(u-r)\,f(u)\,du$$

$$= n(n-1)\int_r^\infty \left[1-e^{-\lambda x}-1+e^{-\lambda u+\lambda r}\right]^{n-2} \lambda e^{-\lambda u+\lambda r}\lambda e^{-\lambda u}du$$

$$= n(n-1)\lambda^2\int_r^\infty e^{-\lambda u(n-2)}\left(-1+e^{\lambda r}\right)^{n-2} e^{-2\lambda u}e^{\lambda r}du$$

$$= n(n-1)\lambda^2\left(-1+e^{\lambda r}\right)^{n-2} e^{\lambda r}\int_r^\infty e^{-\lambda un}du$$

$$= n(n-1)\lambda^2\left(-1+e^{\lambda r}\right)^{n-2} e^{\lambda r}\left[\frac{e^{-\lambda un}}{-n\lambda}\right]_r^\infty$$

$$= (n-1)\lambda\left(-1+e^{\lambda r}\right)^{n-2} e^{\lambda r}e^{-\lambda nr}$$

$$= (n-1)\lambda\left(-1+e^{\lambda r}\right)^{n-2} e^{\lambda r}e^{-\lambda nr} \qquad \text{Q.E.D.}$$

Also

$$(n-1)\lambda\int_{0}^{\infty}\left(-1+e^{\lambda r}\right)^{n-2}e^{-r\lambda n}e^{\lambda r}dr$$

$$(n-1)\int_{1}^{\infty}e^{-\lambda u(n-2)}(t-1)^{n-2}\left(\frac{1}{t}\right)^{n}dt \qquad \because e^{\lambda r}=t,\lambda e^{\lambda r}\,dr=dt$$

$$=(n-1)\int_{1}^{\infty}(t)^{n-2}\left(1-\frac{1}{t}\right)^{n-2}\frac{1}{t^{n}}dt$$

Let $\left(1-\frac{1}{t}\right)=\vartheta$ so that $\left(\frac{1}{t^{2}}\right)dt=dv$

$$I=(n-1)\int_{0}^{1}(\vartheta)^{n-2}dv$$

$$=(n-1)\left[\frac{u^{n-1}}{n-1}\right]_{0}^{1}$$

$$=1$$

Hence,

$$f(r)=(n-1)\lambda\left(e^{-r\lambda}-1\right)^{n-2}e^{-nr\lambda}e^{\lambda r} \qquad 0<r<\infty$$

3.4 Distribution Free Confidence Intervals for Quantiles (Quantiles of a Distribution)

Let X be a continuous random variable with p. d. f. $f(x)$ and c. d. f. $F(x)$. Let p be a positive proper fraction and the equation $F(x)=p$ as a unique solution for x, this unique root is denoted by the symbol ξ_p and is called the quantiles of order p.

Thus,

$$\Pr\left[X\le\xi_p\right]=F\left(\xi_p\right)=p \qquad 0<p<1$$

If $F(x)$ is not strictly increasing, $F(x)=p$ may hold in some interval, in this case any point in the interval would serve as a quantile of order *p*.

Example: The quantile of order 1/2 is the median of the distribution and

$\Pr[X \le \xi_{0.5}] = F(\xi_{0.5}) = 1/2$

Example: *Let* $x_i \,(i=1,2,...,n)$ *be i. i. d. random variable with p. d. f.* $f(x)$ *of the continuous type. If m is the median of the distribution, find the probability that*

i) *All exceed's m*

ii) *The maximum never exceeds m*

***Solution*:** Since *m* is the median of the continuous distribution. Therefore $F(m)=\Pr(X \le m)=1/2$ and $\Pr(X \ge m)=\Pr(X \le m)=1/2$

Now,

i)Pr (all exceed's m)= $\Pr\left(X_{(1)} \ge m\right)$

$$= \int_m^\infty n\{1-F(x)\}^{n-1} f(x)\,dx$$

Let $1-F(x)=t \;\; \Rightarrow -f(x)dx = dt$

so that

Pr (all exceed's m) $= \int_{1/2}^{0} -n\,t^{n-1}\,dt$

$$= \int_0^{1/2} n\,t^{n-1}\,dt = \left[t^n\right]_0^{1/2}$$

$$= \left(\frac{1}{2}\right)^n$$

(ii) Pr (none of the X's exceeds median)= Pr(the maximum never exceeds m)

$$= \Pr\left[X_{(n)} \le m\right]$$

$$= \int_0^m n\left[F(x)\right]^{n-1} f(x)\,dx$$

Let $F(x) = t \quad \Rightarrow f(x)dx = dt$

so that

Pr (none of the *X*'s exceeds *m*) $= \int_0^{1/2} n\,t^{n-1}\,dt = \left[t^n\right]_0^{1/2}$

$$= \left(\frac{1}{2}\right)^n \qquad \text{Q.E.D.}$$

3.4.1 Confidence Interval for Distribution Quantiles

Let $X_1, X_2, ..., X_n$ be a random sample of size n taken from a continuous distribution with distribution function $F(x)$. Let $Y_1 < Y_2 < ... < Y_n$ be the order statistics of the sample. Let $Y_i < Y_j$, we consider the event $Y_i < \xi_p < Y_j$. For the ith order statistic Y_i to be less than ξ_p it must be true that at least *i* of the x values are less than ξ_p. Moreover, for the *j*th order statistic to be greater than ξ_p fewer than *j* of the *x* values are less than ξ_p. That is, if we say that we have a "success" when an individual *x* value is less than ξ_p, then, in the n independent trials, there must be at least *i* success but fewer than *j* success for the event $Y_i < \xi_p < Y_j$ to occur. But since the probability of success on each trial is

$$\Pr\left[X \le \xi_p\right] = F\left(\xi_p\right) = p,$$

the probability of this event is

$$\Pr\left[Y_i < \xi_p < Y_j\right] = \sum_{w=i}^{j-1} \frac{n!}{w!(n-w)!} p^w (1-p)^{n-w}$$

The probability of having at least *i*, but less the *j* success. When particular values of *n*, *i*, and *j* are specified, this probability can be computed. Let this probability be γ

i.e. $\Pr\left[Y_i < \xi_p < Y_j\right] = \gamma$

then we say that the probability is γ that the random interval (Y_i, Y_j) includes the quantile of order p. if the experimental values of Y_i and Y_j are respectively, y_i and y_j, the interval (y_i, y_j) serves as $100\gamma\%$ confidence interval for ξ_p, the quantile of order p.

Theorem. A confidence interval for $q_{p.}$ based on order statistics will be distribution free, *i.e.* if $X_{(r)}$ and $X_{(s)}$ are r^{th} and s^{th} order statistics of a random sample of size n from a continuous distribution, where $1 \le r < s \le n$, and if the interval $(X_{(r)}, X_{(s)})$ covers the population quantile $q_{p.}$ of order p, then the confidence coefficient β will depend only on r, s, n and p, but not on the *cdf* $P(x)$or *pdf* $p(x)$of the population.

Proof*:* Consider the event $\{X_{(r)} \le q_{p.}\}$, which can be expressed as a disjoint union of the two events $\{X_{(r)} \le q_p \cap X_{(s)} \ge q_p\}$ and $\{X_{(r)} \le q_p \cap X_{(s)} < q_p\}$. As we are considering a continuous distribution, $P\{X_{(r)} = q_{p.}\} = 0$ and since the happening of the event $\{X_{(s)} < q_p\}$ implies the event $\{X_{(r)} \le q_{p.}\}$, hence the event $\{X_{(r)} \le q_p \cap X_{(s)} < q_p\}$ is equivalent to the event $\{X_{(s)} < q_p\}$. Thus,

$\{X_{(r)} \le q_{p.}\} = \{X_{(r)} \le q_p \cap X_{(s)} \ge q_p\} \cup \{X_{(s)} < q_p\} = \{X_{(r)} \le q_p \le X_{(s)}\} \cup \{X_{(s)} < q_p\}$.

Hence, $P\{X_{(r)} \le q_{p.}\} = P\{X_{(r)} \le q_p \le X_{(s)}\} + P\{X_{(s)} < q_p\}$.

Or $P\{X_{(r)} \le q_p \le X_{(s)}\} = P\{X_{(r)} \le q_{p.}\} - P\{X_{(s)} < q_p\} = F_r(q_p) - F_s(q_p)$

$$= \sum_{j=r}^{n} \binom{n}{j} P(q_p)^j [1 - P(q_p)]^{n-j} - \sum_{j=s}^{n} \binom{n}{j} P(q_p)^j [1 - P(q_p)]^{n-j}$$

$$= \sum_{j=r}^{n} \binom{n}{j} p^j [1-p]^{n-j} - \sum_{j=s}^{n} \binom{n}{j} p^j [1-p]^{n-j}$$

$$= \sum_{j=r}^{s-1} \binom{n}{j} p^j [1-p]^{n-j} = \beta\,(r, s, n, p) \text{ (confidence coefficient).}$$

Thus the confidence coefficient is a function of r, s, n and p, but not on the *cdf* $P(x)$ or *pdf* $p(x)$ of the population.

Now for constructing a distribution free confidence interval with confidence coefficient greater than or equal to $1 - \alpha$ and given values of n and p, the values of r and s should be so chosen so as to make the confidence coefficient $\beta(r, s, n, p)$ just greater than $1 - \alpha$. Proper choice of r and s is some what arbitrary, but it is reasonable to try to make $s - r$ as small as possible subject to $\beta(r, s, n, p) \geq 1 - \alpha$. If $p = ½$, then the problem is that of construction of a distribution free confidence interval for the median of the population. In this case, it is customary to take $s = n - r + 1$, then the expression of $\beta(r, s, n, p)$ reduces to

$$\beta(r, s, n, p) = \sum_{j=r}^{n-r} \binom{n}{j} \left(\frac{1}{2}\right)^{j} \left(1-\frac{1}{2}\right)^{n-j} = \left(\frac{1}{2}\right)^{n} \sum_{j=r}^{n-r} \binom{n}{j}.$$

Now the problem reduces to that of determination of r, subject to $\beta(r, s, n, p) \geq 1 - \alpha$. Thus for given values of n, p and α, a value of r and s can be selected and then on reducing the sample width $(s - r)$ subject to the above said condition, $X_{(r)}$ and $X_{(s)}$ can be chosen, which is one confidence interval for q_p with confidence coefficient greater than or equal to $1 - \alpha$.

Example: *Find the smallest value of n for which* $\Pr\left[Y_1 < \xi_{0.5} < Y_n\right] \geq 0.99$*, where* $Y_1 < Y_2 < \ldots < Y_n$*, are order statistics of random sample of size n from a distribution of continuous type and* ξ_p *is a quantile of order p.*

Solution: Consider

$$\Pr\left[Y_1 < \xi_{0.5} < Y_n\right] \geq 0.99$$

$$= \sum_{w=1}^{n-1} {}^{n}C_w \left(0.5\right)^{w} \left(1-0.5\right)^{n-w} \geq 0.99$$

so that

$$\sum_{w=1}^{n-1} {}^{n}C_w \left(1/2\right)^{n} \geq 0.99 \qquad (1)$$

Also, we know that

$$\sum_{w=0}^{n} {}^{n}C_w \left(1/2\right)^{n} = 1 = \left(1/2 + 1/2\right)^{n} \qquad \therefore \left(q+p\right)^{n} = \sum_{r=0}^{n} {}^{n}C_r\, p^{r} \left(q\right)^{n-s}$$

$$\left(1/2\right)^{n} + \sum_{w=1}^{n-1} {}^{n}C_w \left(1/2\right)^{n} + \left(1/2\right)^{n} = 1$$

This gives

$$\sum_{w=1}^{n-1} {}^{n}C_{w}\left(1/2\right)^{n} = 1-\left(1/2\right)^{n} \qquad (2)$$

From (1) and (2) we get

$$1-2\left(1/2\right)^{n} \geq 0.99$$

$$1-0.99 \geq 2\left(1/2\right)^{n}$$

$$2\left(1/2\right)^{n} \leq 0.01 \qquad (3)$$

(3) holds for *n*=8, 9, …, hence smallest *n* is 8.

3.5 Distribution Free Tolerance Interval

Let $X_1, X_2, ..., X_n$ denotes a random sample of size n taken from a distribution having a positive and continuous p. d. f. $f\left(x\right)$ if and only if a<x<b. let $F\left(x\right)$ be its distribution function. Consider the random variables $F\left(X_1\right)$, $F\left(X_2\right)$, … $F\left(X_n\right)$. These random variables are mutually stochastically independent and each follows $U\left(0,1\right)$.

Let $Z_1, < Z_2 < ... < Z_n$ be the order statistics of the random sample $F\left(X_1\right)$, $F\left(X_2\right)$,…, $F\left(X_n\right)$. If $Y_1 < Y_2 < ... < Y_n$ are the order statistics of the original sample $X_1, X_2, ..., X_n$ then $Z_1 = F\left(Y_1\right), Z_2 = F\left(Y_2\right), ... Z_n = F\left(Y_n\right)$

Let us consider the difference $Z_j - Z_i = F\left(Y_j\right) - F\left(Y_i\right)$ for every i<j

Now $F\left(Y_j\right) = \Pr\left(X \leq Y_j\right)$

And $F\left(Y_i\right) = \Pr\left(X \leq Y_i\right)$

But $\Pr\left(X = Y_j\right) = \Pr\left(X = Y_i\right) = 0$ (as distribution is continuous)

Thus $Z_j - Z_i = \Pr\left(Y_i < X < Y_j\right)$

Let p be a positive fraction if

$F(Y_j) - F(Y_i) \geq p$

Then at least 100p% of the probability for the distribution of X is between y_i and y_j

Let $\gamma = \Pr\left[F(Y_j) - F(Y_i) \geq p\right]$

$$= \Pr\left[Z_j - Z_i \geq p\right]$$

$$= \int_{0}^{1-p} \int_{p+z_i}^{1} h_{ij}(Z_{i,} Z_j)\, dZ_j dZ_i$$

where $h_{ij}(Z_i, Z_j)$ is joint p. d. f. of Z_i and Z_j.

Then, the random interval (Y_i, Y_j) has probability γ of containing at least 100*p*% of the probability for the distribution of *x* is the tolerance interval of 100*p*% of the probability distribution of *x*. If now y_i and y_j denote respectively, experimental values of Y_i and Y_j, the interval (y_i, y_j) either does or does not contain at least 100*p*% of the probability for the distribution of *x* and y_i and y_j are known as the tolerance limits for 100*p*% of the probability distribution of *x*.

Definition: Tolerance interval is an interval which covers at least a certain proportion of the population with certain probability. Thus if *L* and *U* denotes the lower and upper limits of an interval, then the condition for this interval to be a tolerance interval is that it covers at least a proportion γ of a population with *pdf* $p(x)$ with probability β, where β and γ are pre assigned constants.

Symbolically

$$\mathrm{P}\left[\int_L^U p(x)dx \geq \gamma\right] = \beta. \qquad (1)$$

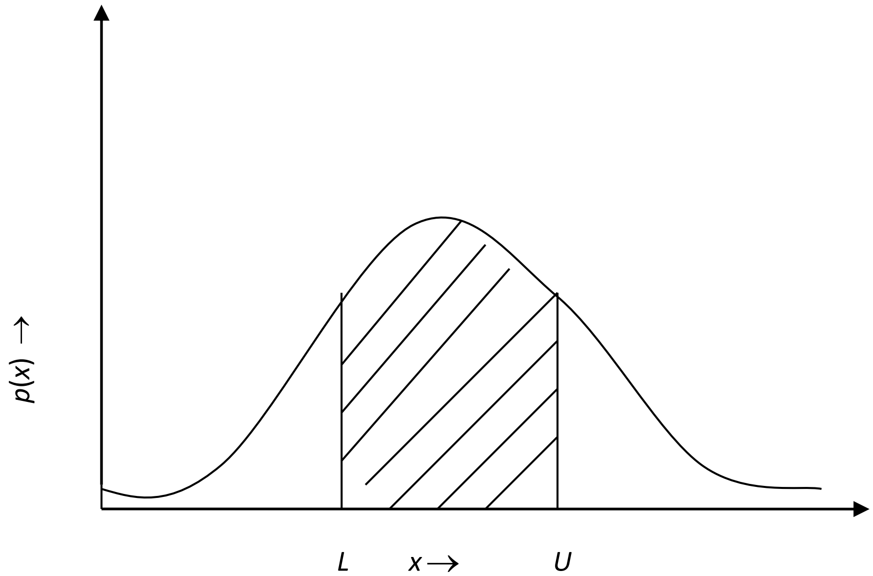


Fig. Graphical Depiction of Tolerance Interval

3.5.1 Difference Between Confidence and Tolerance Intervals

A confidence interval covers a population parameter with a stated confidence, where as a tolerance interval is designed to cover a proportion of the population with a certain probability. As the sample size increases, confidence intervals shrink towards zero, while tolerance intervals tend towards a fixed value.

Theorem. If the limits L and U of a tolerance interval depend on order statistics, then the probability given on the L.H.S. of (1) will be distribution free. In other words, the tolerance interval $(X_{(r)}, X_{(s)})$, where $X_{(r)}$ and $X_{(s)}$ are the r^{th} and s^{th} order statistics of a sample of size n, is distribution free.

Proof*:* Let $X_1, X_2, \ldots, X_n$ be a sample of size n from a continuous distribution with *pdf* $p(x)$, distribution function $P(x)$ and $X_{(r)}$ and $X_{(s)}$ the r^{th} and s^{th} order statistics. If we let $L = X_{(r)}$ and $U = X_{(s)}$, then

$$\int_L^U p(x)dx = P(U) - P(L) = P(X_{(s)}) - P(X_{(r)}). \quad (2)$$

Now making use of *probability integral transformation* $y = P(x)$, which is order preserving, the integral (3.2) is transformed to $Y_{(s)} - Y_{(r)}$, where $Y_{(s)}$ and $Y_{(r)}$ are the s^{th} and r^{th} order statistics from a standard Uniform distribution, $1 \leq r < s \leq n$. The distribution of $Z_{sr} = Y_{(s)} - Y_{(r)}$ was obtained earlier as a β with parameters $s - r$ and $n - s + r - 1$. Hence,

$$\mathrm{P}\left[\int_L^U p(x)dx \geq \gamma\right] = 1 - I_\gamma(s - r, n - s + r + 1), \qquad (3)$$

which does not depend upon $p(x)$ or $P(x)$, where $I_z(a, b) = \frac{1}{\beta(a,b)} \int_0^z y^{a-1}(1-y)^{b-1}dy$ is the incomplete beta function up to the point z.

Now for constructing a distribution free tolerance interval that would cover at least $100\gamma\%$ of the population with probability β, we look for an interval $(X_{(r)}, X_{(s)})$, where $X_{(r)}$ and $X_{(s)}$ are the are the r^{th} and s^{th} order statistics of a sample of size n such that

$$1 - I_\gamma(s - r, n - s + r + 1) = \beta. \qquad (4)$$

Since, r and s are integers, (3.4) can be hardly satisfied. Hence, we look out for such integral values of r and s that satisfy,

$$1 - I_\gamma(s - r, n - s + r + 1) \geq \beta.$$

Thus, for a one sided tolerance interval, we either take $r = 0$ or $s = n + 1$. Then the problem reduces to that of finding the other integer alone.

For a two sided tolerance interval it is usual to take $s = n - r + 1$ and determine that value of r that makes the value of $1 - I_\gamma(s - r, n - s + r + 1)$ as little in excess of β as possible.

Example: *Let Y_1 and Y_n be the smallest (i.e. the first and the n^{th} order statistics of a random sample of size n from the continuous distribution $F(x)$. Find the smallest n such that $P\left[\left\{F(Y_n) - F(Y_1)\right\} \geq 0.5\right]$ is at least 0.95.*

Solution: Consider

$$P\left[\left\{F(Y_n) - F(Y_1)\right\} \geq 0.5\right] \geq 0.95$$

$$1-P\left[\left\{F\left(Y_n\right)-F\left(Y_1\right)\right\}\le 0.5\right]\ge 0.95$$

$$P\left[\left\{F\left(Y_n\right)-F\left(Y_1\right)\right\}\le 0.5\right]\le 0.5$$

$$P\left[0<Z_n-Z_1\le 0.5\right]\le 0.5$$

$$\int_0^{0.5} r_{n-1}\left(\vartheta\right)d\vartheta\le 0.5$$

but $h_{j-i}\left(\vartheta\right)=\begin{cases}\dfrac{\lceil n+1}{\lceil (j-i)\lceil (n-j+i+1)}\vartheta^{j-i-1}\left(1-\vartheta\right)^{n-(j-i)} & ;0<\vartheta<1\\ 0 & ;\text{elsewhere}\end{cases}$

therefore, $h_{n-i}\left(\vartheta\right)=\dfrac{\lceil n+1}{\lceil (n-1)\lceil (n-n+1+1)}\vartheta^{n-2}\left(1-\vartheta\right)^{n-n+1}$

$$=\frac{n!}{(n-2)!}\vartheta^{n-2}\left(1-\vartheta\right)^{1}$$

$$=n\left(n-1\right)\vartheta^{n-2}\left(1-\vartheta\right)$$

Thus,

$$n\left(n-1\right)\int_0^{0.5}\vartheta^{n-2}\left(1-\vartheta\right)d\vartheta\le 0.05$$

$$n\left(n-1\right)\int_0^{0.5}\left(\vartheta^{n-2}-\vartheta^{n-1}\right)d\vartheta\le 0.05$$

$$n\left(n-1\right)\left(\frac{\vartheta^{n-1}}{n-1}-\frac{\vartheta^{n}}{n}\right)_0^{0.5}\le 0.05$$

$$n\left(n-1\right)\left(\frac{(0.5)^{n-1}n-(0.5)^{n}(n-1)}{n\left(n-1\right)}\right)\le 0.05$$

$$\left(\frac{1}{2}\right)^{n-1}\left(\frac{2n-n+1}{2}\right)\le 0.05$$

$$(n+1)\left(\frac{1}{2}\right)^n \le 0.05$$

The smallest *n* satisfying the above equation in *n* = 8.

3.6 Coverage

Let $X_1, X_2, ..., X_n$ denotes a random sample of size n taken from a distribution having a positive and continuous p. d. f. $f(x)$ if and only if a<x<b. let $F(x)$ be its distribution function. Consider the random variables $F(X_1)$, $F(X_2)$, … $F(X_n)$. These random variables are mutually stochastically independent and each follows $U(0,1)$. Thus $F(X_1)$, $F(X_2)$, … $F(X_n)$ is a random sample of size *n* from U(0,1).

Let $Z_1, < Z_2 < ... < Z_n$ be the order statistics of the random sample $F(X_1)$, $F(X_2), \ldots,$ $F(X_n)$. If $Y_1 < Y_2 < ... < Y_n$ are the order statistics of the original sample $X_1, X_2, ..., X_n$ then

$$Z_1 = F(Y_1), Z_2 = F(Y_2), ... Z_n = F(Y_n)$$

Now, consider the random variables

$$C_1 = W_1 = F(Y_1) = Z_1$$

$$C_2 = W_2 = F(Y_2) - F(Y_1) = Z_2 - Z_1$$

$$C_3 = W_3 = F(Y_3) - F(Y_2) = Z_3 - Z_2 .$$

$$\vdots$$

$$C_n = W_n = F(Y_n) - F(Y_{n-1}) = Z_n - Z_{n-1}$$

Then random variable W_1 or C_1 is called the coverage of the random interval $\{x; -\infty < x < Y_1\}$ and the random variable W_i or C_i, *i*=1, 2,..., *n* is called a coverage of random interval $\{x; Y_{i-1} < x < Y_i\}$

Joint p. d. f. of $W_1, W_2, ..., W_n$ or $C_1, C_2, ..., C_n$

We have

$$C_1 = W_1 = Z_1$$

$C_2 = W_2 = Z_2 - Z_1$

$C_3 = W_3 = Z_3 - Z_2$.

$\vdots$

$C_n = W_n = Z_n - Z_{n-1}$

For specific values

$c_1 = w_1 = z_1$

$c_2 = w_2 = z_2 - z_1$

$c_3 = w_3 = z_3 - z_2$.

$\vdots$

$c_n = w_n = z_n - z_{n-1}$

The inverse function of this associated transformation are given by

$z_i = w_1 + w_2 + ... + w_i$

$= c_1 + c_2 + ... + c_i$ for every i=1, 2, …, n

Now jacobian of transformation

$$J = \frac{\partial(z_1, z_2, ..., z_n)}{\partial(w_1, w_2, ..., w_n)}$$

$$= \begin{vmatrix} \frac{\partial z_1}{\partial w_1} & \frac{\partial z_1}{\partial w_2} \cdots & \frac{\partial z_1}{\partial w_n} \\ \frac{\partial z_2}{\partial w_1} & \frac{\partial z_2}{\partial w_2} \cdots & \frac{\partial z_2}{\partial w_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_n}{\partial w_1} & \frac{\partial z_n}{\partial w_2} & \frac{\partial z_n}{\partial w_n} \end{vmatrix}$$

$= 1$

Therefore, mod(J) = 1

Now

$h(w_1, w_2, ..., w_n) = r(c_1, c_2, ..., c_n)$

= (mod J){put $z_1, z_2, ..., z_n$ in terms of $w_1, w_2, ..., w_n$ in the joint p. d. f. of $Z_1, Z_2, ..., Z_n$ }

But $h(z_1, z_2, ..., z_n) = n!$

Thus, $h(w_1, w_2, ..., w_n) = n! \quad ; \; 0 < w_i, i = 1, 2, ..., n; \; w_1 + w_2 + ... + w_n < 1$

$\quad\quad\quad\quad = 0 \quad\quad ;$ elsewhere

Example: *Show that each of the coverages has the beta p. d. f.*

$$k(w) = \begin{cases} n(1-w)^{n-1} & ; 0 < w < 1 \\ 0 & ; elsewhere \end{cases}$$

Solution: Since the joint p. d. f. of the coverages $k(w_1, w_2, ..., w_n)$ is symmetric in $w_1, w_2, ..., w_n$ axis given by

$$k(w_1, w_2, ..., w_n) = \begin{cases} n! & ; 0 < w_i \\ & i = 1, 2, ..., n \\ & w_1 + w_2 + ... + w_n < 1 \\ 0 & ; elsewhere \end{cases}$$

it is evident that the distribution of every sum r, $r < n$ of these coverages $w_1, w_2, ..., w_n$ is exactly the same for fixed value of r.

Consider if $i < j$ and $r = j - i$, the distribution of any sum of j - i coverages

$$Z_j - Z_i = F(Y_j) - F(Y_i) = w_{i+1} + w_{i+2} + ... + w_j$$

$$= (w_1 + w_2 + ... + w_j) - (w_1 + w_2 + ... + w_i)$$

$$= w_{i+1} + w_{i+2} + ... + w_j \qquad\qquad (w_i < w_j)$$

is exactly the same as that of

$$z_{j-1} = F(Y_{j-i}) = w_1 + w_2 + ... + w_{j-i}$$

but we know the p. d. f. of

$$h_{j-i}(v)=\frac{n!}{\{(j-i)-1\}!(n-(j-i))!}v^{\{(j-i)-1\}}(1-v)^{n-(j-i)}$$

$$=\begin{cases}\dfrac{\overline{|n+1}}{\overline{|(j-i)}\,\overline{|(n-j+i+1)}}v^{j-i-1}(1-v)^{n-(j-i)} & ;0<v<1\\ 0 & ;elsewhere\end{cases}$$

Consequently, $Z_j-Z_i=F(Y_j)-F(Y_i)$ has above mentioned p. d. f. Putting r=1 such that *j*=2 and *i*=1, we have p. d. f. of w_1 given by

$$k(w_1)=\frac{\overline{|n+1}}{\overline{|1}\,\overline{|n}}(1-w_1)^{n-1} \qquad 0<w_1<1$$

$$=n(1-w_1)^{n-1}$$

But similarly if *j*=3 and *i*=2

$$k(w_2)=\frac{n!}{1!(n-1)!}(1-w_2)^{n-1} \qquad 0<w_2<1$$

$$=n(1-w_2)^{n-1}$$

Therefore in general, we can say that each of the coverages has the beta p. d. f.

$$k(w)=\begin{cases}n(1-w)^{n-1} & ;0<w<1\\ 0 & ;elsewhere\end{cases}$$

Example: *Let c_i denote the i^{th} coverage, find expectation of c_i.*

Solution: Since each of the coverage c_i, *i* =1, 2,…,, *n* has the beta p. d. f.

$$k(c)=\begin{cases}n(1-c)^{n-1} & ;0<c<1\\ 0 & ;elsewhere\end{cases}$$

and $c_1=Z_1=F(Y_1)$ follows $U(0,1)$. The expectation of each c_i is given by

$$E(c_i)=\int_0^1 nc(1-c)^{n-1}dc$$

$$= n\int_0^1 c(1-c)^{n-1}dc$$

$$= n\left[\left\{\frac{c\left\{-(1-c)^n\right\}}{n}\right\}_0^1 - \int_0^1 \frac{\left\{-(1-c)^n\right\}}{n}dc\right]$$

$$= n\left[0 + \frac{1}{n}\int_0^1 \left\{(1-c)^n\right\}dc\right] = \left[\frac{\left\{-(1-c)^{n+1}\right\}}{n+1}\right]_0^1$$

$$= \frac{1}{n+1}$$

3.7 Summary

This unit provides a thorough understanding of concepts related to Distribution Free Intervals. The concepts of Distribution of range function of order statistics, distribution free confidence intervals for quantiles, distribution free tolerance interval and coverage are described in detail. The learner should try to solve the self-assessment problems given in the next section.

3.8 Self - Assessment Execises

Q1. Derive the distribution of range function of order statistics in case of Uniform distribution.

Q2. Obtain the distribution free confidence intervals for quantiles.

Q3. Describe the concepts of tolerance interval and coverage by giving suitable examples.

UNIT: 4 RANK ORDER STATISTICS

Structure

4.1 Introduction

Statistics are a way of simplifying, summarizing, and interpreting data. Among the various statistical tools and techniques available, Rank Order Statistics stand out as a unique way to understand and interpret data points based on their rank or order rather than their actual values.

Rank Order Statistics are essentially data values arranged in ascending or descending order, providing a ranking for each data point. This allows for a clear distinction in the standing or position of each data value in relation to the others, irrespective of their absolute values. For instance, if we have a data set: {56, 62, 49, 89, 75}, the rank of the value 56 would be 2 (second smallest), while the value 89 would rank 5 (largest).

Rank Order Statistics are crucial in numerous applications:

***Non-Parametric Tests*:** In statistics, not all data conforms to a specific distribution or meets the necessary assumptions for parametric tests. In such cases, non-parametric tests that utilize rank order statistics, like the Mann-Whitney U test or the Kruskal-Wallis test, can be used.

Comparing Variables: It's often important to understand the relative standing of one data point to another, especially when absolute values can be misleading.

Handling Outliers: Rank order statistics can be a valuable tool when dealing with outliers, as ranking reduces the influence of abnormally high or low values.

***Predictive Modelling*:** In fields like machine learning, rank-based methods can be useful for certain types of predictive modelling scenarios, especially in recommendation systems.

Throughout this unit, we will delve deep into the intricacies of Rank Order Statistics, starting from its fundamental definition to its wide range of applications in various fields. The unit will also cover methods to compute ranks, deal with tied ranks, and utilize rank order statistics in practical scenarios.

As we navigate through the world of Rank Order Statistics, we will witness its power and versatility, gaining the skills and knowledge to apply it effectively in our statistical endeavours.

With this introduction as our foundation, we will embark on a detailed journey, starting with the basic definition and computation methods of ranks, progressing to its applications in non-parametric statistical tests and real-world scenarios, and culminating in its importance in the broader realm of data analysis. Rank order statistics provide insight into the distribution and arrangement of data points. In the realm of statistics and probability, there are various techniques and theorems that help us better understand and analyse the order and rank of data. This unit delves into rank order statistics, Dwass' Technique, the Ballot theorem, its generalization, and the implications of these in the realm of random variables.

4.2 Objectives

By the end of this unit, learners will be able to:

- Understand the fundamental concept of rank order statistics.
- Apply Dwass' Technique to statistical data.
- Describe the Ballot theorem and its generalized form.
- Understand the applications and implications of these concepts on fluctuations of sums of random variables.
- Self-assess their knowledge and understanding of the concepts.

4.3 Rank Order Statistics

Rank order statistics deals with the statistics derived from the ranks of sample data. For a sample size n, the smallest observation has a rank of 1, the second smallest has a rank of 2, and so on, with the largest observation having a rank of n

Definition: rank order statistics in the field of Statistics refers to the values in a dataset when it is ordered in ascending or descending order. More formally, given a sample of size n, the r^{th}rank order statistics is the r^{th} smallest value in the sample, often denoted $X_{(r)}$.

Mathematically, let us assume we have a sample $X_1, X_2, \dots, X_n$ from a population. The rank order statistics of the sample are given by:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

Where $X_{(1)}$ is the smallest data value (the minimum) and $X_{(n)}$ the largest data value (the maximum).

Result 1: The expected value of rank order statistics is

$$E\left(X_{(r)}\right) = \int_{-\infty}^{\infty} P\left(X_{(r)} > x\right) dx$$

Proofs for these results are often quite involved and are based on cumulative distribution functions and probability density functions. We will consider the expected value of a rank order statistic as an example:

The cumulative distribution function of the r^{th} order statistic $F_{X_{(r)}}(x)$ is given by:

$$F_{X_{(r)}}(x) = P(X_{(r)} \leq x)$$

Taking derivative with respect to x gives the probability density function as:

$$f_{X_{(r)}}(x) = \frac{d}{dx} F_{X_{(r)}}(x)$$

Then, the expected value of rank order statistics $E(X_{(r)})$ can be found by:

$$E(X_{(r)}) = \int_{-\infty}^{\infty} x f_{X_{(r)}}(x) dx$$

Result 2: Variance of Rank Order Statistics

The variance of the rank order statistics can also be derived using probability density functions, which is beyond the scope of this short text but can be found in various statistical texts.

4.3.1 Applications and Examples

1. Non-parametric Statistical Tests:

Mann-Whitney U Test: Used to determine whether there is a difference between two independent samples.

***Wilcoxon Signed-Rank Test*:** Used to test the median of a single sample or paired samples.

2. Quantile and Percentile Estimation:

Median: $X_{(0.5n)}$ (if n is even) or the average of $X_{(0.5n)}$ and $X_{(0.5n+1)}$ (if n is odd) is an estimator of the median.

3. Outlier Detection:

Identifying extremely low or high rank order statistics helps detect potential outliers in the data.

Examples

Example 1: Sample of test scores: {82, 94, 76, 88, 92}

The rank order statistics of this sample are simply these values ordered:

$X_{(1)} = 76,\ X_{(2)} = 82,\ X_{(3)} = 88,\ X_{(4)} = 92, X_{(5)} = 94$

Example 2: Application in Test

In a psychological test, scores of five participants are: $\{32, 37, 29, 45, 40\}$. We might be interested in the median score to understand the central tendency which can be found using rank order statistics.

Ordering the scores: 29,32,37,40,45

The median (or the 3rd rank order statistic) is 37.

4.3.2 A fundamental Lemma related to Rank Order Statistics

Lemma: Distribution of the r^{th} Rank Order Statistic

Statement: Let $X_1, X_2, \dots, X_n$ be independent and identically distributed random variables with a continuous cumulative distribution function $F(x)$. The probability distribution function of the r^{th} rank order statistic $X_{(r)}$ is given by:

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!\,(n-r)!}[F(x)]^{r-1}[1-F(x)]^{n-r}f(x)$$

Where $f(x)$ is the probability density function of X_i.

Proof: Start by considering the probability that $X_{(r)}$ is less than or equal to some value x. This is the probability that $r-1$ of the $X_i's$ are less than $x\ and\ n-r$ of $X_i's$ are greater than $x, which\ is\ given\ by$:

$$P\left(X_{(r)} \le x\right) = \binom{n}{r-1}[F(x)]^{r-1}[1-F(x)]^{n-r}$$

Where, $\binom{n}{r-1}$ is binomial coefficient representing the number of ways to choose $r-1$ terms out of n.

Differentiating the right-hand side using the product rule and binomial expansion, we arrive at:

$$f_{X_{(r)}}(x) = \frac{n!}{(r-1)!\,(n-r)!}[F(x)]^{r-1}[1-F(x)]^{n-r}f(x)$$

Which completes the proof.

This lemma essentially gives the distribution of the rth smallest value out of a sample of n values, assuming the data follows a known distribution with cumulative distribution function F(x) and probability density function f(x). The lemma is foundational for further analysis of order statistics, especially when deriving properties like expected values and variances for specific rank order statistics.

The proof leverages the combinatorial nature of order statistics — the multiple ways of obtaining specific ranks for the data points — and ties it to their probabilistic interpretation.

4.3.3 Joint Distribution of Two Order Statistics

Statement: Let $X_1, X_2, \dots, X_n$ be independent and identically distributed random variables with a cumulative distribution function $F(x)$. The joint probability density function of the i^{th} and j^{th} rank order statistic $X_{(i)}$ and $X_{(j)}$; $where\ i < j$ wis given by:

$$f_{X_{(i)},X_{(j)}}(x,y) = \frac{n!}{(r-1)!\,(j-i-1)!\,(n-r)!}[F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j}f(x)f(y)$$

Proof: For $x \leq y$

1. The probability that exactly $i-1$ of X_k are less than x is $\frac{n!}{(i-1)!(n-i)!}[F(x)]^{i-1}[1-F(x)]^{n-i+1}$.
2. The probability thsat exactly $j-i-1$ of X_k (not exactly less than x) are between x and y $\binom{n-i+1}{j-i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j+1}$.

3. The joint probability is the product of the above two probabilities, multiplied by the density functions f(x) and f(y) for the two specific values x and y.

Combining the above steps, we derive the stated formula for the joint distribution i.e.

$$f_{X_{(i)},X_{(j)}}(x,y) = \frac{n!}{(r-1)!\,(j-i-1)!\,(n-r)!}[F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j}f(x)f(y)$$

4.3.4 Expectation of the maximum order statistic

Statement: Given $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with a cumulative distribution function $F(x) and\ density\ function\ f(x).$. The expected value of the maximum order statistic $X_{(n)}$ wis given by

$$E\left[X_{(n)}\right] = \int_{-\infty}^{\infty} n[1-F(x)]^{n-1} f(x) x\, dx$$

Proof: the probability density function of $X_{(n)}$ is:

$$f_{X_{(n)}}(x) = n[1-F(x)]^{n-1} f(x)$$

Thus, the expected value of $X_{(n)}$ is given as

$$E\left[X_{(n)}\right] = \int_{-\infty}^{\infty} x f_{X_{(n)}}(x)\, dx$$

Substituting the pdf of largest order statistic, we get

$$E\left[X_{(n)}\right] = \int_{-\infty}^{\infty} n[1-F(x)]^{n-1} f(x) x\, dx$$

These results and their proofs showcase the intricacy of order statistics, blending combinatorics and probability theory. They form the basis for more advanced investigations in the field, such as approximations for large samples and evaluations of moments for order statistics.

4.3.5 Applications

Rank order statistics often have practical applications in day-to-day life that don't necessarily involve direct mathematical operations. Here are some non-mathematical examples related to rank order statistics:

Sports Competitions: In a track and field event, athletes might finish a race in various times. While the exact times might be of interest, often what matters most is their ranking - who came first, second, third, and so on. The athlete with the fastest time is the maximum rank order statistic, while the one with the slowest time is the minimum.

University Rankings: Every year, various organizations release rankings of universities around the world based on a multitude of factors like research output, student satisfaction, and more. While the actual scores might be complex and multifaceted, people often pay attention to the rankings: which university is 1st, which is 2nd, etc.

Job Applicant Screening: Imagine a company has interviewed 10 candidates for a position. Post-interview, the HR might rank them based on their overall performance. While the exact scores on various parameters might matter internally, they might simply offer the job to the top 1 or 2 ranked candidates.

Movie or Restaurant Rankings: Websites might rank movies or restaurants based on user reviews. While each movie or restaurant might have an average rating (like 4.2 out of 5 stars), a list might simply show them in order of rank, from highest to lowest average rating.

Elections: In certain voting systems, rather than just selecting one candidate, voters rank their candidates. The candidate with the most "first choice" votes might get elected, or there might be a more complex system where if no candidate gets over 50% of "first choice" votes, then "second choice" votes (and so on) come into play.

Beauty Pageants: Contestants are often ranked based on various rounds like talent, evening wear, Q&A, etc. In the end, while scores might be given in each round, what's most prominent is who gets crowned as the winner, the first runner-up, the second runner-up, and so on.

Book Bestseller Lists: While books might have various sales numbers, a bestseller list typically just ranks them from the best-selling to the least in that category, providing readers a ranked order statistic of book sales.

Each of these examples incorporates the essence of rank order statistics - focusing on the position or order of items rather than their actual values or scores.

4.4 Dwass' Technique

Dwass' Technique is a method for performing multiple pairwise comparisons on non-parametric data. It's a procedure used after a Kruskal-Wallis test (which is the non-parametric alternative to one-way ANOVA) has determined that there are statistically significant differences among the groups. Meyer Dwass proposed this technique to address the problem of inflating the Type I error rate when multiple comparisons are made. It's essentially a modified Wilcoxon rank-sum test (or Mann-Whitney U test) that adjusts for these multiple comparisons.

4.4.1 The Dwass' Technique

***Rank the Data*:** Just like in the Kruskal-Wallis test, all the data from all groups are ranked together, ignoring the group to which they belong. Ties are handled by assigning the average of the ranks they span.

***Pairwise Comparisons*:** For each pair of groups (A and B, for instance), the test statistic is computed as if performing a regular Wilcoxon rank-sum test.

***Adjust for Multiple Comparisons*:** This is the crucial step. Instead of using the regular critical values for the Wilcoxon rank-sum test, Dwass' method requires more extreme values to declare significance, essentially adjusting for the fact that multiple tests are being performed. The critical values can be found using tables specifically designed for Dwass' method, or more modernly, via simulation or bootstrapping methods.

Determine Significance: If the test statistic for any pair exceeds the adjusted critical value, the difference between those groups is considered statistically significant.

4.4.2 Importance

The Dwass' technique, and other multiple comparison methods, are essential because performing many tests increases the chance of finding at least one significant result just by chance (a false positive). By adjusting the criteria for significance, these methods control the familywise error rate, ensuring that the probability of one or more false positives remains at the desired significance level (typically 0.05).

4.4.3 Limitations

Power: By adjusting for multiple comparisons, these tests are inherently more conservative, reducing the chance of false positives but also potentially increasing the chance of false negatives.

Applicability: Dwass' method is designed specifically for post-hoc comparisons after a Kruskal-Wallis test. It's not suitable for other contexts or for parametric data.

Modern Alternatives: There are newer methods and techniques available for multiple comparison adjustments that might be considered more powerful or flexible than Dwass' method.

This technique is particularly useful in scenarios where non-parametric tests are appropriate due to the data not meeting the assumptions of parametric tests, and where multiple group comparisons are needed. Here are some applications and examples where Dwass' Technique could be employed:

4.4.4 Applications

Medical Research: When comparing the effects of multiple treatments or interventions on a non-normally distributed outcome, such as the number of pain-free days after different therapeutic interventions.

Ecological Studies: Comparing species diversity in various habitats or regions, especially when the data is skewed or has outliers.

***Social Sciences*:** For example, comparing median income or other skewed socio-economic indicators across different regions or groups.

***Market Research*:** When comparing customer satisfaction scores (on a non-parametric scale) across different products or services.

***Educational Studies*:** For instance, comparing the distributions of test scores among students exposed to different teaching methods, especially if scores are not normally distributed.

Examples:

***Clinical Trial*:** Suppose a new drug is being tested in three different dosages against a placebo to check its efficacy in reducing migraine occurrences. After the trial, the median number of migraines in each group over a month is recorded. Given that migraine occurrences may not follow a normal distribution, and there are multiple groups to compare, a Kruskal-Wallis test followed by Dwass' technique can be used.

***Wildlife Conservation*:** A conservationist wants to compare the number of bird species spotted in four different forest conservation areas. Given that such counts can be skewed (some rare species might only appear occasionally), the conservationist decides to employ non-parametric methods. After finding significant differences with the Kruskal-Wallis test, they use Dwass' technique for pairwise comparisons.

***Customer Survey**:* A company launches three different ad campaigns for a product and later surveys customers on their recall of the ad on a scale of 1 to 10. The scores might not be normally distributed since many people might give extreme scores (either 1 or 10). After finding a significant difference among the campaigns using a Kruskal-Wallis test, Dwass' technique is employed to find out which ad campaigns differ significantly from each other.

In each of these examples, Dwass' Technique provides a way to dive deeper into the data after an overall difference is found, helping researchers pinpoint exactly where those differences lie.

4.5 Ballot Theorem and its Generalisation

The Ballot Theorem is a classical result in combinatorial probability, which deals with the chances of one candidate always being ahead of another in a sequence of votes when the final counts are known.

Ballot Theorem Statement: Suppose two candidates, A and B, receive $a\, and\, b$ votes respectively in an election with $a > b$. If the votes are counted in random order, the probability that A is always ahead of B throughout the count is given by: leads throughout

$$P(A\ leads\ throught) = \frac{a-b}{a+b}$$

Proof (Intuitive):

Imagine a path on a coordinate plane that represents the difference in votes between A and B. Each step to the right represents a vote for A, and each step to the left represents a vote for B. The final position on the path will be $a - b$, since A has a votes and B has b votes.

For A to always be leading, the path must never touch the x-axis (because that would mean they have the same number of votes at some point). Now, consider reflecting any portion of the path that goes below the x-axis across the x-axis. This creates a bijection between paths where A leads throughout and paths where B takes the lead at least once.

The key insight is that for every path where A is leading throughout (except for the path where A gets all its votes first), there is a corresponding path where B leads at some point.

The difference in the number of such paths is exactly the paths where A gets all its votes first, which is a. Hence, the probability is: $\frac{a-b}{a+b}$

4.5.1 Generalization

A generalization of the Ballot Theorem is known as Bertrand's Ballot Problem.

***Statement*:** Given two candidates, A and B, receiving b votes respectively, where $a > b$, the probability that A is ahead of B at any randomly chosen point in the count (and not necessarily throughout) is: leads at a point P(A leads at a point)= $\frac{a-b}{a+b}$

This might seem counterintuitive, but the probability remains the same! The key is that being ahead at a random point in the count does not specify a structure to the sequence like leading throughout does.

4.5.2 Applications

The Ballot Theorem and its generalizations can be used in various fields, including economics (random walks in stock market analysis), physics (particle diffusion), and computer science (algorithm analysis).

The generalization of the Ballot theorem explores extensions to more than two candidates and other related scenarios.

The Ballot Theorem and its generalization provide fascinating insights into the behaviour of random processes. Let us explore some examples to make these concepts more tangible.

Example 1: Election Scenario

Setting: Imagine a small town where two candidates, Alice and Bob, are running for mayor. After all votes are counted, Alice has received 7 votes and Bob has received 5 votes. What is the probability that Alice was leading throughout the counting of the votes if they were counted in random order?

Using the Ballot Theorem: leads throughout

P(Alice leads throughout)= $\frac{7-5}{7+5}$= $\frac{2}{12}$

So, there's a 1/6 chance that Alice was leading throughout the count.

Example 2: Stock Market

Setting: Imagine a simplified stock market scenario. A particular stock either goes up by $1 or down by $1 every day. After 10 days, the net change in the stock's value is +$2. What is the probability that the stock was always valued higher during these 10 days compared to its starting value?

Here, we can treat "going up by $1" as a vote for Alice and "going down by $1" as a vote for Bob. After 10 days, Alice has "6 votes" and Bob has "4 votes".

P(Stock always up)= $\frac{6-4}{6+4}$=$\frac{2}{10}$

So, there's a 1/5 chance that the stock was always up relative to its starting value.

Example 3: Watching a Sports Game

Setting: Imagine a basketball game where Team A scores 9 times, and Team B scores 7 times in a match. If you tune in at a random point during the game, what's the probability that Team A is leading?

Using Bertrand's generalization: leads at a point

P(Team A leads at a point)= $\frac{9-7}{9+7} = \frac{2}{16}$

So, there's a 1/8 chance that Team A is leading at a randomly chosen point during the game.

Example 4: Board Game

Setting: Two players, Carla and Dave, are playing a board game. The game is simple: they roll dice, and based on the outcome, they either move one step forward or one step backward on the board. After 20 moves, Carla is 4 steps ahead of the starting point. What is the probability she was always ahead during the game?

This can again be mapped to the Ballot Theorem. Carla has moved forward 12 times (votes for Carla) and backward 8 times (votes for Dave).

always ahead P(Carla always ahead)= $\frac{12-8}{12+8} = \frac{4}{20}$

There is a 1/5 chance that Carla was always ahead during the game.

These examples demonstrate how the Ballot Theorem and its generalization can be applied in various scenarios, offering a glimpse into the probabilistic nature of sequences and events.

4.6 Extension and Application to Fluctuations of Sums of Random Variables

The concepts from rank order statistics and the Ballot theorem can be extended to understand the fluctuations in sums of random variables. Such an understanding is pivotal in areas like stochastic processes and time series analysis.

The Ballot Theorem and Bertrand's generalization find natural extensions in the study of random walks, especially in understanding the fluctuations of sums of random variables. Here's an introduction to this topic:

Random Walks:

Consider a simple random walk on the line, where at each step, you move one unit to the right with probability p or one unit to the left with probability $q=1-p$. This is analogous to counting votes for two candidates in the Ballot Theorem, where a vote for candidate A is a step to the right, and a vote for candidate B is a step to the left.

The position after n steps is given by the sum of those n random variables (each being +1 or -1, depending on the step direction).

Fluctuations:

Now, suppose we are interested in the fluctuations of this random walk, i.e., how the sum of these random variables behaves over time. Two questions of interest are:

What is the probability that the sum is always non-negative?

What is the average value of the sum after n steps?

Using results analogous to the Ballot Theorem, we can tackle the first question. The second question relates to the expected value of the sum of random variables.

Application: Stock Market

A simple model for stock prices is to treat daily changes in stock prices as independent random variables. A stock might go up by some amount with probability *p* or down by some amount with probability *q*.

Over *n* days, the total change in stock price is the sum of these random variables. By understanding the fluctuations in this sum: Investors can model the probability that the stock price remains above a certain level over a period (analogous to always being non-negative). Investors can predict the average change in stock price over a period.

Application: Queuing Theory

In computer networks or service centres, incoming tasks (or packets in a network) can be modelled as random variables, where each task requires a random amount of service time.The total service time required over n tasks is the sum of these random variables. Understanding its fluctuations helps in: Ensuring that queue lengths remain below a certain threshold with high probability. Predicting average waiting times or service delays.

Extension:

One of the fascinating extensions of this idea is the study of Brownian motion, which is a continuous-time version of the random walk. This stochastic process has applications in physics (particle motion in fluids), finance (option pricing), and many other fields.

In the realm of random walks and sums of random variables, understanding fluctuations is crucial. It helps predict and model a wide range of phenomena, from stock prices to queuing delays, and forms the bedrock of many areas in applied probability and statistics.

4.6.1 Foundational Results Related to Random Walks and The Fluctuations of Sums of Random Variables

1. Expected Position:

For a simple random walk where at each step we move one unit to the right with probability p or one unit to the left with probability q=1−p, the expected position

E[Sn] after n steps is: E[Sn]=n(2p−1)

Proof:

The expected value of each step (either +1 or -1) is:

E[S]=p(1)+q(−1)=2p−1

Hence, for n steps:

E[Sn]=nE[S]=n(2p−1)

2. Variance of Position:

The variance of the position Var(Sn) after n steps in the same simple random walk is:

Var(Sn)=4npq

Proof:

The variance for each step is:

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

Since S^2 is always 1 (whether you move left or right), $E[S^2]=1$. Using the earlier result, E[S]=2p−1. Plugging in:

Var(S)=1−(2p−1)^2 =4pq

For n steps:

Var(Sn)=nVar(S)=4npq

3. Reflection Principle:

If a random walk first reaches level k (k positive steps total) at time 2m, then the number of paths that reach level k without ever being negative before time 2m is equal to the number of paths that reach level k and then go to level k−1 at time 2m.

4. Arcsine Law:

For a symmetric random walk (where p=q= 1/2), the probability that the walk stays non-negative up to time $2n$ is $\frac{2}{\pi}\ arcsin\left(\sqrt{\frac{n}{2n}}\right)$

5. Law of the Iterated Logarithm:

This result describes the magnitude of the fluctuations of a random walk. Specifically, for a symmetric simple random walk:

$$\log_{sup_{n\to\infty}} \frac{S_n}{\sqrt{2nlog\log n}} = 1$$

with probability 1.

This last result implies that while a random walk will oscillate and have fluctuations, it won't deviate too wildly from its mean when properly scaled.

These results give a glimpse into the rich tapestry of theorems and insights related to random walks and the behaviour of sums of random variables. They have profound implications in areas ranging from finance to physics.

4.7 Summary

In this comprehensive unit, we began by exploring Rank Order Statistics, which serves as a cornerstone for non-parametric statistical analysis. This method assigns ranks to data points in a dataset, allowing for an analysis that is less influenced by outliers or specific data distributions. The r^{th} rank order statistic, for instance, corresponds to the r^{th} smallest value in a given dataset. This approach's utility becomes especially clear in the context of non-parametric tests, where there is no assumption of a particular underlying data distribution.

Our journey then led us to Dwass' Technique, a specialized non-parametric method tailored for making multiple comparisons of group medians. Especially beneficial when working with three or more independent samples and in situations where sample sizes are small, Dwass' Technique elegantly extends the foundation of the Wilcoxon rank-sum test to encompass multiple comparisons. This ensures that the overall type I error rate is consistently maintained, offering a robust methodology for various analyses.

Further deepening our exploration, we delved into the Ballot Theorem and its subsequent generalization by Bertrand. This theorem offers a fascinating glimpse into the probabilities associated with sequential processes. Specifically, in the case of two candidates, A and B, receiving a and b votes respectively (with a>b), the Ballot Theorem states that the odds of candidate A always leading during a random vote tally is $\frac{a-b}{a+b}$. Bertrand's broader perspective posits that during any random moment within the vote counting, the likelihood of A leading remains the same. While its roots might lie in electoral processes, the applications of this theorem stretch far and wide, encompassing scenarios from stock market shifts to evolving sports scores.

Lastly, the unit pivoted to the realm of Random Walks, providing a window into the fluctuations of sums of random variables. A random walk can be envisaged as a sequence of steps, each determined by a random variable—often represented as a move to the right or left. For a simplistic random walk model, where probabilities p and q dictate right and left moves respectively, it emerges that the expected position after n steps can be described as $n(2p-1)$. Beyond mere academic intrigue, the significance of understanding these fluctuations resonates in real-world phenomena ranging from stock market behaviors, the intricate dance of particles in physics, to the challenges posed in queuing theory. An especially riveting extension of this concept is Brownian motion—a continuous-time interpretation of the random walk, which finds relevance in diverse fields.

4.8 Self-Assessment Questions

Here are some self-assessment questions for the unit on Rank Order Statistics, Dwass' Technique, Ballot Theorem, and the fluctuations of sums of random variables:

Rank Order Statistics:

1. Define rank order statistics.
2. How is the rth rank order statistic represented mathematically?
3. What is the significance of rank order statistics in non-parametric statistical tests?

Dwass' Technique:

4. Explain Dwass' Technique in your own words.
5. How does Dwass' Technique help in multiple comparisons?

Ballot Theorem and its Generalisation:

6. State the Ballot Theorem.
7. How does Bertrand's Ballot Problem generalize the original Ballot Theorem?
8. Provide an example where the Ballot Theorem can be applied in a real-world scenario.

Extension and Application to Fluctuations of Sums of Random Variables:

9. Define a simple random walk. How can it be modeled as a sum of random variables?
10. How does the expected position change in a random walk after n steps if the probability of moving to the right is p and to the left is q?
11. What is the significance of the Reflection Principle in the context of random walks?
12. Describe the Arcsine Law and its relevance to random walks.

Practical Application:

13. In an election, candidate A receives 60 votes and candidate B receives 40 votes. If the votes are counted in a random order, what is the probability that candidate A always leads throughout the count?
14. Consider a simple random walk where you move one step forward with a probability of 0.7 and one step backward with a probability of 0.3. What is the expected position after 100 steps?
15. Given the importance of rank order statistics in non-parametric tests, why might one opt for a non-parametric test over a parametric test in statistical analysis?

16. True or False:
17. In the Ballot Theorem, if two candidates A and B receive equal votes, then the probability that A leads throughout the counting process is 0.5.
18. Dwass' Technique can only be used in pairwise comparisons.
19. The fluctuations in the sums of random variables have no applications in the stock market.

These questions should provide a comprehensive assessment of the learner's understanding of the unit. Properly answering them should demonstrate a solid grasp of the concepts discussed.

1. Witte, R. S., & Witte, J. S. (2019). Statistics (11th ed.). Hoboken, NJ: Wiley.
2. Feller, W. (1968). An introduction to probability theory and its applications (Vol. 1, 3rd ed.). New York: John Wiley & Sons.
3. Bertrand, J. (1887). Calcul des probabilités. Paris: Gauthier-Villars.

4.9 References

- Hollander, M., & Wolfe, D. A. (1999). Nonparametric Statistical Methods (2nd ed.). New York: Wiley-Interscience.
- Dwass, M. (1960). Multiple comparison procedures for discrete test statistics. Journal of the Royal Statistical Society. Series B (Methodological), 22(2), 189-194.
- Grinstead, C. M., & Snell, J. L. (1997). Introduction to probability. Providence, RI: American Mathematical Society.
- Lawler, G. F., & Limic, V. (2010). Random Walk: A Modern Introduction. Cambridge: Cambridge University Press.
- Spitzer, F. (1976). Brownian motion and classical potential theory. New York: Academic Press.
- DeGroot, M. H., & Schervish, M. J. (2012). Probability and statistics (4th ed.). Boston: Pearson.
- Spiegelhalter, D. (2019). The Art of Statistics: Learning from Data. New York: Basic Books.
- Wasserman, L. (2004). All of Statistics: A Concise Course in Statistical Inference. New York: Springer.

4.10 Further Reading

- David, H. A. (1981). Order Statistics. Wiley.
- Dwass, M. (1960). Some k-sample rank-order tests. In Contributions to Probability and Statistics (pp. 198-202). Stanford University Press.
- Bertrand, J. (1887). Calcul des probabilités.
- Feller, W. (1968). An Introduction to Probability Theory and Its Applications (Vol. 1). Wiley.

U.P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 202N/ MASTAT –202N

Non Parametrics

Block: 2 Sequential Analysis

Unit – 5 : Sequential Tests

Unit – 6 : Sequential Estimation

Course Design Committee

Dr. Ashutosh Gupta **Chairman**
Director, School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Anup Chaturvedi **Member**
Ex. Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. S. Lalitha **Member**
Head, Department of Statistics
University of Allahabad, Prayagraj

Prof. Himanshu Pandey **Member**
Department of Statistics,
D. D. U. Gorakhpur University, Gorakhpur

Prof. Shruti **Member-Secretary**
Professor, School of Sciences
U.P. Rajarshi Tandon Open University, Prayagraj

Course Preparation Committee

Prof. Shashi Bhushan **Writer**
Department of Statistics
University of Lucknow, Lucknow

Prof. S. K. Pandey **Editor**
Department of Statistics
University of Lucknow, Lucknow

Prof. Shruti **Course Coordinator**
School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 202N/ MASTAT – 202N **NON PARAMETRICS**

First Edition: ***July 2023***
ISBN :

Printed By:

Block & Units Introduction

The ***Block - 2 – Sequential Analysis*** is the second block with two units.

In ***Unit – 5 – Sequential Tests*** is discussed with SPRT and its properties, Wald's Fundamental identity, OC and ASN functions, Wald's equation.

In ***Unit – 6 – Sequential Estimation*** has been discussed, Cramer Rao Inequality of sequential estimation, Stein's two stage procedure.

At the end of every unit the summary, self assessment questions and further readings are given.

References & Further Readings

- Bagdonavicius, V., Kruopis, J., Nikulin, M.S. (2011). "Non-parametric tests for complete data", ISTE & WILEY: London & Hoboken. ISBN 978-1-84821-269-5.
- Corder, G. W.; Foreman, D. I. (2014). Nonparametric Statistics: A Step-by-Step Approach. Wiley. ISBN 978-1118840313.
- Gibbons, Jean Dickinson; Chakraborti, Subhabrata (2003). Nonparametric Statistical Inference, 4th Ed. CRC Press. ISBN 0-8247-4052-1.
- Hettmansperger, T. P.; McKean, J. W. (1998). Robust Nonparametric Statistical Methods. Kendall's Library of Statistics. Vol. 5 (First ed.). London: Edward Arnold. New York: John Wiley & Sons. ISBN 0-340-54937-8. MR 1604954. also ISBN 0-471-19479-4.
- Hollander M., Wolfe D.A., Chicken E. (2014). Nonparametric Statistical Methods, John Wiley & Sons.
- Sheskin, David J. (2003) Handbook of Parametric and Nonparametric Statistical Procedures. CRC Press. ISBN 1-58488-440-1
- Wasserman, Larry (2007). All of Nonparametric Statistics, Springer. ISBN 0-387-25145-6.

UNIT: 5 SEQUENTIAL TESTS

Structure

5.1 Introduction

In classical Inference the size of the random sample to be drawn from the distribution is a fixed number. We shall consider the situation where the sample number is not fixed but is itself a random variable dependent on the observations. A procedure of making inference about the distribution of one or more variables in which the size of the sample is a random variable is called *sequential procedure.* The principal feature of such a procedure is a sampling scheme which lays down a rule under which one decides at each stage of the sampling whether to stop or to continue sampling, this decision being taken in the light of the observations already obtained.

5.2 Objective

The objective of this unit is to provide a basic understanding of concepts related to Sequential Tests. The concepts of SPRT and its properties, Wald's Fundamental identity, OC and ASN functions, Wald's equation should be clear after reading this material.

5.3 Sequential Procedure

In classical Inference the size of the random sample to be drawn from the distribution is a fixed number. We shall consider the situation where the sample number is not fixed but is itself a random variable dependent on the observations. A procedure of making inference about the distribution of one or more variables in which the size of the sample is a random variable is called *sequential procedure.* The principal feature of such a procedure is a sampling scheme which lays down a rule under which one decides at each stage of the sampling whether to stop or to continue sampling, this decision being taken in the light of the observations already obtained.

Suppose we want to judge whether a coin is unbiased or not. In the usual procedure, we toss the coin a fixed number of times, say n times, and note in how many throws head appears. However, we may as well fix a number, say k, and then go on tossing the coin until k heads appear. In this case, attention is focused on the number of throws needed to get k heads. This number of throws is now a random variable, unlike in the usual type of sampling, and it may take any integral value greater than or equal to k with positive probabilities. Hence the procedure used here is of the sequential kind.

5.3.1 Two Aspects of a Sequential Procedure

A general sequential procedure has two aspects: (I) a stopping rule or a rule which tells us when to stop sampling and (II) an action rule, which tells us what type of inference (or decision) to make after sampling has been stopped.

To visualize the stopping rule, note that we have in view a sequence of random variables, say $X_1, X_2, \dots \dots \dots$, such that for every $m = 1, 2, \dots$ the distribution of $X_m = (X_1, X_2, \dots., X_m)'$ is determined by the value(s) of the parameter(s) θ, where $\theta \epsilon \Theta$, the parameter space.

The mth observation, x_m, is an observation on X_m (which may itself be a vector random variable). Let us write $\mathbb{x}_{\boldsymbol{m}} = (x_1, x_2, \dots., x_m)'$ and let $\varkappa_m$ be the set of all possible values of X_m.

A stopping rule is then equivalent to a specification of sets $A_m \subset \varkappa_m$, for $m = 1, 2, \dots \dots$, such that sampling is to be stopped as soon as $\mathbb{x}_m \epsilon\, A_m$.

Let N be the number of observations actually taken according to a sequential procedure. Then $N = m$ if and only if $\mathbb{x}_j \notin A_j, j = 1, 2, \dots \dots, m-1$, and $\mathbb{x}_m \epsilon\, A_m$.

If the set S_m is defined as

$$S_m = \{\mathbb{x}_m | x_j \notin A_j, j = 1, 2, \dots, (m-1); \mathbb{x}_m \epsilon\, A_m\}; \quad (1)$$

Then we might as well say

$$N = m \Leftrightarrow \mathbb{x}_m \epsilon\, S_m. \quad (2)$$

It may be seen that $S_1, S_2, \dots$. are mutually disjoint sets.

As regards the action rule, it specifies, which decision to take if sampling has been stopped after m observations say and the observed values are $\mathbb{x}_m$.

5.4 Sequential Testing of Hypotheses and SPRT

Suppose our problem is to test a hypothesis H_0 about θ. Then the action rule will dictate which of the two actions, a_0 (denoting acceptance of H_0) and a_1 (denoting rejection of H_0), is to be taken after sampling has been stopped in accordance with the stopping rule.

The action rule may be defined in terms of sets

$A_{0,m}$, and $A_{1,m}$ $(m = 1, 2, \dots)$,

Where $A_{0,m}, A_{1,m}$ are disjoint subsets of $\varkappa_m$. The action rule will be as follows:

(1) Take action a_0 if $\mathbb{x}_m \epsilon A_0, m$

and (2) take action a_1 if $\mathbb{x}_m \epsilon\, A_1, m$.

Note that for a given stopping rule, which specifies S_m, an action is taken at the mth stage iff $\mathbb{x}_m \epsilon\, S_m$. Hence the action rule could be defined only for $A_{0,m} \subset S_m$. However, since the action does not depend on the stopping rule, it is better to define $A_{0,m}$ independently of S_m.

For a given sequential test procedure, let A_i denote the event that the observations result in action a_i being taken $(i = 0, 1)$. Now, the action a_i is taken after m observations iff $\mathbb{x}_m \epsilon\, S_m \cap A_{i,m}$, which event is equivalent to $[N = m] \cap A_i = S_m * \cap A_i$.

Hence

$$P_\theta(A_i) = \Sigma_{m=1}^{\infty} P_\theta([N = m] \cap A_i)$$

$$= \Sigma_{m=1}^{\infty} P_\theta[X_m \epsilon S_m \cap A_i, m], \text{ for } i = 0, 1 \quad (3)$$

We exclude from the above sums the case $m = \infty$, for in that case no action is taken. We have

$$P_\theta(A_0) + P_\theta(A_1) \leq 1 \tag{4}$$

and equality if the experiment terminates with probability 1, i.e. if
$P_\theta[N < \infty] = 1$.

5.5 OC and ASN Function

In order to judge the merits of sequential tests, we use the OC function and the ASN function.

Operating characteristics (OC) function of a test the probability of the hypothesis H_0 being accepted when θ is the true value of the parameter, regarded as a function of θ, denoted by $L(\theta)$ such that

$$L(\theta) = P_\theta(A_0) \tag{5}$$

The OC function is closely related to the notion of the power function in the function is the probability of rejecting H_0 when θ is the true value.

An OC function is considered the more favorable the higher the value of $L(\theta)$ for θ consistent with H_0 and the lower the value of $L(\theta)$ for θ not consistent with H_0.

The number of observations N required by a sequential procedure to reach a decision is not predetermined but is a random variable. A sequential procedure preferable if it requires a small value of N on the average. This average value of N is called the average sample number (or ASN) of sequential procedure denoted by $E_\theta(N)$. The smaller the value of $E_\theta(N)$ the better is the sequential procedure. The OC function describes how well the procedure achieves its objective of making correct decisions, while the ASN function represents the price one has to pay to reach a decision, in terms of the number of observations required by the test.

5.6 Wald's SPR Test

The sequential probability ratio (SPR) test was developed by Wald. Suppose the random variables $X_1, X_2, \ldots \ldots$ are independently and identically distributed with the common p.m.f. or p.d.f. f_θ. Further suppose that there are just two values of θ interest to us, say θ_0 and θ_1 and that we have two hypotheses $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$.

For any positive integer m, the probability or probability density that the observations $x_1, x_2, \ldots\ldots, x_m$ are obtained is given by

$$f_{0,m} = \Pi_{i=1}^{m} f_{\theta_0}(x_i) \tag{6}$$

When H_0 is true and by

$$f_{1,m} = \Pi_{i=1}^{m} f_{\theta_1}(x_i) \tag{7}$$

When H_1 is true.

The SPR test for testing H_0 against H_1 is defined in terms of the ratio $\frac{f_{1,m}}{f_{0,m}}$ as follows: Specify two constants A and B such that

$$0 < B < 1 < A \tag{8}$$

Continue taking observations as long as

$$B < \frac{f_{1,m}}{f_{0,m}} < A \tag{9}$$

Stop taking observations as soon as one of inequalities is violated and

$$\text{Accept } H_0, \frac{f_{1,m}}{f_{0,m}} \leq B \tag{10}$$

$$\text{And reject } H_0 \ \frac{f_{1,m}}{f_{0,m}} \geq A. \tag{11}$$

It is convenient to deal with the logarithm of the ratio $f_{1,m}/f_{0,m}$ than with the ratio itself. Therefore,

$$\log\left(\frac{f_{1,m}}{f_{0,m}}\right) = \Sigma_{i=1}^{m} \log\left(\frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)}\right)$$

$$= \Sigma_{i=1}^{m} z_i \tag{12}$$

Where

$$z_i = \log\left(\frac{f_{\theta_1}(x_i)}{f_{(\theta_0)(x_i)}}\right)$$

And sums are easier to deal with than products.

Using the quantities z_i $(i = 1, 2, \ldots.)$, the procedure is continue taking observations as long as

$$logB < \Sigma_{i=1}^{m} \tag{13}$$

Stop taking observations as soon as one of the inequalities is violated and

$$\text{Accept } H_0 \text{ if } \Sigma_{i=1}^{m} z_i \leq logB \tag{14}$$

And reject H_0 if $\Sigma_{i=1}^{m} z_i \geq logA$

Determination of A and B

Now, we need to determine the constant A and B in the SPR test for the event that sampling is stopped exactly after m observations,

$$S_m = \left\{ \mathbb{x}_m | B < \frac{f_{1,j}}{f_{0,j}} < A, j = 1, 2, \dots \dots, (m-1); \frac{f_{1,m}}{f_{0,m}} \leq B \; or \geq A \right\} \tag{15}$$

Also, for the event that sampling is terminated at the mth stage with the acceptance of H_0,

$$A_{0,m} = \left\{ \mathbb{x}_m | B < \frac{f_{1,j}}{f_{0,j}} < A, j = 1, 2, \dots \dots, (m-1); \frac{f_{1,m}}{f_{0,m}} \leq B \; or \geq A \right\} \tag{16}$$

And, for the event that experiment is terminated at the mth stage with the rejection of H_0

$$A_{1,m} = \left\{ X_m | B < \frac{f_{1,j}}{f_{0,j}} < A, j = 1, 2, \dots \dots, (m-1); \frac{f_{1,m}}{f_{0,m}} \geq A \right\} \tag{17}$$

Clearly, $A_{0,m}$ and $A_{1,m}$ from a portion of S_m

If, we denote by A_0 the event that ultimately H_0 is accepted, then we have

$$P_{\theta_0}(A_0) = \Sigma_{1 \leq m \leq \infty} P_{\theta_0}(A_{0,m})$$

$$= \Sigma_{1 \leq m \leq \infty}{}_{A_{0,m}}\int f_{0,m} dx_m$$

$$\geq \Sigma_{1 \leq m \leq \infty}{}_{A_{0,m}}\int \frac{1}{B} f_{1,m} dx_m = \frac{1}{B} \Sigma_{1 \leq m < \infty} P_{\theta_1}(A_{0,m}) = \frac{1}{B} . P_{\theta_1}(A_0),$$

$$\text{i.e. } P_{\theta_0}(A_0) \geq \tfrac{1}{B} . P_{\theta_1}(A_0) \tag{18}$$

Similarly, denoting by A_1 the event that ultimately accepted,

$$P_{\theta_0}(A_1) = \Sigma_{1 \leq m < \infty} P_{\theta_0}(A_{1,m})$$

$$= \Sigma_{1 \leq m < \infty} \int_{A_{1,m}} f_{0,m} dx_m \qquad using \text{ (17)}$$

$$\leq \Sigma_{1 \leq m < \infty} \int_{A_{1,m}} \frac{1}{A} f_{1,m} dx_m = \frac{1}{A} \Sigma_{1 \leq m < \infty} P_{\theta_1}(A_{1,m}) = \frac{1}{A} P_{\theta_1}(A_1)$$

$$\text{i.e. } P_{\theta_0}(A_1) \leq \tfrac{1}{A} P_{\theta_1}(A_1) \tag{19}$$

from (18) and (19) the upper bounds to the probabilities of wrong decisions

$P_{\theta_1}(A_0) \leq B$ (ignoring $P_{\theta_0}(A_0) \leq 1$)

$P_{\theta_0}(A_1) \leq \frac{1}{A}$ (ignoring $P_{\theta_1}(A_1) \leq 1$)

Also, from (18) and (19), we have

$$1 - P_{\theta_0}(A_1) \geq \frac{1}{B} P_{\theta_1}(A_0) \qquad (20)$$

$$\text{And } P_{\theta_0}(A_1) \leq \frac{1}{A}\left[1 - P_{\theta_1}(A_0)\right] \qquad (21)$$

Let

$\alpha' = P(Rejecting\ H_0|H_0) = P_{\theta_0}(A_1)$ and $\beta' = P(accepting\ H_o|H_1\ ) = P_{\theta_1}(A_0)$, so that (20) and (21) give

$$B \geq \frac{\beta'}{1 - \alpha'} \qquad (21a)$$

$$\text{And } A \leq \frac{1-\beta'}{\alpha'} \qquad (21b)$$

$$\text{Or } \alpha' + \frac{1}{B}.\beta' \leq 1 \qquad (21c)$$

$$\text{And } \alpha' + \frac{1}{A}.\beta' \leq \frac{1}{A}$$

Suppose consecutive values of $f_{1,m}/f_{0,m}$ do not differ too much from each other. It will then be possible to replace the above inequalities by equalities.

To examine what the consequences of this substitution will be, let us define α and β by the equations

$$\alpha + \frac{1}{B}.\beta = 1$$

$$\text{And } \alpha + \frac{1}{A}.\beta = \frac{1}{A}$$

Subtracting these, we have

$$\left(\frac{1}{B} - \frac{1}{A}\right)\beta = 1 - \frac{1}{A},$$

or $$\beta = \frac{B(A-1)}{A-B},$$

and $$\alpha = 1 - \frac{A-1}{A-B} = \frac{1-B}{A-B}.$$

In other words,

$$B = \frac{\beta}{1-\alpha}, \quad A = \frac{1-\beta}{\alpha}. \tag{22}$$

Equation (22), determine the constants B and A in terms of the error probabilities, α and β. It is tacitly assumed that $\alpha < \frac{1}{2}, \beta < \frac{1}{2}$. The actual error probabilities α' and β', resulting from the use of constants so determined will, however, be different, although the differences are not likely to be serious.

From (21a) and (21b),

$$\left.\begin{array}{r} \beta' \leq \dfrac{\beta}{1-\alpha}(1-\alpha') \\ and \;\; \alpha' \leq \dfrac{\alpha}{1-\beta}(1-\beta') \end{array}\right\} \tag{23}$$

Or

$$\left.\begin{array}{r} (1-\alpha)\beta' \leq \beta(1-\alpha') \\ and \;\; (1-\beta)\alpha' \leq \alpha(1-\beta') \end{array}\right\} \tag{24}$$

Adding these two inequalities, we get

$$\alpha' + \beta' \leq \alpha + \beta \tag{25}$$

Further,

$$\alpha' \leq \frac{\alpha}{1-\beta} \tag{26}$$

$$and\; \beta' \leq \frac{\beta}{1-\alpha} \tag{27}$$

As is obvious from (23).

The chosen values of α and β will usually be small in applications of the sequential procedure. Quite often they will vary between 0.01 and 0.05.

In other words, for all practical purposes one may take

$$A = \left(1 - \frac{\beta}{\alpha}\right), B = \left(\frac{\beta}{1-\alpha}\right).$$

When the prescribed levels of the probabilities of wrong decision are α andβ.

Example: Let X be a discrete random variable having the p.m.f.

$$f_\theta(x) = \begin{cases} \theta^x (1-\theta)^{1-x} & if \;\; x = 0, 1 \\ 0 & otherwise \end{cases}$$

In other words,

$$X = \begin{cases} 1 & with\ probability\ \theta \\ 0 & with\ probability\ 1 - \theta, \end{cases}$$

Where $0 < \theta < 1$.

If $X_1, X_2, \ldots.$ are independent and identically distributed random variables having the same distributed X has, then for testing $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$, we have

$$f_{1,m} = \theta_1^{s_m} (1 - \theta_1)^{m - s_m}$$

And

$$f_{0,m} = \theta_0^{s_m} (1 - \theta_0)^{m - s_m}$$

Where $\qquad s_m = \Sigma_{i=1}^m x_i;$

i.e. the number of 1's among the first $m$observations. Thus

$$\frac{f_{1,m}}{f_{0,m}} = \left(\frac{\theta_1}{\theta_0}\right)^{s_m} \left(\frac{1-\theta_1}{1-\theta_0}\right)^{m-s_m}$$

Or $\log\left(\frac{f_{1,m}}{f_{0,m}}\right) = s_m . \log\left(\frac{\theta_1}{\theta_0}\right) + (m - s_m). \log\left(\frac{1-\theta_1}{1-\theta_0}\right)$ (28)

Hence in case the prescribed levels of the probabilities of wrong decision are α and β, the SPR test procedure will be as follows:

Continue taking observations as long as

$$\log\left(\frac{\beta}{1-\alpha}\right) < \log\left(\frac{f_{1,m}}{f_{0,m}}\right) < \log\left(\frac{1-\beta}{\alpha}\right)$$

Stop taking further observations as soon as one of the inequalities is violated, and

$$\text{Accept } H_0 \text{ if } \log\left(\frac{f_{1,m}}{f_{0,m}}\right) \leq \log\left(\frac{\beta}{1-\alpha}\right)$$

$$\text{reject } H_0 \text{ if } \log\left(\frac{f_{1,m}}{f_{0,m}}\right) \geq \log\left(\frac{1-\beta}{\alpha}\right)$$

By virtue of (28), the procedure may be laid down as follows:

Continue taking observations as long as

$$a_m < s_m < r_m$$

Stop taking further observations as soon as one of inequalities is violated, and

$$\text{Accept } H_0 \text{ if } s_m \leq a_m$$

$$\text{And reject } H_0 \text{ if } s_m \geq r_m$$

$$\text{Here } a_m = \frac{\log\left(\frac{\beta}{1-\alpha}\right)}{\log\left(\frac{\theta_1}{\theta_0}\right)-\log\left(\frac{1-\theta_1}{1-\theta_0}\right)} + m\frac{\log\left(\frac{1-\theta_0}{1-\theta_1}\right)}{\log\left(\frac{\theta_1}{\theta_0}\right)-\log\left(\frac{1-\theta_1}{1-\theta_0}\right)} \tag{29}$$

$$and\ r_m = \frac{\log\left(\frac{1-\beta}{\alpha}\right)}{\log\left(\frac{\theta_1}{\theta_0}\right)-\log\left(\frac{1-\theta_1}{1-\theta_0}\right)} + m\frac{\log\left(\frac{1-\theta_0}{1-\theta_1}\right)}{\log\left(\frac{\theta_1}{\theta_0}\right)-\log\left(\frac{1-\theta_1}{1-\theta_0}\right)} \tag{30}$$

Example: Suppose the random variables $X_1, X2$ ... are independently and identically distributed $N(\theta, \sigma^2)$, with unknown θ and known σ^2.Suppose we want to test $H_0: \theta = \theta_0$ against$H_1: \theta = \theta_1$.

Here

$$f_{1,m} = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^m} \exp\left[-\frac{\Sigma_{i=1}^m (x_i - \theta_1)^2}{2\sigma^2}\right]$$

And

$$f_{0,m} = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^m} \exp\left[-\frac{\Sigma_{i=1}^m (x_i - \theta_0)^2}{2\sigma^2}\right]$$

So that, since

$$\Sigma_{i=1}^m (x_i - \theta)^2 = m(\bar{x}_m - \theta)^2 + \Sigma_{i=1}^m (x_i - \bar{x}_m)^2$$

$$= m(\bar{x}_m^2 - 2\theta\ \bar{x}_m + \theta^2) + \Sigma_{i=1}^m (x_i - \bar{x}_m)^2$$

$$\frac{f_{1,m}}{f_{0,m}} = \exp\left[\frac{1}{2\sigma^2}\{\Sigma_{i=1}^m (x_i - \theta_0)^2 - \Sigma_{i=1}^m (x_i - \theta_1)^2\}\right]$$

$$= \exp\left[\frac{m}{\sigma^2}\ \ \bar{x}_m(\theta_1 - \theta_0) - \frac{m}{2\sigma^2}(\theta_1^2 - \theta_0^2)\right]$$

And $\log\left(\frac{f_{1,m}}{f_{0,m}}\right) = \frac{1}{\sigma^2}\Sigma_{i=1}^m x_i(\theta_1 - \theta_0) - \frac{m}{2\sigma^2}(\theta_1^2 - \theta_0^2)$

$$= \frac{\theta_1 - \theta_0}{\sigma^2}\left[s_m - \frac{m(\theta_1 + \theta_0)}{2}\right] \quad (31)$$

Where $\qquad s_m = \Sigma_{i=1}^m x_i.$

SPR test procedure may be defined as follows:

Continue taking observations as long as

$$\frac{\sigma^2}{\theta_1 - \theta_0} log \frac{\beta}{1 - \alpha} + \frac{m}{2}(\theta_1 + \theta_0) < s_m < \frac{\sigma^2}{\theta_1 - \theta_0}.\log\frac{1 - \beta}{\alpha} + \frac{m}{2}(\theta_1 + \theta_0).$$

Stop taking observations as soon as one of the inequalities is violated, and

Accept H_0 if $s_m \leq \frac{\sigma^2}{\theta_1 - \theta_0}.\log\frac{\beta}{1-\alpha} + \frac{m}{2}(\theta_1 + \theta_0)$

reject H_0 if $s_m \geq \frac{\sigma^2}{\theta_1 - \theta_0}.\log\frac{1-\beta}{\alpha} + \frac{m}{2}(\theta_1 + \theta_0)$.

5.7 Wald's Fudamental Identity and Equation

The sample number N in a sequential procedure is a random variable that can take any positive integral value. If the probability for N being finite be less than 1, it will mean that the procedure may not terminate, i.e., no decision may be reached in following this procedure. However, that for the SPR test this is not the case.

Theorem: Let $Z_1, Z_2, ...$ be independently and identically distributed random variables such that

$$P[Z_1 = 0] < 1$$

Further, let a and b be real numbers such that $b < a$, and let N denote the least positive integer m such that

$$\Sigma_{i=1}^{m} Z_i \leq b \text{ or } \Sigma_{i=1}^{m} Z_i \geq a$$

Then

(i) $$P[N < \infty] = 1$$

(ii) $$E(N^k) < \infty \text{ for } k = 1, 2, \ldots$$

(iii) $$E(e^{tN}) < \infty, \text{ for some } t > 0$$

Proof: Let

$$Q_s = P[|\Sigma_{i=1}^{s} Z_i| < a - b]$$

We shall show that s can be so chosen that $Q_s < 1$.

By virtue of our assumption regarding the $Z's$,

$P[Z_1 < 0] > 0$ and / or $P[Z_1 > 0] > 0$.

Assume, without any loss of generality, that

$$P[Z_1 > 0] > 0$$

Then by choosing a sufficient large s, we can have

$$P\left[Z_1 > \frac{a-b}{s}\right] > 0$$

For such an s,

$$1 - Q_s = P[|\Sigma_{i=1}^{s} Z_i \geq a - b|]$$

$$\geq P[\Sigma_{i=1}^{s} Z_i \geq a - b]$$

$$\geq P[Z_i \geq \frac{a-b}{s}, \qquad all\ i = 1, 2, \ldots., s]$$

$$= \Pi_{i=1}^{s} P\left[Z_i \geq \frac{a-b}{s}\right] \text{ since the } Z's \text{ are mutually independent.}$$

$$= \left\{P\left[Z_i \geq \frac{a-b}{s}\right]\right\}^s \text{since the } Z's \text{ are identically distributed.}$$

$$> 0$$

$\Rightarrow$ $$Q_s < 1.$$

We may now proceed to prove the theorem.

(i) Let us take fixed s such that $Q_s < 1$. For any positive integer r , we have

$$P[N > r + s] = P\left[b < \Sigma_{i=1}^{j} Z_i < a, \quad all \quad j = 1,2, \dots, r + s\right]$$

$$\leq P\left[b < \Sigma_{i=1}^{j} Z_i < a, \quad all \quad j = 1,2, \dots, r; b < \Sigma_{i=1}^{r+s} Z_i < a\right]$$

$$\leq P\left[b < \Sigma_{i=1}^{j} Z_i < a, \quad all \quad j = 1,2, \dots, r; b - a < \Sigma_{i=r+1}^{r+s} Z_i < a - b\right]$$

$$= P\left[b < \Sigma_{i=1}^{j} Z_i < a, \quad all \quad j = 1,2, \dots, r\right] * P[b - a < \Sigma_{i=r+1}^{r+s} Z_i < a - b]$$

$$= P\left[b < \Sigma_{i=1}^{j} Z_i < a, \quad all \quad j = 1,2, \dots, r\right] * P[b - a < \Sigma_{i=1}^{r+s} Z_i < a - b]$$

$$= P[N > r]Q_s.$$

Hence for any positive integer r, we have

$$P[N > r + s] \leq P[N > r]Q_s.$$

Now, putting $r = (k - 1)s$, where $k \geq 2$,

$$P[N > ks] \leq P[N > (k - 1)s]Q_s$$

$$\leq P[N > (k - 2)s]Q_s^2$$

$$\leq P[N > s]Q_s^{k-1}$$

$$\leq Q_s^{k-1}$$

$\Rightarrow$ for $k \geq 1$

$$P[N > ks] \leq Q_s^{k-1}$$

Again,

$$Q_s < 1 \Rightarrow \lim_{k \to \infty} Q_s^{k-1} = 0,$$

So that

$$\lim_{k \to \infty} P[N > m] = 0$$

$$\Rightarrow \quad P(N < \infty) = 1$$

Let us first prove the part (iii),

(iii) Assume $t > 0$. Then

$$E(e^{tN}) = \Sigma_{m=1}^{\infty} P[N = m]e^{tm}$$

$$= \Sigma_{m=1}^{s} P[N = m]e^{tm} + \Sigma_{m=s+1}^{2s} P(N = m)e^{tm}$$

$$+ \cdots$$

$$\leq P[0 < N \leq s]e^{st} + P[s < N \leq 2s]e^{2st}$$

$$+ \cdots$$

$$= \Sigma_{k=0}^{\infty} P[ks < N \leq (k+1)s]e^{(k+1)st}$$

$$\leq \Sigma_{k=0}^{\infty} P[N > ks]e^{(k+1)st}$$

$$\leq e^{st} + \Sigma_{k=1}^{\infty} Q_s^{k-1} e^{(k+1)st}$$

$$Q_s e^{st} < 1 \Rightarrow \Sigma_{k=1}^{\infty} (Q_s e^{st})^k < \infty$$

But by a suitable choice of t (i.e. by taking t small enough), we can make $Q_s e^{st} < 1$ because $Q_s < 1$. This means that, for such that a t,

$$E(e^{tN}) < \infty$$

(ii) This result follows from (iii).

$$\frac{x^k}{k!} < e^x \text{ if } x > 0 \qquad | \quad e^x = \frac{\Sigma_{k=0}^{\infty} x^k}{x!}$$

$$\Rightarrow \frac{E(tN)^k}{k!} \leq E(e^{tN}) < \infty \text{ for some } t > 0$$

$\Rightarrow E(N^k) < \infty$

5.7.1 OC Function of the SPR Test

The SPR procedure is a test for the simple hypothesis $H_0: \theta = \theta_0$ against the simple alternative $H_0: \theta = \theta_1$. However, it may be that the true value of θ is neither θ_0 nor θ_1. Hence it is of interest to study the nature of the OC function $L(\theta)$ of the test over the whole parameter space.

Lemma 1. Let Z be a random variable with distribution function F. (I) Let $E(e^{tZ})$ exist for all real t and be denoted by $\phi(t)$ and (II) let $P[Z > 0] > 0, P[Z < 0] > 0$. Then we have

(A) $E(|Z|^k) < \infty$ for all $k = 1, 2, \ldots \ldots;$

(B) The equation $\phi(t) = 1$ has the only real root $t = 0$ if $E(Z) = 0$ and exactly one real root $h \neq 0$ if $E(Z) \neq 0$ such that h and $E(Z)$ are of opposite signs.

Proof: We first show that under condition (II), derivatives of $\phi(t)$ of all orders exist and that

$$\phi^{(k)}(t) = \int_{-\infty}^{\infty} z^k e^{tZ} dF(z)$$

$$\text{(i. e. differentiation under the integration sign is permissible), which will imply result (A).}$$

For this it is sufficient to prove that for any t_0, $\exists\ g(z)$ such that

$$|t - t_0| < \delta \Rightarrow |z^k e^{tz}| \leq g(z)$$

Where

$$\int_{-\infty}^{\infty} g(z) dF(z) < \infty$$

Since $|t - t_0| < \delta$,

$$e^{tz} < \max\left[e^{(t_0+\delta)z} + e^{(t_0-\delta)z}\right]$$

$$< e^{(t_0+\delta)z} + e^{(t_0-\delta)z}$$

Also,

$$|z^k| = C\frac{\delta^k |z|^k}{k!} < Ce^{\delta|z|} < C\left(e^{\delta z} + e^{-\delta z}\right),$$

Where $C = \frac{k!}{\delta^k}$

From these two inequalities, we have the new inequality

$$|z^k e^{tz}| < C\left[e^{\delta z} + e^{-\delta z}\right]\left[e^{(t_0+\delta)z} + e^{(t_0-\delta)z}\right]$$

Let us define $g(z)$ to be the expression on the right hand side of the inequality. Then $g(z)$ has the property that $\int_{-\infty}^{\infty} g(z)dF(z) < \infty, g(z)$ being integrable because $\phi(t)$ has been assumed to exist for all real t.

In order to prove result (B), we note that the result on $\phi^{(k)}(t)$ gives, in particular,

$$\phi'(0) = E(Z)$$

$$\text{And } \phi''(t) = \int_{-\infty}^{\infty} z^2 e^{tz} dF(z)$$

$$> 0 \text{ for all real } t,$$

By condition (II). So the function $\phi(t)$ is convex (with increasing first derivative).

Hence if $E(Z) = 0, \phi(t)$ has a unique minimum at $t = 0$. If $E(Z) > 0,$ then $\phi(t)$ is strictly greater than $\phi(0) = 1$ for $t > 0$. To show that the equation $\phi(t) = 1$ has a root $h < 0$, it is sufficient to prove that $\phi(t) \to \infty$ as $t \to -\infty$. Similarly, to show that the equation has a root $h > 0$ when $E(Z) < 0$, it is enough to prove that $\phi(t) \to \infty$ as $t \to \infty$.

We prove the second result, the proof of the first being similar.

By condition (II), $\exists\ \eta > 0$ such that

$$P[Z > \eta] > 0, P[Z < -\eta] > 0.$$

Suppose, for definiteness, that $t > 0$. Then

$$\phi(t) = \int_{-\infty}^{\infty} e^{tz}\, dF(z)$$

$$> \int_{\eta}^{\infty} e^{tz}\, dF(z) \text{ since } e^{tz} > 0 \text{ even if } tz > 0$$

$$\geq e^{t\eta} \int_{\eta}^{\infty} dF(z) = e^{t\eta} P[Z > \eta].$$

As $t \to \infty$,

$$e^{t\eta} P[Z > \eta] \to \infty \ \ (\text{since } P[Z > \eta] > 0)$$

$$\Rightarrow \phi(t) \to \infty,$$

This completes the proof of the lemma.

Let us consider the problem of determining $L(\theta)$. Taking

$$Z = \log \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}$$

Where f_θ is the p.d.f. of a continuous random variable X, we have

$$\phi(t) = \int_{-\infty}^{\infty} \left[\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}\right]^t f_\theta(x)dx.$$

If it is assumed that

$$E_\theta \left[\log \frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}\right] \neq 0$$

And that the conditions of the lemma are satisfied, then follows that $\exists\ h(\theta) \neq 0$ such that

$$\int_{-\infty}^{\infty} \left[\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}\right]^{h(\theta)} f_\theta(x)dx = 1.$$

So that the integrand has the properties of a p.d.f; i.e.,

$$g_\theta(x) = \left[\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}\right]^{h(\theta)} f_\theta(x)$$

May be regarded as a p.d.f.

Since $h(\theta) \neq 0$, either $h\theta) > 0$, or $h(\theta) < 0$. First let us consider the case $h(\theta) > 0$.

We shall denote by H the hypothesis that f_θ is the true p.d.f. of X and by H^* the hypothesis that g_θ is the true p.d.f. We may consider the SPR test for testing H against H^* defined as follows :

Continue taking observations as long as

$$B^{h(\theta)} < \prod_{i=1}^{m} \frac{g_\theta(x_i)}{f_\theta(x_i)} < A^{h(\theta)} \tag{32}$$

Stop taking observations as soon as one of the inequalities is violated,

Accept H if $\prod_{i=1}^{m} \frac{g_\theta(x_i)}{f_\theta(x_i)} \leq B^{h(\theta)}$ (33)

And reject H (i.e. accept H^*) if $\prod_{i=1}^{m} \frac{g_\theta(x_i)}{f_\theta(x_i)} \geq A^{h(\theta)}$ (34)

Since

$$\frac{g_\theta(x)}{f_\theta(x)} = \left[\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}\right]^{h(\theta)}$$

And $h(\theta) > 0$, the inequalities (32), (33) and (34) are equivalent, respectively, to

$$B < \Pi_{i=1}^{m}\left(\frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)}\right) < A$$

$$\Pi_{i=1}^{m}\left(\frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)}\right) \leq B$$

$$\text{And} \quad \Pi_{i=1}^{m}\left(\frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)}\right) \geq A$$

These inequalities are identical with those which define the SPR test for H_0 against H_1, when the constant A and B are used. It is thus seen that the SPR test for H_0 against H_1 leads to the acceptance or rejection of H_0 according as the SPR test for H against H^* leads to the acceptance or rejection of H. It follows that the probability of accepting H_0 in using the former test when θ is the true value of the parameter is the same as the probability of accepting H in using the latter test when f_θ is the true p.d.f. of X.

Let $\alpha' = P(rejecting\ H|H)$ & $\beta' = P(accepting\ H|H^*)$

Applying the inequalities (21a) and (21b) to this test, we then have

$$A^{h(\theta)} \leq \frac{1-\beta'}{\alpha'} \tag{35}$$

$$B^{h(\theta)} \geq \frac{\beta'}{1-\alpha'} \tag{36}$$

When the excess of $\prod_{i=1}^{\infty}\frac{g_\theta(x_i)}{f_\theta(x_i)}$ over the boundaries A and B at the termination of experiment is negligible, the equality sign holds

$$A^{h(\theta)} \simeq \frac{1-\beta'}{\alpha'}$$

$$\text{And } B^{h(\theta)} \simeq \frac{\beta'}{1-\alpha'}$$

$$\Rightarrow \alpha' \simeq \frac{1-B^{h(\theta)}}{A^{h(\theta)}-B^{h(\theta)}}$$

Since $\alpha' = 1 - L(\theta)$, this means

$$L(\theta) \simeq \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}} \tag{37}$$

These results hold good even when $h(\theta) < 0$, only A and B are to be interchanged.

Example: Consider the first example of previous section

Here the equation giving $h(\theta)$ is

$$\Sigma_x \left[\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}\right]^{h(\theta)} \qquad f_\theta(x) = 1,$$

Which can be written

$$\theta \left(\frac{\theta_1}{\theta_0}\right)^{h(\theta)} + (1-\theta)\left(\frac{1-\theta_1}{1-\theta_0}\right)^{h(\theta)} = 1$$

For determining the OC function $L(\theta)$, it is not necessary to solve this equation w.r.t. $h(\theta)$. We may just regard $h \equiv h(\theta)$ as a parameter and solve the equation w.r.t. θ. This root is

$$\theta = \frac{1 - \left(\frac{1-\theta_1}{1-\theta_0}\right)^h}{\left(\frac{\theta_1}{\theta_0}\right)^h - \left(\frac{1-\theta_1}{1-\theta_0}\right)^h} \tag{38}$$

Now putting $A = (1-\beta)/\alpha$ and $B = \beta/(1-\alpha)$, we have from (37)

$$L(\theta) \simeq \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h} \tag{39}$$

For any arbitrarily chosen value of h, the point with co-ordinates θ and $L(\theta)$, given by (38) and (39), respectively, will be a point on the OC curve. So the OC curve can be drawn by plotting a sufficiently large number of such points corresponding to different values of h.

Example: Consider second example of previous section**,** in this case $h(\theta)$, for any given θ, is the non-zero root of the equation

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp[-(x-\theta)^2/2\sigma^2].\left[\frac{\exp\left[-\frac{(x-\theta_1)^2}{2\sigma^2}\right]}{\exp\left[-\frac{(x-\theta_0)^2}{2\sigma^2}\right]}\right]^{h(\theta)} dx = 1 \qquad (40)$$

Since the integrand equals

$$\frac{1}{\sigma\sqrt{2\pi}}.\exp\left[-\frac{1}{2\sigma^2}\{x^2-2x[\theta+h(\theta)(\theta_1-\theta_0)]+[\theta^2-h(\theta)(\theta_1^2-\theta_0^2)]\}\right]$$

$=\frac{1}{\sigma\sqrt{2\pi}}.\exp\left[-\frac{1}{2\sigma^2}\{x-\theta-h(\theta)(\theta_1-\theta_0)\}^2\right]*\exp\left[-\frac{1}{2\sigma^2}(\theta_1-\theta_0)\{h(\theta)(\theta_1+\theta_0-2\theta)-\right.$

$\left.h^2(\theta)(\theta_1-\theta_0)\}\right]$

Equation (40) reduces to

$$\exp\left[-\frac{(\theta_1-\theta_0)h(\theta)\{(\theta_1+\theta_0-2\theta)-h(\theta)(\theta_1-\theta_0)\}}{2\sigma^2}\right]=1$$

Or to $\Rightarrow\ (\theta_1+\theta_0-2\theta)-h(\theta)(\theta_1-\theta_0)=0$

$$\Rightarrow h(\theta)=\frac{\theta_1+\theta_0-2\theta}{\theta_1-\theta_0} \qquad (41)$$

For any given θ, we can approximation get the corresponding value of $L(\theta)$ by substituting $(\theta_1+\theta_0-2\theta)/(\theta_1-\theta_0)$ for $h(\theta)$ in formula (39). The OC curve may be drawn by taking a sufficient number of values of θ and plotting the points $(0, L(\theta))$ on graph paper.

5.7.2 Wald's Equation

Theorem 2:

Suppose

(a) $Z_1, Z_2, \ldots ..$ are identically distributed random variables;

(b) $E(Z_1)$ exists (and is finite);

(c) N is a random variable whose values are the positive integers, and

(i) the event $[N \le j]$ and the random variable Z_k are independent for $j < k$

(ii) $E(N)$ is finite.

Then the expectation$E(\Sigma_{i=1}^{N} Z_i)$ exists and

$$E(\Sigma_{i=1}^{N} Z_i) = E(N)E(Z_1)$$

Proof: Let us define the random variables Y_j $(j = 1, 2, \dots.)$ by

$$Y_j = \begin{cases} 1 & if\ N \geq j \\ 0 & otherwise \end{cases} \tag{42}$$

Then Y_j is the indicator function of the event $[N \geq j]$, or one minus the indicator function of the event $[N \leq j-1]$. By condition (i) of (c), Y_j and Z_j are mutually independent. Also, condition (ii) of (c) implies that N is finite with probability one. Hence $\Sigma_{j=1}^{N} Z_j$ is also defined (i.e. is convergent) with probability one.

Also, because (42),

$$\Sigma_{j=1}^{N} Z_j = \Sigma_{j=1}^{\infty} Y_j Z_j \tag{43}$$

Now, we have

$$E\left(\Sigma_{j=1}^{\infty} |Y_j Z_j|\right) = \Sigma_{j=1}^{\infty} E\left(Y_j\right) E|Z_j|) \text{ since } Y_j \text{ and } Z_j \text{ are independent}$$

$$= \Sigma_{j=1}^{\infty} P[N \geq j] E(|Z_j|)$$

$$= E(N)E(|Z_j|)$$

$$< \infty$$

By (b) and (ii) of (c), noting that

$$E(N) = \Sigma_{j=1}^{\infty} jP[N = j]$$

$$= \Sigma_{j=1}^{\infty} P[N \geq j]$$

Hence from (43), it follows that $E\left(\Sigma_{j=1}^{N} Z_j\right)$ exists and

$$E\left(\Sigma_{j=1}^{N} Z_j\right) = E\left(\Sigma_{j=1}^{\infty} Y_j Z_j\right)$$

$$= \Sigma_{j=1}^{\infty} E(Y_j Z_j)$$

$$= \Sigma_{j=1}^{\infty} E(Y_j) E(Z_j)$$

$$= \Sigma_{j=1}^{\infty} P[N \geq j] E(Z_1)$$

$$= E(N)E(Z_1)$$

From this theorem, we have the result that

$$E_\theta(N) = \frac{E_\theta\left(\Sigma_{j=1}^{N} Z_j\right)}{E_\theta(Z_1)}, \tag{44}$$

Provided $E_\theta(Z_1) \neq 0$, where E_θ denotes expectation when θ is true value of the parameter.

If the excess of the probability ratio $f_{1,m}/f_{0,m}$ over the boundaries A and B at the termination of sampling is neglected, then the random variable $\Sigma_{j=1}^{N} Z_j$ may be supposed to take only the values $\log A$ and $\log B$ with probabilities $1 - L(\theta)$ and $L(\theta)$, respectively. Hence

$$E_\theta\left(\Sigma_{j=1}^{N} Z_j\right) \simeq L(\theta) \log B + [1 - L(\theta)] \log A\,,$$

So that we obtain the following approximate formula for the ASN function:

$$E_\theta(N) = \frac{L(\theta) \log B + [1 - L(\theta)] \log A}{E_\theta(Z_1)} \tag{45}$$

Example: Consider the first example of previous section**,** we have already obtained an approximate formula concerning $L(\theta)$. Hence to obtain an approximate formula for $E_\theta(N)$ we need only compute $E_\theta(Z_1)$.

$$f_\theta(x) = \theta^x (1-\theta)^{1-x}, \text{ for } x = 0{,}1,$$

$$\Rightarrow \quad E_\theta(Z_1) = E_\theta\left[\log \frac{f_{\theta_1}(X_1)}{f_{\theta_0}(X_1)}\right]$$

$$= \theta \log \frac{f_{\theta_1}(1)}{f_{\theta_0}(1)} + (1-\theta) \log \frac{f_{\theta_1}(0)}{f_{\theta_0}(0)}$$

$$= \theta \log \frac{\theta_1}{\theta_0} + (1-\theta) \log \frac{1-\theta_1}{1-\theta_0}$$

Example: Consider the second example of previous section.

$$f_\theta(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2}\right]$$

$$\Rightarrow Z_1 = \log \frac{f_{\theta_1}(X_1)}{f_{\theta_0}(X_1)} = \frac{1}{2\sigma^2}\left[2X_1(\theta_1 - \theta_0) + \theta_0^2 - \theta_1^2\right]$$

Since $E_\theta(X_1) = \theta$, we then have

$$E_\theta(Z_1) = \frac{1}{2\sigma^2}\left[2\theta(\theta_1 - \theta_0) + \theta_0^2 - \theta_1^2\right]$$

5.7.3 Efficiency of the SPR Test

Consider sequential tests of the simple hypothesis $H_0\colon \theta = \theta_0$ against the simple hypothesis $H_1\colon \theta = \theta_1$.

Let us restrict ourselves to the class of sequential tests of a given strength (α, β). Then a test may be regarded as preferable to another test of the class if the former requires a smaller number of observations, on the average, than the latter. Hence if a test exists in the class for which $E_{\theta_0}(N)$ and $E_{\theta_1}(N)$ are smaller than the corresponding numbers for any other test of the class, then the former test may be called an optimum test.

We shall denote by $N_0(\alpha, \beta)$ the minimum value of $E_{\theta_0}(N)$ in the class and by $N_1(\alpha, \beta)$ the minimum value of $E_{\theta_1}(N)$. The efficiency of a given test in the class under H_0 we mean the ratio.

$$\frac{N_0(\alpha, \beta)}{E_{\theta_0}(N)}$$

And by its efficiency under H_1 we mean the ratio

$$\frac{N_0(\alpha, \beta)}{E_{\theta_1}(N)}$$

Clearly, both under H_0 and H_1, the efficiency of the given test lies between 0 and 1.

5.8 SPR Test for a Composite Hypothesis

Let us consider the simple case where θ is a single parameter and the simple hypothesis $H_0: \theta = \theta_0$ is tested against a one-sided composite alternative, say $H: \theta > \theta_0$. The zone of preference for acceptance of H_0 may be said to consist of the single value θ. The degree of preference for rejection will generally increase with increasing θ in the domain $\theta > \theta_0$. It will then be possible to find a value $\theta_1 > \theta_0$ such that the acceptance of H_0 is considered an error of practical importance whenever $\theta \geq \theta_1$, while for $\theta_0 < \theta < \theta_1$ the acceptance of H_0 is an error of no particular importance. Thus $\theta \geq \theta_1$ may be said to constitute the zone of preference for rejection while $\theta_0 < \theta < \theta_1$constitutes the zone of indifference.

The following restrictions may be imposed on the OC function:

$$L(\theta_0) = 1 - \alpha$$

And

$$L(\theta) \leq \beta \text{ for } \theta \geq \theta_1$$

In most cases, the SPR test of strength (α, β) for testing $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$ will satisfy these requirements, since $L(\theta)$ will be monotonically decreasing in θ for $\theta \geq \theta_1$. Hence in all such cases the SPR test for $H_0: \theta = \theta_0$ against a properly chosen $H_1: \theta = \theta_1$ provides a satisfactory solution to our problem.

The case where $H_0: \theta = \theta_0$ is to be tested against the composite alternative $H: \theta < \theta_0$ may be similarly treated.

Again, if our problem is to test $H_0': \theta \le \theta_0$ against $H_1': \theta \ge \theta_1$ (where $\theta_0 < \theta_1$), and we imposed the restrictions

$1 - L(\theta) \le \alpha$ if $\theta \le \theta_0$

and $L(\theta) \le \beta$ if $\theta \ge \theta_1$

Then too the SPR test for $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$ will generally provide a satisfactory solution to our problem.

5.8.1 Wald's Approach

Assume that we have a sequence of random variables, $X_1, X_2, \ldots \ldots$, such that $f_{\theta,n}$ is the joint p.d.f. of $X_1, X_2, \ldots \ldots, X_n$. Suppose our problem is to find a sequential test for $H_0: \theta \epsilon \Theta_0$ against $H_1: \theta \epsilon \Theta_1$, where Θ_0 and Θ_1 are mutually exclusive sets of the parameter space Θ.

Let us introduce a weight function (a distribution function) ζ_0 for $\theta \epsilon \Theta$ and a second weight function ζ_1 for $\theta \epsilon \Theta_1$.

Hence

$$\zeta_0(\theta) \ge 0 \text{ for } \theta \epsilon \Theta_0 \text{ and } \int_{\Theta_0} d\zeta_0(\theta) = 1; \tag{46}$$

$$\zeta_1(\theta) \ge 0 \text{ for } \theta \epsilon \Theta_1 \text{ and } \int_{\Theta_1} d\zeta_1(\theta) = 1; \tag{47}$$

Consider $g_{0,n}$ and $g_{1,n}$ defined for $n = 1, 2, \ldots$ as follows :

$$\text{Then } g_{0,n} = \int_{\Theta_0} f_{\theta,n} f \zeta_0(\theta),\ \ g_{1,n} = \int_{\Theta_1} f_{\theta,n} d\zeta_1(\theta). \tag{48}$$

Then $g_{0,n}$ and $g_{1,n}$ may themselves be looked upon as p.d.f.'s (under H_0 and H_1, respectively) of the joint distribution of the first n random variables in the sequence $X_1, X_2, \ldots$ ....

Wald's suggestion is to choose the weight functions ζ_0 and ζ_1 in a suitable way and then to test the simple hypothesis H_0^* that the joint density (of $X_1, X_2, \ \ldots \ldots, X_n$ for $n = 1, 2, \ldots \ldots$) is $g_{0,n}$ against the simple alternative H_1^* that the joint density is $g_{1,n}$.

If α be the probability according to this test of rejecting H_0^* when it is true, while $\alpha(\theta)$ is the probability of H_0 being rejected when the true value of the parameter is θ $(\epsilon \Theta_0)$, then

$$\int_{\Theta_0} \alpha(\theta) d\zeta_0(\theta) = \alpha \tag{49}$$

Similarly, if β be the probability according to this test of accepting H_0^* when H_1^* is true, while $\beta(\theta)$ is the probability of H_0 being accepted when the value of the parameter is $\theta(\epsilon\Theta_1)$, then

$$\int_{\Theta_1} \beta(\theta) d\zeta_1(\theta) = \beta \tag{50}$$

Hence α and β may be looked upon as the averages of the error probabilities of the original problem under ζ_0 and ζ_1, respectively. When the desired values of α and β are given, we may for all practical purposes take $A = (1-\beta)/\alpha$ and $B = \beta/(1-\alpha)$.

Example: Let X be distributed as $N(\mu, \sigma^2)$, where μ and σ^2 are both unknown. Consider the hypothesis $H_0: \mu = \mu_0$. If the true value μ differs only slightly from μ_0, i.e. if $|\mu - \mu_0|/\sigma$ is small, then the acceptance of H_0 will not be a serious error. The importance of the error committed in accepting H_0 will increase with $\frac{|\mu-\mu_0|}{\sigma}$. We may thus find a value $\delta > 0$ such that the acceptance of H_0 is considered as error of practical importance only when $\frac{|\mu-\mu_0|}{\sigma} \geq \delta$. Accordingly, we may say that our problem is to test $H_0: \theta\epsilon\Theta_0$ against $H_1: \theta\epsilon\Theta_1$, where $\Theta_0 = \{(\mu, \sigma^2) | \mu = \mu_0\}$ and $\Theta_1 = \{(\mu, \sigma^2) || \mu - \mu_0| = \delta\sigma\}$

Wald choose as the weight functions under H_0 and H_1, respectively, the distribution functions ζ_0 and ζ_1 of σ such that

$$d\zeta_0(\sigma) = \begin{cases} \dfrac{1}{c} & if \quad \sigma \leq c \\ 0 & otherwise, \end{cases}$$

And

$$d\zeta_1(\sigma) = \begin{cases} \dfrac{1}{2c} & if \quad \sigma \leq c \\ 0 & otherwise, \end{cases}$$

Where $c \to \infty$.

For fixed c, the probability ratio is

$$\frac{g_{1,n}}{g_{0,n}} = \frac{\frac{1}{2}\int_0^c \frac{1}{\sigma^n}\left[e^{-\Sigma_{i=1}^n (x_i-\mu_0-\delta\sigma)^2/2\sigma^2} + e^{-\Sigma_{i=1}^n (x_i-\mu_0+\delta\sigma)^2/2\sigma^2}\right] d\sigma}{\int_0^c \frac{1}{\sigma^n} e^{-\Sigma_{i=1}^n (x_i-\mu_0)^2/2\sigma^2}\, d\sigma}$$

when $c \to \infty$, the ratio becomes in the limit

$$\frac{g_{1,n}}{g_{0,n}} = \frac{\frac{1}{2}\int_0^\infty \frac{1}{\sigma^n}\left[e^{-\Sigma_{i=1}^n (x_i-\mu_0-\delta\sigma)^2/2\sigma^2} + e^{-\Sigma_{i=1}^n (x_i-\mu_0+\delta\sigma)^2/2\sigma^2}\right]d\sigma}{\int_0^\infty \frac{1}{\sigma^n} e^{-\Sigma_{i=1}^n (x_i-\mu_0)^2/2\sigma^2}\,d\sigma}$$

Wald shows that this ratio is a strictly increasing function of

$$\sqrt{n}|\bar{x} - \mu_0|/s$$

The test procedure, called the *sequential t-test*, may be enunciated as follows: Continue taking observations as long as

$$B < \frac{g_{1,n}}{g_{0,n}} < A;$$

Stop taking observations as soon as this fails to hold and

Accept H_0 if $\frac{g_{1,n}}{g_{0,n}} \leq B$

And reject H_0 if $\frac{g_{1,n}}{g_{0,n}} \geq A$

It has been shown that in order to make $\alpha(\theta) \leq \alpha$ in Θ_0 and $\beta(\theta) \leq \beta$ in Θ_1, we may take $A = \frac{1-\beta}{\alpha}$, $B = \frac{\beta}{1-\alpha}$.

5.9 Summary

This unit provides a thorough understanding of concepts related to Sequential Tests. The concepts of sequential procedure, sequential testing of hypotheses and SPRT, OC and ASN functions, Wald's Fundamental Identity and Equation, SPR test for a composite hypothesis are described in detail. The learner should try to solve the self-assessment problems given in the next section.

5.10 Self-Assessment Exercises

Q1. Describe sequential probability ratio test, OC and ASN functions.

Q2. State and prove Wald's Fundamental Identity and Equation.

Q3. Describe the procedure of SPR test for a composite hypothesis.

Q4. Determine the SPR test for testing $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1 (\theta_1 > \theta_0)$, where θ is the parameter of a Poisson distribution. Find approximations to the OC function and the ASN function of the test.

Q5. Consider the SPR test for testing $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$, where θ is the mean of a normal distribution with known variance. If $\alpha = \beta$(in usual notation), prove that the ASNs under H_0 and H_1 are equal.

Q6. Let X be a random variable having the normal distribution $N(\mu, \theta)$, where μ is known. Consider the problem of testing $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1 (\theta_1 > \theta_0)$. For type I and type II error probabilities α and β, respectively, and using the approximations for B and A, show that the SPR test is as follows:

Continue taking observations as long as

$$\frac{\theta_0\theta_1\left[2\log\frac{\beta}{1-\alpha} + m\log\frac{\theta_1}{\theta_0}\right]}{\theta_1 - \theta_0} < \sum_{i=1}^{m}(x_i - \mu)^2 < \frac{\theta_0\theta_1\left[2\log\frac{1-\beta}{\alpha} + m\log\frac{\theta_1}{\theta_0}\right]}{\theta_1 - \theta_0}$$

Stop taking observations the first time this fails to hold, and

accpet H_0 if $\sum_{i=1}^{m}(x_i - \mu)^2 \leq \frac{\theta_0\theta_1\left[2\log\frac{\beta}{1-\alpha} + m\log\frac{\theta_1}{\theta_0}\right]}{\theta_1 - \theta_0}$

and reject H_0 if $\sum_{i=1}^{m}(x_i - \mu)^2 \geq \frac{\theta_0\theta_1\left[2\log\frac{1-\beta}{\alpha} + m\log\frac{\theta_1}{\theta_0}\right]}{\theta_1 - \theta_0}$

Q7. (Continuation) Show that the OC function L(θ)) is given by

$$L(\theta) \simeq \frac{\left(\frac{1-\beta}{\alpha}\right)^h - 1}{\left(\frac{1-\beta}{\alpha}\right)^h - \left(\frac{\beta}{1-\alpha}\right)^h},$$

Where $h \equiv h(\theta)$ is such that

$$\theta = \frac{\theta_0\theta_1\left[1 - \left(\frac{\theta_0}{\theta_1}\right)^{2h}\right]}{h(\theta_1 - \theta_0)}$$

Q8. (Continuation) Show that the ASN function $E_\theta(N)$ is given by

$$E_\theta(N) \simeq \frac{2\theta_0\theta_1\left[L(\theta)\log\frac{\beta}{1-\alpha} + \{1 - L(\theta)\}\log\frac{1-\beta}{\alpha}\right]}{(\theta_1 - \theta_0)\theta + \theta_0\theta_1\log\left(\frac{\theta_0}{\theta_1}\right)}.$$

UNIT 6: SEQUENTIAL ESTIMATION

6.1 Introduction

A sequential procedure may also be used for the estimation of a function of the parameters. However, it has not been possible to develop a general theory of sequential estimation. In general, it is not easy to determine the most appropriate estimator or to formulate a rule for the termination of sampling.

One principle is to do just enough sampling to be able to obtain an estimate that would have a preassigned degree of precision (independent of all the parameters being estimated). In the case of point estimation, the precision may be measured in terms of variance or mean square error of the estimator. In interval estimation, this may be expressed by the length of the interval with a given confidence coefficient.

6.2 Objective

The objective of this unit is to provide a basic understanding of concepts related to Sequential Estimation. The concept of the Cramer Rao Inequality of sequential estimation, Stein's two stage procedure should be clear after study of this material.

6.3 Sequential Cramer-Rao Inequality

A sequential estimation procedure may be said to be defined by a sequence $\delta =\{ \Psi_n, T_n\}$, where Ψ_n represents the stopping rule and T_n the estimator to be used in case the experiment stops at the *n*th stage.

More precisely,

$$\Psi_n(X_1, X_2,\ldots\ldots, X_n) = \begin{cases}1, & if\ N = n\\ 0, & if\ N \neq n\end{cases}$$

and in case $N = n$, the procedure requires that the parametric function γ(θ) be estimated by $t_n = T_n\ (x_1, x_2,\ldots\ldots, x_n)$.

We shall assume that $\sum_{n=1} \Psi_n = 1$; i.e., the procedure will be supposed to terminate with probability one, for all θ ε Θ. Further, let T_n be an unbiased estimator of γ(θ). We make the following assumption:

(I) Θ is a real non-degenerate open interval;

(II) $\frac{\partial}{\partial\theta}\sum_{n=1}^{\infty}\int \Psi_n\, f_{\theta,n}\, dx_n$ exists

$= \sum_{n=1}^{\infty}\int \Psi_n \frac{\partial f_{\theta,n}}{\partial\theta}\, dx_n$;

(III) $\frac{\partial}{\partial\theta}\sum_{n=1}^{\infty}\int \Psi_n\, t_n\, f_{\theta,n}\, dx_n$ exists

$= \sum_{n=1}^{\infty}\int \Psi_n\, t_n\, \frac{\partial f_{\theta,n}}{\partial\theta}\, dx_n$;

(IV) $\gamma'(\theta)$ exists for all $\theta\ \varepsilon\ \Theta$;

(V) $E_\theta\left[\frac{\partial}{\partial\theta} log f_{\theta,N}(X_1, X_2, \ldots\ldots, X_n)\right]^2 > 0$.

Unser these assumptions,

Theorem: For all θ ε Θ, we have

$$var_\theta\ (T_N) \geq \frac{[\gamma'(\theta)]^2}{E_\theta\left[\frac{\partial}{\partial\theta}\log f_{\theta,N}\ (X_1, \ldots\ldots, X_n)\right]^2}$$

Proof : We have $\sum_{n=1}^{\infty}\int \Psi_n\, f_{\theta,n} dx_n = 1$

$\Rightarrow \frac{\partial}{\partial\theta}\sum_{n=1}^{\infty}\int \Psi_n\, f_{\theta,n}\, dx_n = 0$

$$\Rightarrow \sum_{n=1}^{\infty}\int \Psi_n \left(\frac{\partial}{\partial\theta}\log f_{\theta,n}\right) f_{\theta,n}\, dx_n = 0 \tag{50}$$

Again,

$$E_\theta(T_n) \;=\; \sum_{n=1}^{\infty} \int \Psi_n t_n\, f_{\theta,n}\; d\mathrm{x}_\mathrm{n} = \gamma(\theta)$$

$$\Rightarrow \frac{\partial}{\partial\theta} \sum_{n=1}^{\infty} \int \Psi_n\, t_n f_{\theta,n} d\mathrm{x}_\mathrm{n} = \gamma'(\theta)$$

$$\Rightarrow \sum_{n=1}^{\infty} \int \Psi_n t_n \left(\frac{\partial}{\partial\theta} \log f_{\theta,n}\right) f_{\theta,n} d\mathrm{x}_\mathrm{n} = \gamma'(\theta) \tag{51}$$

Multiplying (50) by $\gamma(\theta)$, subtracting the product from (51) and lastly squaring the difference, we have

$$\left[\sum_{n=1}^{\infty} \int \Psi_n [t_n - \gamma(\theta)] \left(\frac{\partial}{\partial\theta} \log f_{\theta,n}\right) f_{\theta,n} d\mathrm{x}_\mathrm{n}\right]^2 = [\gamma'(\theta)]^2$$

Noting that left-hand side equals

$$\left[E_\theta\big(T_N - \gamma(\theta)\big)\left(\frac{\partial}{\partial\theta} \log f_{\theta,N}\right)\right]^2$$

and is, by Schwarz's Inequality, not greater than

$$E_\theta[T_N - \gamma(\theta)]^2 \left[E_\theta\left(\frac{\partial}{\partial\theta} \log f_{\theta,N}\right)\right]^2$$

we have finally

$$var_\theta\ (T_N) \geq \frac{[\gamma'(\theta)]^2}{E_\theta\left[\frac{\partial}{\partial\theta} \log f_{\theta,N}\right]^2}$$

Special Case*:* Consider now the case where the random variables $\mathrm{X}_1, \mathrm{X}_2, \dots \dots, \mathrm{X}_\mathrm{n}$ (for each n) are independently and identically distributed with common p.d.f. f_θ. Then

$$f_{\theta,n} = \prod_{i=1}^{n} f_\theta(x_i)$$

$$\Rightarrow \log f_{\theta,n} = \sum_{i=1}^{n} \log f_\theta(x_i)$$

Further, putting $Z_i = \frac{\partial}{\partial\theta} \log f_\theta\,(X_i)$ and assuming $E_\theta(Z^2)$ exists, we get

$$E_\theta\left[\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\log f_\theta\,(X_i)\right]^2 = E_\theta(N)E_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta\,(X)\right]^2$$

Hence the above inequality takes the form

$$var_\theta\,(T_N) \geq \frac{[\gamma'(\theta)]^2}{E_\theta(N)E_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2}$$

Example: Suppose $X_1, X_2, \ldots \ldots$ are independent random variables having the common distribution with p.m.f.

$$f_\theta(x) = f(x) = \begin{cases} \theta^x(1-\theta)^{(1-x)} & if\ x = 0,1 \\ 0 & otherwise, \end{cases}$$

Where $0 < \theta < 1$.

for $x = 0,1$,

$$\frac{\partial}{\partial\theta}\log f_\theta(x) = x\frac{1}{\theta} + (1-x)\left(-\frac{1}{1-\theta}\right)$$

$$= -\frac{1}{1-\theta} + \frac{x}{\theta(1-\theta)}$$

$$E_\theta\left[\frac{\partial}{\partial\theta}\log f_\theta(X)\right]^2 = (1-\theta)\left(-\frac{1}{1-\theta}\right)^2 + \theta\left(\frac{1}{\theta}\right)^2$$

$$= \frac{1}{\theta(1-\theta)}$$

Here the regularity conditions are satisfied, so that for any unbiased sequential estimator T_N of $\gamma(\theta)$, we have

$$var_\theta(T_N) \geq \theta(1-\theta)\,\frac{[\gamma'(\theta)]^2}{E_\theta(N)}.$$

Equality holds

$$T_N - \gamma(\theta) = c(\theta)\sum_{i=1}^{N}\frac{\partial}{\partial\theta}\log f_\theta\,(X_i)$$

$$= c_1(\theta)[S_N - \theta N], \tag{52}$$

say,

where $S_N = X_1 + X_2 + X_3 + \cdots ... + X_N$

***Case 1*:** Let $N = n$, a constant.

To have equality, we must have

$$T_N = a + b\frac{S_n}{n},$$

where a and b are constant.

$$\gamma(\theta) = E_\theta(T_N) = a + b\theta \tag{53}$$

***Case 2*:** Let N be the least n such that $S_n = k$. It can be proved that this sequential procedure terminates with probability one.

Again, to have equality, we must have

$$T_N = a + b_N$$

and the only parametric functions for which MVB estimators exist are of the from

$$\gamma(\theta) = a + bE_\theta(N).$$

From (52)

$$b[n - E_\theta(N)] = c_1(\theta)[k - \theta N],$$

and this must be an identity in N. Such, equating coefficients of powers of N, one gets

$$b = -c_1(\theta)\theta \text{ or } c_1(\theta) = -\frac{b}{\theta}$$

and $$-bE_\theta(N) = c_1(\theta)k = -\frac{b}{\theta}k\ ,$$

so that

$$E_\theta(N) = \frac{k}{\theta}.$$

Hence in this case the parametric functions $\gamma(\theta)$ with MVB estimators are of the form $\gamma(c) = a + b\frac{k}{\theta}$.

Case 3: Let N be the least n such that $n - S_n = k$ (a positive integer). Again, it can be proved that the sequential procedure terminates with probability one.

Proceeding as, we find $E_\theta(N) = \frac{k}{1-\theta}$ and that the functions with MVB estimators have the form

$$\gamma(\theta) = a + \frac{bk}{1-\theta}$$

The sequential sampling procedures used in *Case 2* and *3* have been called *inverse sampling*, since they yield, MVB estimators of the reciprocals of θ and $1 - \theta$.

6.4 Stein's Two-Stage Sampling

Let X be normally distributed with unknown mean μ and unknown variance σ^2. If it is our aim to find a confidence interval for μ at a given confidence level α in such a way that the interval will be of a specified length, then it is not possible to achieve this on the basis of a single sample of fixed size. For, the length of the interval will vary from one sample to another and so it is not possible to ensure that its length for a given sample will have the specified value. Hence Stein suggests a two-stage sampling procedure which will enable us to fulfill this condition.

The sampling procedure is as follows:

Draw a first sample size *m*, say $X_1 + X_2 + X_3 + \cdots ... + X_m$, and let

$$\bar{X}_m = \frac{1}{m}\sum_{i=1}^{m} X_i,\ S^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X}_m)^2. \tag{54}$$

Define

$$N=N(S) \tag{55}$$

as the smallest integer *n* such that

$$n \geq \max\{m, \frac{b^2}{c^2}S^2\}, \tag{56}$$

where b^2 is the variance of the t-distribution with $df = (m-1)$ and *c* is a preassigned constant.

With *n* so defined, now take a second sample of $(N - m)$ observations, say, $X_{m+1} + X_{m+2} + \cdots + X_N$ and calculate

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

Before proving that the related point estimator or test or confidence interval has the stated properties, we shall prove two lemmas:

Lemma: $P_\sigma[N(S) > n] \to 0$ as $n \to \infty$ for all σ, which implies that the procedure terminates with probability one.

Proof: By definition, $N(S) =$ least integer greater than or equal to the larger of m and $\frac{b^2}{c^2}S^2$

Since finally we shall let $n \to \infty$, consider those values of n which are greater than m.

For such n,

$$N(S) > n \Leftrightarrow \frac{b^2}{c^2}S^2 > n,$$

So that

$$P_\sigma[N(S) > n] = P_\sigma\left[\frac{b^2}{c^2}S^2 > n\right]$$

$$= P_\sigma\left[\chi^2 > \frac{n(m-1)c^2}{b^2\sigma^2}\right]$$

where $\chi^2 > \frac{(m-1)S^2}{\sigma^2}$ has $df(m-1)$.

However,

$$P_\sigma\left[\chi^2 > \frac{n(m-1)c^2}{b^2\sigma^2}\right] \to 0 \; as \; n \to \infty,$$

and so the result is established.

Lemma: The random variable

$$T = \sqrt{N}(\bar{X}_N - \mu)/S$$

has Student's *t*-distribution with $df(m-1)$.

***Proof*:** Let ϑ be a fixed integer $\geq m$. Consider the random variable $\sqrt{\vartheta}(\overline{X_\vartheta} - \mu)/\sigma$, which is distributed as $N(0,1)$.

Now,

$$\vartheta\bar{X}_\vartheta = \sum_{i=1}^{m} X_i + \sum_{i=m+1}^{\vartheta} X_i$$

$$= m\bar{X}_m + \sum_{i=m+1}^{\vartheta} X_i.$$

$X_{m+1}, X_{m+2}, \ldots, X_\vartheta$ are independent of $X_1, X_2, \ldots, X_m$ and $\bar{X}_m$ is independent of S^2. Hence $\bar{X}_\vartheta$ is independent of S^2. As such,

$$\sqrt{\vartheta}(\bar{X}_\vartheta - \mu)/S$$

has Student's t-distribution with $df(m-1)$ when ϑ is fixed. This is the conditional distribution of T, given $N = \vartheta$. But since the conditional distribution does not depend on ϑ, the unconditional distribution of T is also of Student's form with $df(m-1)$.

***Theorem*:** $\bar{X}_N$ is an unbiased estimator of μ with variance $\leq c^2$

Proof: Since $\sqrt{N}(\bar{X}_N - \mu)/\sigma$ is distributed as $N(0,1)$,

$$E_\theta\left[\frac{\sqrt{N}(\bar{X}_N - \mu)}{\sigma}\right] = 0.$$

Also,

$$E_\theta(\bar{X}_N - \mu) = E_\theta\left[\frac{\sqrt{N}(\bar{X}_N - \mu)}{\sigma} . \frac{\sigma}{\sqrt{N}}\right]$$

$$= E_\theta\left[\frac{\sqrt{N}(\bar{X}_N - \mu)}{\sigma} . \Psi(S^2, \sigma)\right], \text{ where } \Psi(S^2, \sigma) = \frac{\sigma}{\sqrt{N}}$$

$$= E_\theta\left[\frac{\sqrt{N}(\bar{X}_N - \mu)}{\sigma}\right] E_\theta[\Psi(S^2, \sigma)]$$

$$= 0 \text{ for all } \theta.$$

Hence

$$E_\theta(\bar{X}_N) = \mu \text{ for all } \theta.$$

Furthermore,

$$E_\theta(\bar{X}_N - \mu)^2 = E_\theta\left[\frac{N(\bar{X}_N - \mu)^2}{S^2} . \frac{S^2}{N}\right]$$

$$\leq E_\theta\left[\frac{N(\bar{X}_N-\mu)^2}{S^2}\cdot\frac{c^2}{b^2}\right]$$

$$=\frac{c^2}{b^2}E_\theta\left[\frac{N(\bar{X}_N-\mu)^2}{S^2}\right]$$

$$=\frac{c^2}{b^2}.b^2$$

$$=c^2,$$

i.e., $var_\theta(\bar{X}_N)\leq c^2$for all θ.

Theorem: Let $2l$ be the specified length of the confidence interval estimator of θ. Then the confidence interval estimator of θ. Then the confidence coefficient of the estimator $\bar{X}_N\pm l$ is not less than $1-\alpha$ if we take $c=bl/t_{\frac{\alpha}{2},m-1}$.

Proof:

$$P_\theta(|\bar{X}_N-\mu|\leq l)=P_\theta\left[\frac{\left|\sqrt{N}(\bar{X}_N-\mu)\right|}{S}\leq\frac{\sqrt{N}l}{S}\right]$$

$$\leq P_\theta\left[\frac{\left|\sqrt{N}(\bar{X}_N-\mu)\right|}{S}\leq\frac{bl}{c}\right]$$

since $\frac{S}{\sqrt{N}}\leq\frac{c}{b}$.

However, since $\sqrt{N}(\bar{X}_N-\mu)/S$ has the *t*-distribution with df (m-1) and $c=bl/t_{\frac{\alpha}{2},m-1}$, i.e., $\frac{bl}{c}=$ $t_{\frac{\alpha}{2},m-1}$,

$$P_\theta\left[\frac{\left|\sqrt{N}(\bar{X}_N-\mu)\right|}{S}\leq\frac{bl}{c}\right]=1-\alpha.$$

Hence

$$P_\theta[|(\bar{X}_N-\mu)|\leq l]\geq 1-\alpha,$$

and this is true for all θ.

6.5 Summary

This unit provides a thorough understanding of concepts related to Sequential Estimation. The concepts of sequential Cramer-Rao Inequality and Stein's two-stage sampling procedure are described in detail. The learner should try to solve the self-assessment problems given in the next section.

6.6 Self-Assessment Exercises

Q1. State and prove Sequential Cramer-Rao Inequality.

Q2. Describe Stein's two-stage sampling procedure in detail.

U.P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 202N/ MASTAT –202N
Non Parametrics

Block: 3 Nonparametric Tests and Inference

Course Design Committee

Course Preparation Committee

Prof. Shashi Bhushan **Writer**
Department of Statistics
University of Lucknow, Lucknow

Prof. S. K. Pandey **Editor**
Department of Statistics (Unit-7 & 8)
University of Lucknow, Lucknow

Prof. Shruti **Course Coordinator**
School of Sciences (Unit-9)
U. P. Rajarshi Tandon Open University, Prayagraj

Prof. Shruti **Course Coordinator**
School of Sciences
U. P. Rajarshi Tandon Open University, Prayagraj

MScSTAT – 202N/ MASTAT – 202N **NON PARAMETRICS**
©UPRTOU
First Edition: ***July 2023***
ISBN:

Printed By:

Block & Units Introduction

The ***Block - 3 – Nonparametric Tests and Inference*** has three units.

Unit – 7 – One-Sample and Two-Sample Location Tests dealt with One and two sample location tests, Sign test. Wilcoxon test, Median test.

Unit – 8 – Other Non- Parametric Tests dealt with Mann- Whitney U- Test, Application of U-statistic to rank tests. One sample and two sample Kolmagorov-Smirnov tests. Run tests.

Unit – 9 – Non-Parametric Inference, The Kruskal-Wallis one way ANOVA Test, Friedman's two-way analysis of variance by ranks, efficiency criteria and theoretical basis for calculating ARE, Pitman ARE.

At the end of every block/unit the summary, self assessment questions and further readings are given.

References & Further Readings

- Bagdonavicius, V., Kruopis, J., Nikulin, M.S. (2011). "Non-parametric tests for complete data", ISTE & WILEY: London & Hoboken. ISBN 978-1-84821-269-5.
- Corder, G. W.; Foreman, D. I. (2014). Nonparametric Statistics: A Step-by-Step Approach. Wiley. ISBN 978-1118840313.
- Gibbons, Jean Dickinson; Chakraborti, Subhabrata (2003). Nonparametric Statistical Inference, 4th Ed. CRC Press. ISBN 0-8247-4052-1.
- Hettmansperger, T. P.; McKean, J. W. (1998). Robust Nonparametric Statistical Methods. Kendall's Library of Statistics. Vol. 5 (First ed.). London: Edward Arnold. New York: John Wiley & Sons. ISBN 0-340-54937-8. MR 1604954. also ISBN 0-471-19479-4.
- Hollander M., Wolfe D.A., Chicken E. (2014). Nonparametric Statistical Methods, John Wiley & Sons.
- Sheskin, David J. (2003) Handbook of Parametric and Nonparametric Statistical Procedures. CRC Press. ISBN 1-58488-440-1
- Wasserman, Larry (2007). All of Nonparametric Statistics, Springer. ISBN 0-387-25145-6.

UNIT:7 ONE-SAMPLE AND TWO-SAMPLE LOCATION TESTS

Structure

7.9 Introduction

The parametric inferential methods are based on stringent assumptions about the probability distribution of the parent population like the form of probability distribution is apriori available, availability of observations either on ratio scale or atleast on interval scale etc. However, these assumptions may not be satisfied in many practical situations. For instance, the measurement on the units under study is often made on nominal or ordinal scale owing to practical difficulties. Usually, we do not know the distribution characterizing the phenomena of the experiment. However, we can often choose a sufficiently large class of distributions $\{F_\theta(x)\}$ invariably indexed by an unknown parameter θ. The range of θ is Ω which is called the parameter space. The statistician has to decide upon the particular probability distribution which explains most the phenomena of the experiment. That is, the statistician has to make a decision about the value of the parameter, by means of the observable random variable X. However, in many situations the outcome X is a complicated set of numbers. If at all feasible, he would like to condense his data and come out with a magic number which contains all the relevant information about the parameter θ. In such situations where the stringent assumptions of the parametric inferential methods are not satisfied, we resort to non-parametric methods. The non-parametric methods rely on relatively mild assumptions about the probability distribution of the parent population.

The statistical methods which are not concerned with estimation of testing for parameter(s) of probability distribution functions are known as NON-PARAMETRIC METHODS. Nonparametric statistical procedures are widely used due to their simplicity, applicability under fairly general assumptions and robustness to outliers in the data. Hence they are popular statistical tools in industry, government and various other disciplines. Also, there an extensive amount of literature is available on nonparametric statistics ranging from theory to applications.

The term non-parametric is sometimes synonymously used with distribution free methods as if both have the same meaning. There is a slight difference between the two methods. The statistical inferential procedures whose validity does not depend on the form of probability distribution of the population from which the sample has been drawn are known as DISTRIBUTION FREE METHODS. The distribution free procedures are primarily devised for non-parametric problems; hence the two terms are used interchangeably. Also, the non-parametric methods are devised for no parameter problems.

7.10 Objectives

The objective of this unit is to provide a basic understanding of concepts related to One-sample Location Tests. The concept of the one and two sample location tests, Sign test. Wilcoxon test, Median test should be clear after study of this material.

7.11 Non-Parametric Inferences

The classical statistical inference techniques are based on the assumptions regarding the nature of the population distribution from which the samples are drawn. i.e. form of the population distribution and the parameters of the population distribution. The exact sample tests are based on the assumption that the parent population is normal. Most of the standard statistical techniques are based on the assumptions of normality, independence and homoscedasticity.

Remark:

The statistical methods which remain valid under violation of assumptions of normality, independence & homoscedasticity are called ‘robust’.

Parametric Test:

The parametric tests are those tests in which certain conditions are imposed about the parameters of the population from which the samples are drawn.

Ex- t-test, F-test.

General Assumption of Parametric Tests:

The parent population from which the samples are drawn is assumed to be normal.

The form of the basic distribution is always known.

Non-Parametric Test:

The non-parametric test are those tests in which no assumption, regarding the test of the population from which the samples are drawn is made.

The non-parametric tests are the tests for a hypothesis which is not a statement about the parameter values. Here, the hypothesis is concerned with either form of the population (e.g-goodness of fit) or with some characteristic of the probability distribution of the sample data (e.g- test of randomness).

General Assumption of Non-Parametric Tests:

1. The parent population is continuous.
2. The sample observations are independent.
3. The distribution of the parent population is symmetrical.
4. The lower order moments exist.

7.12 Advantages and Dis-Advantages of Non- Parametric Methods

Advantages of Non-Parametric Tests:

1. Non-parametric tests are quick and easy to apply and do not require complicated sample theory.
2. No assumption is required about the form of the distribution of the parent population from which the samples are drawn.
3. Non-parametric tests can be used even in the situations where actual measurements are unavailable and the data are obtained only as ranks. i.e. if measurements scale is nominal or ordinal, non-parametric methods can be used.
4. The probability statements obtained from most of the non-parametric tests are exact probabilities.
5. Non-parametric tests are used in the situation where sample data are taken from several different populations.
6. With non-accurate and dirty data (e.g: contaminated observations, outliers etc.), many non-parametric methods are appropriate.
7. Non-parametric tests require no. of minimum sample size for valid and reliable results.
8. Non-parametric tests require minimal calculation.

Disadvantages of Non-Parametric Tests:

1. If all the assumptions of a statistical model are satisfied by the data and if the observations are of required strength, then non-parametric tests are wasteful of time and data.

2. Non-parametric tests are designed to test the statistical hypothesis only and not for estimating the parameters.

3. Power efficiency on non-parametric tests are always less than parametric tests.

4. No non-parametric test exists for testing interactions in the analysis of variance model unless specific assumptions about the additivity of the model are made.

5. It is not possible to determine the actual power of non-parametric test due to want of actual situation or actual probability distribution.

7.13 Non-Parametric Test for Location

The following are the non-parametric test for location parameter of a population or the non-parametric tests for the location parameters of two populations.

In non-parametric theory, the most frequently used measure of location is "population median" M or $K_{0.5}$, which is the unique real solution of the equation.

$$F(M) = \frac{1}{2} \text{ or } F(K_{0.5}) = \frac{1}{2}$$

7.13.1 SIGN TEST

The sign test is a non-parametric test for the location parameter median M of a population.

7.13.1.1 Hypothesis and Assumptions

In this test, we make the assumption of independence and homoscedasticity but the assumption of normality for the parent population is not required.

We wish to test the null hypothesis

$H_0 : M = M_0$ (a given value)

against

(i) a one-sided alternative

$H_1 : M < M_0$ (left tailed test)

$H_1 : M > M_0$ (right tailed test)

(ii) a two-sided alternative

$H_1 : M \neq M_0$

7.13.1.2 Test Procedure

Let $X_{(1)},\ldots,X_{(n)}$ be the order statistics corresponding to a random sample $X_1,\ldots..X_n$ of size n drawn from the population having distribution function F with unknown median M, where F is assumed to be continuous in the neighborhood of M so that $P(X = M) = 0$.

By definition of median, we have

$$P(X > M) = P(X < M) = \frac{1}{2}$$

If the sample data are consistent with the hypothetical value of median M_0, then on the average half of the sample observations will be greater than M_0. We replace each observation greater than M_0 by a plus sign (+) and each observation smaller than M_0 by a minus sign (-). Further, we count the numbers of plus signs and the minus signs and denoted it by r and s respectively, with $r + s \leq n$. The number of plus signs (r) may be used to test H_0.

When the population is dichotomized, the sampling distribution of r given $(r+s)$ is binomial with parameter $p = P(X > M_0) = \frac{1}{2}$. Thus, the testing of H_0 becomes an equivalent testing for the hypothesis that the binomial parameter p has the value $\frac{1}{2}$ .i.e. $H_0 : p = \frac{1}{2}$

The critical region for

$H_0 : M = M_0$ or $H_0 : p = \frac{1}{2}$

against $H_1 : M \neq M_0$ or $H_1 : p \neq \frac{1}{2}$

for α level of significance is given by

$r \geq r_{\alpha/2}$ and $r \leq r_{\alpha/2}$.

where $r_{\alpha/2}$ is the smallest integer such that

$$\sum_{k=r_{\alpha/2}}^{r+s} {}^{r+s}C_k . p^{r+s} \le \frac{\alpha}{2}$$

i.e. $$\sum_{k=r_{\alpha/2}}^{r+s} {}^{r+s}C_k . \left(\frac{1}{2}\right)^{r+s} \le \frac{\alpha}{2}$$

and $r'_{\alpha/2}$ is the largest integer such that

$$\sum_{k=0}^{r'_{\alpha/2}} {}^{r+s}C_k . p^{r+s} \le \frac{\alpha}{2}$$

i.e. $$\sum_{k=0}^{r'_{\alpha/2}} {}^{r+s}C_k . \left(\frac{1}{2}\right)^{r+s} \le \frac{\alpha}{2}$$

For testing $H_0 : M = M_0$ or $H_0 : p = \frac{1}{2}$

against $H_1 : M > M_0$ or $H_1 : p > \frac{1}{2}$

the critical region for α level of significance is given by

$$r \ge r_\alpha$$

where r_α is the smallest integer such that

$$\sum_{k=r_\alpha}^{r+s} {}^{r+s}C_k . p^{r+s} \le \alpha$$

i.e. $$\sum_{k=r_\alpha}^{r+s} {}^{r+s}C_k . \left(\frac{1}{2}\right)^{r+s} \le \alpha$$

In this alternative hypothesis the sample will have excess of plus signs.

For testing $H_0 : M = M_0$ or $H_0 : p = \frac{1}{2}$

against $H_1 : M < M_0$ or $H_1 : p < \frac{1}{2}$

the critical region for α level of significance is given by

$$r \geq r_{\alpha}^{'}$$

where $r_{\alpha}^{'}$ is the largest integer such that

$$\sum_{k=0}^{r_{\alpha}^{'}} {}^{r+s}C_k . p^{r+s} \leq \alpha$$

i.e. $\sum_{k=0}^{r_{\alpha}^{'}} {}^{r+s}C_k . \left(\frac{1}{2}\right)^{r+s} \leq \alpha$

In this alternative hypothesis the sample will have less plus signs.

7.13.1.3 Large Sample Test

If $(r+s) > 25$, then we use the normal approximation to the binomial to perform the test.

In this case, under H_0

$$Z = \frac{r - E(r)}{\sqrt{V(r)}} \to N(0,1)$$

since $E(r) = \frac{(r+s)}{2}$ and $V(r) = \frac{(r+s)}{4}$

Hence, under H_0

$$Z = \frac{r - \left(\frac{(r+s)}{2}\right)}{\sqrt{\frac{(r+s)}{4}}} = \frac{r-s}{\sqrt{(r+s)}} \to N(0,1)$$

Example1. Test the null hypothesis that the median length θ of ear-head of a variety of wheat is $\theta_0 = 9.9$ cm. against the alternative that $\theta_0 \neq 9.9$ cm., with $\alpha = 0.05$, on the basis of the following 25 ear-head measurements:

9.5	8.9	10.5	11.5	8.5	9.4	10.6	8.8	11.7	10.5

11.2	9.2	9.8	9.5	9.9	10.9	10.2	9.1	10.8	9.4
11.6	8.7	8.3	11.3	8.1					

Solution: First, we determine the signs of all measurement and replace each measurement greater than θ_0 by $+$ sign and each measurement less than θ_0 by $-$ sign. Measurement which is equal to θ_0 is ignored.

9.5 $(-)$	8.9 $(-)$	10.5 $(+)$	11.5 $(+)$	8.5 $(-)$	9.4 $(-)$	10.6 $(+)$	8.8 $(-)$	11.7 $(+)$	10.5 $(+)$
11.2 $(+)$	9.2 $(-)$	9.8 $(-)$	9.5 $(-)$	9.9 ignored	10.9 $(+)$	10.2 $(+)$	9.1 $(-)$	10.8 $(+)$	9.4 $(-)$
11.6 $(+)$	8.7 $(-)$	8.3 $(-)$	11.3 $(+)$	8.1 $(-)$					

From the above table, we observe that no. of plus signs $r = 11$ and the no. of minus signs $= s = 13$ and one observation is ignored.

So we have to test whether $r = 11$ support the hypothesis $H_0 : \theta_0 = 9.9$, or equivalently to judge how likely are 11 successes (the number of plus signs) to occur in 24 trials from a binomial distribution with $p = 0.5$. The critical region for the level α two-sided test is given by

$$r \geq r_{\alpha/2} \text{ and } r \leq r'_{\alpha/2},$$

where $r_{\alpha/2}$ is the smallest and $r'_{\alpha/2}$ is the largest integer such that

$$\sum_{r_{\alpha/2}}^{n} {}^nC_x \left(\frac{1}{2}\right)^n \leq \alpha/2$$

and $$\sum_{0}^{r'_{\alpha/2}} {}^nC_x \left(\frac{1}{2}\right)^n \leq \alpha/2$$

From binomial tables, we find that $r_{0.025} = 18$ and $r'_{0.025} = 6$ for $n = 24$ and $p = 0.5$. Thus, for $r = 11$ null hypothesis is to be accepted.

Note: The critical region for one-sided test alternative

$$\sum_{r_\alpha}^{24} {}^{24}C_x \left(\frac{1}{2}\right)^{24} \leq 0.05$$

since under the alternative hypothesis the sample will have an excess of plus signs. In the case of the other one-sided alternative, viz.

$$H : \theta < 9.9 \text{cm. or } H : p < 0.5$$

The critical region for the level α will be $r \leq r'_\alpha$, where r'_α is the largest integer such that

$$\sum_{0}^{r_\alpha} {}^{24}C_x \left(\frac{1}{2}\right)^{24} \leq \alpha$$

Example. The weights of 12 persons before they are subjected to a change of diet and after a lapse of six months are recorded below:

S. No.	Weight (in kg.)	
	Before	After
1	57	62
2	48	55
3	55	62
4	45	53
5	62	59
6	42	45
7	49	45
8	60	55
9	65	64
10	51	55
11	46	50
12	58	66

Test whether there has been any significant gain in weight as a result of the change of diet.

Solution: Let y and x be the weight of a person before and after the change of diet, then the hypothesis to be tested is $H_0 : \theta = 0$ and the alternative is $H : \theta > 0$, where θ is the median of the distribution of differences d_i. The gain in weight (d_i) for 12 persons are:

$$+5, +7, +7, +8, -3, +3, -4, -5, -1, +4, +6, -8$$

Here, the no. of plus signs $= 7$ and the no. of minus signs $= 5$. Under the null hypothesis, the expected number of plus signs among the differences in a sample of 12 pairs is 6. The sampling distribution of the number of plus signs is the binomial distribution with probability of plus signs 0.5. From table, we find that the probability of 7 or more plus signs is 0.387. So the null hypothesis is accepted at the 5% level.

7.13.1.4 Merits and Demerits

Merits

- It is very simple to calculate.
- It requires minimum effort for calculation.

Demerits:

- The disadvantage of the sign test is that, although it takes account of signs of the deviations, it makes no allowance for their magnitudes.

7.13.2 WILCOXON TEST

Now discuss

7.13.2.1 One Sample Wilcoxon Signed-Rank Test

It is a non-parametric test for the location parameter (median) of a population.

7.13.2.2 Hypothesis and Assumptions

In this test, we make the assumption of independence and homoscedasticity but do not assume normality for the parent population. Also, if we assume that the parent population is continuous and symmetric, the Wilcoxon signed rank test is more efficient than the sign test for testing median of the population, since it takes into account both the magnitudes and signs of the deviations.

We wish to test the null hypothesis

$H_0 : M = M_0$ (a given value)

against

a: one-sided alternative

$H_1 : M < M_0$ (left tailed test)

$H_1 : M > M_0$ (right tailed test)

or b: two-sided alternative

$H_1 : M \neq M_0$

7.13.2.3 Test Procedure

Let $x_1, x_2, ..., x_n$ be a random sample of size n drawn from a population which is continuous and symmetric about median *M*. Then, under H_0 , the differences $D_i = X_i - M_0, \ \forall i = 1, 2, ..., n$ are symmetrically distributed about zero, so that the positive and negative differences of the equal absolute value have the same probability of occurrence. Thus,

$$P(D_i \geq C) = P(D_i \geq -C)$$

or $P(D_i \geq C) = 1 - P(D_i \leq C)$

Suppose we order these absolute differences $|D_1|, |D_2|, ..., |D_n|$ from smallest to largest and assign them ranks $i = 1, 2, ..., n$. Let T^+ be the sum of ranks of the positive D_i and T^- be the sum of the ranks of the negative D_i.

If H_0 is true (i.e. M_0 is the true median of the symmetrical population), then expectation of T^+ equals the expectation of T^-. Since the sum of all the ranks is a constant given by

$$T^+ + T^- = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

The tests based on T^+, T^- and T^-, T^+ will be equivalent (since they are linearly related). In practice, the minimum of T^+ and T^- is used as the test statistic.

Let us define a new random variable:

$$D_{(i)} = \begin{cases} 1 & \text{if } \mathrm{D_i} > 0 \, \mathrm{for} \, \mathrm{i}^{\mathrm{th}} \text{ smallest} |\mathrm{D_i}| \\ 0 & \text{if } \mathrm{D_i} < 0 \, \mathrm{for} \, \mathrm{i}^{\mathrm{th}} \text{ smallest} |\mathrm{D_i}| \end{cases}$$

$D_{(i)}$ are independent Bernoulli random variables but are not identically distribute such that

$E\left[D_{(i)}\right] = p_i$

$V\left[D_{(i)}\right] = p_i\left(1 - p_i\right)$

and $\operatorname{cov}\left[D_{(i)}, D_{(j)}\right] = 0,\ i \neq j$

We can write

$T^+ = \sum_{i=1}^{n} iD_{(i)}$ and $T^- = \sum_{i=1}^{n} i\left[1 - D_{(i)}\right]$

Thus $E\left[T^+\right] = \sum_{i=1}^{n} iE\left[D_{(i)}\right] = \sum_{i=1}^{n} ip_i$

and $V\left[T^+\right] = \sum_{i=1}^{n} i^2 V\left[D_{(i)}\right] = \sum_{i=1}^{n} i^2 p_i\left(1 - p_i\right)$

under H_0, i.e. when $p_i = \frac{1}{2}$

$E\left[T^+\right] = \sum_{i=1}^{n} i\left(\frac{1}{2}\right) = \frac{1}{2}\sum_{i=1}^{n} i = \frac{n(n+1)}{4}$

and $V\left[T^{+}\right]=\sum_{i=1}^{n} i^{2}\left(\frac{1}{2}\right)\left(1-\frac{1}{2}\right)=\frac{1}{4} \sum_{i=1}^{n} i^{2}=\frac{n(n+1)(2n+1)}{24}$

Similarly, for T^{-},

Let $T=\min\left[T^{+}, T^{-}\right]$ and T_{α} be such that $P\left[T \leq T_{\alpha}\right]=\alpha$.

Then the critical regions for α level of significance for testing $H_0: M=M_0$ against different types of alternatives are given as

Alternative Hypothesis	Critical Region
$H_1: M>M_0$	$T^{-} \leq T_{\alpha}$
$H_1: M<M_0$	$T^{+} \leq T_{\alpha}$
$H_1: M \neq M_0$	$T^{+} \leq T_{\alpha/2}$ or $T^{-} \leq T_{\alpha/2}$

If $n>25$, then distribution of T is taken to be approximation normal i.e. under H_0 we have

$$Z=\frac{T-E[T]}{\sqrt{V[T]}} \rightarrow N(0,1)$$

where $T=\min\left[T^{+}, T^{-}\right]$ and

$$E[T]=\frac{n(n+1)}{4}$$

$$V[T]=\frac{n(n+1)(2n+1)}{24}$$

Also, the sample size n is adjusted to include only non-zero differences.

Example. Test the null hypothesis that the median length θ of ear-head of a variety of wheat is $\theta_0=9.9$ cm. against the alternative that $\theta_0 \neq 9.9$ cm., with $\alpha=0.05$, on the basis of the following 25 ear-head measurements:

9.5	8.9	10.5	11.5	8.5	9.4	10.6	8.8	11.7	10.5
11.2	9.2	9.8	9.5	9.9	10.9	10.2	9.1	10.8	9.4
11.6	8.7	8.3	11.3	8.1					

Solution: First, we determine

S.no.	x_i	$d_i = x_i - \theta_0$	Rank of $\|d_i\|$
1	9.5	-0.4	3.5
2	8.9	-1	13.5
3	10.5	0.6	7.5
4	11.5	1.6	20.5
5	8.5	-1.4	18.5
6	9.4	-0.5	5.5
7	10.6	0.7	9.5
8	8.8	-1.1	15
9	11.7	1.8	23.5
10	10.5	0.6	7.5
11	11.2	1.3	17
12	9.2	-0.7	9.5
13	9.8	-0.1	1
14	9.5	-0.4	3.5
15	9.9	0	
16	10.9	1	13.5
17	10.2	0.3	2
18	9.1	-0.8	11
19	10.8	0.9	12
20	9.4	-0.5	5.5
21	11.6	1.7	22
22	8.7	-1.2	16
23	8.3	-1.6	20.5
24	11.3	1.4	18.5
25	8.1	-1.8	23.5

Here $T^+ = 153.5$, $T^- = 146.5$, so that $T = 150$ From table, for $n = 24$ and $\alpha = 0.05$, we have $T_\alpha = 81$. Since T^+ and T^- are both greater than T_α, there is not sufficient evidence to reject H_0.

In the case of the one-sided alternative $H : \theta < 9.9$ cm. ($H : \theta > 9.9$ cm.), we shall compare $T^+ = 153.5$ $\left(T^- = 146.5\right)$ with the critical value $T_\alpha = 81$, at $\alpha = 0.025$, and arrive at same conclusion that there is no ground for rejecting H_0 (in favour if the appropriate on-sided alternative) since $T > T_\alpha$.

Example. The weights of 12 persons before they are subjected to a change of diet and after a lapse of six months are recorded below:

S.no.	Weight (in kg.)	
	Before	After
1	57	62
2	48	55
3	55	62
4	45	53
5	62	59
6	42	45
7	49	45
8	60	55
9	65	64
10	51	55
11	46	50
12	58	66

Test whether there has been any significant gain in weight as a result of the change of diet.

Solution: Let y and x be the weight of a person before and after the change of diet, then the hypothesis to be tested is $H_0 : \theta = 0$ and the alternative is $H : \theta > 0$, where θ is the median of the distribution of differences d_i. The gain in weight (d_i) and the absolute rank for 12 persons are:

S.no.	Weight (in kg.)		$d_i = x_i - y_i - \theta_0$	Rank of $\lvert d_i \rvert$
	y_i	x_i		
1	57	62	+5	6.5
2	48	55	+7	9.5
3	55	62	+7	9.5
4	45	53	+8	11.5
5	62	59	-3	2.5
6	42	45	+3	2.5
7	49	45	-4	4.5
8	60	55	-5	6.5
9	65	64	-1	1
10	51	55	+4	4.5
11	46	50	+6	8
12	58	66	-8	11.5

Here, $T^{+} = 52$ and $T^{-} = 26$; here T^{-} will be used. From table, we have, for $n = 12$ and $\alpha = 0.01$ (one-sided), $T_{\alpha} = 10$. Since $T^{-} > T_{\alpha}$, therefore we conclude that there is no sufficient evidence to reject the null hypothesis that there is no effect of diet in favour of the alternative hypothesis at the 1% level.

7.13.2.4 Merits and Demerits

The Wilcoxon Signed rank test takes into account the magnitude of the deviations.

As one of the assumptions made here is that intendance of observations continuity everywhere and symmetry which is not practically possible all the time.

7.13.2.5 Comparision of Sign Test and Wilcoxon Signed Rank Test

1. In sign test, the assumptions required are independence of observations and the population is continuous at media. In Wilcoxon signed rank test, the assumptions required are the population is continuous everywhere and it is symmetric about median.

2. In sign test, we consider only the directions of the deviations while in Wilcoxon signed rank test, we consider directions of the deviations as well as the magnitudes of the directions. Thus, Wilcoxon signed rank test is more efficient than the sign test.

3. Both the tests are useful generally for the same type of problem. But only Wilcoxon signed test is suitable for a test of symmetry as well.

7.13.3 MEDIAN TEST

If $N = m + n$ is even then

Median $=$ any number between $\frac{N}{2}$th and $\left(\frac{N+2}{2}\right)$th order statistic

Let U be the number of X sample observations that are less than the sample median for the combined sample.

The test based on U , the number of observations from X sample median which are less than the combined sample median, is called the sample median. Then

$$t = \begin{cases} \frac{N-1}{2}, & if \text{ N is odd} \\ \frac{N}{2}, & if \text{ N is even} \end{cases}$$

The probability distribution of U for fixed t is

$$f(u) = \frac{{}^{m}C_{u}\,{}^{n}C_{t-u}}{{}^{m+n}C_{t}};\ u = 0,1,2,........,t$$

where $t = \frac{N}{2}$.

If H_0 is true, then $P(X < M) = P(X > M); \forall M$ and here M is combined sample median. i.e. the two populations have a common median which is estimated by M .

7.13.3.1 Hypothesis and Assumptions

The general location alternative is

$$H_L : F_X(x) = F_Y(x-\theta); \text{ for some } x \;\&\; \theta \neq 0$$

if U is too large, then

$H_L : F_X(x) \geq F_Y(x)$; if $\theta > 0$ and $\forall x$

i.e. $H_L : F_X(x) > F_Y(x)$; if $\theta > 0$ and for some x

i.e. the median of the X population is smaller than the median of Y population.

If U is too small, then

$H_L : F_X(x) \leq F_Y(x)$; if $\theta < 0$ and $\forall x$

i.e. $H_L : F_X(x) < F_Y(x)$; if $\theta < 0$ and for some x

i.e. the median of the X population is greater than the median of Y population.

The critical region for α level of significance is given as

Alternative Hypothesis	Critical Region
$\theta > 0$ or $M_X < M_Y$	$u \geq c'_{\alpha}$
$\theta < 0$ or $M_X > M_Y$	$u \leq c_{\alpha}$
$\theta \neq 0$ or $M_X \neq M_Y$	$u \leq c_{\frac{\alpha}{2}}$ or $u \geq c'_{\frac{\alpha}{2}}$

7.13.3.2 Test Procedure

1. Consider the observations in the order in which they are obtained.
2. Determine the median of those observations i.e. determine the sample median M.
3. For each observation note that whether it is above or below the sample median. Denote the observation below the sample median M by B or (-) sign and observations above the sample median M by A or (+) sign . The zero values will be ignored.
4. Denote the number of minus signs or the numbers by B's by n_1 and the number of plus signs or the number of A's by n_2 .
5. Count the number of runs and denote this number by R.

6. Reject the null hypothesis H_0 the sample is random if

$R \geq R_1$ or $R \leq R_u$.

where R_1 and R_u are critical of R to be determined from the distribution of R n_1 and n_2. The critical values of R required for significance have been have been tabulated.

Example: Suppose in a random sample of size 30, there 12 runs above and below the sample median where n_1=number of minus (-) sings=10

n_2= number of minus (+) sings= 20

Test the hypothesis the sample is random.

Solution

R= Number of runs above and below the sample median =12

n_1=number of minus (-) sings=10

n_2= number of minus (+) sings= 20

from table the lower critical value of R, R_1=9

the upper critical values of R, R_u=20

since $9 < R < 20$

the hypothesis of randomness is accepted at 5% level of signifance. i.e. sample is random.

7.13.3.3 Merits and Demerits

Median test when the sample observations are divided into two types on the basis of deviations from sample median.

7.14 Summary

This unit provides a thorough understanding of concepts related to Location Tests pertaining to one sample. The concepts of non- parametric methods, sign test, Wilcoxon test, median test are described in detail. The learner should try to solve the self-assessment problems given in the next section.

7.15 Self-Assessment Exersises

Q1. What do understand by sign test. Describe the procedure by giving an example.

Q2. Describe Wilcoxon test by giving a suitable example.

Q3. Explain the procedure median test stating clearly the assumptions. Give an example for illustration.

UNIT:8 OTHER NON- PARAMETRIC TESTS

8.1 Introduction

This section disucsses some other non parameteric test like one sample and two sample location test, one sample and two sample Kolmagorov-Smirnov tests and run test.

8.2 Objectives

The objective of this unit is to provide a basic understanding of concepts related to some Other non- parametric tests. The concept of the Mann- Whitney U- Test, Application of U-statistic to rank tests. One sample and two sample Kolmagorov-Smirnov tests. Run tests should be clear after study of this material.

8.3 One Sample Location Tests

There are some following tests are given.

8.3.1 MANN-WHITNEY U TEST

Now Discuss

8.3.1.1 Introduction and Assumptions

Mann Whitney U test is a non-parametric test for testing that the two populations differ in their location. It is useful to the t -test, if the assumption of t -test are violated, we use Mann Whitney U test. We assume that the two samples are drawn from continuous populations.

Let we have two populations X and Y with cumulative distribution functions F_X and F_Y respectively. A random sample of size m is drawn from the X population and another random sample of size n is drawn from the Y population, denoted as

$$X_1, X_2, \ldots. X_m \text{ and } Y_1, Y_2, \ldots. Y_n$$

These $N = m + n$ observations drawn from two populations are arranged in order of magnitude from smallest to largest.

Like run test, this test is based on the idea that the particular pattern is exhibited when m observations of X random variable and n observations of Y random variables are arranged together in increasing order of magnitude.

The test criterion is based on the positions of Y's in the combined ordered sequence. A sample pattern where most of the Y's are greater than the most of the X's or vice-versa can be used as statistical criteria for rejection of null hypothesis of identical distribution.

Since, in this case, we see that there is no random missing in the sample observation. The Mann Whitney U statistic is defined as the number of times Y proceeds X in the combined ordered arrangement of two independent random samples.

8.3.1.2 Test Procedure

If mn random variable is defined as

$$D_{ij} = \begin{cases} 1 & if\ Y_j < X_i; \quad \forall i = 1, 2, \ldots.., m \\ 0 & if\ Y_j > X_i; \quad \forall j = 1, 2, \ldots.., n \end{cases}$$

Thus Mann-Whitney U statistic is defined as

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij}$$

We wish to test the null hypothesis

$$H_0 : F_X(x) = F_Y(x) \ ; \ \forall x$$

i.e. two samples are drawn from the identical populations.

The general location alternative is

$$H_L : F_X(x) = F_Y(x - \theta); \ \text{ for some } x \ \& \ \theta \neq 0$$

If U is too large, then

$$F_X(x) \geq F_Y(x); \ \forall x \text{ and if } \theta > 0$$

i.e. $F_X(x) > F_Y(x)$; for some x if $\theta > 0$

If U is too large, then

$$F_X(x) \leq F_Y(x); \ \forall x \text{ and if } \theta < 0$$

i.e. $F_X(x) < F_Y(x)$; for some x if $\theta > 0$

We define,

$$\pi = P(D_{ij} = 1) = P(Y < X)$$

$$= P[-\infty < X < \infty, -\infty < Y < X]$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{x} f(y) f(x) \, dy \, dx$$

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

Under H_0, i.e. $H_0 : F_X(x) = F_Y(x)$

Then

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

For solving above integration, let $F_X(x) = v$ and differentiate this equation w.r.t x, we get $f(x) = dv$. Also, the limits changes as $x = -\infty \Rightarrow F(-\infty) = 0 = v$ and $x = \infty \Rightarrow F(\infty) = 1 = v$

Therefore, integral reduces to

$$\pi = \int_0^1 v \ dv = \left[\frac{v^2}{2}\right]_0^1 = \frac{1}{2}$$

Hence $H_0 : F_X(x) = F_Y(x)$ or $H_0 : \pi = \frac{1}{2}$

Also $H_L : F_Y(x) \geq F_X(x)$; $\forall x$

Is equivalent to $H_L : \pi \geq \frac{1}{2}$; $\forall x$

i.e. $H_L : F_Y(x) > F_X(x)$; for some x

is equivalent to $H_L : \pi > \frac{1}{2}$; for some x

and $H_L : F_Y(x) \leq F_X(x)$; $\forall x$

is equivalent to $H_L : \pi \leq \frac{1}{2}$; $\forall x$

i.e. $H_L : F_Y(x) < F_X(x)$; for some x

is equivalent to $H_L : \pi < \frac{1}{2}$; for some x

The mn random variables D_{ij} are Bernoulli variables, with parameter π. i.e.

$$E\left[D_{ij}\right] = E\left[D_{ij}^2\right] = \pi$$

$$V\left[D_{ij}\right] = \pi(1-\pi)$$

We define the parameters π_1 and π_2 as,

$$\pi_1 = P\left(Y_j < X_i \cap Y_k < X_i\right) = \int_{-\infty}^{\infty} \left[F_Y(x)\right]^2 f(x)dx$$

and

$$\pi_2 = P\left(X_i > Y_j \cap X_h < X_j\right) = \int_{-\infty}^{\infty} \left[1 - F_Y(x)\right]^2 f(y)dy$$

Since $U = \sum_{i=1}^{m}\sum_{j=1}^{n} D_{ij}$

Therefore, mean and variance of U are defined as

$$E[U] = \sum_{i=1}^{m}\sum_{j=1}^{n} E\left[D_{ij}\right] = \sum_{i=1}^{m}\sum_{j=1}^{n} \pi$$

$$E[U] = mn\pi$$

and

$$V[U] = V\left(\sum_{i=1}^{m}\sum_{j=1}^{n} D_{ij}\right)$$

$$= mn\pi(1-\pi) + mn(n-1)(\pi_1 - \pi^2) + mn(m-1)(\pi_2 - \pi^2)$$

$$= mn\left[\pi - \pi^2 + (n-1)(\pi_1 - \pi^2) + (m-1)(\pi_2 - \pi^2)\right]$$

$$= mn\left[\pi - \pi^2 + (m+n-1) + (n-1)\pi_1 + (m-1)\pi_2\right]$$

$$V[U] = mn\left[\pi - \pi^2 + (N-1) + (n-1)\pi_1 + (m-1)\pi_2\right]$$

as $m,n \to \infty$

$E[U/mn] = \pi$ and $V[U/mn] \to 0$

Hence U/mn is a consistent estimator of π .

If we define another random variable

$$U' = \sum_{i=1}^{m}\sum_{j=1}^{n}(1 - D_{ij})$$

The critical region for α level of significance is given as

Alternative hypothesis	Critical region
$\pi \le \frac{1}{2}$ or $F_Y(x) \le F_X(x)$	$U \le C_\alpha$
$\pi \ge \frac{1}{2}$ or $F_Y(x) \ge F_X(x)$	$U' \le C_\alpha$
$\pi \ne \frac{1}{2}$ or $F_Y(x) \ne F_X(x)$	$U \le C_{\frac{\alpha}{2}}$ or $U' \le C_{\frac{\alpha}{2}}$

Under H_0, i.e. $H_0 : F_X(x) = F_Y(x)$

Then $\pi = \frac{1}{2}$

and $\pi_1 = \pi_2 = \dfrac{1}{3}$

Thus $E[U] = \dfrac{mn}{2}$

$$V[U] = mn\left[\frac{1}{2} - \frac{1}{4}(N-1) + \frac{(n-1)}{3} + \frac{(m-1)}{3}\right]$$

$$V[U] = mn\left[\frac{1}{12} - \frac{N}{4} + \frac{1}{3}(n+m)\right]$$

$$= mn\left[\frac{1}{12} - \frac{N}{4} + \frac{N}{3}\right]$$

$$V[U] = \frac{mn(N+1)}{12}$$

If N is large, then under H_0

$$Z = \frac{U - E[U]}{\sqrt{V[U]}} = \frac{U - \dfrac{mn}{2}}{\sqrt{\dfrac{mn(N+1)}{12}}} \to N(0,1)$$

Example: The following are the marks secured by two batches of salesmen in the final test taken after completion of training. Use the U-test with $\alpha = 0.02$ for the null hypothesis that the samples are drawn from identical distributions against the alternative that the distributions differ in location only.

Batch A: 28, 25, 27, 29, 25, 19, 23, 26, 30, 22, 21, 28

Batch B: 20, 24, 25, 26, 18, 28, 23

Solution: Here $n_1 = 7$, $n_2 = 12$ and $N = n_1 + n_2 = 12$

$U = 51$, $U^{'} = 26$

where U is the number of times x_i precedes y_j among all (x_i, y_j) pairs and $U^{'}$ is the number of times y_j precedes x_i among all (x_i, y_j) pairs assuming no $x = y$ ties. From table, we find that for two-tail test $n_1 = 7$ and $n_2 = 12$ at the level 0.02, the critical value is 14. Since 20(the smaller of U and U') is greater than 14, so we have no reason to believe that the samples are not drawn from identical distribution.

8.3.1.3 Merits and Demerits

It is a good substitute for t-test when the conditions imposed on parent populations are not met.

8.3.1.4 Application of U-Statistic to Rank Tests

Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be two random samples from two populations with continuous cdf $F_X(x)$ and $F_Y(x)$ respectively. Now to test **H₀**: $F_X(x) = F_Y(x)\ \forall\ x$, the two samples are combined and arranged in an ascending order of magnitude. The ranks of the *X* values in the combined ordered arrangement of the two samples would generally be larger than the ranks of the *Y* values if the median of the *X* population exceeds the median of the *Y* population. Hence Wilcoxon (1945) proposed a test, known as *Wilcoxon rank sum test*, where the null hypothesis is rejected if the sum of the ranks of the *X* values is too large if the alternative is $\mathbf{H_1}$: $F_X(x) > F_Y(x)$ for some *x*, or if the sum of the ranks of the *X* values is too small if the alternative is $\mathbf{H_1}$: $F_X(x) < F_Y(x)$ for some *x*, or if the sum of the ranks of the *X* values is too large or too small if the alternative is a two- sided, *i.e.* $\mathbf{H_1}$: $F_X(x) \neq F_Y(x)$ for some *x*. The test statistic of this test for $N = m + n$, is defined as

$W_N = \sum_{i=1}^{N} iZ_i$,

where $Z_i, i = 1, 2, \cdots, N$, are indicator random variables defined as

$$Z_i = \begin{cases} 1 \text{ if the } i^{th} \text{random variable in the combined ordered sample is } X \\ 0 \text{ if the } i^{th} \text{random variable in the combined ordered sample is } Y. \end{cases}$$

This test is actually same as the *Mann-Whitney U test*, since a linear relationship exists between the two test statistics. As *U* is defined as the number of times a *Y* value precedes an *X* value, we have

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij} = \sum_{i=1}^{m} (D_{i1} + D_{i2} + \cdots + D_{in}),$$

where $D_{ij} = \begin{cases} 1 \text{ if } Y_j < X_i \\ 0 \text{ if } Y_j > X_i \end{cases}$ for all $i = 1,2,\cdots,m$ and $j = 1,2,\cdots,n$.

Then $\sum_{j=1}^{n} D_{ij}$ is the number of values of j for which $Y_j < X_i$ or the rank of X_i reduced by n_i, the number of X values which are less than or equal to X_i. For example consider the sequence $YYXXY$, where m =2 and n =3. Here, D_{11}= 1, D_{12}= 1, D_{13} = 0, D_{21} = 1, D_{22}= 1, D_{23}= 0.

$\therefore \sum_{j=1}^{3} D_{1j}$ = 1 + 1 + 0 = 2 = rank$(X_1) - 1$ = 3 – 1 = 2 (here $n_1 = 1$)

and $\sum_{j=1}^{3} D_{2j}$= 1 + 1 + 0 = 2 = rank$(X_2) - 2$ = 4 – 2 = 2 (here $n_2 = 2$).

Thus, if $r(X_i)$ denotes the rank of X_i, then

$$U = \sum_{i=1}^{m}[r(X_i) - n_i] = \sum_{i=1}^{m} r(X_i) - (n_1 + n_2 + \cdots + n_m)$$

$$= \sum_{i=1}^{N} iZ_i - \frac{m(m+1)}{2} = W_N - \frac{m(m+1)}{2}.$$

Hence, all the properties of the tests are the same, including consistency and the minimum asymptotic relative efficiency relative to t-test. Also, we have

$E(U|\mathrm{H}_0) = mn/2$ and $V(U|\mathrm{H}_0) = mn(N + 1)/12$. Hence,

$E(W_N|\mathrm{H}_0) = E(U|\mathrm{H}_0) + \frac{m(m+1)}{2}$ and $V(W_N|\mathrm{H}_0) = V(U|\mathrm{H}_0) = mn(N + 1)/12$.

Just as U, the statistic W_N is also symmetric about its mean.

8.3.1.5 Test of Goodness of Fit

This type of test is designed for a null hypothesis which is a statement about the form of the cumulative distribution function or probability function of the parent population from which the sample is drawn.

Let a random sample of size n is drawn from a population with unknown cumulative distribution function say F. We want to test the null hypothesis

$$H_0 : F(x) = F_0(x)\ ;\ \forall x$$

against the alternative hypothesis

$$H_1 : F(x) \neq F_0(x)\ ;\ \text{for some } x$$

If F_0 is specified with all its parameters, then H_0 is a simple hypothesis. If F_0 is not completely specified, then H_0 is a composite hypothesis and the unknown parameters are to be estimated from the sample data in order to perform any test. The alternative hypothesis in both the cases will be composite therefore rejection of H_0 does not provide any result.

8.3.2 One Sample Kolmogorov-Smirnov (K-S) Test

Goodness of fit tests are used when only the form of the population is in question, with the hope that the null hypothesis will be found accepted. The two types of goodness of fit tests are:

1. Chi Square goodness of fit test
2. Kolmogrov Siminirov test

8.3.2.1 Chi Square Goodness of Fit Test

Hypothesis and Assumptions:

If a random sample of size n is drawn from a population with unknown cumulative distribution function F.

We wish to test the null hypothesis

$H_0 : F(x) = F_0(x)\ ;\ \forall x$

against the alternative hypothesis

$H_1 : F(x) \neq F_0(x)\ ;\ \text{for some } x$

In order to apply the chi-square test in continuous distribution, the sample data must be grouped according to some scheme in order to form a frequency distribution.

Assuming that the population distribution F_0 is completely specified by the null hypothesis H_0, we can obtain the probability p_i that a random observation will be classified in the i^{th} category $(i = 1, 2, \ldots, k)$.

These probabilities multiplied by n, the sample size, give the expected frequencies under H_0. i.e.

$$E_i = np_i,\ (i = 1, 2, \ldots, k)$$

Let the n observations have been grouped into k mutually exclusive categories, O_i and E_i are the observed and expected frequencies respectively, for the i^{th} group $(i = 1, 2, \ldots, k)$.

We use the test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(O_i - E_i\right)^2}{E_i} \tag{1}$$

with $\sum_{i=1}^{k} O_i = \sum_{i=1}^{k} E_i$

The exact sampling distribution of this test statistic is complicated. But for large samples, it has χ^2 distribution with $(k-1)$ degree of freedom. This approximation is good for every $E_i \geq 5$. For $E_i < 5$, we combine the adjacent categories till the expected frequency in the combined category is at least 5.

If

$cal\,\chi^2 > tab\,\chi^2_{\alpha,(k-1)}$

then H_0 is rejected at α level of significance.

If F_0 is completely specified, then H_0 is a composite hypothesis and the unknown parameters are to be estimated from the sample data in order to perform the test.

In this case, the test statistic described by (1) has χ^2 distribution with $(k-r-1)$ degree of freedom, where r is the number of independent parameters of F_0 estimated from the sample data.

Thus $mS_m(x)$ is the number of X sample observations that are less than or equal to x. And $nT_n(x)$ is the number of Y sample observations that are less than or equal to x.

For large m and n, the deviations between two empirical distribution functions, $\left|S_m(x) - T_n(x)\right|$ should be small for all values of x.

Thus, the test statistic

$$D_{m,n} = \max_x \left|S_m(x) - T_n(x)\right|$$

is called Kolmogrov-Smirnov two sample test statistic.

The probability distribution of $D_{m,n}$ does not depend upon F_X and F_Y as long as F_X and F_Y are continuous.

Therefore, $D_{m,n}$ may be called a distribution free statistic.

The directional deviations are defined as

$$D_{m,n}^{+} = \max_{x} \left| S_m(x) - T_n(x) \right|$$

$$D_{m,n}^{-} = \max_{x} \left| T_n(x) - S_m(x) \right|$$

$D_{m,n}^{+}$ and $D_{m,n}^{-}$ are called one-sided kolmogrov-smirnov statistic. These are also distribution free.

We wish to test the null hypothesis

$$H_0 : F_X(x) = F_Y(x) \;\; ; \;\; \forall x$$

i.e. under H_0, the population distributions are identical and we have two samples from the sample population.

against

(i) One sided alternative

$H_1 : F_X(x) > F_Y(x) \;\; ; \;\; \forall x$ (right tailed test)

The appropriate test statistic is

$$D_{m,n}^{+} = \max_{x} \left| S_m(x) - T_n(x) \right|$$

or

$H_1 : F_X(x) < F_Y(x) \;\; ; \;\; \forall x$ (left tailed test)

The appropriate test statistic is

$$D_{m,n}^{-} = \max_{x} \left| T_n(x) - S_m(x) \right|$$

8.3.2.2 Comparison of Chi-square test with Kolmogrov-Siminirov test for Goodness of Fit

1. Both types of tests are distribution free because the sampling distribution of the test statistic does not depend on the cumulative distribution function.

2. The chi-square tests are specially designed for use with categorical data, while K-S tests are for random samples from the continuous populations.

3. The chi-square test is sensitive to vertical deviations between the observed and expected histograms, whereas the K-S test is based on vertical deviations between the observed and expected cumulative distribution functions.

4. K-S test can be applied for any sample size, while chi-square test can be applied for large sample size when each expected cell frequency is not too small.

5. The advantage of K-S test is that the exact sampling distribution of K-S test statistic is known and tabulated, whereas the sampling distribution of chi-square test statistic is approximately chi-square for finite sample size.

6. When H_0 is composite, the chi-square test is easily modified by reducing the number of degrees of freedom (as some parameters are estimated) while K-S test can't be modified in the situation.

7. The K-S test is more powerful and more flexible than the chi-square test.

8. The chi-square test also comes in the category of parametric tests whereas K-S test is only a non-parametric.

9. In K-S test, we can use one side test also which is not possible chi-square test.

8.3.3 RUN TEST

If we are given an ordered sequence of two or more types of symbols, a run is defined to be a succession of one or more identical symbol which are followed and proceed by a different symbol or no symbol at all.

In any situation, if the sample observations may not behave random, it is necessary to test the randomness of the sequence before the usual statistical methods based on randomness are applied.

Too few runs, too many runs, a run of excessive length or too many runs of excessive length etc. can be used as statistical criteria for rejection of the null hypothesis of randomness, since these situations should occur rarely in a truly random sequence.

A null hypothesis of randomness would be rejected if the total number of runs is either too small or too large.

8.3.3.1 Advantages

1. Test of randomness are an important addition to the statistical theory, because almost all the classical statistical techniques are based on the assumption of a random sample.
2. The run tests are applicable to either qualitative or quantitative data.

8.3.3.2 Distribution of Runs

Let us suppose an ordered sequence of n elements of two types, n_1 of the first type i.e. values of x and n_2 of the second type i.e. the values of y such that

$n_1 + n_2 = n$.

If $r_1 = \text{number of runs of type 1st elements i.e. X's}$

$r_2 = \text{number of runs of type 2nd elements i.e. Y's}$

The total number of runs in this sequence is

$r_1 + r_2 = r \quad ; \quad r \le n$

The probability distribution of the random variable 'R' is obtained as follows:

We can select n_1 positions for the n_1 values of X from $(n_1 + n_2)$ positions in ${}^{n_1+n_2}C_{n_1}$ ways.

The probability of each arrangement $= \dfrac{1}{{}^{n_1+n_2}C_{n_1}}$

Now, we have to determine how many of these arrangements yield $R = r$. Here, two cases arise:

Case (i): When r is odd i.e. $r = 2k + 1$; $k \in I^+$ i.e. there are $(k+1)$ runs of ordered values of X and k runs of ordered values of Y or vice-versa.

First we consider the number of ways of obtaining $(k+1)$ runs of n_1 values of X. This can be done in ${}^{n_1-1}C_k$ ways.

Similarly, we consider the number of ways of obtaining k runs of n_2 values of Y. This can be done in ${}^{n_2-1}C_{k-1}$ ways.

The joint operation can be performed in $\left({}^{n_1-1}C_k\right)\left({}^{n_2-1}C_{k-1}\right)$ ways.

Secondly, considering the number of ways of obtaining $(k+1)$ runs of n_2 values of Y. This can be done in ${}^{n_2-1}C_k$ ways.

Similarly, we consider the number of ways of obtaining k runs of n_1 values of X. This can be done in ${}^{n_1-1}C_{k-1}$ ways.

The joint operation can be performed in $\left({}^{n_1-1}C_{k-1}\right)\left({}^{n_2-1}C_k\right)$ ways.

Thus,

$$P(r=2k+1)=\frac{\left({}^{n_1-1}C_k\right)\left({}^{n_2-1}C_{k-1}\right)+\left({}^{n_1-1}C_k\right)\left({}^{n_2-1}C_{k-1}\right)}{\left({}^{n_1+n_2}C_{n_1}\right)}$$

Case (ii): When r is even i.e. $r=2k\ ;\ k\in I^+$ i.e. there are k runs of ordered values of X and k runs of ordered values of Y or vice-versa.

First we consider the number of ways of obtaining k runs of n_1 values of X. This can be done in ${}^{n_1-1}C_{k-1}$ ways.

Similarly, we consider the number of ways of obtaining k runs of n_2 values of Y. This can be done in ${}^{n_2-1}C_{k-1}$ ways.

The joint operation can be performed in $\left({}^{n_1-1}C_{k-1}\right)\left({}^{n_2-1}C_{k-1}\right)$ ways.

Secondly, considering the number of ways of obtaining k runs of n_2 values of Y. This can be done in ${}^{n_2-1}C_{k-1}$ ways.

Similarly, we consider the number of ways of obtaining k runs of n_1 values of X. This can be done in ${}^{n_1-1}C_{k-1}$ ways.

The joint operation can be performed in $\left({}^{n_1-1}C_{k-1}\right)\left({}^{n_2-1}C_{k-1}\right)$ ways.

Thus, $P(r=2k)=\dfrac{2\left({}^{n_1-1}C_{k-1}\right)\left({}^{n_2-1}C_{k-1}\right)}{\left({}^{n_1+n_2}C_{n_1}\right)}$

Thus the probability distribution of R, the total number of runs of $n_1 + n_2 = n$ objects, n_1 of type 1st and n_2 of type 2nd, is given as:

$$f(x) = \begin{cases} \dfrac{2\left({}^{n_1-1}C_{r/2-1}\right)\left({}^{n_2-1}C_{r/2-1}\right)}{\left({}^{n_1+n_2}C_{n_1}\right)} & ;\ \text{if } r \text{ is even} \\ \dfrac{\left({}^{n_1-1}C_{r-1/2}\right)\left({}^{n_2-1}C_{r-3/2}\right)+\left({}^{n_1-1}C_{r-3/2}\right)\left({}^{n_2-1}C_{r-1/2}\right)}{\left({}^{n_1+n_2}C_{n_1}\right)} & ;\ \text{if } r \text{ is odd} \end{cases}$$

where $r = 2, 3, \ldots, n_1 + n_2$

8.3.3.3 Test of Randomness

Sometimes, it is desirable to test whether the sample observations can be regarded as random or not. To test the randomness of the sample observations, we use run test.

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n taken from continuous distribution. In the given sequence $X_1, X_2, \ldots, X_n$ for each observation we note whether it is above or below the sample median.

8.3.3.4 Hypothesis and Assumptions

Run test is used for examining whether or not a set of observations constitutes a random sample from an infinite population. Test for randomness is of major importance because the assumption of randomness underlies statistical inference. In addition, tests for randomness are important for time series analysis. Departure from randomness can take many forms.

H_0: Sample values come from a random sequence

H_1: Sample values come from a non-random sequence.

8.3.3.5 Test Procedure

Let r be the number of runs (a run is a sequence of signs of same kind bounded by signs of other kind). For finding the number of runs, the observations are listed in their order of occurrence. Each observation is denoted by a '+' sign if it is more than the previous observation and by a '-' sign if it is less than the previous observation. Total number of runs up (+) and down (-) is counted. Too few runs indicate that the sequence is not random (has persistency) and too many runs also indicate that the sequence is not random (is zigzag).

Critical Value: Critical value for the test is obtained from the table for a given value of n and at desired level of significance (α). Let this value be r_c.

Decision Rule: If r_c (lower) $\leq r \leq$ r_c (upper), accept H_0. Otherwise reject H_0.

Tied Values: If an observation is equal to its preceding observation denote it by zero. While counting the number of runs ignore it and reduce the value of n accordingly.

Large Sample Sizes: When sample size is greater than 25 the critical value r_c can be obtained using a normal distribution approximation.

The critical values for two-sided test at 5% level of significance are

r_c (lower) $= \mu - 1.96\sigma$

r_c (upper) $= \mu + 1.96\sigma$

For one-sided tests, these are

r_c (left tailed) $= \mu - 1.65\sigma$, if $r \leq r_c$, reject H_0

r_c (right tailed) $= \mu + 1.65\sigma$, if $r \geq r_c$, reject H_0,

where

$$\mu = \left(\frac{2n-1}{3}\right) \text{ and } \sigma = \sqrt{\left(\frac{16n-29}{90}\right)}$$

Example: Data on value of imports of selected agricultural production inputs from U.K. by a county (in million dollars) during recent 12 years is given below: Is the sequence random?

5.2	5.5	3.8	2.5	8.3	2.1	1.7	10.0	10.0	6.9	7.5	10.6

Solution:

H_0: the sequence is random.

H_1: the sequence is not random.

5.2	5.5	3.8	2.5	8.3	2.1	1.7	10.0	10.0	6.9	7.5	10.6
	+	-	-	+	-	-	+	**0**	-	+	+

Here n = 11, the number of runs r = 7. Critical n values for α = 5% (two-sided test) from the table are r_c (lower) = 4 and r_c (upper) = 10.

Since r_c (lower) $\le r \le$ r_c (upper), i.e., observed r lies between 4 and 10, H_0 is accepted. The sequence is random.

8.3.3.6 Merits and Demerits

- The number of runs a sequence indicative of randomness.
- any set patterns of symbols in a sequence shows lack of randomness.
- Too many or too less runs show lack of randomness.

8.4 Two Sample Problem

In two sample problem, we are concerned with the data which consists of two independent random samples; i.e. random samples are drawn independently from each of two populations. Not only the elements within each sample are independent, but also every element in the first sample is independent of every element in the second sample.

We have two populations called as X and Y populations, with cumulative distribution functions F_X and F_Y respectively.

A random sample $X_1, X_2, ...X_m$ of size m is drawn from the population X and another random sample $Y_1, Y_2, ...Y_n$ of size n is drawn from the population Y.

Generally, the hypothesis of interest in two sample problem is that the two samples are drawn from the identical populations. i.e.

$$H_0 : F_X(x) = F_Y(x) \;\; ; \;\; \forall x$$

We shall discuss three types of alternatives:

(a) In the first type of alternative, we consider the alternative hypothesis that the two populations differ in any manner i.e. the two populations may differ in location or in dispersion or in skewness or in kurtosis etc.

(i) The two-sided alternative is

$$H_1 : F_X(x) \neq F_Y(x) \;\; ; \;\; \text{for some}\, x$$

(ii)

(iii) A one-sided alternative is

$$H_1 : F_X(x) \leq F_Y(x) \;\; ; \;\; \forall\, x$$

i.e. $H_1 : F_X(x) < F_Y(x) \;\; ; \;\; \text{for some}\, x$

i.e. the variable is stochastically larger than the variable Y.

or

$$H_1 : F_X(x) \geq F_Y(x) \;\; ; \;\; \forall\, x$$

$$H_1 : F_X(x) > F_Y(x) \;\; ; \;\; \text{for some}\, x$$

i.e. the variable is stochastically smaller than the variable Y.

For this type of problem, we shall discuss the following tests:

1. Wald-Wolfowitz Run Test
2. Kolmogrov-simirnov two sample Test

(b) In the second type of alternatives, we consider the alternative hypothesis that the two populations differ in location only, this type of alternative is called the location alternative.

$$H_L : F_X(x) = F_Y(x - \theta); \;\; \text{for some}\, x \;\&\; \theta \neq 0$$

i.e. the cumulative distribution function of Y is shifted to left if $\theta < 0$

i.e. $F_X(x) \leq F_Y(x); \; \forall x$ or $F_X(x) < F_Y(x); \text{for some}\, x$

and

the cumulative distribution function of Y is shifted to right if $\theta > 0$

i.e. $F_X(x) \geq F_Y(x); \; \forall x$ or $F_X(x) > F_Y(x); \text{ for some}\, x$

For this type of problem, we shall discuss the following tests:

1. Median Test
2. Mann-Whitney U Test
3. Wilcoxon Test

(c) In the third type of alternative hypothesis, we consider the alternative hypothesis that the two populations differ in scale parameter only, this type of alternative is called the scale alternative.

$$H_S : F_X(\theta x) = F_Y(x); \text{ for some } x \text{ \& } \theta \neq 1$$

i.e. the cumulative distribution function of Y is with compressed scale if $\theta > 1$ and the cumulative distribution function of Y is with enlarged scale if $\theta < 1$.

For this type of problem, we shall discuss the following tests:

1. Mood Test
2. Sukhatme Test

8.4.1 WALD-WOLFOWITZ RUN TEST

This two-sample test is based on the assumption that the populations under consideration are continuous.

We wish to test the hypothesis that the two independent samples have been drawn from the identical populations against the alternative that the two populations differ in any manner i.e. in location, in dispersion, in skewness or in kurtosis etc.

Let $X_1, X_2, ...X_m$ and $Y_1, Y_2, ...Y_n$ be two random samples of sizes m and n respectively drawn from two populations. These $N = m + n$ observations drawn from two populations are arranged in order of magnitude from smallest to largest, keeping in view which of the observations correspond to the X sample and which to Y sample.

For example, with $m = 4$ & $n = 5$, the arrangements might be

$X\,Y\,Y\,X\,X\,Y\,X\,Y\,Y$, $m + n = 9$.

We have 6 runs, 3 runs of $X's$ and $Y's$.

The total number of runs in the ordered pooled sample is indicative of the degree of random mixing. We wish to test the null hypothesis

$H_0 : F_X\left(x\right) = F_Y\left(x\right)$; $\forall x$

against $H_1 : F_X\left(x\right) \neq F_Y\left(x\right)$; for some x

where F_X & F_Y are the cumulative distribution functions of the populations.

Let r be the total number of runs in the group of N observations.

A run is defined to be a succession of one or more identical symbols which are followed and proceed by a different symbol or no symbol at all.

Under H_0, the two samples are drawn from the same population. i.e. Under H_0, the two samples are expected to be well mixed and r is expected to be large.

But r is small, if the two samples come from the different populations. i. e. if H_0 is fase.

If all the values of Y are greater than all the values of X (or vice-versa), then there will be only two runs.

Since too few runs will provide the critical region (or rejection region for null hypothesis H_0).

The Wald-Wolfowitz run test for α level of significance has the critical region $r \leq r_\alpha$ where r_α is the largest integer such that

$$P\left[r \leq r_\alpha / H_0\right] \leq \alpha$$

If H_0 is true, then all the ${}^{m+n}C_n = {}^{m+n}C_m$ different possible arrangements of m $X's$ and n $Y's$ in a line are equally likely.

When r is odd i.e. $r = 2k+1$; $k \in I^+$ .i.e. there are $(k+1)$ runs of ordered values of X and k runs of ordered values of Y or vice-versa. Then,

$$P\left[r = 2k+1 / H_0\right] = \frac{\left({}^{m-1}C_k\right)\left({}^{n-1}C_{k-1}\right) + \left({}^{m-1}C_{k-1}\right)\left({}^{n-1}C_k\right)}{{}^{m+n}C_m}$$

When r is even i.e. $r = 2k$; $k \in I^+$.

i.e. there are k runs of ordered values of X and k runs of ordered values of Y or vice-versa. Then,

$$P\left[r=2k / H_0\right]=\frac{2\left({}^{m-1}C_k\right)\left({}^{n-1}C_{k-1}\right)}{{}^{m+n}C_n}$$

Under H_0, the mean and variance of r are given as

$$E\left[r\right]=\frac{2mn}{m+n}+1$$

$$V\left[r\right]=\frac{2mn\left(2mn-m-n\right)}{\left(m+n\right)^2\left(m+n-1\right)}$$

For large m,n under H_0

$$Z=\frac{r-E\left[r\right]}{\sqrt{V\left[r\right]}} \sim N\left(0,1\right)$$

Note: It is the test for equality of distributions based on runs.

8.4.1.1 Rank Order Statistics

If the rank order statistics of a random sample $X_1, X_2, \ldots\ldots, X_n$ are denoted by $r\left(x_1\right), r\left(x_2\right), \ldots\ldots, r\left(x_n\right)$.

The i^{th} rank order statistic $r\left(x_i\right)$ is called the rank of the i^{th} observation in the unordered sample.

Ex: $r\left(x_i\right)=i$

The functional definition of the rank of any x_i in a set of n observations is given as,

$$r\left(x_i\right)=\sum_{j=1}^{n} S\left(x_i-x_j\right)$$

where $S\left(u\right)=\begin{cases}1 & ;\text{if } u \geq 0 \\ 0 & ;\text{if } u \geq 0\end{cases}$

8.4.1.2 Linear Rank Statistics

If the two independent random samples $X_1, X_2, ..., X_m$ and $Y_1, Y_2, ..., Y_n$ are drawn from the two populations with cumulative distribution functions F_X and F_Y respectively.

We consider the null hypothesis

$H_0 : F_X(x) = F_Y(x)$; $\forall x, F$ unknown

The set of $m+n=N$ observations are assigned ranks $1, 2, ..., N$.

The functional definition of the rank of observations in the combined sample (with no ties) is given as,

$$r(x_i) = \sum_{j=1}^{m} S(x_i - x_j) + \sum_{j=1}^{n} S(x_i - y_j)$$

$$r(y_i) = \sum_{j=1}^{n} S(y_i - y_j) + \sum_{j=1}^{n} S(y_i - x_j)$$

where $S(u) = \begin{cases} 1 & ;\text{if } u \geq 0 \\ 0 & ;\text{if } u \geq 0 \end{cases}$

we denote the combined ordered sample by a vector of indicator random variables as follows:

Let $Z = (z_1, z_2, ..., z_N)$ be the combined ordered sample. Then we describe

$$z_i = \begin{cases} 1 & ;\text{if i}^{\text{th}}\text{random variable in the combined ordered sample is X} \\ 0 & ;\text{if i}^{\text{th}}\text{ random variable in the combined ordered sample is} Y \end{cases} \; ; \; \forall i = 1, 2, ..., N$$

The vector Z indicates the rank order statistics of the combined samples. The linear rank order statistics is defined as

$$T_N = \sum_{i=1}^{N} a_i z_i$$

Where a_i are given numbers or weights.

Note: under H_0

$$E(z_i) = \frac{m}{N}$$

$$V(z_i) = \frac{mn}{N^2}$$

$$\text{cov}\left(z_i, z_j\right) = \frac{-mn}{N^2\left(N-1\right)} \quad , \ \forall i, j = 1, 2, ..., N$$

8.4.2 Two Sample Kolmogorov-Smirnov Test

Hypothesis and Assumptions:

Suppose a random variable is continuously distributed in each of two populations, the distribution functions being denoted by F and G. Further, suppose that independent random samples, say

$x_1, x_2, x_3, ..., x_m$ and $y_1, y_2, y_3, ..., y_n$ have been drawn from the two continuous distributions F_m and G_n respectively.

Here our problem is to test the hypothesis that the to distribution are identical i.e.

$$H_0 : F(\theta) = G(\theta)$$

against $$H_1 : F(\theta) \neq G(\theta); \qquad \forall t$$

Then an appropriate test criterion for testing hypothesis is K-S statistic which is as follows

$$D_{mn} = \max_{-\infty < t < \infty} \left| F_m(\theta) - G_n(\theta) \right|$$

If the hypothesis is true, one expects the value of D_{mn} to be small, while a large value of D_{mn} may be taken as an indication that the parent distributions are not identical.

8.4.3 MOOD Test for Dispersion

If we have two populations called as X and Y with cumulative distribution functions F_X and F_Y respectively. A random sample of size m is drawn from X population and another random sample of size n is drawn from Y population denoted as:

$X_1, X_2, ..., X_m$ and $Y_1, Y_2, ..., Y_n$

These $m + n = N$ observations drawn from the two populations are arranged in order of magnitude from smallest to largest.

In this combined ordered sample of N observations (with no ties), the average rank is the mean of first N integer. i.e. $\left(\frac{N+1}{2}\right)$.

The deviation of the i^{th} ordered variable about its mean rank is $\left[1-\left(\frac{N+1}{2}\right)\right]$. The amount of deviation is an indication of the relative spread.

In linear rank statistic, we may take weights either the absolute value of the deviations or the squared values of the deviations to measure the relative spread.

In Mood test, we take weights as the squared values of the deviations. We define the Mood Test Statistic as

$$M_N = \sum_{i=1}^{N}\left[i-\frac{N+1}{2}\right]^2 z_i$$

It gives the sum of squares of the deviations of the X ranks from the average combined rank.

We wish to test the null hypothesis that the two samples are drawn from the identical populations.

$$H_0 : F_Y(x) = F_X(x) \ ; \ \forall x$$

The general scale alternative is

$H_s : F_Y(x) = F_X(\theta x) \ ; \ \forall x$ and $\theta \neq 1$

If M_N is too small, then

$H_s : F_Y(x) \geq F_X(\theta x) \ ; \ \forall x$ and $\theta > 1$

i.e. $H_s : F_Y(x) > F_X(\theta x) \ ; \ \forall x$ and $\theta > 1$

If M_N is too large, then

$H_s : F_Y(x) \leq F_X(\theta x) \ ; \ \forall x$ and $\theta < 1$

$H_s : F_Y(x) < F_X(\theta x) \ ; \ \forall x$ and $\theta < 1$

Since,

$$M_N = \sum_{i=1}^{N} \left[i - \frac{N+1}{2} \right]^2 z_i$$

Then mean and variance of Mood's test statistic is

$$E[M_N] = \frac{m(N^2 - 1)}{12}$$

Also variance is obtained as

$$V[M_N] = E\left[M_N - E[M_N]\right]^2$$

By solving it, we get

$$V[M_N] = \frac{mn(N+1)(N^2 - 4)}{180}$$

When m, n are large, then under H_0

$$Z = \frac{M_N - E[M_N]}{\sqrt{V[M_N]}} \sim N(0,1)$$

8.4.4 Sukhatme Test for Dispersion

If we have two populations called as X and Y with cumulative distribution functions F_X and F_Y respectively. A random sample of size m is drawn from X population and another random sample of size n is drawn from Y population denoted as:

$X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots\ldots, Y_n$

These $m + n = N$ observations drawn from the two populations are arranged in order of magnitude from smallest to largest.

Here the X and Y populations have or can be adjusted to have equal medians; without loss of generality, we assume that this common median is zero.

In this case, we arrange the observations such that most of the negative Y's should proceed negative X's, and most of the positive Y's should follow positive X's, if Y's have a larger spread than X's.

If mn indicator random variables are defined as

$$D_{ij} = \begin{cases} 1 & \text{if } Y_j < X_i < 0 \ or \ 0 < X_i < Y_j; \quad \forall i = 1,2,.....,m \\ 0 & \text{otherwise}; \quad \forall j = 1,2,.....,n \end{cases}$$

Thus, Sukhatme test statistic is defined as

$$T = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij}$$

i.e.

$$\pi = \int_{-\infty}^{0} \left[F_Y(x) - F_X(x) \right] f(x)dx + \int_{0}^{\infty} \left[F_X(x) - F_Y(x) \right] f(x)dx$$

$$+ \int_{-\infty}^{0} F_X(x) \, f(x)dx - \int_{0}^{\infty} F_X(x) \, f(x)dx + \int_{0}^{\infty} f(x)dx$$

$$\pi = \int_{-\infty}^{0} \left[F_Y(x) - F_X(x) \right] f(x)dx + \int_{0}^{\infty} \left[F_X(x) - F_Y(x) \right] f(x)dx + \frac{1}{4}$$

Under H_0, $\pi = \frac{1}{4}$

Hence $H_0 : F_Y(x) = F_X(x)$ or $H_0 : \pi = \frac{1}{4}$

The mn random variable D_{ij} are Bernoulli variables, with parameter π .i.e.

$$E\left[D_{ij} \right] = E\left[D_{ij}^2 \right] = \pi$$

$$V\left[D_{ij} \right] = \pi(1 - \pi)$$

We define the parameters π_1 and π_2 as

$$\pi_1 = P\left[\left(Y_j < X_i < 0 \ or \ 0 < X_i < Y_j \right) \cap \left(Y_k < X_i < 0 \ or \ 0 < X_i < Y_k \right) \right]$$

$$= P\left[\left(Y_j < X_i < 0 \right) \cap \left(Y_k < X_i < 0 \right) + \left(0 < X_i < Y_j \right) \cap \left(0 < X_i < Y_k \right) \right]$$

$$\pi_1 = \int_{-\infty}^{0} \left[F_Y(x) \right]^2 f(x)dx + \int_{0}^{\infty} \left[1 - F_Y(x) \right]^2 f(x)dx$$

and

$$\pi_2 = P\left[\left(Y_j < X_i < 0 \; or \; 0 < X_i < Y_j\right) \cap \left(Y_j < X_h < 0 \; or \; 0 < X_h < Y_j\right)\right]$$

$$= P\left[\left(Y_j < X_i < 0\right) \cap \left(Y_j < X_{hi} < 0\right) + \left(0 < X_i < Y_j\right) \cap \left(0 < X_h < Y_j\right)\right]$$

$$\pi_2 = \int_{-\infty}^{0} \left[\frac{1}{2} - F_X\left(y\right)\right]^2 f(y)dy + \int_{0}^{\infty} \left[F_X\left(y\right) - \frac{1}{2}\right]^2 f(y)dy$$

Since $\quad T = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij}$

Then mean and variance of T is defined as

$$E\left[T\right] = \sum_{i=1}^{m} \sum_{j=1}^{n} E\left(D_{ij}\right) = \sum_{i=1}^{m} \sum_{j=1}^{n} \pi = mn\pi$$

and $V\left[T\right] = V\left(\sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij}\right)$

$$V\left[T\right] = mn\left[\pi - \pi^2\left(N-1\right) + \left(n-1\right)\pi_1 + \left(m-1\right)\pi_2\right]$$

As $m,n \to \infty$

$$E\left[T/mn\right] = \pi$$

$$V\left[T/mn\right] \to 0$$

Hence T/mn is an unbiased ad consistent estimator of π .

If we define $T^{'}$ as

$$T^{'} = \sum_{i=1}^{m} \sum_{j=1}^{n} D_{ij}^{'}$$

where

$$D_{ij}^{'} = \begin{cases} 1 & \text{if } X_i < Y_j < 0 \; or \; 0 < Y_j < X_i \\ 0 & \text{otherwise} \end{cases}$$

The critical region for α level of significance is given as

Alternative Hypothesis Critical Region

$\pi < \frac{1}{4} \ (\theta > 1)$ $\quad T \le C_{\alpha}$

$\pi > \frac{1}{4} \ (\theta < 1)$ $\quad T' \le C'_{\alpha}$

$\pi \ne \frac{1}{4} \ (\theta \ne 1)$ $\quad T \le C_{\frac{\alpha}{2}} \ or \ T' \le C'_{\frac{\alpha}{2}}$

Under H_0, i.e. $H_0 : F_Y(x) = F_X(x)$

Then $\pi = \frac{1}{4}$

and $\pi_1 = \pi_2 = \frac{1}{12}$

Thus $E[T] = \frac{mn}{4}$

$$V[U] = mn\left[\frac{1}{2} - \frac{1}{4}(N-1) + \frac{(n-1)}{3} + \frac{(m-1)}{3}\right]$$

$$V[U] = \frac{mn(N+7)}{48}$$

If N is large, then under H_0

$$Z = \frac{T - E[T]}{\sqrt{V[T]}} \to N(0,1)$$

i.e. $Z = \dfrac{U - \frac{mn}{4}}{\sqrt{\frac{mn(N+7)}{48}}} \to N(0,1)$

8.4.5 Contingency Table

A contingency table is an array of natural numbers in matrix from where those natural numbers represent counts, or frequencies. For example, an entomologist observing insects may say he observed 37 insects, or he may say he observed

Moths	Grasshoppers	others	Total
12	22	3	37

using 1×3 (one by three) contingency table. This is one-way contingency table because it has only one row.

The entomologist may wish to be more specific and use a 2×3 contingency table, as follows.

	Moths	**Grasshoppers**	**others**	**Total**
Alive	3	21	3	27
Dead	9	1	0	10
Total	12	22	3	37

The totals, consisting of two row totals, three column totals, and grand total. It is a two way contingency table and may be extended to include several rows (r) and several columns (c) as an $r \times s$ contingency table.

8.4.5.1 The 2×2 Contingency Table

In general $r \times c$ contingency table is an array of natural numbers arranged in to r rows and c columns and thus has rc *cells* or places for the numbers. This section is concerned only with the case where r = 2 and c = 2, the 2×2 contingency table, because there are four cells, 2×2 contingency table is also called the *fourfold* contingency table.

One application of the 2×2 contingency table arise when N objects (or persons), possible selected at random from some population, are classified in to one of two categories before a treatment is applied or an event takes place. After the treatment is applied the same N object are again examined and classified in to two categories. The question to be answered is, "Does the treatment significantly alter the proportion of object in each of two categories?" The appropriate statistical procedure was seen to be a variation of the sign test known as the McNemar test. The McNemar test is often able to detect subtle differences, primarily because the same sample is used in the two situations (such as "before" and "after"). Another way of testing the same hypothesis tested with the McNemar test is by drawing a random sample from the population before the treatment and then comparing it with another random sample drawn from the population after the treatment. The additional variability introduced by using to different random sample is undesirable because it tends to obscure the changes in the population caused by the treatment. However, there are times when it is not practical, or even possible, to use the same sample twice. Then the procedures to be described in the section may be used.

In the first procedure, two random samples are drowned, one from each of two populations, two test the null hypothesis that the probability of event A (some specified event) is the same for both populations. The null hypothesis may also be stated as "the proportion of the population with characteristic A is same for both populations."

8.4.5.2 The Chi-Squared Test for Differences in Probabilities, 2×2

A random sample of n_1 observations is drawn from one population (or before a treatment is applied) and each observation is classified in to either class 1 or class 2, the total numbers in the two classes being o_{11} and o_{12} respectively,

Where $o_{11}+o_{12}=n_1$. A second random sample of n_2 observations is drawn from a second population (or the first population after some treatment is applied) and the number of population in class 1 or class 2 is o_{21} or o_{22} respectively, where $o_{21}+o_{22}=n_2$. The data are arranged in to the following 2×2 contingency table.

Assumptions

1. Each sample is a random sample.
2. The two sample are mutually independent.
3. Each observation may be categorized in to class 1 or class 2.

Test Statistic: If any column total is zero, the test statistic is defined as $T_1=0$. Otherwise,

$$T_1=\frac{\sqrt{N}(O_{11}O_{22}-O_{12}O_{21})}{\sqrt{n_1 n_2 C_1 C_2}} \qquad (1)$$

Null distribution the exact distribution of T_1 is difficult to tabulate because of all the different combination of values possible for o_{11} , o_{12} , o_{21} and o_{22} . Therefore, the large sample approximation is used, which is the standard normal distribution whose quintiles are given in Table.

Hypothesis: Let the probability that a randomly selected element will be in class 1 be denoted by p_1 in population 1 and p_2 in population 2. Note that it is not necessary for p_1 and p_2 to be known. The hypotheses merely specify a relationship between them.

A. (Two-Tailed Test)

$H_0 : p_1 = p_2$

$H_1 : p_1 \neq p_2$

Reject H_0 at the approximate level α if T_1 is less than the $\alpha / 2$ quintile of a standard normal random variable Z, or if T_1 is grater then the 1-$\alpha / 2$ quintile of Z, where the quintiles of Z are given in table.

The p-value is twice the smaller of the probabilities that Z is less then the observed value of T_1 or grater then the observed value of T_1, from table.

Note that for the above hypotheses, T_1^2 is often use instead of T_1 as the test statistic. Then the rejection region is the upper tail of the chi-squared distribution with 1 degree of freedom given in table.

B. (Lower-Tailed Test)

$$H_0 : p_1 \geq p_2$$

$$H_1 : p_1 < p_2$$

Reject H_0 at the approximate level α if T_1 is less than the α quintile of a standard normal random variable Z, where the quintiles of Z are given in table.

The p-value is the probability that Z is less than the observed value of T_1, obtained from table.

C. **(Upper-Tailed Test)**

$$H_0 : p_1 \leq p_2$$

$$H_1 : p_1 < p_2$$

Reject H_0 at the approximate level α if T_1 is greater than the $1-\alpha$ quintile of a standard normal random variable Z, where the quintiles of Z are given in table.

The p-value is the probability that Z is greater than the observed value of T_1, obtained from table.

Example: Two Carloads of manufactured items are sampled randomly to determine if the proportion of defective items is different for the two carloads. From the first carload 13 of the 86 items were defective. From the second carload 17 of the 74 items were considered defective.

	Defective	**Non defective**	**Totals**
Carload 1	13	73	86
Carload 2	17	57	74
Totals	30	130	160

The assumptions are met, and so the two-tailed test is use to test

H_0: The proportion of defective is equal in two carloads using the test statistic

$$T_1 = \frac{\sqrt{N}(O_{11}O_{22} - O_{12}O_{21})}{\sqrt{n_1 n_2 C_1 C_2}}$$

$$= \frac{\sqrt{160}((13)(57) - (73)(17))}{\sqrt{(86)(74)(30)(130)}}$$

$$= -1.2695$$

The 0.975 quintile of a standard normal random variable is found from Table A1 to be 1.9600. therefore, the rejection region of approximate size 0.05 consist of all value of T_1 grater then 1.9600, or less then -1.9600. The observed value is -1.2695, so the null hypothesis is accepted at the $\alpha = 0.05$ level of significance.

The p-value is twice the probability of Z being less then the observed value -1.2695, which is found from the table as 0.102, so the p-value is approximately 0. 204.Therefore the decision to accept H_0 seems to be a fairly safe one.

The following example illustrates the use of one-tailed test.

Example: - At the U.S. Naval Academy, a new lighting system was installed throughout the midshipmen's living quarters. It was claimed that the new lighting system resulted in poor eyesight due to continual strain on the eyes of the midshipmen. Consider a (fictitious) study to test the null hypothesis,

H_0: The probability of good vision is less now than it was

Let p_1 be the probability that a randomly selected graduating midshipman had good vision under the old lighting system and let p_2 be the corresponding probability with the new light. Then the preceding hypotheses may be restated as

$$H_0 : p_1 \leq p_2$$

$$H_1 : p_1 > p_2$$

Which matches the set C of hypotheses. The random sample are taken to be the entire graduation class just prior to the installation class to spend 4 years using the new light system for population 2. it is hoped that these sample will behave the same as would random samples from the entire population of graduating seniors, real and potential.

Suppose the results were as fallows.

	Good vision	**Poor vision**	
Old lights	O_{11} =714	O_{12} =111	n_1 =825
New Lights	O_{21} =662	O_{22} =154	n_2 =816
Totals	1376	265	1641

Decision rule C defines the critical region $\alpha = 0.05$ to be all values of T_1 greater than 1.6449 from table. Computation of T_1 gives

$$T_1 = \frac{\sqrt{N}(O_{11}O_{22} - O_{12}O_{21})}{\sqrt{n_1 n_2 C_1 C_2}}$$

$$= \frac{\sqrt{1641}((714)(154) - (111)(662))}{\sqrt{(825)(816)(1376)(265)}}$$

$$= 2.982$$

So the null hypotheses is clearly rejected. From Table we see that the null hypotheses could have been rejected at a level of significance as small as about 0.002, so that p-value is 0.002.

We may there for conclude that the population represented by the two graduation classes do differ with respect to the proportions having poor eyesight, and the direction predicted. That is, population 2 (with the new light) has poor eyesight then population 1 (with the old light). Whether the poorer eyesight is result of the new lights has not been shown. However, an association of poor eyesight with the new lights has been shown in this hypothetical example.

Data: The N observations in the data are summarized in a 2×2 contingency table as previously both of the row totals, r and N-r and both of the column totals, c and N-c, are determined beforehand and are therefore fixed not random.

	Col 1	**Col 2**	
Row 1	x	$r-x$	r
Row 2	$c-x$	$N-r-c+x$	$N-r$
Total	c	$N-c$	N

8.4.6 Fisher's Exact Test

Now Discuss

8.4.6.1 Assumptions

1. Each observation is classified into exactly one cell.
2. The row and column totals are fixed, not random. (However see the comment at the end for random totals in rows, columns, or both.)

8.4.6.2 Test Statistic

The test statistic T_2 is the number of observations in the cell in row 1, column 1.

8.4.6.3 Null Distribution

The exact distribution of T_2 when H_0 is true is given by the hyper geometric distribution

$$P(T_2 = x) = \frac{\binom{r}{x}\binom{N-r}{c-x}}{\binom{N}{c}} \qquad x = 0,1..........,\min(r,c)$$

$$= 0 \qquad \textit{for all other values of } x \qquad (1)$$

For a large approximation use

$$T_3 = \frac{x - \frac{rc}{N}}{\sqrt{\frac{rc(N-r)(N-c)}{N^2(N-1)}}}$$

which has the standard normal distribution given in table as an approximation. If row totals or column totals, or both, are random it is more accurate to use T_1 given by

$$T_1 = \frac{\sqrt{N}(O_{11}O_{22} - O_{12}O_{21})}{\sqrt{n_1 n_2 C_1 C_2}}$$

in the large sample approximation.

8.4.6.4 Hypotheses

Let p_1 be the probability of an observation in row 1 being classified into column 1. Let p_2 be the probability of an observation in row 2 being classified in column 1. Let t_{obs} be the observed value of T_2.

A. (Two-tailed test)

$$H_0 : p_1 = p_2$$
$$H_1 : p_1 \neq p_2$$

First find the p- value using equation (1). The p-value is twice the smaller of $P(T_2 \leq t_{obs})$ or $P(T_2 \geq t_{obs})$. Reject H_0 at the level of significance α if the p-value is less than or equal to α.

B. (Lower-tailed test)

$$H_0 : p_1 \geq p_2$$
$$H_1 : p_1 < p_2$$

Find the p- value $P(T_2 \leq t_{obs})$ using equation (1). Reject H_0 at the level of significance α if $P(T_2 \leq t_{obs})$ is less than or equal to α.

C. (Upper-tailed test)

$$H_0 : p_1 \leq p_2$$
$$H_1 : p_1 > p_2$$

Find the p- value $P(T_2 \geq t_{obs})$ using equation (1). Reject H_0 at the level of significance α if $P(T_2 \geq t_{obs})$ is less than or equal to α.

Example: Fourteen newly hired business majors, 10 males and 4 females, all equally qualified, are being assigned by the bank president to their new jobs. Ten of the new jobs are as tellers , and four are as account representatives. The null hypothesis is that males and females have equal chances at getting the more desirable account representative jobs. The one-sided alternative of interest is that females are more likely than males to get the account representative jobs.

Only one female is assigned a teller position. Can the null hypothesis be rejected? The information given is sufficient to fill in the following 2×2 contingency table, because the row totals and column totals are already known.

	Account representative	**Teller**	
Males	1	9	10
Females	3	1	4
Total	4	10	N =14

$$H_0 : p_1 \geq p_2$$
$$H_1 : p_1 < p_2$$

The exact lower-tailed p-value is given by Equation (1) as

$$P(T_2 \le 1) = P(T_2 = 0) + P(T_2 = 1)$$

$$P(T_2 = x) = \frac{\binom{10}{0}\binom{4}{4}}{\binom{14}{4}} + \frac{\binom{10}{1}\binom{4}{3}}{\binom{14}{4}}$$

$$= \frac{1}{1001} + \frac{40}{1001} = 0.041$$

The null hypothesis is rejected at $\alpha = 0.05$.

8.5 Summery

This unit provides a thorough understanding of concepts related to non- parametric tests. The concepts of Mann-Whitney U Test, U-Statistic and Rank Tests, One Sample Kolmogorov-Smirnov Test, Two Sample Kolmogorov-Smirnov Test, Run Test, Wald-Wolfowitz Run Test, Mood Test for Dispersion, Sukhatme Test for Dispersion, Contingency Table. are described in detail. The learner should try to solve the self-assessment problems given in the next section.

8.6 Self-Assessment Exercises

Q1. Describe the utility of U-Statistic and Rank Tests.

Q2. Explain the following test:

a. Mann-Whitney U Test,

b. One Sample Kolmogorov-Smirnov Test,

c. Two Sample Kolmogorov-Smirnov Test,

d. Run Test,

e. Wald-Wolfowitz Run Test,

f. Mood Test for Dispersion, and

g. Sukhatme Test for Dispersion.

UNIT:9 NON-PARAMTRIC INFERENCE

Structure:

9.1 Introduction

This section provides a deeper understanding of non parametric inference by exposing to the concept of one way and two-way analysis of variance along with the concept of Pitman ARE.

9.2 Objectives

The objective of this unit is to provide a basic understanding of concepts related to Nonparametric Inference. The concept of the Kruskal-Wallis one way ANOVA Test, Friedman's two-way analysis of variance by ranks, efficiency criteria and theoretical basis for calculating ARE, Pitman ARE should be clear after study of this material.

9.3 Asymptotic Relative Efficiency (ARE)

In point estimation, the efficiency of two unbiased estimators for a parameter is defined as the inverse ratio of their variances. In the case of tests, the power efficiency; is defined as follows:

Definition **(Pitman)**

Let T_n, and T_n^*, $n = l, 2, \ldots$ be two sequences of test statistics of the same null hypothesis H_0, at the same significance level α. Let the distributions be indexed by a real parameter, so that $\delta = 0$ gives a distribution in H_0, and other 's correspond to distributions under the alternative hypothesis. Let us consider a sequence of alternatives δ_i. If, for the same power with respect to the same alternative δ_i, the test T_n, requires n_i observations and the test T_n^* requires n_i^* observations, then the relative efficiency of test T_n, with respect to test T_n^* is given by the ratio

$$e = \frac{n_i^*}{n_i}. \tag{1}$$

In general, e depends on α, δ_i, and n_i^*. The evaluation of $e = e(\alpha, \delta_i, n_i^*)$ as a function of the three arguments is not simple. Some asymptotic value of $e(\alpha, \delta_i, n_i^*)$ may be computed keeping one argument constant and letting the others approach some suitable limits. If the tests T_n, and T_n^* are consistent, then their powers will approach 1 with increasing sample size. Now let the sequence δ_i be such that $\delta_i \to 0$ as $i \to \infty$, and such that the power of each test lies in the open interval $(\alpha,\ 1)$ for finite sample sizes and approaches some limit between α and 1. We thus have the following definition of ARE:

Definition: Let T_n and T_n^* be two sequences of level α tests of a null hypothesis $\delta = 0$ against an alternative hypothesis δ_i. The asymptotic relative efficiency (ARE) of the test T_n relative to test T_n^* is the limiting value of (1) while simultaneously $n_i^* \to \infty$ and $\delta_i \to 0$, i.e.

$$ARE = \lim_{\substack{n_i^* \to \infty \\ \delta_i \to 0}} \frac{n_i^*}{n_i} = \lim_{\substack{n_i^* \to \infty \\ \delta_i \to 0}} e(\alpha, \delta_i, n_i^*), \tag{2}$$

if this limit exists and is constant for all increasing sequences of sample sizes n_i, n_i^*.

In many Important applications (2) does not depend on α. The ARE is also called the *local asymptotic efficiency* since it is the large-sample power in the vicinity of H_0.

There is difficulty of us exact power calculation with non-parametric tests, the above two criteria have found use in non-parametric inference. Also, in some cases (2) provides an approximation to the exact efficiency. There are some theorems for the evaluation of ARE.

Let $E(T_n)$ and $var(T_n)$ denote the mean and variance of T_n under an appropriate distribution.

The following regularity assumptions will be made for the sequences of tests T_n and T_n^*:

I. $dE(T_n)/d\delta$ exists and is non-zero and continuous at $\delta = 0$.

II. There exists a positive constant c such that

$$\lim_{n\to\infty} \frac{dE(T_n)/d\delta|_{\delta=0}}{\sqrt{n\, var(T_n)|_{\delta=0}}} = c.$$

III. For the sequence of alternatives $\delta_n = d/\sqrt{n}$,

(a) $\lim_{n\to\infty} \frac{dE(T_n)/d\delta|_{\delta=\delta_n}}{dE(T_n)d\delta|_{\delta=0}} = 1$

(b) $\lim_{n\to\infty} \frac{\sqrt{var(T_n)}|_{\delta=\delta_n}}{\sqrt{var(T_n)}|_{\delta=0}} = 1$

IV. Corresponding to $\delta_n = d/\sqrt{n}$, T_n is asymptotically normal with mean $E(T_n)$ and variance $var(T_n)|_{\delta=\delta_n}$.

Now consider the case of a one-sided alternative $H: \delta_n > 0$.

Theorem 1 Under the above regularity conditions, the limiting power of the test T_n is $1 - \Phi(\tau_\alpha - dc)$, where Φ is the standard normal distribution function.

***Proof*:** The limiting power of the one-sided test T_n is

$$\lim_{n\to\infty} P\left[T_n \geq T_{n;\alpha} \middle| \delta = \delta_n\right], \text{ where } T_{n;\alpha} \text{ is the level } \alpha \text{ critical value}$$

$$= \lim_{n\to\infty} P\left[\left.\frac{T_n - E(T_n)}{\sqrt{var(T_n)}}\right|_{\delta=\delta_n} \geq \left.\frac{T_{n,\alpha} - E(T_n)}{\sqrt{var(T_n)}}\right|_{\delta=\delta_n}\right]$$

$$= 1 - \Phi(\tau), say, due\ to\ regularity\ condition\ IV.$$

Now,

$$\tau = \lim_{n\to\infty} \frac{T_{n,\alpha} - E(T_n)}{\sqrt{var(T_n)}}\bigg|_{\delta=\delta_n} = \lim_{n\to\infty}\left[\frac{T_{n,\alpha} - E(T_n)|_{\delta=\delta_n}}{\sqrt{var(T_n)|_{\delta=0}}}.\frac{\sqrt{var(T_n)|_{\delta=0}}}{\sqrt{var(T_n)|_{\delta=\delta_n}}}\right]$$

$= \lim_{n\to\infty} \frac{T_{n,\alpha}-E(T_n)|_{\delta=\delta_n}}{\sqrt{var(T_n)|_{\delta=0}}}$, due to regularity condition III(b),

Since, by Taylor's expansion,

$$E(T_n)|_{\delta=\delta_n} = E(T_n)|_{\delta=0} + \delta_n.\frac{dE(T_n)}{d\delta}_{|\delta=0} + \in,$$

where $\lim_{n\to\infty} \in = 0$ due to regularity condition III, we have

$$\tau = \lim_{n\to\infty}\left[\frac{T_{n,\alpha} - E(T_n)|_{\delta=0}}{\sqrt{var(T_n)|}_{\delta=0}} - \frac{\delta_n[dE(T_n)/d\delta]|_{\delta=0} + \in}{\sqrt{var(T_n)}|_{\delta=0}}\right]$$

$= \lim_{n\to\infty}\left[\frac{T_{n,\alpha}-E(T_n)|_{\delta=0}}{\sqrt{var(T_n)|}_{\delta=0}}\right] - dc$, due to regularity condition II

$$= \tau_\alpha - dc$$

Thus the limiting power is $1 - \Phi(\tau_\alpha - dc)$.

Theorem 2 **(Pitman)** If T_n and T_n^* are two sequences of tests satisfying the four regularity conditions, then ARE of T_n relative to T_n^* is

$$\lim_{n\to\infty}\left[\frac{\frac{dE(T_n)}{d\delta}}{\frac{dE(T_n^*)}{d\delta}}\Bigg|_{\delta=0}\right]^2 \frac{var(T_n^*)}{var(T_n)}\bigg|_{\delta=0}$$

Proof: From ***Theorem 1***, the limiting powers of the tests T_n and T_n^* are, respectively,

$$1 - \Phi(\tau_\alpha - dc)$$

And $$1 - \Phi(\tau_\alpha - d^*c^*).$$

The tests will have the same limiting power if $dc = d^*c^*$, i.e. if

$$\frac{d^*}{d} = \frac{c}{c^*}.$$

From regularity condition III, the sequences of alternatives will be the same if

$$\frac{d}{\sqrt{n}} = \frac{d^*}{\sqrt{n^*}}.$$

It follows, then, that the two tests will have the same limiting power iff

$$\frac{n^*}{n} = \left(\frac{d^*}{d}\right)^2 = \left(\frac{c}{c^*}\right)^2$$

$$\lim_{n\to\infty}\left[\left.\frac{\frac{dE(T_n)}{d\delta}}{\frac{dE(T_n^*)}{d\delta}}\right|_{\delta=0}\right]^2 = \left.\frac{var(T_n^*)}{var(T_n)}\right|_{\delta=0}$$

Definition: The efficacy of a test of the hypothesis $H_0 = \delta = 0$ based on the test statistic T_n is defined as

$$\left.\frac{\left[\frac{dE(T_n)}{d\delta}\right]^2}{var(T_n)}\right|_{\delta=0}$$

Thus, under the regularity conditions, the ARE of T_n relative to T_n^* is the ratio of their efficacies.

The ARE does not depend on the significance level or power of the test when the regularity conditions are satisfied. The above theorem is also true if both T_n, and T_n^* are two-sided tests with the same values α_1, and α_2, for the sizes of the left- and right-hand critical regions, with $\alpha_1 + \alpha_2 = \alpha$.

Example: Let T denote the test statistic using sign test for testing **H$_0$:** μ_e = 0 against **H$_1$:** μ_e = 1,where μ_e is the median of the population, using a sample of size n from a normal distribution with mean μ and variance unity and T^* the test statistic using normal theory for testing **H$_0$:** μ= 0 against **H$_1$:** μ = 1 using n^* observations. Then the above hypothesis sets are identical as for a normal distribution, the mean and median coincide. Suppose we are interested in obtaining the power efficiency of sign test relative to normal test for a power of γ = 0.90 with a significance level α = 0.05.In the normal theory, the test based on n^* observations,

$\Pr[T^* > T_\alpha|\mathrm{H}_0]$ = 0.05

$\Rightarrow \frac{\bar{X}-0}{1/\sqrt{n^*}} > 1.64 \Rightarrow \sqrt{n^*}\bar{X} > 1.64.$

Setting the power γ equal to 0.90, n^* is found as follows.

Pr$[T^* \geq \ T_\alpha|\mathrm{H}_1]$ = Pr$\left[\sqrt{n^*}\bar{X} \geq 1.64|\ \mu = 1\right]$ = 0.90.

Or Pr$\left[\sqrt{n^*}(\bar{X}-1) \geq 1.64-\sqrt{n^*}\right]$ = 0.90

$\Rightarrow 1-$ Pr$\left[\sqrt{n^*}(\bar{X}-1) < 1.64-\sqrt{n^*}\right]$ = 0.90Type equation here.

$\Rightarrow$ Pr$\left[\sqrt{n^*}(\bar{X}-1) < 1.64-\sqrt{n^*}\right]$ = 0.10

$\Rightarrow \Phi\left(1.64-\sqrt{n^*}\right)$ = 0.10 $\Rightarrow\ 1.64-\sqrt{n^*}$ = -1.28 $\Rightarrow n^* \approx 9.$

In case of sign test,

Pr$[T > \ T_\alpha|\mathrm{H}_0]$ = 0.05 gives

$\sum_{r=r_\alpha}^{n}\binom{n}{r}\left(\frac{1}{2}\right)^n = \alpha,$ (1)

where r is the number of positive observations X_i and r_α is the critical value for rejection of the null hypothesis. The power of the test T is then given by

Pr$[T > \ T_\alpha|\mathrm{H}_1] = \gamma$

$\Rightarrow \sum_{r=r_\alpha}^{n}\binom{n}{r}p^r(1-p)^{n-r} = \gamma,$ (2)

where p = Pr$[X > 0|\mu_e\ =\ 1] = 1-$ Pr$[X \leq 0|\mu_e\ =\ 1]$

$= 1-$ Pr$\left[Z = \frac{X-\mu_e}{\sqrt{Var(X)}} \leq \frac{0-\mu_e}{\sqrt{Var(x)}}|\mu_e\ =\ 1\right]$ (since mean and median coincide for normal distribution)

$= 1-$ Pr$[Z \leq -1|\mu_e\ =\ 1] = 1-\ \Phi(-1)$ =0.8413.

The number n and r_α will be those values which satisfy (1) and (2) simultaneously for α = 0.05 and γ = 0.90. If p is rounded off to 0.85, then ordinary tables of binomial distribution can be used. On solving this, the value of n turns out to be 14 or 15. Thus the normal test requires only nine observations to be as powerful as a sign test using 14 or 15 observations, so that the power efficiency is around 0.60 or 0.64. This value of efficiency applies only for the particular value of α = 0.05 and γ = 0.90 and therefore is not in any sense a general comparison. More over the hypotheses under consideration were simple hypotheses. General conclusions for composite hypotheses and any values of α and γ are certainly impossible to obtain.

In many cases the limit of the ratio n^*/n may not be a function of α and γ or even the parameter value when it is in the neiborhood of the hypothesized value. In such cases, the asymptotic relative efficiency is a more suitable criterion, as it leads to a single number.

9.4 One-Way ANOVA and Kruskal-Wallis Test

Let the data consist of $N = \sum_{j=1}^{k} n_j$ observations, with n_j observations from the jth treatment, $j = 1, \dots, k$.

Treatments			
1	2	…	k
X_{11}	X_{12}		X_{1k}
X_{21}	X_{22}	…	X_{2k}
:	:		:
$X_{n_1 1}$	$X_{n_2 2}$	…	$X_{n_k k}$

9.4.1 Assumptions

A1. The N random variables $\{X_{1j}, X_{2j}, \dots, X_{nj}\}, j = 1, \dots, k$, are mutually independent.

A2. For each fixed $j \in \{1, \dots, k\}$, the n_j random variables $\{X_{1j}, X_{2j}, \dots, X_{nj}\}$ are a random sample from a continous distribution function F_j.

A3. The distribution functions $F_1, \dots, F_k$ are connected through the relationship

$$F_j(t) = F(t - \tau_j), \qquad -\infty < t < \infty, \tag{1}$$

For $j = 1, \dots, k$, where F is a distribution function for a continous distribution with unknown median θ and τ_j is the unknown treatment effect for the jth population.

We note that the assumptions $A1 - A3$ correspond directly to the usual one-way layout model commonly associated with normal theory assumptions; that is, Assumptions $A1 - A3$ are equivalent to the representation

$$X_{ij} = \theta + \tau_j + e_{ij}, \quad i = 1, \dots, n_j, \quad j = 1, \dots, k,$$

where θ is the overall median, τ_j is the treatment j effect, and the $N\ e's$ from a random sample from a continous distribution with median 0. (Under the additional assumptions of normality, the medians θ and 0 are, of course, also the respective means.)

The null hypothesis of interest is that of no differences among the treatment effects $\tau_1, \dots, \tau_k$, namely,

$$H_0 : [\tau_1 = \dots = \tau_k] \tag{2}$$

This null hypothesis asserts that each of the underlying distributions $F_1, \dots, F_k$ is the same , corresponding to $F_1 \equiv F_2 \equiv \dots \equiv F_k \equiv F$ in (1).

9.4.2 KRUSKAL-WALLIS TEST

We present a procedure for testing H_0 (2) against the general alternative that at least two of the treatment effects are not equal, namely,

$$H_1 : [\tau_1, \dots, \tau_k \ not\ all\ equal] \tag{3}$$

To compute the Kruskal-Wallis statistic, H, we first combine all N observations from the k samples and order them from least to greatest. Let r_{ij} denote the rank of X_{ij} in this joint ranking and set

$$R_j = \sum_{i=1}^{n_j} r_{ij} \quad and \quad R_{.j} = \frac{R_j}{n_j}, \quad j = 1, \dots, k. \tag{4}$$

Thus, for example, R_1 is the sum of the joint ranks received by the treatment 1 observations and $R_{.1}$ is the average rank for these same observations. The Kruskal-Wallis statistic H is then given by

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} n_j \left(R_{.j} - \frac{N+1}{2}\right)^2$$

$$= \left(\frac{12}{N(N+1)} \sum_{\mathrm{j=1}}^{\mathrm{k}} \frac{R_j^2}{n_j} \right) - 3(N+1) \qquad (5)$$

where $(N+1)/2 = \ (\sum_{j=1}^{k} \sum_{i=1}^{n_j} r_{ij}/N)$ is the average rank assigned in the joint ranking.

To test

$$H_0: [\tau_1 = \cdots = \tau_k]$$

versus the general alternative

$$H_1: [\tau_1 , \ldots , \tau_k \ not\ all\ equal],$$

At the α level of significance,

$$\text{Reject } H_0 \ if\ H \geq h_\alpha; Otherwise\ do\ not\ reject \qquad (6)$$

where the constant h_α is choosen to make the type I error probability equal to α. The constant h_α is the upper α percentile for the null $(\tau_1 = \cdots = t_k)$ distribution of H. When H_0 is true the statistic H has, as $\min(n_1, \ldots, n_k)$ tends to infinity , an asymptotic Chi-square(χ^2) distribution with $(k-1)$ degrees of freedom. The chi-square approximation for procedure (6) is

$$\text{Reject } H_0 \text{ if } H \geq \chi^2_{k-1\,,\alpha} \ ; \ Otherwise\ do\ not\ reject, \qquad (7)$$

where $\chi^2_{k-1,\alpha\,,}$ is the upper α percentile point of a chi-square distribution with $k-1$ degrees of freedom.

9.5 Two-Way ANOVA and Friedman Test

The procedure is associated with within-blocks rankings (known as the *Friedman ranks*).

	Treatments			
Blocks	1	2	...	k
1	X_{111}	X_{121}	...	X_{1k1}
	$\vdots$	$\vdots$	...	$\vdots$
2	$X_{11C_{11}}$	$X_{12C_{12}}$	...	$X_{1kC_{1k}}$
	X_{211}	X_{221}	...	X_{2k1}
	$\vdots$	$\vdots$	...	$\vdots$

	$X_{21C_{21}}$	$X_{22C_{22}}$	...	$X_{2kC_{2k}}$
⋮	⋮	⋮	⋮	⋮
n	X_{n11}	X_{n21}	...	X_{nk1}
	⋮	⋮	...	⋮
	$X_{n1C_{n1}}$	$X_{n2C_{n2}}$	...	$X_{nkC_{nk}}$

The data consist of $N = \sum_{i=1}^{n}\sum_{j=1}^{k} c_{ij}$ observations, with c_{ij} observations from the combination of the ith block with the jth treatment (i.e., the (i, j)th cell), for $i = 1, \ldots\ldots, n$ and $j = 1, \ldots, k$.

9.5.1 Assumptions

A1. The N random variables $\left\{\left(X_{ij1}, \ldots, X_{ijC_{ij}}\right), i = 1, \ldots, n \text{ and } j = 1, \ldots., k\right\}$ are mutually independent.

A2. For each fixed (i, j) with $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, k\}$, the c_{ij} random variables $\left(X_{ij1}, \ldots, X_{ijc_{ij}}\right)$ are a random sample from a continuous distribution with distribution function F_{ij}.

A3. The distribution functions $F_{11}, \ldots, F_{1k}, \ldots, F_{n1}, \ldots, F_{nk}$ are connected through the relationship

$$F_{ij}(u) = F\left(u - \beta_i - \tau_j\right), -\infty < u < \infty, \tag{1}$$

for $i = 1, \ldots, n$ and $j = 1, \ldots, k$, where F is a distribution function for a continuous distribution with unknown median θ, β_i is the unknown additive effect contributed by the block i, and τ_i is the unknown additive treatment effect contributed by the jth treatment.

We note that Assumption A1-A3 correspond directly to the usual two-way layout *additive* (See Comment 6) model associated with normal theory assumptions; that is, Assumption A1-A3 are equivalent to the representation

$$X_{ijt} = \theta + \beta_i + e_{ijt}, \quad i = 1, \ldots, n; j = 1, \ldots, k; t = 1, \ldots, c_{ij,}$$

where θ is the overall median, τ_i is the treatment j effect, β_i is block i effect, and the N e's form a random sample from continuous distribution with median 0. (Under the additional assumption of normality, the medians θ and 0 are, of course, also the respective means.)

The null hypothesis of interest is that of no difference among the additive treatment effects $\tau_1, \ldots, \tau_k$, namely,

$$H_0: [\tau_1 = \cdots = \tau_k]. \tag{2}$$

The null hypothesis asserts that the underlying distributions $F_{i1}, \ldots, F_{ik}$ within block i are the same, for each fixed $i = 1, \ldots, n$; that is, $F_{i1} \equiv F_{i2} \equiv \ldots \equiv F_{ik} \equiv F_i$, for $i = 1, \ldots, n$, in (1).

We consider the special case of one observation per treatment-block combination (commonly known as a *randomized complete block design*), corresponding to $c_{ij} = 1$ for every $i = 1, \ldots, n$ and $j = 1, \ldots, k$. For ease of notation in these five sections, we drop the third subscript on the X variables, since it is always equal to 1 in this setting.

9.5.2 FRIEDMAN TEST

We present a procedure for testing H_0 (2) against the general alternative that at least two of the treatment effects are not equal, namely,

$$H_1: [\tau_1, \ldots, \tau_k \text{ not all equal}], \tag{3}$$

when $c_{ij} \equiv 1$, for $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

To compute the Friedman statistic S, we first order the k observations from least to greatest separately within each of the n blocks. Let r_{ij} denote the rank of X_{ij} in the joint ranking of the observations $X_{i1}, \ldots, X_{ik}$ in the ith block and set

$$R_j = \sum_{i=1}^{n} r_{ij} \text{ and } R_{.j} = \frac{R_j}{n}. \tag{4}$$

Thus, for example, R_2 is the sum (over the n blocks) of the within-blocks ranks received by the treatment 2 observations and $R_{.2}$ is the average within-blocks rank for these same observations. The Friedman statistic S is then given by

$$S = \frac{12n}{k(k+1)} \sum_{j=1}^{k} \left(R_{.j} - \frac{k+1}{2} \right)^2$$

$$= \left[\frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 \right] - 3n(k+1), \tag{5}$$

where $(k+1)/2 = \sum_{i=1}^{n} \sum_{j=1}^{k} r_{ij}/nk$ is the average rank assigned via this within-blocks ranking scheme.

To test

$$H_0 = [\tau_1 = \ldots = \tau_k]$$

versus

$$H_1: [\tau_1, \ldots, \tau_k \text{ not all equal}],$$

at the α level of significance,

$$\text{Reject } H_0 \text{ if } S \geq s_\alpha; \quad \text{otherwise do not reject,} \tag{6}$$

where the constant s_α is chosen to make the type I error probability equal to α. The constant s_α is the upper α percentile for the null $(\tau_1 = \ldots = \tau_k)$ distribution of S.

9.6 Large-Sample Approximation

When H_0 is true, the statistics S has, as n tends to infinity, an asymptotic chi-square (χ^2) distribution with $k-1$ degree of freedom. The chi-square approximation for procedure (6) is

$$\text{Reject } H_0 \text{ if } S \geq \chi^2_{k-1,\alpha}; \quad \text{otherwise, do not reject,} \tag{7}$$

where $\chi^2_{k-1,\alpha}$ is the upper α percentile point of a chi-square distribution with $k-1$ degree of freedom.

9.7 Tukey's Test for Non-Additivity

In the analysis of variance with one observation per cell (fixed effects model), we assumed the interaction between the row and column to be absent, as it cannot be estimated with only one observation. But in case of any doubt regarding the presence of interaction, Tukey has proposed a test under the following set up.

The model for a two-way lay-out with interaction effect is given by

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ij}\ ;\ i = 1, \ldots, p;\ j = 1, \ldots, q; \qquad (1)$$

where α_i is the additive effect due to i^{th} row, β_j is the additive effect due to j^{th} column, γ_{ij} is the additive effect due to the interaction of i^{th} row and j^{th} column and e_{ij} is the random component which is assumed to be iid and distributed as N(0, σ_e^2). The side conditions are

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_i \gamma_{ij} = \sum_j \gamma_{ij} = 0.$$

If we have model (1), then we have $p - 1$ degrees of freedom for rows, $q - 1$ d.f. for columns, $(p - 1)(q - 1)$ d.f. for interaction and we are left with no d.f. for the error, since the total d.f. available with us is

$$pq - 1 = (p - 1) + (q - 1) + (p - 1)(q - 1).$$

Therefore we write down the model (1) in another form as follows :

$$y_{ij} = \mu + \alpha_i + \beta_j + \lambda\, \alpha_i \beta_j + e_{ij}\ ;\ i = 1, \ldots, p;\ j = 1, \ldots, q; \qquad (2)$$

where λ is a constant quantity, the side conditions are $\sum_i \alpha_i = \sum_j \beta_j = 0$ and e_{ij}'s are independent normal variates with zero mean and unknown variance σ_e^2.

In model (2) we have expressed γ_{ij} to be equal to $\lambda\, \alpha_i \beta_j$ and we are justified in expressing it in this way as shown below.

Since γ_{ij} is the interaction effect due to i^{th} row and j^{th} column, therefore it will be a function of α_i and β_j and let us assume that γ_{ij} is a function of α_i and β_j upto the second-degree term, then

$$\gamma_{ij} = \mathrm{A} + \mathrm{B}\,\alpha_i + \mathrm{C}\,\beta_j + \mathrm{D}\,\alpha_i^2 + \lambda\,\alpha_i \beta_j + \mathrm{H}\,\beta_j^2 . \qquad (3)$$

Summing (3) over j and dividing by q, we have

$$\gamma_{i.} = \mathrm{A} + \mathrm{B}\,\alpha_i + \mathrm{D}\,\alpha_i^2 + \mathrm{H}\theta, \quad \text{where } \theta = \sum_{j=1}^{q} \frac{\beta_j^2}{q} \qquad (4)$$

$= 0$ since $\sum_j \gamma_{ij} = 0$.

Similarly summing (3) over *i* and dividing by *p*, we have

$$\gamma_{.j} = \text{A} + \text{C}\,\beta_j + \text{D}\varphi + \text{H}\,\beta_j^2, \text{ where } \quad \varphi = \sum_{i=1}^{p} \frac{\alpha_i^2}{p} \tag{5}$$

$= 0$.

Hence, from (4), we have

$$\text{B}\,\alpha_i + \text{D}\,\alpha_i^2 = -\text{A} - \text{H}\theta \tag{6}$$

and from (5),

$$\text{C}\,\beta_j + \text{H}\,\beta_j^2 = -\text{A} - \text{D}\varphi. \tag{7}$$

$\therefore$ from (3),(6) and (7), we get

$$\gamma_{ij} = \text{A} - \text{A} - \text{H}\theta - \text{A} - \text{D}\varphi + \lambda\,\alpha_i\,\beta_j$$

$$= -\text{A} - \text{H}\theta - \text{D}\varphi + \lambda\,\alpha_i\,\beta_j. \tag{8}$$

Summing (8) over *j*, we have

$$0 = -\text{A} - \text{H}\theta - \text{D}\varphi.$$

Hence, from (8), we arrive at

$$\gamma_{ij} = \lambda\,\alpha_i\,\beta_j.$$

The null hypothesis for testing interaction will be **H₀ :** $\lambda = 0$. In this case we cannot apply the usual mean square theory because the expectations are non-linear, but however, we find from (2) that

$\bar{y}_{i.} = \mu + \alpha_i + \bar{e}_{i.}$, where $\bar{y}_{i.} = \frac{1}{q}\sum_{j=1}^{q} y_{ij}$ is the i^{th} row mean and $\bar{e}_{i.} = \frac{1}{q}\sum_{j=1}^{q} e_{ij}$ is the corresponding value of the error. Similarly, the j^{th} block mean is given by

$$\bar{y}_{.j} = \mu + \beta_j + \bar{e}_{.j}, \quad \text{where } \bar{y}_{.j} = \frac{1}{p}\sum_{i=1}^{p} y_{ij} \text{ and } \bar{e}_{.j} = \frac{1}{p}\sum_{i=1}^{p} e_{ij}.$$

Finally the overall mean is given by

$$\bar{y}_{..} = \mu + \bar{e}_{..}, \quad \text{where } \bar{y}_{..} = \frac{1}{pq}\sum_{i=1}^{p}\sum_{j=1}^{q} y_{ij} \text{ and } \bar{e}_{..} = \frac{1}{pq}\sum_{i=1}^{p}\sum_{j=1}^{q} e_{ij}.$$

Hence, $E(\bar{y}_{i.} - \bar{y}_{..}) = E(\cancel{\mu} + \alpha_i + \bar{e}_{i.} - \cancel{\mu} - \bar{e}_{..}) = \alpha_i$.

In other words, we can say that $\bar{y}_{i.} - \bar{y}_{..}$ is an unbiased estimate of α_i. Similarly,

$E(\bar{y}_{.j} - \bar{y}_{..}) = E(\cancel{\mu} + \beta_j + \bar{e}_{.j} - \cancel{\mu} - \bar{e}_{..}) = \beta_j \Rightarrow \bar{y}_{.j} - \bar{y}_{..}$ is an unbiased estimate of β_j, $E(\bar{y}_{..}) = E(\mu + \bar{e}_{..}) = \mu \Rightarrow \bar{y}_{..}$ is an unbiased estimate of μ, and finally,

$$E(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}) = E(\cancel{\mu} + \alpha_i + \beta_j + \lambda\,\alpha_i\,\beta_j + e_{ij} - \cancel{\mu} - \alpha_i - \bar{e}_{i.} - \cancel{\mu} - \beta_j - \bar{e}_{.j} + \cancel{\mu} + \bar{e}_{..})$$

$$= E(\lambda\,\alpha_i\,\beta_j + e_{ij} - \bar{e}_{i.} - \bar{e}_{.j} + \bar{e}_{..}) = \lambda\,\alpha_i\,\beta_j.$$

$\Rightarrow y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$ is an unbiased estimate of $\lambda\,\alpha_i\,\beta_j$.

If we assume that μ, α_i and β_j are known, then from (2), we find that it is linear in λ and we can find the estimate of λ by making use of the usual least square procedure, which is by obtaining the residual sum of squares

$$\text{SSE} = \sum_{i=1}^{p}\sum_{j=1}^{q}(y_{ij} - \mu - \alpha_i - \beta_j - \lambda\alpha_i\beta_j)^2.$$

Differentiating this with respect to λ and equating to zero, we have

$$\frac{\delta SSE}{\delta\lambda} = 0 \Rightarrow \sum_{i=1}^{p}\sum_{j=1}^{q}\alpha_i\beta_j(y_{ij} - \mu - \alpha_i - \beta_j - \lambda\alpha_i\beta_j) = 0$$

$$\Rightarrow \sum_{i=1}^{p}\sum_{j=1}^{q}\alpha_i\beta_j(y_{ij} - \mu - \alpha_i - \beta_j) = \sum_{i=1}^{p}\sum_{j=1}^{q}\alpha_i^2\beta_j^2.$$

Or $$\lambda^* = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q}\alpha_i\beta_j(y_{ij} - \mu - \alpha_i - \beta_j)}{\sum_{i=1}^{p}\sum_{j=1}^{q}\alpha_i^2\beta_j^2}.$$

When μ, α_i and β_j are not known, then they are replaced by their unbiased estimates and we obtain

$$\lambda^{**} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{.j} - \bar{y}_{..})(y_{ij} - \bar{y}_{..} - \bar{y}_{i.} + \bar{y}_{..} - \bar{y}_{.j} + \bar{y}_{..})}{\sum_{i=1}^{p}(\bar{y}_{i.} - \bar{y}_{..})^2 \sum_{j=1}^{q}(\bar{y}_{.j} - \bar{y}_{..})^2}$$

$$= \frac{pq\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{.j} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})}{S_A S_B},$$

where $S_A = \sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{i.} - \bar{y}_{..})^2 = q\sum_{i=1}^{p}(\bar{y}_{i.} - \bar{y}_{..})^2$

and $S_B = \sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{.j} - \bar{y}_{..})^2 = p\sum_{j=1}^{q}(\bar{y}_{.j} - \bar{y}_{..})^2$.

Now, $\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{.j} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$

$$= \sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{.j} - \bar{y}_{..})y_{ij} - \sum_{i=1}^{p}\bar{y}_{i.}(\bar{y}_{i.} - \bar{y}_{..})\sum_{j=1}^{q}(\bar{y}_{.j} - \bar{y}_{..})$$

$$- \sum_{i=1}^{p}(\bar{y}_{i.} - \bar{y}_{..})\sum_{j=1}^{q}\bar{y}_{.j}(\bar{y}_{.j} - \bar{y}_{..}) + \bar{y}_{..}\sum_{i=1}^{p}(\bar{y}_{i.} - \bar{y}_{..})\sum_{j=1}^{q}(\bar{y}_{.j} - \bar{y}_{..})$$

$$= \sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{.j} - \bar{y}_{..})y_{ij}$$

Hence, $\lambda^{**} = \dfrac{pq\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{i.} - \bar{y}_{..})(\bar{y}_{.j} - \bar{y}_{..})y_{ij}}{S_A S_B}$.

Thus, if our null hypothesis $\mathbf{H_0}$ is true, then $E(\lambda^{**}|\alpha_i\beta_j) = 0$

And $V(\lambda^{**}|\alpha_i\beta_j) = \dfrac{p^2q^2}{S_A^2 S_B^2}\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{y}_{i.} - \bar{y}_{..})^2(\bar{y}_{.j} - \bar{y}_{..})^2 \sigma^2 = \dfrac{pq}{S_A S_B}\sigma^2$.

$$\therefore \frac{\lambda^{**2}}{\frac{pq}{S_A S_B}\sigma^2} \sim \chi_1^2 .$$

Also we know that $\dfrac{\sum_{i=1}^{p}\sum_{j=1}^{q}(y_{ij}-\overline{y}_{i.}-\overline{y}_{.j}+\overline{y}_{..})^2}{\sigma^2} \sim \chi^2_{(p-1)(q-1)}$

and $\dfrac{pq[\sum_{i=1}^{p}\sum_{j=1}^{q}(\overline{y}_{i.}-\overline{y}_{..})(\overline{y}_{.j}-\overline{y}_{..})(y_{ij}-\overline{y}_{..}-\overline{y}_{i.}+\overline{y}_{..}-\overline{y}_{.j}+\overline{y}_{..})]^2}{S_A S_B} = \dfrac{SSN}{\sigma^2} \sim \chi^2_1$,

where SSN = $\dfrac{pq[\sum_{i=1}^{p}\sum_{j=1}^{q}(\overline{y}_{i.}-\overline{y}_{..})(\overline{y}_{.j}-\overline{y}_{..})(y_{ij}-\overline{y}_{..}-\overline{y}_{i.}+\overline{y}_{..}-\overline{y}_{.j}+\overline{y}_{..})]^2}{S_A S_B}$.

$\therefore$ SSE – SSN = SS due to residuals (SSR) and SSR will have (*p-1*)(*q-1*) -1 *df.* In other words, $\dfrac{SSR}{\sigma^2} \sim \chi^2_{(p-1)(q-1)-1}$. Hence the test statistic for testing H_0 is

$$F = \frac{SSN/1}{SSR/[(p-1)(q-1)-1]} = \frac{MSN}{MSR} \sim F_{1,(p-1)(q-1)-1}.$$

Analysis of Variance Table

Source of Variation	d.f.	SS	MSS	Variance Ratio
Rows	p – 1	$S_A = \sum\sum(\overline{y}_{i.}-\overline{y}_{..})^2$	$MS_A = \dfrac{S_A}{p-1}$	
Columns	q – 1	$S_B = \sum\sum(\overline{y}_{.j}-\overline{y}_{..})^2$	$MS_B = \dfrac{S_B}{q-1}$	
Non-additivity	1	$\left.\begin{matrix} SSN \\ SSR \end{matrix}\right\\} SSE$	MSN = SSN	$F = \dfrac{MSN}{MSR}$
Residuals	(p -1)(q – 1) – 1		$MSR = \dfrac{SSR}{(p-1)(q-1)-1}$	
Total	pq - 1	$TSS = \sum\sum(y_{ij}-\overline{y}_{..})^2$		

9.8 Summary

This unit provides a thorough understanding of concepts related to Nonparametric Inference. The concepts of Asymptotic Relative Efficiency, One Way ANOVA and Kruskal-Wallis Test, Two-way ANOVA and Friedman Test. are described in details. The learner should try to solve the self-assessment problems given in the next section.

9.9 Self-Assessment Exercises

Q1. What do you understand by Asymptotic Relative Efficiency by Pitman.

Q2. Describe the procedure of One Way ANOVA and Kruskal-Wallis Test by clearly stating the assumptions.

Q3. Explain the procedure of Two-way ANOVA and Friedman Test.