PART II
7.1 Expected Value Method of Probability Plotting:

7.1.1 In order to test whether a given sample belongs to a certain pre-assigned parent population, Folks and Blankenship have proposed a practical method called Probability Plotting Method with the use of an ordinary graph paper, for Normal, Exponential and Weibull populations. According to this method, if \((y_1, y_2, ..., y_n)\) is a random sample from a population, if \(y_{(i)}\), \(i=1,2,...,n\), are these observations when ordered such that \(y_{(1)} \leq y_{(2)} \leq ... \leq y_{(n)}\) and if \(X_{(i)}\) are the expected values of \(y_{(i)}\) in the particular population, \(X_{(i)} = E(y_{(i)})\), then, the \(n\) points \([X_{(i)}, y_{(i)}]\), \(i=1,2,...,n\), when plotted on an ordinary graph paper, should fall almost along a straight line. We shall refer to this method as the Expected Value Method or Method I.

7.1.2 For the Normal population \(N(0,1)\), the expected values of ordered observations, also called order statistics, have been
calculated for different values of \( n \) and are available in Table XX of Fisher and Yates [25], or in Table 10.B.1 of Sarhan and Greenberg [53], or in Table 28 of Pearson and Hartley [49] or in Table 9.1 of Rao, Mitra and Mathai [50]. If \( X(i) \) is the expected value of the \( i \)th ordered observation from \( \mathcal{N}(0,1) \), the expected value of the \( i \)th ordered observation \( y(i) \) from \( \mathcal{N}(\mu, \sigma^2) \) will be

\[
E(y(i)) = \mu + \sigma X(i)
\]

This of course is true for any population with location and scale parameters \( \mu \) and \( \sigma \). If the points \([X(i), y(i)]\) are plotted on an ordinary graph paper, they should lie almost along a straight line and the slope of the fitted straight line and its intercept on the \( y \)-axis will provide estimates of the parameters \( \sigma \) and \( \mu \).

7.1.3 For the standard Exponential population \( dF = e^{-y}dy, 0 < y < \infty \), the following expressions for the expected value of \( i \)th order statistic and its variance have been obtained by Epstein and Sobel [23];

\[
E(y(i)) = \frac{1}{\sum_{j=1}^{i} \frac{1}{n-j+1}}, \tag{7.1}
\]

\[
\text{Var}(y(i)) = \frac{1}{\sum_{j=1}^{i} \frac{1}{(n-j+1)^2}}, \tag{7.2}
\]

\( n \) being the size of the sample. It appears that these results
were earlier established by Gumbel [30]. In Section 1.4 below, we shall derive a similar expression for the $r$th cumulant $k_r$ of the distribution of the $i$th order statistic in the standard Exponential distribution, from which the above results follow. It is easy to calculate $X_{(i)}$ from (7.1). However, for $n \leq 10$, ready-made tables, e.g. Table 11.A.1 of Sarhan and Greenberg [53] are available, which can be used. For the general case, $f(x) = \frac{1}{\alpha} \exp \left( \frac{x-A}{\alpha} \right)$, $x > A$, $\alpha > 0$, if $y_{(i)}$ is the $i$th ordered observation in a random sample of size $n$ from this population, its expected value will be equal to $A + \alpha X_{(i)}$, where $X_{(i)}$ is the expected value of the $i$th ordered observation in a sample of size $n$ from the standard Exponential population, which can be calculated by using (7.1). Here also, if the points $[X_{(i)}, y_{(i)}]$, $i = 1, 2, \ldots, n$ are plotted on an ordinary graph paper, the $n$ points would lie almost along a straight line and, the slope of the fitted straight line and its intercept on the $y$-axis would provide estimates of the parameters $\alpha$ and $A$.

7.1.4 We shall now obtain an expression for $k_r [y_{(i)}]$ the $r$th cumulant of the $i$th order statistic in sample of size $n$ from the standard Exponential population. Let $y$ be the standard exponential variate with frequency density function $f(y) = e^{-y}$, $0 < y < \infty$. Then $F(y) = \int_0^y e^{-y} \, dy = (1-e^{-y})$. If $y_{(i)}$ is the $i$th order statistic in this distribution, the distribution of $y_{(i)}$
will be given by:
\[
\begin{align*}
\frac{dF}{(i-1)!(n-i)!} & \int [F(y(i)) - F(y(i))] [1-F(y(i))] \text{ d}y(i)
\end{align*}
\]
\[
= \frac{\int(n+1)}{\int(i)\int(n-i+1)} \left[1-e^{-y(i)}\right] \left[e^{-y(i)}\right]^{n-i+1} \text{ d}y(i).
\]

The moment generating function \( \varphi(t) \) of \( y(i) \) will be
\[
\varphi(t) = \int_0^\infty \frac{\int(n+1)}{\int(i)\int(n-i+1)} \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \frac{1}{n-i-t+j+1} \text{ d}y(i).
\]

Now
\[
\frac{\int(n-i-t+1)\int(i)}{\int(n-t+1)} = B(n-i-t+1,i) = \int_0^1 (1-x)^{i-1} x^{n-i-t} \text{ d}x
\]
\[
= \sum_{j=0}^{i-1} (-1)^j \binom{i-1}{j} \frac{1}{n-i-t+j+1}.
\]

Hence
\[
\varphi(t) = \frac{\int(n+1)}{\int(i)\int(n-i+1)} \cdot \frac{\int(n-i-t+1)\int(i)}{\int(n-t+1)}
\]

If the cumulant generating function is \( K(t) = \log \varphi(t) \), the
rth cumulant \( k_r \) of the distribution of \( y(i) \) is given by,
\[
k_r = \left[ \frac{d^r}{dt^r} K(t) \right]_{t=0} = \left[ \frac{d^r}{dt^r} \left\{ \log \int(n-i-t+1) - \log \int(n-t+1) \right\} \right]_{t=0}
\]

Now
\[
\frac{d}{dz} \log \int z = \Psi(z) = -\gamma(z-1) \sum_{j=0}^{\infty} \frac{1}{(j+1)(z+j)}
\]
\[
\]
Hence \( \frac{d}{dz} \psi(z) = \sum_{j=0}^{\infty} \left[ \frac{-1}{(j+1)(z+j)^2} + \frac{1}{(j+1)(z+j)} \right] \)

\[ = \sum_{j=0}^{\infty} \frac{1}{(z+j)^2}, \]

and \( \frac{d^r}{dz^r} \log |z| = \frac{d^{r-1}}{dz^{r-1}} \psi(z) = \sum_{j=0}^{\infty} \frac{(-1)^r(r-1)!}{(z+j)^r} \)

We therefore get:

\[ \frac{d^r}{dt^r} \left[ \log \left( n-i-t+1 \right) - \log \left( n-t+1 \right) \right] \]

\[ = (r-1)! \sum_{j=0}^{\infty} \left[ \frac{1}{(n-i-t+j+1)^r} - \frac{1}{(n-t+j+1)^r} \right] \]

Hence \( k_r = (r-1)! \sum_{j=0}^{\infty} \left[ \frac{1}{(n-i+j+1)^r} - \frac{1}{(n+j+1)^r} \right] \)

\[ = (r-1)! \left[ \frac{1}{(n-i+1)^r} + \frac{1}{(n-i+2)^r} + \ldots + \frac{1}{(n-1)^r} + \frac{1}{n^r} + \frac{1}{(n+1)^r} + \frac{1}{(n+2)^r} + \ldots + \frac{1}{(n+3)^r} + \ldots \right] \]

\[ = (r-1)! \left[ \frac{1}{(n-i+1)^r} + \frac{1}{(n-i+2)^r} + \ldots + \frac{1}{n^r} \right] \]

i.e. \( k_r = (r-1)! \sum_{j=1}^{\infty} \frac{1}{(n-j+1)^r} \)
Since \( k_1 = \mu_1, k_2 = \mu_2, k_3 = \mu_3, k_4 = \mu_4 - 3/\mu_2^2 \) etc., we can get the moments of the \( i \)th order statistic in the standard exponential population, and in particular:

\[
\begin{align*}
\mu'_1 (y(i)) &= \sum_{j=1}^{i} \frac{1}{(n-j+1)}, \\
\mu_2 (y(i)) &= \sum_{j=1}^{i} \frac{2!}{(n-j+1)^2}, \\
\mu_3 (y(i)) &= \sum_{j=1}^{i} \frac{3!}{(n-j+1)^3} \text{ etc.}
\end{align*}
\]

7.2 **Quantile Method of Probability Plotting:**

7.2.1 In order to test whether a given sample belongs to a Normal population, Rao, Mitra and Mathai [50] have proposed a Probability Plotting Method with the use of an ordinary graph paper, by using quantiles (also called fractiles or partition values). In this method, if \((y_1, y_2, \ldots, y_n)\) is a random sample from a Normal population, if \(y_{(i)}, i=1,2,\ldots,n\) are these observations when ordered such that \(y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}\) and if \(x_{p_i}\) are certain quantiles in the standard Normal population \(N(0,1)\), \(p_i\) to be obtained from \(i\) and \(n\) by \(p_i = \frac{i}{n}\), then the points \([x_{p_i}, y(i)]\), \(i=1,2,\ldots,n\) when plotted on an ordinary graph paper, should fall almost along a straight line, if the sample comes from a Normal population. A similar method
has been given by Epstein [22] for the Exponential population.
We shall refer to this method as the Quantile Method or Method II.

7.2.2 We may mention here a very simple but a very important property of a partition value namely that, 'it survives a transformation'. Let $x$ and $u$ be two random variables such that $x$ is a certain transformation of $u$ i.e. $x = \phi(u)$. Let $\phi(u)$ be a monotonic increasing function of $u$ and let the inverse function $x = \phi^{-1}(u)$ be single valued in the range of $u$. If the probability differential of $x$ is $f(x)dx$, the probability differential of $u$ is $f[\phi(u)]\phi'(u)du$. Hence

$$\int_{-\infty}^{\mu_p} f[\phi(u)]\phi'(u)du = p$$

is equivalent to

$$\int_{-\infty}^{\mu_p} f(x)dx = p,$$

where

$$\mu_p = \phi(u_p) \text{ or } u_p = \phi^{-1}(\mu_p).$$

This means that, if $u_p$ is a certain partition value in the distribution of $u$, such that $p$ proportion of observations are less than or equal to $u_p$, the corresponding partition value $x_p$ in the distribution of $x$ such that $p$ proportion of observations are less than or equal to $x_p$, is given by $x_p = \phi(u_p)$. We say that a partition value survives a transformation. Means, moments etc. do not enjoy this property.
7.2.5 If $y_p$ is a certain partition value in a particular population with location and scale parameters $\mu$ and $\sigma$, if $x_p$ is the corresponding partition value in this population when standardized, then $y_p = \mu + \sigma x_p$. Hence, in order to test whether a given sample belongs to a particular population or not, we can have a modified Probability Plotting Method with the use of an ordinary graph paper, by using a certain number of partition values, which may not necessarily be "equi-distant" quantiles in which $p_2-p_1 = p_3-p_2 = \cdots = p_r-p_{r-1}$.

Let $y_{p_1}, i=1,2,\ldots,r$, be certain $r$ partition values, calculated from the given $n$ observations in the random sample of size $n, n > r$, from a certain population with location and scale parameter $\mu$ and $\sigma$ and let $x_{p_1}$ be the corresponding partition values in this population when standardized. If these $r$ points $[x_{p_i}, y_{p_i}], i=1,2,\ldots,r$, are plotted on an ordinary graph paper, they should fall almost along a straight line and the slope of the fitted straight line and its intercept on the $y$-axis would provide estimates of the parameters $\sigma$ and $\mu$.

This is a more general approach to the problem. But there are various problems which will have to be solved. What value of $r$ should be selected i.e. how many partition values should be used? How to obtain the $p_i$'s from $i$ and $n$? We do not propose to discuss these questions here. Here, we shall consider only the Quantile Method proposed by Rao, Mitra and Mathai for the
Normal and by Epstein for the Exponential populations, in which instead of a few partition values all the observations in the sample are directly used.

7.2.4 As stated earlier, in the Quantile Method, the ordered observations $y_i$, $i=1,2,\ldots,n$ obtained from the random sample of size $n$ namely $(y_1,y_2,\ldots,y_n)$, are plotted against $x_{p_i}$, which are certain quantiles of the standard parent population, where $p_i$ are to be obtained from $i$ and $n$ in a certain way. Rao, Mitra and Mathai have proposed to take $p_i=i/n$. But, for $i=n$, this will present difficulty and we consider $p_i=i/n$ unsuitable for our purpose. The ordered observations $y_i$, $i=1,2,\ldots,n$ may be looked upon as $n$ quantiles partitioning the whole area into $n+1$ equal parts, just as the five sextiles partition the whole area under the frequency curve into six equal parts.

Instead of $p_i=i/n$, it is preferable to take $p_i=i/(n+1)$. But, there are other values for $p_i$ which may be preferable. For our investigation, however, we shall consider only three values of $p_i$ to choose from, namely (i) $p_i=i/(n+1)$, (ii) $p_i=(i-\frac{1}{2})/n$ and (iii) $p_i=(i-3/8)/(n+\frac{1}{8})$. Taking $p_i=i/(n+1)$ appears to have been first proposed by Weibull (1939) [57] and recommended by Gumbel (1954) [31]. Taking $p_i=(i-\frac{1}{8})/n$ appears to have been first discussed by Bliss (1937) [2] and by Ipsen and Jerne (1944) [39].
7.2.5 We shall give some details of Blom's work \[3\] on this problem, which throws light on the two Probability Plotting Methods being considered here and which also suggests a link between them. Blom has proposed a general rule for \( p_i \) namely

\[
p_i = \frac{i - \alpha}{n - \alpha - \beta + 1}, \quad (\alpha, \beta \leq 1)
\]

where \( \alpha \) and \( \beta \) may be so chosen that

\[
E(Y_{(1)}) \sim x_{p_i}.
\]

He has stated that, in the Normal distribution, the rule

\[
p_i = \frac{i - 3/8}{n + \frac{3}{8}}
\]

i.e. taking \( \alpha = \beta = 3/8 \), leads to a practically unbiased estimate of \( \sigma \) with a mean square deviation about \( \sigma \) which is about the same as that of the unbiased best linear estimate and that the rule given by Chernoff and Lieberman (1954) \[8\], namely,

\[
p_i = \frac{i - \frac{3}{8}}{n}
\]

i.e. taking \( \alpha = \beta = \frac{3}{8} \) leads to a biased estimate of \( \sigma \) with nearly minimum mean square deviation about \( \sigma \). Taking \( \alpha = \beta \) has been recommended as a working rule for any probability plotting by Bernard and Bos-Levenbach (1953) \[1\], who proposed to take \( \alpha = \beta = 0.3 \) irrespective of the shape of the distribution.

A similar problem is found in the work of Ogawa \[45,46\] in connection with the spacing of quantiles and he suggests
Equi-Probable and Intuitive Plausible Spacings of quantiles
in which $p_i = i/(n+1)$ and $p_i = (i - \frac{1}{2})/n$ are taken. Dave [15] also
has come across a similar situation in his thesis on
"Transformations of Variates" in which he wants to fit a
polynomial transformation of the standard normal variate on
non-normal data by using certain partition values which should
be "optimally spaced". The problem of choosing $p_i$ is an
important problem which needs further investigation. The
working rules suggested for $p_i$ need also confirmation. In our
present investigation into this problem, we have confined
ourselves to the comparison of only three values of $p_i$ namely
$p_i = i/(n+1)$, $p_i = (i - \frac{1}{2})/n$ and $p_i = (i - 3/8)/(n+\frac{1}{2})$, considering
Normal as well as Exponential populations and considering
complete as well as censored samples.

7.2.6 Thus in the Quantile Method, to determine whether a
sample $(y_1, y_2, \ldots, y_n)$ comes from a Normal population $N(\mu, \sigma^2)$,
corresponding to the $i$th ordered observation $y_i$, we obtain
$x_{p_i}$ such that

$$\int_{-\infty}^{x_{p_i}} \frac{1}{\sqrt{2\pi}} e^{-\frac{3}{2}x^2} dx = p_i$$

...(7.3)

where $p_i$ may be obtained from $i$ and $n$ in any one of the ways
mentioned in Section 7.2.4 or any other way and $x_{p_i}$ then can
be calculated by using Normal Tables like Table I of Fisher
and Yates [25] or Table 4 of Pearson and Hartley [49].
If the $n$ points $[x_{p_i}, y(i)]$ lie almost along a straight line, it is concluded that the sample comes from a Normal population and the slope of the fitted straight line and its intercept on the $y$-axis provide estimates for $\sigma$ and $\mu$. For the Exponential case, the method is similar. Here $x_{p_i}$ are to be obtained from

$$
\int_0^{x_{p_i}} e^{-x} \, dx = p_i
$$

We easily see that

$$
x_{p_i} = \log \frac{1}{1-p_i}, \quad \ldots (7.4)
$$

where $p_i$'s have to be suitably chosen. If we take $p_i = i/(n+1)$, $x_{p_i} = \log (n+1)/(n-i+1)$ and so on.

7.3 Censored Sampling:

7.3.1 While distinguishing between censored sampling and sampling from a truncated population, Gupta [32] has defined two types of censored sampling; Type I when observations in the sample below (or above) a given value $t_o$ of the variate (also called a truncation point) may be censored, and Type II when the $r_1$ smallest (or $r_2$ greatest) observations out of a sample of size $n$ in the sample may be censored. In Type I censoring, the number of observations censored is a random variable and the variate value, $t_o$, below (or above) which the
observations are censored is fixed; whereas in Type II censoring, the number of observations censored is fixed and the variate value of the largest (or smallest) censored observation is a random variable. Estimation of the parameters from censored samples has been considered by Gupta [32], Ipsen [38], Cohen [8,9,12], Hald [34,35,36], Halperin [37(a), 37(b)], Epstein [21], Epstein and Sobel [23], Sarhan and Greenberg [51,52] and many others.

It would not be out of way if some examples of censored sampling are cited here. Some examples from [52] are quoted here. In Experimental Biology, a known number of individuals might be exposed to a stimulus and the responses of some may fall outside the limits. Thus, if $n$ animals are injected with the same dose of an antigen and blood samples from each animal are tested for antibody response after a period of time, there may be only $n-r_1$ of the animals with measurable amounts, as $r_1$ of the animals develop the antibody to a level (say some fixed level) which cannot be measured by the prevailing technique. Thus, of the $n$ items, the smallest $r_1$ observations are censored because of fixed bounds. The individual values of the variate of $n-r_1$ largest observation are available for statistical study such as estimation of the parameters of the population. This sample is called a singly censored sample from the left and is of Type I.
One may have $n$ items drawn at random from a population and to save time and expense, the experiment is discontinued before all items have actually developed the phenomenon being observed. Such a decision to cut off the experiment is made as soon as the first $n-r_2$ experimental units have responded and the censoring is based upon fixed proportion of the observations. For example a biologist may perform an experiment on $n$ animals to determine the effect of exposure to a drug by noting reaction time. Some animals may require an extremely long time to react. The experiment might be stopped when a fixed percentage have reacted. Thus, one would have the exact data on the smallest $n-r_2$ items. This sample is termed a singly censored sample from the right and is of Type II. This situation may occur in life testing, incubation period, fatigue testing etc.

Furthermore, the above two situations may occur jointly such that there are $r_1$ smallest observations in a sample of $n$ items which are missing plus $r_2$ largest observations which are censored. This is termed as a doubly censored sample. For example, in certain studies of blood clotting, the speed of the reaction is such that $r_1$ animals may respond almost spontaneously before individual measurements can be taken on them whereas some animals barely respond and may require a long waiting period. In such a situation, censoring on the
left is by Type I whereas that on the right is by Type II. The case where observations are missing from both extremes is the most general one and the first two illustrations are special instances of it.

Sarhan and Greenberg [51] and Davis [16] have stated that the common situations of censored samples encountered in practice are those which occur with samples drawn from either an Exponential or a Normal population. Hence in our investigation, we have studied only the two cases of Normal and Exponential populations. But other populations can be studied on the lines of this investigation.

7.3.2 Consider a doubly censored sample of size \( n \) with \( r_1 \) smallest observations and \( r_2 \) largest observations missing, and denote by \( y_{(i)} \) the \( i \)th observation in ascending order of magnitude \( i = r_1 + 1, \ldots, n - r_2 \). Lloyd [41] has proposed best linear unbiased estimates of the location parameter \( \mu \) and the scale parameters \( \sigma \) based on ordered observations of complete samples. It is well reproduced by him in Chapter 3 of [53]. Later Sarhan and Greenberg [51,52,53] and others used this technique of estimation for censored samples also. These estimates are linear functions of the available ordered observations \( y_{(i)} \), namely,
\[ \hat{\mu} = \frac{\sum_{i=r_1+1}^{n-r_2} w_i y(i)}{n-r_2} \]
\[ \hat{\sigma} = \frac{\sum_{i=r_1+1}^{n-r_2} w_i y(i)}{n-r_2} \]

They have given tables for \( w_i \) and \( w_i' \) when parent population is Normal, Table 10.C.1 [53] and when it is Exponential [52], etc., for reasonable values of \( n \) and for \( r_1 = 0, 1, \ldots, n \) and \( r_2 = 0, 1, 2, \ldots, n \). It may be noted that the sample ceases to be a censored one for \( r_1 = r_2 = 0 \).

7.4 The Present Investigation:

7.4.1 The purpose of the present investigation, the results of which are incorporated in this chapter, is mainly to compare the Expected Value and the Quantile Methods of Probability Plotting. In the first two sections of this chapter, we have described these methods, with some comments and additions of our own. Before we go to the next section on the objectives and the plan of the present investigation, we would like to add a few more observations on the Probability Plotting Methods and our present investigation into them.

7.4.2 Probability Plotting Methods are simple graphical methods, which can be used quickly and by workers who may not have knowledge of Statistical Theory or Mathematics. It is true
that these methods provide only rough approximations or tentative conclusions. But, in these days when statistical methods are applied in many fields by an increasing number of laymen in Statistics, such methods are in demand and are useful. They are useful even to statisticians, who may be generally using accurate statistical methods, by providing them quickly with tentative or preliminary conclusions, which often provide a basis for getting more accurate results. A quick test, whether a sample comes from a Normal or an Exponential population has often to be made and hence the two Probability Plotting Methods considered here are useful in practical applications of Statistics. It is necessary that such simple types of statistical tools be studied and improved upon by research workers.

7.4.3 It may be argued by some that these methods are only rough methods and they cannot be compared. Some may think that both these methods are equally good or equally bad. This cannot be true. Both the methods may or may not be equally efficient and an investigation can be undertaken to compare them, of course on certain assumptions and under certain conditions. A theoretical investigation is difficult and we have made this investigation in a practical way, by using simulated sampling.
7.4.4 A fundamental difficulty in this comparison is that the fitting of the straight line is supposed to be done by eye and objective evaluation of merits and demerits of such a procedure cannot be made. But, in Probability Plotting Methods, the fitting of a straight line need not be continued to be done by eye. Workers are now-a-days acquainted with more and more of statistical methods and the ancillary mathematical computation and many of them know how to compute means, standard deviations, correlation coefficients, fitting of straight line in the simple case when the observations are assumed to be independent with equal variances, etc. The present methods can be improved and if the points fall almost along a straight line, the straight line can be fitted, instead of by eye, by simple Least Squares Method, obtaining estimates of location and scale parameters more accurately. At least, for comparison purposes, we may do the fitting by the simple Least Squares Method. It is true that the expected values or the quantiles are correlated and the simple Least Squares Method is not adequate. It is possible to evolve and use a more exact method. In our present investigation by using simulated sampling, the use of an elaborate method would have involved us into a heavy amount of computational work and we have used only the simple Least Squares Method. It is generally the experience of workers that the improvement made by using Least Squares Method for the correlated case instead of the
simple Least Squares Method is generally slight, e.g. Hahn and Shapiro [33] and the use of simple Least Squares in our investigation has, therefore, some support.

7.4.5 Even if we fit a straight line by simple Least Squares Method, we have to alternatives; Minimising \( \Sigma(y-\hat{y})^2 \) in which the sum of squares of vertical deviations is minimised, or minimising \( \Sigma(x-\hat{x})^2 \) in which the sum of squares of horizontal deviations is minimised. It is shown by Eisenhart [20] that when \( x \) is the independent variable, we should minimise \( \Sigma(y-\hat{y})^2 \). But it has also been shown by Krutchkoff [40] that this is not so in his problem. We therefore decided to fit the straight line in both of these ways and see which way gives better results in our problem.

7.4.6 An important feature of these two Probability Plotting Methods is that they are applicable to censored samples. In our investigation, therefore, we have compared the results given by these methods in the case of censored samples of Type II from Normal as well as Exponential populations. These methods are applicable to other types of populations. But we have limited our investigation to the Normal and Exponential cases only.

7.4.7 We are here concerned with the comparison of the two Probability Plotting Methods. Methods, which are better and
easier than Probability Plotting Methods are of course available. For example, from the given set of ordered observations, it is possible to obtain best linear estimates for location and scale parameters, using certain ready-made tables of Sarhan and Greenberg, as cited in Section 3.2. We shall therefore compare estimates obtained by Probability Plotting Methods with those obtained by Sarhan and Greenberg.

7.5 The Objectives and the Plan of the Present Investigation:

7.5.1 The objectives of the present investigation are:

(a) To compare the results due to taking
   (i) \( p_i = i/(n+1) \), (ii) \( p_i = (i - \frac{1}{2})/n \) and (iii) \( p_i = (i - 3/8)(n + \frac{1}{2}) \) in the Quantile Method.

(b) To compare the Expected Value and the Quantile Methods of Probability Plotting.

(c) To compare the results obtained by minimising
   (i) \( \sum (y - \hat{y})^2 \), the sum of squares of vertical deviations
   and (ii) \( \sum (x - \hat{x})^2 \), the sum of squares of horizontal deviations.

7.5.2 The main points of the plan of the present investigation are given below:

(a) The investigation is confined to the Normal and Exponential populations. In the case of Exponential, the probability density function with only one parameter, namely, the scale parameter, called \( \alpha \), is considered.
(b) For simulation, samples from three Normal and three exponential populations are taken. These populations are: Normal, \(N(80,3^2), N(80,6^2)\) and \(N(80,8^2)\), Exponential, \(E(\alpha=50), E(\alpha=125)\) and \(E(\alpha=250)\).

(c) 100 samples each of size 10 and 20 are drawn from each of these six populations. In the case of the Normal population \(N(\mu,\sigma^2)\), the random normal variate \(x\) is given by \(\sigma R + \mu\) where \(R\) is a random normal deviate from \(N(0,1)\) which can be readily obtained from Table A.2 of Dixon and Massey [18]. In the case of the Exponential Population, the random exponential variate \(x\) having p.d.f. \(f(x; \alpha)= \frac{1}{\alpha} e^{-x/\alpha}\), is obtained by using the relation \(x= -\alpha \log(R/10000)\) where \(R\) is the usual random number with four digits. For this purpose usual tables of random numbers are used.

(d) Complete samples as well as censored samples of Type II, (as defined in Section 3 of this Chapter) are drawn from these populations. Two sizes of the complete samples namely 10 and 20 are considered. Censoring is done by taking the first 60% of the ordered observations of the complete samples. Thus, in the notations explained in Section 3.1 of this Chapter we have the right censored samples of Type II with
\[ r_1=0, \ r_2=4, \text{ for } n=10 \]
and
\[ r_1=0, \ r_2=8, \text{ for } n=20. \]
(e) As the samples are actually drawn from particular populations, plotting on a graph paper (merely to verify whether the points cluster along a straight line) is not done. Location and scale parameters are estimated not graphically, but by the use of simple Least Squares Method, in the following two ways:

(i) Minimising $\Sigma (y-\hat{y})^2$ and taking the fitted straight line as $\hat{y}=\bar{y} + b_{yx}(x-\bar{x})$ [obtaining Minimum Vertical Deviation (MVD) Estimates]:

MVD Estimate of Scale parameter namely $\hat{\sigma}_1$ (or $\hat{\lambda}_1$)

$= \text{Slope of the fitted straight line}$

$= b_{yx}$

$= \frac{\Sigma (y-\bar{y})(x-\bar{x})}{\Sigma (x-\bar{x})^2} = \frac{\Sigma xy - n\bar{y}\bar{x}}{\Sigma x^2 - n\bar{x}^2}$ ... (7.5)

MVD Estimate of location parameter namely $\hat{\mu}_1$

$= \text{Intercept on y axis by the fitted straight line}$

$= \bar{y} - b_{yx} \bar{x}$ ... (7.6)

(ii) Minimising $\Sigma (x-\hat{x})^2$ and taking the fitted straight line as $\hat{x}=\bar{x} + b_{xy}(y-\bar{y})$ [obtaining Minimum Horizontal Deviation (MHD) Estimates]:

MHD Estimate of scale parameter namely $\hat{\sigma}_2$ (or $\hat{\lambda}_2$)

$= \text{Reciprocal of the slope of the fitted straight line}$

$= b_{xy}^{-1}$

$= \frac{\Sigma (y-\bar{y})^2}{\Sigma (x-\bar{x})(y-\bar{y})} = \frac{\Sigma y^2 - n\bar{y}^2}{\Sigma xy - n\bar{x}\bar{y}}$ ... (7.7)
MHD Estimate of location parameter namely $\hat{\mu}_2$

$= \text{Intercept on y-axis by the fitted straight line}$

$= \bar{y} - b_{xy} \bar{x}$ \hspace{1cm} \ldots (7.8)

In the above expression $x$ stands for $X_{(i)}$ in Method I and for $x_{p_i}$ in Method II whereas $y$ stands for $y_i$ in both the Methods.

(f) From each set of 100 samples, we calculate the estimates of location and scale parameters and calculate further the mean and variance of these estimates, which we shall call the simulated mean and simulated variance of the particular parameter for the particular population. In the tables of our results, we shall give the simulated variances and in bracket the simulated means.

(g) We first take up the scale parameter. For each set of 100 samples from the three Normal and the three Exponential populations, we have calculated the MVD and MHD estimates for the scale parameter $\sigma$ (or $\lambda$), using Method I which gives only one and Method II which gives three estimates corresponding to the three values of $p_i$ considered here, considering each complete as well as censored sample. Thus for a given complete as well as a censored sample, there are eight estimates of the scale parameter. Furthermore, the mean and variance of these estimates of scale parameters,
based on 100 estimates, have been calculated. Then for each set of 100 samples from a population, we have eight simulated means and eight simulated variances for the scale parameter. Simulated means and variances for complete samples from Normal and Exponential populations respectively for the scale parameter are given in columns (4) through (11) of Tables 7.1 and 7.3 (with means given in brackets). Similar quantities for the censored samples are given in columns (4) through (11) of Tables 7.2 and 7.4.

For the location parameter, MVD and MHE estimates are calculated only in the case of censored samples from the Normal Population. Simulated means and variances for this parameter (based on 100 estimates) are obtained and are given in columns (4) through (11) of Table 7.5. It may be noted that, as $x$'s are expected values or quantiles in the standard population and $\bar{x}=0$ for a complete sample from a standard Normal Population, all eight estimates of location parameter will be identically equal to the sample mean, which is evident from the expressions (7.6) and (7.8) given in (e) of this Section. Hence simulated means and variances based on complete samples for location parameter have not been calculated. In the case of the Exponential population, the problem of estimation of location parameter does not arise, as we are considering this population only with one parameter, namely, the scale
parameter. Table 7.5 therefore is only for the location parameter of the Normal population in censored samples.

(h) Variance of the best linear unbiased estimate for scale parameter for both the types of populations, from complete samples as well as censored samples are calculated by using tables of Sarhan and Greenberg; Table 10.C.2 of [53] in the case of Normal and Table II of Sarhan and Greenberg [52] in the case of the Exponential, are used for this purpose. However, in the case of the Exponential population considered here, it may be noted, that this variance can be readily obtained from $\sigma^2/n$ in the case of complete samples and $\sigma^2/(n-r_2)$ in the case of censored samples. This simplicity is due to the fact that we are considering the exponential population depending upon only one parameter namely the scale parameter and right censoring of Type II. The above variances are given in column (12) of Tables 7.1 through 7.4.

(i) Variance of the best linear unbiased estimate of the location parameter of Normal Population based on censored sample also has been calculated by using Table 10.C.2 of Sarhan and Greenberg [53]. It is given in column (12) of Table 7.5.

(j) The maximum likelihood estimate of the scale parameter $\sigma$ of the Normal population and its variance are
given by
\[ \hat{\sigma} = s = \sqrt{\frac{\sum(x-x)^2}{n}} \]

and
\[ \nu(s) = \left[ 2(n-1) - 2nc_2^2 \right] \frac{\sigma^2}{2n} \]

where
\[ c_2 = \frac{\left( \frac{2}{n} \right) \left( \frac{n-2}{2} \right)!}{\left( \frac{n-3}{2} \right)!} \]

Since \( E(s) = c_2 \sigma \), \( s \) is a biased estimate of \( \sigma \). We have calculated \( \nu(s) \), \( \nu(s/c_2) \) and \( E(s-\sigma)^2 = \nu(s) + (\text{Bias})^2 \). These quantities are given respectively in columns (13), (14) and (15) of Table 7.1.

(k) Asymptotic variances of the maximum likelihood estimate of the scale and the location parameters of the Normal population based on censored sample are calculated by using the Table given by Gupta [32] and are given in Column (13) of Tables 7.2 and 7.5. These could have been also calculated from Cohen's right hand part/Table 3 of [12].

(l) In the case of the Exponential population under study, best linear unbiased estimate and maximum likelihood estimate of the scale parameter are identical for both the complete and censored samples. Hence no additional calculations are needed and columns (12) and (13) of Tables 7.3 and 7.4 are the same.
Table of $X(i)$ and $x_{p_i}$ are prepared for ready reference for $i=1,2,\ldots,n$, $n=10(1)30$. These tables are prepared for both the Normal and the Exponential populations and are given in the Appendix (Table A.1, Table A.2). In the case of the Normal population the values of $X(i)$ are quoted from Table 28 of Pearson and Hartley $[49]$ and the values of $x_{p_i}$ are calculated by referring to Table 4 of Pearson and Hartley $[49]$. It may, however, be noted that the values of $X(i)$ for $n=27$ and 29 are not available. In the case of the Exponential population, the values of $X(i)$ and $x_{p_i}$ are computed by using (7.1) and (7.4).

We may explain the details of calculation by an example. Table 7.6 gives details of calculations for one specimen of a complete sample with $n=20$ from the Normal population $N(80,8^2)$ and one specimen of the censored sample for $n=20$, $r_1=0$, $r_2=8$ from the Exponential population with $\alpha=50$. The observations are presented in ascending order. Along with them are the columns of $X(i)$ and $x_{p_i}$ obtained from Table A.1 and A.2. In the case of censored sample from the Exponential population we have to refer to Table A.2 for $n=20$ and pick out the values of $X(i)$ and $x_{p_i}$ corresponding to the ordered values available in the sample (in this case, the first 12 values).

We should mention here one practical point regarding the calculation of $X(i)$ and $x_{p_i}$ for the case when a sample
contains two or more observations of equal value. To illustrate this point, let us take the values of the sample from the Normal population rounded off to the nearest one place of decimal. It is revealed from the column (3) of this table that the values 67.2 and 91.5 are repeated twice. The former corresponds to 3rd and 4th ordered observations and the latter corresponds to 17th and 18th ordered observations. Then, for \( y=67.2 \) we proposed to take \( x \) equal to the average of \( X(3) \) and \( X(4) \). Referring to Table A.1 we find that this average is \(-1.026\). Similarly for \( y=91.5 \) we take \( x=+1.026 \) averaging \( X(17) \) and \( X(18) \). Thus the total number of pairs \((X(i), Y(i))\) to be plotted in this case would be 18 instead of 20.

Let us now consider the determination of \( x_{p_1} \) in the case of ties. It may be observed, from the definition of \( x_{p_1} \), that \( p_1 \) represents the probability that a random variable \( x \) is less than or equal to \( x_{p_1} \). We find in our sample that there are 4 observations which are less than or equal to 67.2. Therefore \( p_1 \) and then \( x_{p_1} \) may be determined by substituting \( i=4 \) in the expressions of \( p_1 \) and \( x_{p_1} \). The values of \( x_{p_1} \) then, for \( y(i) = 67.2 \), are \(-0.878\), \(-0.935\), \(-0.919\) for \( p_1 = i/(n+1) \), \( p_1 = (i-\frac{3}{8})/n \) and \( p_1 = (i-\frac{3}{8})/(n+\frac{1}{2}) \). Similarly for \( y(i) = 91.5 \) the values of \( x_{p_1} \) are \(1.067\), \(1.150\), \(1.126\) obtained by substituting \( i=18 \) in the corresponding expressions. Thus in this
case too, there will be 18 pairs instead of 20 pairs for further considerations such as plotting on the graph, calculation of $b_{yx}$ etc. Under such circumstances $\sum X(i)$ and $\sum x_i$ need not be zero and the estimate of location parameter need not be equal to the sample mean, even if the sample is complete.

Estimates of the scale parameter and location parameter for the illustrative example are calculated and are given in Table 7.7. Estimates of the scale parameter of the Exponential population are also calculated from the example and are given in the same table.

It may be noted that in our work, we have taken $y$'s correct to three places of decimals to avoid two or more $y$'s with equal value. Hence our sample sizes are always 20 or 10 in complete samples.

7.6 Discussion and Conclusions:

7.6.1 MVD and MHD Estimates:

(a) Variance of the Scale Parameter: Let us first compare the MVD and the MHD estimates of the scale parameter from the variance point of view. On comparing the entries of simulated variances of the estimates of the scale parameters given in columns (8), (9), (10) and (11) with the corresponding entries given in columns (4), (5), (6) and (7), in
Table 7.1 through 7.4, we find that, the simulated variance of the MHD estimates of the scale parameter is always greater than the corresponding simulated variance of the MVD estimates of the scale parameter, in the cases of both the Normal and the Exponential populations, in complete as well as censored samples. We therefore conclude that the MVD estimate of the scale parameter is better than the MHD estimate of that parameter from the variance point of view, in both Normal and Exponential populations.

(b) Bias of the scale parameter: (i) Let us now compare the MVD and the MHD estimates of the scale parameter from the 'bias' point of view. On comparing the entries of the simulated means of the estimates of the scale parameter given in brackets in columns (8), (9), (10) and (11) with the corresponding entries given in columns (4), (5), (6) and (7), in Table 7.1 through 7.4, we find the same trend, which we observed in variances and we find that, the simulated mean of the MHD estimates of the scale parameter is always greater than the corresponding simulated mean of the MVD estimates of the scale parameter. We find this trend not only in the entries of the simulated means given in our Tables, but also in the detailed individual estimates of the scale parameter from which these simulated means are calculated, in the case of each sample. Though this result is an interesting one, it does not lead to any conclusion about bias.
(ii) On comparing the absolute values of the bias, obtained from the simulated means, in Tables 7.1 and 7.2 for the Normal population, we find that, barring only one exception in Table 7.1 for $g=8$, $n=20$, the MVD estimate of the scale parameter gives less bias than the MHD estimate of the scale parameter, in the cases of $p_i = (i-3/8)/(n+1/2)$, $p_i = i/(n+1)$ and Method I; but for $p_i = (i-3/8)/n$, the MVD estimate of the scale parameter gives greater bias than the MHD estimate of that parameter. In the Normal population, we further find from Tables 7.1 and 7.2 that the MHD estimates and the MVD estimate in the case of $p_i = i/(n+1)$ generally over-estimate the scale parameter, giving a positive bias, while the MVD estimates for $p_i = (i-3/8)/n$, $p_i = (i-3/8)/(n+1/2)$ and Method I under-estimate the scale parameter, giving a negative bias, which is greater in the case of $p_i = (i-3/8)/n$ than in the other two cases of $p_i = (i-3/8)/(n+1/2)$ and Method I, where the magnitudes of bias are small and nearly equal.

(iii) Referring to Tables 7.3 and 7.4 for comparing the bias in the MVD and MHD estimates of the scale parameter of the Exponential population, we find the same trends as those found in the case of Normal population, but less strongly and with several exceptions, with the notable difference that the MVD estimate even for $p_i = (i-3/8)/n$ gives, in general, less bias than the MHD estimate for the scale
parameter of the exponential population. In the Exponential population, therefore, the MVD estimate of the scale parameter has generally less bias than the MHD estimate.

(c) Location Parameter in Normal Population: Let us now compare the MVD and the MHD estimates of the location parameter for which purpose we have to compare the entries given in columns (4), (5), (6), and (7) with those given in columns (8), (9), (10) and (11) in Table 7.5 only. Among the simulated variances of the location parameter in Type II censored samples from Normal population, we find generally the same trend: the simulated variance of the MVD estimates is less than that of the MHD estimates, in four rows out of the six rows in Table 7.5; but for $6=6$, $n=20$, $r_1=0$, $r_2=8$ and for $6=3$, $n=10$, $r_1=0$, $r_2=4$, the simulated variance of the estimates of location parameter of the Normal population is slightly less for the MHD estimates than for the MVD estimates. If we compare the simulated means for bias, we find a similar trend but with more exceptions.

(d) Conclusion: The results mentioned above suggest that, in general, we may prefer MVD estimates to MHD estimates, for the estimation of both the location and the scale parameters of the Normal and the scale parameter of the Exponential populations, in complete as well as censored samples.
7.6.2 Variance of Estimates in Quantile Methods

(a) Referring to columns (4) through (7) of the Table 7.1 through 7.5, we find that the simulated variance given in column (4) in which we take \( p_i = \frac{i}{n+1} \), is always greater than the corresponding variance given in columns (5) to (7), in which we use \( p_i = \frac{(i-\frac{3}{8})}{n+\frac{1}{4}} \), \( p_i = \frac{(i-3/8)}{(n+\frac{1}{4})} \) and the Expected Value Method i.e. Method I. Thus we find that the Quantile Method i.e. Method II, with \( p_i = \frac{i}{n+1} \) gives less efficient estimates for both location and scale parameters, in the cases of the Normal as well as the Exponential populations, in complete as well as censored samples. Hence, from variance point of view, the rule \( p_i = \frac{i}{n+1} \) has to be discarded in favour of one of the other two rules for \( p_i \) or Method I.

(b) Comparing column (5) with columns (4) and (6) in all the Tables 7.1 and 7.5, to study the differences due to taking \( p_i = \frac{(i-\frac{3}{8})}{n} \) and \( p_i = \frac{(i-3/8)}{(n+\frac{1}{4})} \), we find that the simulated variance given in column (5) is always less than the simulated variance given in columns (4) and (6), except in one case of censored sample in Table 7.5 for \( \delta = 3, n = 20, r_1 = 0, r_2 = 8 \), where the simulated variance under column (5) is 0.56327 which exceeds the simulated variance under column (6) namely 0.56298 by a small quantity equal to 0.00029.
(c) We may, therefore, conclude that, among the three values of \( p_i \) considered here for calculation of \( x_{p_i} \), \( p_i = (i - \frac{3}{8}) / n \) leads to an estimate with the least variance for both the location and scale parameters in the Normal population and the scale parameter in the exponential population, for complete as well as censored samples.

### 7.6.3 Bias in the Estimates in Quantile Methods:

(a) **Normal Population**: On comparing the simulated means given in brackets in the column (5) with the corresponding means in the columns (4) and (6) in Tables 7.1, 7.2, and 7.5 for the Normal population, we find that the bias in the estimates of both the location and scale parameters generally increases as we take \( p_i = (i - \frac{3}{8}) / (n + \frac{1}{2}) \), \( p_i = (i - \frac{1}{2}) / n \) and \( p_i = i / (n + 1) \), except in a few cases. In the exceptional cases what we find is that the bias decreases as we pass from \( p_i = (i - \frac{3}{8}) / n \) to \( p_i = i / (n + 1) \). But the property that the bias increases as we pass from \( p_i = (i - \frac{3}{8}) / (n + \frac{1}{2}) \) to \( p_i = (i - \frac{1}{2}) / n \) holds good for all the entries under consideration in the tables 7.1, 7.2 and 7.5. In the exceptional cases, the bias in taking \( p_i = i / (n + 1) \) is never found less than that found in taking \( p_i = (i - \frac{3}{8}) / (n + \frac{1}{2}) \). Hence we may conclude that, for Normal population \( p_i = (i - \frac{3}{8}) / (n + \frac{1}{2}) \) gives estimates with the least bias, while \( p_i = (i - \frac{1}{2}) / n \) gives estimates with greater bias than those given by \( p_i = (i - \frac{3}{8}) / (n + \frac{1}{2}) \) (but as seen above in
Section 6.2 of this Chapter with the least variance) both for scale and location parameters.

Blom [3] had observed this for the scale parameter of Normal population. We confirm Blom and further find the result to be true for the location parameter of the Normal population also.

We find that taking \( p_i = i/(n+1) \) is to be discarded, in favour of one of the other two values of \( p_i \), from bias as well as variance points of view, for the estimation of any parameter of Normal population.

(b) Exponential Population: On comparing the entries of simulated means given in brackets in the columns (5) with those in columns (4) and (6) in Table 7.3 and 7.4, for the scale parameter of the Exponential Population, we find that, for complete samples, \( p_i = i/(n+1) \) gives estimates for scale parameter with greater bias than those given by the other values of \( p_i \), but \( p_i = (i-\frac{1}{2})/n \) generally gives the estimate for scale parameter of the Exponential population with less bias than that given by \( p_i = (i-3/8)/(n+\frac{1}{4}) \), except for \( \alpha = 125, n=10 \). Hence from both bias and variance points of view, for complete samples from the Exponential population, \( p_i = (i-\frac{3}{8})/n \) appears to give the best estimate of the scale parameter. For censored samples, this trend cannot be
confirmed by our results as far as bias is concerned, but from variance point of view, \( p_i = (i-\frac{3}{8})/n \) has been already found to be preferable. Hence, for the estimation of the scale parameter of the Exponential population by Quantile Methods, we may generally prefer \( p_i = (i-\frac{3}{8})/n \) to the other two values of \( p_i \).

7.6.4 Methods I and II:

(a) In the case of the Normal Population, Method I and Method II with \( p_i = (i-3/8)/(n+\frac{1}{2}) \) give very similar results, which can be seen very clearly in the entries in columns (6) and (7) of Tables 7.1, 7.2, and 7.5. Thus Method II with \( p_i = (i-3/8)/(n+\frac{1}{2}) \) gives estimates for location and scale parameters of Normal population which are equally good as those given by Method I, from the points of view of bias as well as variance, for complete as well as censored samples.

(b) In the case of comparison of Methods I and II, the conclusions for the Exponential population are not the same as those for the Normal population. Comparing the simulated variance for the estimates of the scale parameter of the Exponential population given in columns (6) and (7) of Table 7.4, we find that, for censored sample, Method I is better than Method II with \( p_i = (i-3/8)/(n+\frac{1}{2}) \) and further comparing columns (5) and (7) of this table we find that
Method I and Method II with $p_i=(i-\frac{1}{2})/n$ are equally efficient from both bias and variance points of view. We may conclude that for the Exponential population the rule $p_i=(i-\frac{1}{2})/n$ generally gives as good results as those given by the Expected Value Method.

7.6.5 Summary of the Conclusions:

(a) MVD and MHD Estimates: We may prefer in general MVD estimates to MHD estimates for both the location and scale parameters of the Normal population and the scale parameter of the Exponential population for both complete as well as censored samples.

(b) Quantile Methods; Normal Population:

(i) Among the three values of $p_i$ considered here for the Quantile Method, $p_i=i/(n+1)$ should be discarded for the estimation of both the location and scale parameters of the Normal population for both complete as well as censored samples, from both bias and variance points of view.

(ii) $p_i=(i-3/8)/(n+\frac{1}{2})$ gives estimates for both the location and scale parameters of the Normal population for both complete as well as censored samples with the least bias; but $p_i=(i-\frac{1}{2})/n$ gives these estimates with the least variance.
(c) Quantile Methods: Exponential Population:

(i) Among the three values of $p_i$, $p_i = i/(n+1)$ should be discarded for the scale parameter of the Exponential Population in complete as well as censored samples.

(ii) $p_i = (i - \frac{1}{2})/n$ gives better estimates for the scale parameter of the Exponential population in complete samples, than those given by the other two values of $p_i$, from both the points of view of bias and variance. For censored samples, in the Exponential population, $p_i = (i - \frac{1}{2})/n$ is preferable, at least from variance point of view. Generally, for the estimation of the scale parameter of the Exponential population by Quantile Methods, we may use $p_i = (i - \frac{1}{2})/n$.

(d) Comparison of the two Methods:

(i) For both the location and scale parameters of the Normal population, in complete as well as censored samples, the Expected Value Method gives as good results as the Quantile Method with $p_i = (i - 3/8)/(n + \frac{1}{4})$.

(ii) For the scale parameter of the Exponential population, the Quantile Method with $p_i = (i - \frac{3}{8})/n$ generally gives results, which are equally good as those given by the Expected Value Method.
7.7 A REMARK:

Comparing columns (12), (13), (14) and (15) of Table 7.1 an interesting property of the usual sample standard deviation, s, as an estimate of the scale parameter of the Normal population is observed. We note that these entries are not based on simulation, their basis being already described in Section 5.2 of this Chapter. Now we find that \( V(s) \) is less than the variance of the best linear estimate for all values of \( n \) and \( \sigma^2 \) considered here. Furthermore, neither \( V(s/\sigma^2) \) nor \( \mathbb{E}(s-\sigma)^2 \) exceeds the variance of the best linear estimate. Thus, this usual estimate is better than the best linear estimate for estimating the scale parameter of the Normal population based on complete sample. Here, \( s \) is a nonlinear estimate. By an example, Godambe and Joshi [29] have shown that the general feeling that corresponding to every nonlinear estimate there exists a linear estimate having a smaller variance than the nonlinear estimate is wrong. Here, perhaps is another example, supporting the statement of Godambe and Joshi [29].

7.8 An Application to a Geological Problem:

Situations arise in practice where Method II may be found more suitable. Geologists are often interested in determining the distribution of size of sediments. An outline of data sheet for size analysis by sieving is of the following type:
The pattern of cumulative weight given in column (7) changes as the material to be sieved changes. The usual practice is to plot $\phi$ (which is log mesh-size) and the percent cumulative weight on Normal probability paper with $\phi$ on arithmatic scale and the percent cumulative weight on probability scale. If the points lie close to a straight line further statistical constants such as quartile, mean etc. are obtained from this graph. The above data sheet, the description of the experiment and the determination of statistical constants are given by Robert L. Folk [26].

In statistical language what follows from the above experiment is that, $\phi$ has a normal distribution. Furthermore, what is directly observed is the cumulative probability. In the notations used here, column (7) of the above data sheet gives $p_i$. The experimenter observes $p_i$ and plots $(\phi_i, p_i)$ with $\phi_i$ on arithmatic scale and $p_i$ on the probability scale of the special normal probability paper. If now one wants to use ordinary paper, one has only to determine $x_{p_i}$ such that

<table>
<thead>
<tr>
<th>Material held on</th>
<th>Raw weight</th>
<th>Percent aggregate</th>
<th>Correlated weight in percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mesh Mm $\phi$</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td></td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
</tr>
</tbody>
</table>
As explained earlier $x_{p_i}$ can readily be read from Table 4 of Pearson and Hartley [49]. Hence if we plot $(x_{p_i}, \varphi_i)$ with $x_{p_i}$ along x-axis and $\varphi_i$ along y-axis, this graph too will help in determining whether the pattern of variation is normal and the usual statistical constants.
Table 7.1

Simulated Mean and Variance of the Estimates of Scale Parameter (σ) of the Normal Population based on 100 Samples.
(Nature of the Sample: Complete)

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Hypo. Size of Sample</th>
<th>MLE Estimate (Method I)</th>
<th>Expected Value</th>
<th>MLE Estimate (Method II)</th>
<th>Expected Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Quantile Method (Method I)</td>
<td></td>
<td>Quantile Method (Method II)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$1 - \frac{1}{n}$</td>
<td>$1 - \frac{2}{n}$</td>
<td>$1 - \frac{3}{n}$</td>
<td>$1 - \frac{1}{n}$</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>20</td>
<td>0.26487</td>
<td>0.21459</td>
<td>0.22814</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.19341)</td>
<td>(2.86239)</td>
<td>(2.95563)</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>20</td>
<td>1.17791</td>
<td>0.99994</td>
<td>0.95874</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>20</td>
<td>1.56690</td>
<td>1.25256</td>
<td>1.33729</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>10</td>
<td>0.60081</td>
<td>0.42285</td>
<td>0.46664</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(3.79400)</td>
<td>(2.83916)</td>
<td>(2.89773)</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>10</td>
<td>2.96925</td>
<td>2.09854</td>
<td>2.32264</td>
</tr>
</tbody>
</table>
Simulated Mean and Variance of the Estimates of Scale Parameter \( \sigma \) of the Normal Population Based on 100 Samples.

(More of the Sample? Censored.)

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Hypothesis Size of the sample</th>
<th>MVD Estimate</th>
<th>Quantile Method (Method II)</th>
<th>Expected Value</th>
<th>MHD Estimate</th>
<th>Quantile Method (Method I)</th>
<th>Expected Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \frac{1}{\sqrt{n}} )</td>
<td></td>
<td></td>
<td>( \frac{1}{\sqrt{n}} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \frac{1}{n} )</td>
<td></td>
<td></td>
<td>( \frac{1}{n} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \frac{1}{n^2} )</td>
<td></td>
<td></td>
<td>( \frac{1}{n^2} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \frac{1}{n^4} )</td>
<td></td>
<td></td>
<td>( \frac{1}{n^4} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \frac{1}{n^6} )</td>
<td></td>
<td></td>
<td>( \frac{1}{n^6} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \frac{1}{n^8} )</td>
<td></td>
<td></td>
<td>( \frac{1}{n^8} )</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.2

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Hypothesis Size of the sample</th>
<th>MVD Estimate</th>
<th>Quantile Method (Method II)</th>
<th>Expected Value</th>
<th>MHD Estimate</th>
<th>Quantile Method (Method I)</th>
<th>Expected Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \frac{1}{\sqrt{n}} )</td>
<td></td>
<td></td>
<td>( \frac{1}{\sqrt{n}} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \frac{1}{n} )</td>
<td></td>
<td></td>
<td>( \frac{1}{n} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \frac{1}{n^2} )</td>
<td></td>
<td></td>
<td>( \frac{1}{n^2} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \frac{1}{n^4} )</td>
<td></td>
<td></td>
<td>( \frac{1}{n^4} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \frac{1}{n^6} )</td>
<td></td>
<td></td>
<td>( \frac{1}{n^6} )</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \frac{1}{n^8} )</td>
<td></td>
<td></td>
<td>( \frac{1}{n^8} )</td>
<td></td>
</tr>
</tbody>
</table>

Variance Asymptotic

Estimate of the variance linear estimate of MLE
<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Size of sample (n)</th>
<th>Simulated Mean of Quantile Method</th>
<th>Variance of Quantile Method</th>
<th>Simulated Mean of Median Method</th>
<th>Variance of Median Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>305.208</td>
<td>123.83</td>
<td>305.208</td>
<td>123.83</td>
</tr>
<tr>
<td>2</td>
<td>125</td>
<td>1478.670</td>
<td>1182.393</td>
<td>1478.670</td>
<td>1182.393</td>
</tr>
<tr>
<td>3</td>
<td>250</td>
<td>6964.798</td>
<td>5531.710</td>
<td>6964.798</td>
<td>5531.710</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>721.905</td>
<td>509.248</td>
<td>721.905</td>
<td>509.248</td>
</tr>
<tr>
<td>5</td>
<td>125</td>
<td>2945.168</td>
<td>2206.848</td>
<td>2945.168</td>
<td>2206.848</td>
</tr>
<tr>
<td>6</td>
<td>250</td>
<td>14763.440</td>
<td>10526.844</td>
<td>14763.440</td>
<td>10526.844</td>
</tr>
</tbody>
</table>

Table 7.3
Simulated Mean and Variance of the Estimates of Scale Parameter \( \alpha \) of the Exponential Population Based on 100 Sampler
(Nature of the Sample: Complete)
Table 7.4

Simulated Mean and Variance of the Estimates of Scale Parameter ($\alpha$) of the Exponential Population based on 100 Samples.
(Nature of the Sample: Censored)

| Sr. No. | Hypothesis Size | Method | Method I | Method II | Method I | Method II | Variance of the | Variance of the |
|---------|-----------------|--------|----------|-----------|----------|-----------| best estimate   | best estimate   |
|         | Size of Censored |        | Expected | Expected |         | Expected |                |                |
|         | Sample Value of $\alpha$ |        | Value | Value |         | Value |                |                |
|         | (1)             | (2)    | (4)     | (5)      | (6)      | (8)      | (10)            | (11)            |
| 1       | 50 n=20 r=0, r'=0 | 314.654 | 291.789 | 297.322 | 291.762 | 352.124 | 326.167 | 332.464 | 326.235 | 208.33 | 208.33 |
| 2       | 125 n=20 r=0, r'=0 | 1540.252 | 1428.313 | 1455.595 | 1428.311 | 1623.892 | 1504.596 | 1533.343 | 1504.684 | 1302.08 | 1302.08 |
| 3       | 250 n=10 r=0, r'=0 | 6623.285 | 6146.944 | 6262.270 | 6145.364 | 7736.862 | 7161.022 | 7309.365 | 7163.437 | 5208.33 | 5208.33 |
| 4       | 50 n=20 r=0, r'=0 | 487.977 | 420.277 | 436.323 | 421.369 | 592.388 | 510.462 | 529.747 | 511.449 | 416.66 | 416.66 |
| 5       | 125 n=10 r=0, r'=0 | 2976.201 | 2568.766 | 2664.122 | 2572.630 | 3653.420 | 3149.863 | 3267.837 | 3155.573 | 2604.17 | 2604.17 |
| 6       | 250 n=10 r=0, r'=0 | 12128.35 | 10480.41 | 10863.69 | 10494.54 | 14735.57 | 12684.24 | 13166.99 | 12709.35 | 10416.7 | 10416.7 |
Table 7.5
Simulated Mean and Variance of the Estimates of Location Parameter (μ) of the Normal Population Based on 100 Samples
(Nature of the Sample: Censored)

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Hypo-Size of the sample</th>
<th>MVD Estimate (Method I)</th>
<th>Expected Estimate (Method I)</th>
<th>MHD Estimate (Method I)</th>
<th>Expected Estimate (Method I)</th>
<th>Variance of Sample Mean</th>
<th>Asymptotic linear estimate of WLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>r₁=0, r₂=8</td>
<td>0.56720</td>
<td>0.56720</td>
<td>0.56720</td>
<td>0.56720</td>
<td>0.6045</td>
<td>0.6100</td>
</tr>
<tr>
<td></td>
<td>(80.0143) (79.9822) (80.0004) (80.0035) (80.1278) (80.1027) (80.1221) (80.1230)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>n=12</td>
<td>2.29373</td>
<td>2.29373</td>
<td>2.29373</td>
<td>2.29373</td>
<td>2.27175</td>
<td>2.21558</td>
</tr>
<tr>
<td></td>
<td>(79.9871) (79.9886) (80.3420) (80.2088) (80.2487) (80.2507)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(79.8316) (79.6775) (79.7251) (79.7275) (80.1724) (80.0011) (80.0526) (80.0549)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>r₁=0, r₂=4</td>
<td>1.03557</td>
<td>1.03557</td>
<td>1.03557</td>
<td>1.03557</td>
<td>1.05289</td>
<td>1.01296</td>
</tr>
<tr>
<td>5</td>
<td>n=8</td>
<td>5.24055</td>
<td>5.24055</td>
<td>5.24055</td>
<td>5.24055</td>
<td>5.64997</td>
<td>5.46253</td>
</tr>
<tr>
<td></td>
<td>(79.8177) (79.3970) (80.3866) (80.7204) (80.5846) (80.6290) (80.6372)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(79.8187) (79.6565) (79.7130) (79.7229) (80.3588) (80.1866) (80.2426) (80.2531)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>$Y(1)$</td>
<td>$Y(1)$</td>
<td>$\frac{1}{n+1}$</td>
<td>$\frac{1}{n}$</td>
<td>$\frac{1}{n+2}$</td>
<td>$X(1)$</td>
<td>$Y(1)$</td>
</tr>
<tr>
<td>----</td>
<td>--------</td>
<td>--------</td>
<td>-----------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>-------</td>
<td>--------</td>
</tr>
<tr>
<td>1</td>
<td>65.256</td>
<td>65.3</td>
<td>-1.665</td>
<td>-1.960</td>
<td>-1.866</td>
<td>0.115</td>
<td>0.0488</td>
</tr>
<tr>
<td>2</td>
<td>65.424</td>
<td>65.4</td>
<td>-1.311</td>
<td>-1.440</td>
<td>-1.405</td>
<td>0.1705</td>
<td>0.1001</td>
</tr>
<tr>
<td>3</td>
<td>67.184</td>
<td>67.2</td>
<td>-0.878</td>
<td>-0.935</td>
<td>-0.919</td>
<td>0.315</td>
<td>0.1542</td>
</tr>
<tr>
<td>4</td>
<td>67.208</td>
<td>67.2</td>
<td>-0.913</td>
<td>-0.755</td>
<td>-0.745</td>
<td>0.415</td>
<td>0.3113</td>
</tr>
<tr>
<td>5</td>
<td>72.448</td>
<td>72.4</td>
<td>-0.565</td>
<td>-0.598</td>
<td>-0.589</td>
<td>0.655</td>
<td>0.3365</td>
</tr>
<tr>
<td>6</td>
<td>75.544</td>
<td>73.5</td>
<td>-0.432</td>
<td>-0.454</td>
<td>-0.444</td>
<td>0.855</td>
<td>0.4055</td>
</tr>
<tr>
<td>7</td>
<td>75.548</td>
<td>75.5</td>
<td>-0.303</td>
<td>-0.319</td>
<td>-0.313</td>
<td>1.005</td>
<td>0.4797</td>
</tr>
<tr>
<td>8</td>
<td>75.348</td>
<td>78.5</td>
<td>-0.179</td>
<td>-0.189</td>
<td>-0.187</td>
<td>1.150</td>
<td>0.5597</td>
</tr>
<tr>
<td>9</td>
<td>78.800</td>
<td>78.8</td>
<td>-0.060</td>
<td>-0.063</td>
<td>-0.063</td>
<td>2.081</td>
<td>0.6467</td>
</tr>
<tr>
<td>10</td>
<td>79.776</td>
<td>79.8</td>
<td>0.060</td>
<td>0.063</td>
<td>0.063</td>
<td>27.374</td>
<td>0.7421</td>
</tr>
<tr>
<td>11</td>
<td>80.406</td>
<td>80.4</td>
<td>0.179</td>
<td>0.189</td>
<td>0.187</td>
<td>42.772</td>
<td>0.8475</td>
</tr>
<tr>
<td>12</td>
<td>83.576</td>
<td>83.6</td>
<td>0.303</td>
<td>0.319</td>
<td>0.313</td>
<td>61.150</td>
<td>1.067</td>
</tr>
<tr>
<td>13</td>
<td>84.155</td>
<td>84.1</td>
<td>0.432</td>
<td>0.454</td>
<td>0.448</td>
<td>1.150</td>
<td>1.150</td>
</tr>
<tr>
<td>14</td>
<td>85.232</td>
<td>85.2</td>
<td>0.565</td>
<td>0.598</td>
<td>0.589</td>
<td>1.408</td>
<td>1.408</td>
</tr>
<tr>
<td>15</td>
<td>86.024</td>
<td>86.0</td>
<td>0.713</td>
<td>0.753</td>
<td>0.745</td>
<td>1.867</td>
<td>1.867</td>
</tr>
</tbody>
</table>

**NORMAL POPULATION**

**EXPONENTIAL POPULATION**
Table 7.7
Estimates of the Scale and Location Parameters of the Normal and Exponential Populations, for specimen Examples.

Nature and Size of Sample from Normal: Complete, \( n=20 \), Actual Number of pairs = 18

Nature and Size of Sample from Exponential: Censored, \( n=20 \), \( r_1=0 \), \( r_2=8 \). Actual No. of pairs=12.

<table>
<thead>
<tr>
<th>Population Parameter</th>
<th>Hypothetical value of the parameter</th>
<th>MVD ESTIMATES</th>
<th>Expected value method</th>
<th>MHD ESTIMATES</th>
<th>Expected value method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Quantile Method (Method II)</td>
<td></td>
<td>Quantile Method (Method II)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \frac{1}{n+1} )</td>
<td>( \frac{1}{n} )</td>
<td>( \frac{1-3/8}{n+1} )</td>
<td>( \frac{1}{n+1} )</td>
</tr>
<tr>
<td></td>
<td>Location parameter (( \mu ))</td>
<td>80</td>
<td>79.291</td>
<td>79.291</td>
<td>79.289</td>
</tr>
<tr>
<td>Exponential</td>
<td>Scale parameter (( \lambda ))</td>
<td>50</td>
<td>44.37</td>
<td>43.25</td>
<td>43.16</td>
</tr>
</tbody>
</table>