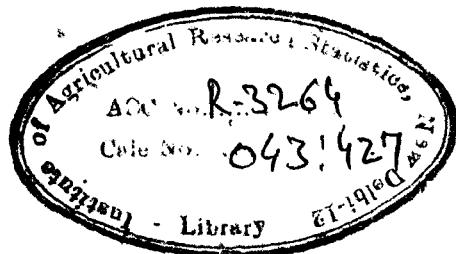


✓ B.4

✓ 95

EFFECT OF NON - NORMALITY
ON RESPONSE TO SELECTION

M. GOPINATH RAO



Diploma submitted in fulfilment of the
requirements for the award of Diploma
in Agricultural Statistics of the
Institute of Agricultural
Research Statistics
(I.C.A.R.)

INSTITUTE OF AGRICULTURAL RESEARCH STATISTICS
(I.C.A.R.)
LIBRARY AVENUE, NEW DELHI - 12

ACKNOWLEDGEMENTS

I have great pleasure in expressing my deep sense of gratitude to Dr. J.P. Jain, Associate Professor, Institute of Agricultural Research Statistics (I.A.R.S.) , for his inspired guidance, keen interest and constant encouragement throughout the period of investigation and immense help in writing-up this thesis.

I am thankful to Mr. R. Gopalan whose helpful suggestions and criticisms were invaluable throughout the course of investigation.

It is my earnest and pleasant duty to place on record my gratefulness to Dr. Dargha Singh, Director, Institute of Agricultural Research Statistics, who provided the necessary facilities for this work.

My acknowledgement will not be complete unless I mention the Indian Council of Agricultural Research for extending the financial assistance in the form of fellowship; Mr. Gaya Prashad Pal (I.A.R.S.) , who typed this dissertation in time inspite of several problems.

M. Gopinath Rao

(M. GOPINATH RAO)

CHAPTER OF CONTENTS

Chapter	Page
1 INTRODUCTION	1 - 7
2 REVIEW OF LITERATURE	8 - 11
3 INTENSITY OF SELECTION IN NORMAL POPULATIONS	12 - 23
3.1 Large Populations	
3.1.1 General distribution	
3.1.2 Normal distribution	
3.2 Small Populations	
3.2.1 General distribution	
3.2.2 Normal distribution	
3.3 Classification of Population	
4 INTENSITY OF SELECTION IN NON-NORMAL LARGE POPULATIONS	24 - 39
4.1 Pearson's Type I (First Main Type) Distribution	
4.1.1 Beta distribution	
4.2 Pearson's Type III (Transition Type) Distribution	
4.2.1 Gamma distribution	
4.2.2 Chi-square distribution	
4.3 Exponential Distribution	
4.4 Log-normal Distribution	
5 INTENSITY OF SELECTION IN NON-NORMAL SMALL POPULATIONS	40 - 54
5.1 Beta Distribution	
5.2 Gamma Distribution	
5.3 Exponential Distribution	
5.4 Log-normal Distribution	

Chapter	Page
6 EFFECT OF NON-NORMALITY ON RESPONSE TO SELECTION IN LARGE POPULATIONS	55 - 60
6.1 Theoretical Procedure	
6.2 Discrepancy in Response to Selection	
6.2.1 Normal vs. beta distribution	
6.2.2 Normal vs. gamma distribution	
6.2.3 Normal vs. exponential distribution	
6.2.4 Normal vs. log-normal distribution	
7 EFFECT OF NON-NORMALITY ON RESPONSE TO SELECTION IN SMALL POPULATIONS	61 - 65
7.1 Theoretical Procedure	
7.2 Discrepancy in Response to Selection	
7.2.1 Normal vs. beta distribution	
7.2.2 Normal vs. gamma distribution	
7.2.3 Normal vs. exponential distribution	
8 SUMMARY	66 - 67
REFERENCES	

CHAPTER - I

INTRODUCTION

Statistical geneticists are concerned with methods for changing the genetic specifications of plants and animals to make them more useful to man. This can be achieved through formulation of efficient breeding programmes on scientific lines and their implementation. Any programme of breeding involves : (i) the choice of mating system with reference to the particular population and (ii) the selection of individuals to be used as parents. The choice of mating system depends upon the genetic variation present in the population that is to be improved. If this variation is mostly additive, selection of parents combined with random mating is to be preferred but if the variation is mostly non-additive, selection from among two populations one each for either sex combined with their cross breeding would be adopted.

For a given system of mating, the main tool the breeder has in his control of improving the performance of stock is selection. By selection is meant ' causing or permitting some kinds of individuals to produce more offspring than other kinds do ' (Lush, 1945). It is the number raised and added to the breeding population rather than the number born which matters since those that are born but do not get a chance to reproduce cannot affect the composition of future population. The primary effect of selection is to change the gene frequencies increasing thereby the frequencies of gametes carrying certain gene combinations. In other words, a proper selection

produces two phenotypically discernible effects, namely (i) change in the mean value of a population by increasing the frequency of desirable genes and hence genotypes and (ii) shift in the original range of expression of the trait under selection by the preegration of new genetic combination. The change produced by selection that chiefly interests a breeder is the change of population mean which is known as response to selection or genetic gain per generation. The genetic gain ΔG_s is the difference of mean value between the offspring of the selected parents and whole of the parental generation before selection. The amount of selection applied can be expressed in terms of selection intensity or the percentage of population permitted to reproduce itself. The most useful measure of the intensity of selection actually practised is the difference between the average of those selected to be parents (\bar{x}_s) and the average of the whole population (\bar{x}_o) in which they were born. This superiority of selected parents (i.e., $(\bar{x}_s - \bar{x}_o)$) is often spoken as phenotypic selection differential (P.S.D.) and the standardised selection differential, $(\bar{x}_s - \bar{x}_o) / s_x$, is called the intensity of selection symbolised by I , s_x being the phenotypic standard deviation.

The genetic advance is estimated from the phenotypic selection differential by the following relation (Falconer, 1965)

$$\Delta G = b_{op} (\bar{x}_s - \bar{x}_o) \quad (1.1)$$

that. The intensity of selection will not be known, until the effect of selection, accuracy of selection and genotypic standard deviation of the traits the genetic gain depends upon three factors namely intensity of selection, the trait in question and σ_G is the genotypic standard deviation of that trait. It is the correlation between the selected criterion and genotype of where RGI is the intensity of selection in standard deviation units.

$$(1.3) \quad \Delta G = RGI \cdot \sigma_G$$

be shown to be equal

In general, for any criterion of selection I the genetic gain can

values of the character and is known as the accuracy of selection.

where σ_{Gx} is the correlation between the genotypic and phenotypic

$$(1.2) \quad \Delta G = \sigma_{Gx} \cdot D_x$$

$$= (\sigma_G / \sigma_x) \cdot D_x$$

$$= (\sigma_G^2 / \sigma_x^2)^{1/2}$$

$$\Delta G = b_{op}^2 (\bar{x}^2 - \bar{x}_0^2)$$

that of this the genetic gain is given by

ratio of additive genetic variation to the total phenotypic variation. In this

it can be shown that $b_{op}^2 = 1 - \frac{\sigma_g^2}{\sigma_p^2}$ is equal to the heritability, h^2 , which is the

where b_{op}^2 is the regression coefficient of offspring on midparent.

selection among the parental generation has actually been made. To be able to predict further ahead we need to know the magnitude of i , given the proportion, p_s , of the population saved and the phenotypic standard deviation, s_x , of the character. The expression of i in terms of these parameters can be obtained if the form of the frequency distribution of the character is known. The phenotypic values of the metric character are usually assumed to be normally distributed. The assumption of normality seems justified on the ground that the quantitative characters are governed by large number of factors called polygenes, having small and cumulative effects and superimposed upon them the effects of non-heritable factors. For this distribution, the intensity of selection, i , can be determined by the relation

$$i = \frac{z}{p} \quad (1.4)$$

where p is the proportion of population beyond the point of truncation and z is the height of the standard normal ordinate corresponding to the proportion, p . The relation between i and p given in equation (1.4), holds strictly only to a large sample, among which the selection is to be made. When selection is made out of a small number of measured individuals, the mean deviation of the selected group is a little less and can be found from tables of deviations of ranked data of Fisher and Yates (1938).

So, the two ways open to the breeder for improving the rate of response to selection are : one by increasing the accuracy of selection

and the other by reducing the proportion selected and thereby increasing the intensity of selection. Since in case of mass or individual selection the accuracy of selection is the square-root of the heritability of the character, there is very little one can do to increase the accuracy of selection if the heritability is high. For characters with low heritability the accuracy of selection can be enhanced if in selection in addition to individual's own performance appropriate attention is paid to the phenotype of its ancestors or progeny or collaterals. Such schemes of selection have been considered by Lush (1947), Lerner (1950), Osborne (1957 a,b), Jardine (1958), Le Rey (1958), Skjervold and Odegard (1959), Young (1961) and Jain and Narain (1974).

Reducing the proportion selected seems at first sight to be straightforward means of improving the response, but there is a lower limit to the number of individuals to be used as parents. For example, breeding dairy cattle for milk production the proportion of females culled in any generation cannot be more than one-third owing to their slow rate in reproduction and consequently involving a higher replacement rate. On the other hand the selection in males can be made more rigorous as very few bulls are needed for breeding purposes.

Now the next question that arises is how far the observed or realised responses agree with those expected. Almost all the selection experiments conducted specifically for the purpose have shown that the genetic change

which should be expected from the selection practice is not always realised (Dickerson, 1951, 1955 ; Falconer, 1953 ; Clayton, Morris and Robertson, 1957). The discrepancy is generally attributed to one or more of the following causes (Lerner, 1958 ; Jain and Gopalan, 1976) :

- (i) The criterion of selection may not be normally distributed ;
- (ii) Differential viability and fertility of parents resulting in unequal number of progeny per sire and dam family ;
- (iii) Use of non-robust estimates of genetic parameters ;
- (iv) Use of σ_G estimated from single population is not valid if the intensities of selection in the two sexes are different ;
- (v) No modification is incorporated for the selection practised at different stages in calculating the expected genetic gain ;
- (vi) The genetic model used for the analysis may not be appropriate ;
- (vii) Deterioration of environment may occur either independently of the selection practised or because of it ;
- (viii) The recurrent mutation may act in the direction opposite to that of selection ;
- (ix) Reduction of selection intensity because of selection for many traits at once is common ;
- (x) Antagonism between natural and artificial selection .

Not all the disturbing factors enumerated above are mutually exclusive. In fact, they overlap each other and sometimes represent different aspects of the same phenomenon. The intervention of one or more

of these factors produces a decrease in selection response either directly (reduction in effective heritability / accuracy of selection) or by lowering selection intensity. All these causes need thorough probing so that the formula used for predicting genetic gain is perfected. It is the purpose of this thesis to investigate the effect of one of the disturbing causes on response to selection, namely, the effect of normality assumption when in effect the criterion of selection follows a non-normal distribution.

The expressions for intensity of selection have been worked out for a number of non-normal distributions and compared with the corresponding expression for normal distribution for both large and small populations.

CHAPTER - II

REVIEW OF LITERATURE

For studies on prediction of response to selection it is usually assumed that the metric character or characters under study follow a normal distribution. Empirical studies, however, have borne out that the traits may depart from the normal form. Especially in the field of animal husbandry and poultry, it has been known for sometime past that the various characters related to milk yield and egg production, do not follow the normal law. Consequently the prediction of response to selection on the assumption of normality of the character may not be justified.

Although quite a few research workers, while discussing the realised genetic gain vis-a-vis that expected on the assumption of normality, have emphasized the need for using the selection intensity appropriate to the nature of the distribution (Lush, 1948 ; Henderson, 1963 ; Burrows, 1972 ; Falconer, 1975 ; Jain and Gopalan, 1976) but practically no attempt has been made to derive the expression of selection intensity for any continuous distribution other than the normal excepting that by Burrows (1972) who derived approximations for expected selection differentials for normally and exponentially distributed test scores as applicable to small populations. Several studies, have, however, been made to study the distributions of characters in the field of animal husbandry. These studies shew that Pearsonian system of curves (1895) provide good fit for most

of the characters, where the normal distribution fails. A brief resume of the same is as follows.

Quantitative characters like size of the foreheads of crabs (Kapteyn, 1903) and aflatoxin in peanut (Quesenberry, Whitaker and Dickens, 1976) have been usefully represented by lognormal distribution.

Pearl and Miner (1919) have made a study of the variation in milk yield and fat content of Ayrshire cows from the records of Ayrshire's Cattle Milk Records Committee for the years 1908 and 1909. The investigation was confined to those records of cows which had been 32 weeks or more in milk. The frequency distributions of milk yield and fat content of cows of the same age for each year of age fitted were of Pearson Type I, II, III and IV. Normal curves were also found to fit some of the distributions.

Gewen (1920) in a memoir giving the results of his analysis of the records of a Jersey herd over a number of years has reported Pearsonian Types I, II, III, IV and V for milk yield and butter fat.

Techer (1928) studied milk yield data of 4912 Ayrshire cows having complete records and also having a second calf within 60 weeks of previous calving. He found that age of cow followed Pearsonian distribution of Type I. Type IV was fitted for milk yield for all ages. This was not a good fit owing to the heterogeneity of material with respect to age. Pearsonian Type IV were good fits for milk yield for each year of age.

Om Prakash and Mahajan (1959) have studied frequency distribution of lactation yield, lactation length, age at first calving and calving interval for different order of lactations in four leading Indian herds - Red Sindhi and Kangayam herds of Hosur, Red Sindhi of Bangalore and Tharparkar herd of Patna. Among the 38 sets of data studied, 12 were found to be Pearsonian Type I, two to be J-shaped (exponential), 15 to be Type IV, three to be Type VI, one each of Type V and VII. The remaining four were found to be normally distributed.

Malhotra (1974) studied the distribution of scores of 1587 birds for rate of lay. The data came from the Regional Poultry Farm, Bhopal for the laying periods of 1971 and 1972. Selection scores were worked out for each bird using three different methods of selection viz., (a) selection on the basis of an index with optimum weights attached to the individual's performance and its full-sib family averages, (b) selection on the basis of an index with optimum weights attached to the individual's performance and its half-sib family averages and (c) selection on the basis of an index with optimum weights attached to individual's performance, full-sib family averages and half-sib family averages. Pearson Type I was found to be the appropriate distribution when the scores were based on the methods (a) and (b), whereas for the scores based on the method (c), normal distribution was appropriate. He found that there is a tendency in the distribution of selection scores to tend to normality when information on

more than two scores are combined.

The studies carried out at the Institute of Agricultural Research Statistics, on the data for the years 1972-76 pertaining to the Regional Poultry Farm, Bhopal showed that selection scores of birds for rate of lay followed Pearson Type I distribution. Pearson Type IV distribution fitted well for the character egg weight.

From the foregoing studies on the nature of frequency distribution of selection criteria it is seen that Peaysenian Types I and III provided adequate representation to many types of data in animal husbandry and poultry. Pearson Types II, IV and V have also fitted well in some of the cases. The expressions of selection intensity have been derived for Pearson Type I including its derivative namely beta distribution. Type III along with its derived distributions namely gamma, chi-square and exponential and for log normal distribution as applicable to both large and small populations.

CHAPTER - III

INTENSITY OF SELECTION IN NORMAL POPULATIONS

The expressions for selection intensity as applicable to both large and small populations when the character under study is distributed normally are given in this chapter.

3.1 Large Populations

3.1.1 General distribution. Let x denote the character on which selection is based. Further, let \bar{x}_g be the mean of the entire population of x and \bar{x}_s the average of the selected individuals. Then $(\bar{x}_s - \bar{x}_g)$ is defined to be the 'phenotypic selection differential' (P.S.D.) or 'reach'. The largest phenotypic selection differential is achieved through 'truncated selection' i.e. all values of $x < c$, a given point of truncation are chosen. In this case we could speak of selecting a fraction ' p ' with the highest value of x . Corresponding to ' p ' there is a point ' c ' and an \bar{x}_s which can be calculated if the distribution of the phenotypic values in the population is known. If, for instance, $f(x) dx$ is the probability that the phenotype of a random individual lies within the range $x, x + dx$ then, c , the point of truncation is defined to be the solution of the equation

$$p = \int_c^{\infty} f(x) dx \quad (3.1.1.1)$$

and \bar{x}_s is obtained as

$$\bar{x}_s = \frac{1}{p} \int_c^{\infty} x f(x) dx \quad (3.1.1.2)$$

The phenotypic selection differential which is the average superiority of the selected individuals over the mean of the population is obtained as

$$P.S.D. = \frac{1}{p} \int_c^{\infty} (x - \bar{x}_o) f(x) dx \quad (3.1.1.3)$$

Since, for different characters with differing distributions same p may result in different amounts of P.S.D. it is necessary to standardize P.S.D. by dividing by σ_x , the standard deviation of x . The standardized P.S.D. is taken as the measure of the intensity of selection, i.e.,

$$t = \frac{P.S.D.}{\sigma_x} = \frac{(\bar{x}_s - \bar{x}_o)}{\sigma_x} \quad (3.1.1.4)$$

where $(\bar{x}_s - \bar{x}_o)$ is given by (3.1.1.3)

3.1.2 Normal distribution. In the case of normal distribution the frequency function is given by

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left[-\frac{(x - \bar{x}_o)^2}{2\sigma_x^2} \right]$$

where \bar{x}_o is the population mean and σ_x^2 is the phenotypic variance.

The proportion, p , of individuals retained corresponding to the truncation point, c (Figure 1) is given by

$$p = \int_c^{\infty} \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left[-\frac{(x - \bar{x}_o)^2}{2\sigma_x^2} \right] dx$$

Using (3.1.1.3), the P.S.D. can be written as

$$P.S.D. = \frac{1}{p} \left[\frac{1}{\sigma_x \sqrt{2\pi}} \int_c^{\infty} (x - \bar{x}_o) \exp \left(-\frac{(x - \bar{x}_o)^2}{2\sigma_x^2} \right) dx \right] \quad (3.1.2.1)$$

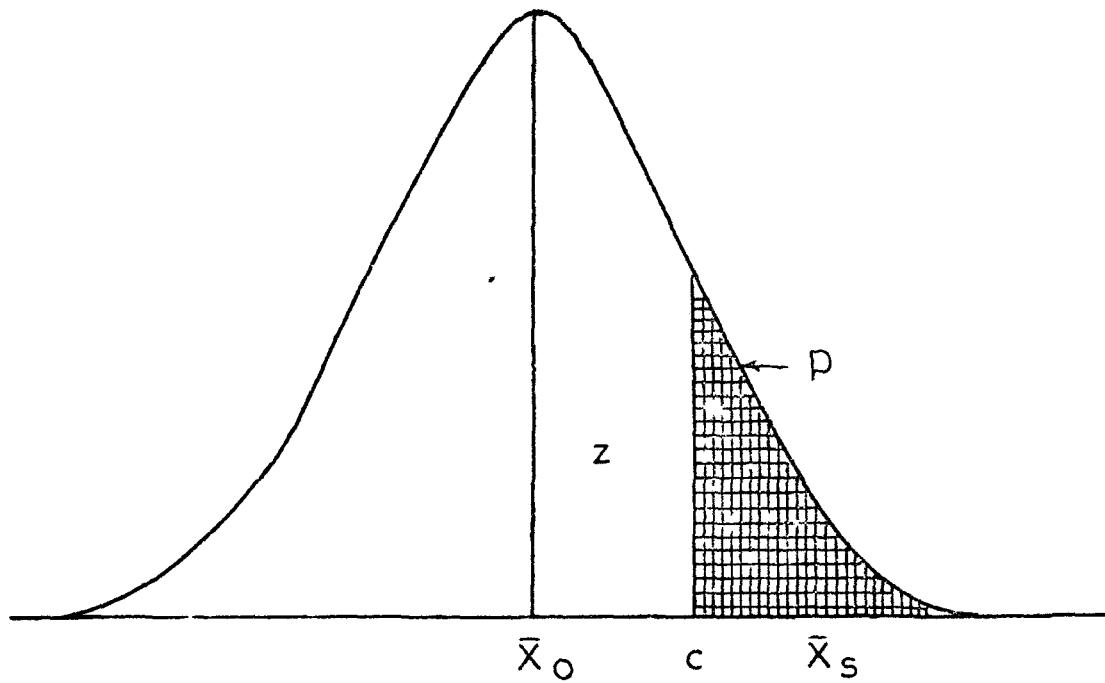


Fig.1 The individuals in the shaded area to the right of c , the point of truncation are retained for further breeding; the ones to the left are culled.

In order to evaluate (3.1.2.1), we make the following transformation

$$\begin{aligned} \frac{(x - \bar{x}_0)}{\sigma_x} &= y, \quad \frac{(c - \bar{x}_0)}{\sigma_x} = d, \quad \frac{dx}{\sigma_x} = dy, \text{ then} \\ \text{P.S.D.} &= \frac{\sigma_x}{p \sqrt{2\pi}} \int_c^{\infty} y e^{-y^2/2} dy \\ &= \frac{\sigma_x}{p \sqrt{2\pi}} e^{-d^2/2} \\ &= \frac{\sigma_x}{p} z \end{aligned}$$

where z is the ordinate at the cut-off point corresponding to the proportion p which can be read directly from Table II of Part II of Karl Pearson's (1931) tables.

The selection intensity, i , is then given by

$$i = \frac{\text{P.S.D.}}{\sigma_x} = \frac{z}{p} \quad (3.1.2.2)$$

This is a well-known expression used for obtaining the intensity of selection corresponding to a fraction, p , saved (Kempthorne, 1957; Lerner, 1958; Crow and Kimura, 1970; and Falconer, 1975).

3.2 Small Populations

3.2.1 General distribution. For small populations, the intensity of selection obtained by 'truncation selection' discussed earlier will be an over-estimate because of the discontinuity of the frequency distribution. The scheme of selection in small populations consists essentially in ranking all the n

individuals of the population in order of their magnitudes and selecting the desired number of x individual with the highest scores. The intensity of selection $t_{x/n}$, corresponding to the selection of x top-ranking individuals out of n , is thus defined as the average of the expected values of the first x standardized largest deviates (in phenotypic standard deviation units). Symbolically,

$$t_{x/n} = \frac{t_{1,n} + t_{2,n} + \dots + t_{x,n}}{x} \quad (3.2.1.1)$$

where

$$t_{x,n} = \frac{(\mu_{x,n} - \bar{x}_e)}{\sigma_x}, \quad x = 1, 2, 3, \dots, n \quad (3.2.1.2)$$

$\mu_{x,n} = E(x_{x,n})$, the expected value of the x -th largest order statistic in sample of size n , \bar{x}_e is the mean of the population and σ_x is the standard deviation of x .

In case of ties, i.e., when two individuals are adjudged to be equal, the average of the expected values of these observations is taken as the expected value for these observations (Fisher and Yates, 1938). Thus, in small populations the problem reduces in effect of obtaining the expected values of $\mu_{x,n}$'s ($x = 1, 2, \dots, n$) i.e., the expected values of the largest, second largest, third largest etc., down to the x -th largest order statistic in a sample of size n . These can be calculated if the distribution $f(x)$ of the phenotypic values in the population is known.

Following the approach given by David (1970), the n observations X_1, X_2, \dots, X_n when arranged in descending order of magnitude

are denoted as $X_{(1)}, X_{(2)}, \dots, X_{(n)}$. Let $f(x)$ and $F(x)$ denote the probability density function and cumulative distribution function of the character x . Further, let $F_r(x)$ ($r = 1, 2, \dots, n$) denote the cumulative distribution function of the r -th largest order statistic. Thus the cumulative distribution function of the largest statistic is given by

$$\begin{aligned} F_1(x) &= \Pr(X_{(1)} \leq x) \\ &= \Pr(\text{all } X_i \leq x) = \left[F(x) \right]^n \quad (3.2.1.3) \end{aligned}$$

The cumulative distribution function of the smallest statistic is

$$\begin{aligned} F_n(x) &= \Pr(X_{(n)} \leq x) = 1 - \Pr(X_{(n)} > x) \\ &= 1 - \Pr(\text{all } X_i > x) = 1 - \left[1 - F(x) \right]^n \quad (3.2.1.4) \end{aligned}$$

These are important special cases of the general result for $F_r(x)$

$$\begin{aligned} F_r(x) &= \Pr(X_{(r)} \leq x) \\ &= \Pr(\text{at least } (n-r+1) \text{ of the } X_i \text{'s are} \\ &\quad \text{less than or equal to } x) \\ &= \sum_{i=r}^n \binom{n}{i} F^i(x) \left[1 - F(x) \right]^{n-i} \quad (3.2.1.5) \end{aligned}$$

since the term in the summand is the binomial probability that exactly i of X_1, X_2, \dots, X_n are less than or equal to x . From the well-known relation between binomial sums and incomplete beta function, (3.2.1.5) can written as

$$F_r(x) = I_{F(x)}(n-r+1, r) \quad (3.2.1.6)$$

where the function

$$I_{F(x)}(a, b) = \int_0^{F(x)} t^{a-1} (1-t)^{b-1} dt / B(a, b)$$

Let $f_x(x)$ denote the probability density function of $X_{(x)}$, then we have from (3.2.1.5)

$$\begin{aligned} f_x(x) &= \frac{1}{B(n-r+1, r)} \frac{d}{dx} \int_0^{F(x)} t^{n-r} (1-t)^{r-1} dt \\ &= \frac{1}{B(n-r+1, r)} F^{n-r}(x) \int_{1-F(x)}^1 t^{r-1} dF(x) \end{aligned} \quad (3.2.1.7)$$

Hence the mean or the expected value of the r -th largest order statistic is given by the equation

$$\mu_{r,n} = \frac{n!}{(n-r)! (r-1)!} \int_{-\infty}^{\infty} x \left[F(x) \right]^{n-r} \left[1-F(x) \right]^{r-1} f(x) dx \quad (3.2.1.8)$$

After solving (3.2.1.8) for $\mu_{r,n}$, the expected value of the r -th largest deviate in standard deviation units i.e., $I_{r,n}$ given by (3.2.1.2) is obtained. The intensity of selection corresponding to the selection of r top-ranking individuals is then obtained by substituting the values of $I_{r,n}$ in equation (3.2.1.1).

3.2.2 Normal distribution. Without loss of generality, let us consider the standard normal population (mean zero and unit variance) with frequency function $f(x)$ given by

$$f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

The expected value of the r -th largest observation in a sample of size n from a standard normal population will then be given by equation (3.2.1.8) i.e.,

$$\mu_{r,n} = \frac{n!}{(n-r)! (r-1)!} \int_{-\infty}^{\infty} x^r F(x) \bar{F}^{n-r} \bar{F}^{r-1} f(x) dx \quad (3.2.2.1)$$

where

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-x^2/2} dx$$

Since in the standard normal distribution case $i_{r,n} = \mu_{r,n}$, the intensity of selection $i_{r,n}$ given by (3.2.1.1) will be computed as

$$i_{r,n} = \frac{\mu_{1,n} + \mu_{2,n} + \dots + \mu_{r,n}}{r} \quad (3.2.2.2)$$

An exact solution of the expression (3.2.2.1) is rather difficult to obtain. A number of attempts have been made to compute the values of $\mu_{r,n}$'s for different values of r and n by numerical integration. A brief resume of the work of some authors is discussed in what follows.

Fisher and Yates (1938, Table XX) published a two-decimal-place table of the values of $\mu_{r,n}$'s for samples of size $n = 2 (1) 50$. Their values are correct except for four errors of a unit in the last place, due to rounding.

Hastings, Mosteller, Tukey and Winsor (1947) computed the values of $\mu_{r,n}$'s upto five-decimal-places for samples of size $n = 2 (1) 10$. Their values are correct except when $n = 10$, where there are errors of 1 to 7 units in the last place.

Gedwin (1949) obtained more accurate values of $\mu_{r,n}$'s for samples of size $n = 2 (1) 10$. He tabulated the values upto seven-decimal-places.

E.S. Pearson and Hartley (1954, Table 28) published a table of $\mu_{r,n}$'s to three-decimal-places for $n = 2 (1) 20$ and to two-decimal-places

for $n = 21(1) 26(2) 50$; values for $n = 2(1) 10$ were compiled from Gedwin's table, those for $n = 11(1) 20$ were freshly computed, while those for $n > 20$ were taken from the table of Fisher and Yates. These values are correct except for three errors of a unit in the last place, due to rounding.

Federer (1951) used a somewhat different approach than did most of the aforementioned authors, who depended largely on numerical integration for the determination of tabular values. Federer used Tippett's (1925) table of expected values of the range to derive the expected value of the largest deviate which is one-half of the range in the case of normal distribution. The expected values of the second, third, fourth and fifth largest deviate were computed using recursion formulas which relate lower deviates with the next higher deviates of different sample sizes. Since we will be making use of this recursion formula for obtaining the standardized deviates $\lambda_{x,n}$ for different non-normal distributions in a later chapter, the set-up of the same in terms of $\mu_{x,n}$'s is given below:

The expected value of the largest deviate from a sample of size n from (3.2.1.8) is

$$\mu_{1,n} = n \int x \bar{F}(x) \bar{x}^{n-1} f(x) dx$$

which can be written as

$$\begin{aligned} \mu_{1,n} &= n \int x \bar{F}(x) \bar{x}^{n-2} f(x) dx - n \int x \bar{F}(x) \bar{x}^{n-2} \bar{F}(1-x) \bar{f}(x) dx \\ &= \frac{n}{n-1} \mu_{1,n-1} - \frac{1}{n-1} \mu_{2,n} \end{aligned}$$

Therefore, the expected value of the second largest member from a

sample of size n may be obtained from the largest members of samples n and $(n-1)$, i.e.,

$$\mu_{2,n} = n \mu_{1,n-1} + (n-1) \mu_{1,n}$$

The recursion formula for obtaining the next lower largest deviates can be found by the same device to be

$$\mu_{3,n} = \frac{1}{2} \left[n \mu_{2,n-1} + (n-2) \mu_{2,n} \right]$$

$$\mu_{4,n} = \frac{1}{3} \left[n \mu_{3,n-1} + (n-3) \mu_{3,n} \right]$$

$$\mu_{5,n} = \frac{1}{4} \left[n \mu_{4,n-1} + (n-4) \mu_{4,n} \right]$$

.....

.....

or in general,

$$\mu_{r,n} = \frac{1}{r-1} \left[n \mu_{r-1,n-1} + (n-r+1) \mu_{r-1,n} \right] \quad (3.2.2.3)$$

It can easily be seen that the recursion formula (3.2.2.3) holds in terms of standardized deviates as well and also for any general distribution, i.e.,

$$\begin{aligned} l_{r,n} &= \frac{\mu_{r,n} - \bar{x}_o}{\sigma_x} \\ &= \frac{1}{\sigma_x(r-1)} \left[n \mu_{r-1,n-1} + (n-r+1) \mu_{r-1,n} - (r-1) \bar{x}_o \right] \quad (\text{using } (3.2.2.3)) \\ &= \frac{1}{(r-1)} \left[- \frac{n(\mu_{r-1,n-1} - \bar{x}_o)}{\sigma_x} - \frac{(n-r+1)(\mu_{r-1,n} - \bar{x}_o)}{\sigma_x} \right] \\ &= \frac{1}{(r-1)} \left[n l_{r-1,n-1} + (n-r+1) l_{r-1,n} \right] \quad (3.2.2.4) \end{aligned}$$

Federer computed three-decimal-place values of the three largest normal deviates for sample of size $n = 41 (1) 200$ and two-decimal-place values of the fourth largest normal deviate for $n = 41 (1) 200$ and of the fifth largest normal deviate for $n = 41 (1) 100$. Because of loss of accuracy with repeated application of the recurrence formula, some of the Federer's values are in error by from 1 to 3 units in the last place. The recurrence formula is in the form

$$\mu_{r-1, n-1} = \frac{1}{n} \left[-(r-1) \mu_{r, n} + (n-r+1) \mu_{r-1, n} \right] \quad (3.2.2.5)$$

suggested in the paper of Harter (1961) is superior to the one used by Federer in the sense that it gives no serious accumulation of rounding errors.

Harter (1961) computed the values of $\mu_{r, n}$'s for samples of size $n = 2 (1) 100 (25) 250 (50) 400$ using the expression (3.2.2.1) in a slightly different way

$$\mu_{r, n} = \frac{n!}{(n-r)! (r-1)!} \int_{-\infty}^{\infty} x \left[\frac{1}{2} - F'(x) \right]^{r-1} \left[\frac{1}{2} + F'(x) \right]^{n-r} f(x) dx$$

where

$$F'(x) = \int_0^x f(x) dx$$

The values tabulated by Harter are accurate to within a unit in fifth - decimal - place.

Thus from the foregoing, it is seen that for greater accuracy and for samples upto 400 from normal distributions, the values tabulated by Harter (1961) can be used in computing the selection intensity $\frac{r}{n}$ corresponding to the selection of r individuals with the highest scores out of n .

3.3 Classification of Population

A pertinent question that arises in selection studies is that for what value of n a given population can be considered as large? As mentioned earlier the selection intensity in small populations obtained by truncation selection i.e., by using the relationship $I = n/p$ will be an overestimate. The criterion for treating the material as coming from a large population would obviously depend upon the magnitude of overestimation that can be tolerated. For this purpose, it seems reasonable to assume an overestimation of about 2 percent or less as negligible. With this yardstick, it can be seen from Table 1 which gives the intensity of selection for samples of different sizes that a sample of 100 or more measured units can treated as large. Even for a sample of size 50, the overestimation does not exceed 2 percent except for extremely heavy and extremely light cullings i.e., for values $p \leq 0.2$ and for $p \geq 0.8$. Thus when n is above 50, the assumption that the population is infinite in size is accurate for most purposes in computing the intensity of selection except in cases of extremely heavy or extremely light culling.

Selection intensities for different values of p in samples
of different size for normal distribution

Proportion selected,	Size of sample								
	50	100	150	200	250	300	350	400	450
p									
0.1	1.7e5	1.730	1.738	1.742	1.745	1.746	1.748	1.749	1.755
0.2	1.372	1.386	1.398	1.393	1.394	1.395	1.396	1.396	1.399
0.3	1.139	1.149	1.152	1.154	1.155	1.156	1.156	1.156	1.159
0.4	0.951	0.958	0.961	0.962	0.963	0.963	0.964	0.964	0.966
0.5	0.786	0.792	0.794	0.795	0.795	0.796	0.796	0.796	0.798
0.6	0.634	0.639	0.640	0.641	0.642	0.642	0.642	0.643	0.644
0.7	0.488	0.492	0.494	0.495	0.495	0.495	0.495	0.496	0.497
0.8	0.343	0.346	0.348	0.348	0.349	0.349	0.349	0.349	0.349
0.9	0.189	0.192	0.193	0.194	0.194	0.194	0.194	0.194	0.195

CHAPTER - IV

INTENSITY OF SELECTION IN NON - NORMAL LARGE POPULATIONS

In this chapter following the approach as outlined in section 3.1.1, we derive expressions for intensity of selection, i.e. when the quantitative character x under selection follows a continuous non-normal distribution. The different distributions considered are the Pearson's Type I, beta, Pearson's Type III, gamma, chi-square, exponential and lognormal.

4.1 Pearson's Type I (First Main Type) Distribution

When the character under consideration has the following probability density function

$$f(x) = y_0 \left[1 + \frac{x}{a_1} \right]^{n_1} \left[1 - \frac{x}{a_2} \right]^{n_2} \quad (4.1.1)$$

$$\text{where } -a_1 < x < a_2, \frac{n_1+1}{a_1} = \frac{n_2+1}{a_2} \text{ and } y_0 = \frac{\frac{n_1}{a_1} \frac{n_2}{a_2}}{(a_1+a_2)^{n_1+n_2+1} B(n_1+1, n_2+1)}$$

then it is said to follow Pearson's Type I distribution.

The mean and variance of the distribution are given by (Elderton and Johnson, 1969)

$$\bar{x}_0 = 0$$

and $\sigma_x^2 = \frac{(a_1+a_2)^2 (n_1+1) (n_2+1)}{(n_1+n_2+2)^2 (n_1+n_2+3)}$

The proportion of selected parents, p , corresponding to the truncation point c is

$$p = \int_c^{a_2} y_0 \left[1 + \frac{x}{a_1} \right]^{n_1} \left[1 - \frac{x}{a_2} \right]^{n_2} dx$$

$$\begin{aligned}
 &= \frac{y_o (a_1 + a_2)^{n_1 + n_2 + 1}}{a_1^{n_1} a_2^{n_2}} \int_d^1 z^{n_1} (1-z)^{n_2} dz, \\
 &\text{where } z = \frac{x+a_1}{a_1+a_2} \text{ and } d = \frac{c+a_1}{a_1+a_2} \\
 &= \int_d^1 \frac{z^{n_1} (1-z)^{n_2} dz}{B(n_1+1, n_2+1)} \\
 &= 1 - \int_0^d \frac{z^{n_1} (1-z)^{n_2} dz}{B(n_1+1, n_2+1)} \\
 &= 1 - I_d(n_1+1, n_2+1) \tag{4.12}
 \end{aligned}$$

where $I_x(p, q) = B_x(p, q) / B(p, q)$ and $B_x(p, q)$ is the incomplete beta function defined as

$$B_x(p, q) = \int_0^x u^{p-1} (1-u)^{q-1} du$$

The function $I_x(p, q)$ has been tabulated by Karl Pearson (1934) for different values of $x = (0.00(0.01)1)$, $p = (0.5(0.5)11.0(1), 50)$, and $q = (0.5(0.5)11.0(1)50)$ with $p \geq q$.

The mean value of the group selected is

$$\begin{aligned}
 \bar{x}_s &= \frac{y_o}{p} \int_c^{a_2} x \left[1 + \frac{x}{a_1} \right]^{-n_1} \left[1 - \frac{x}{a_2} \right]^{-n_2} dx \\
 &= \frac{y_o}{p} \int_c^{a_2} \left(x + a_1 \right)^{-n_1} \frac{(x+a_1)^{n_1} (a_2-x)^{n_2}}{a_1^{n_1} a_2^{n_2}} dx + \frac{y_o}{p} \int_c^{a_2} \frac{a_1 (a_1+x)^{n_1} (a_2-x)^{n_2}}{a_1^{n_1} a_2^{n_2}} dx \\
 &= \frac{y_o (a_1 + a_2)^{n_1 + n_2 + 2}}{p a_1^{n_1} a_2^{n_2}} \int_d^1 z^{n_1+1} (1-z)^{n_2} dz - \frac{y_o (a_1 + a_2)^{n_1 + n_2 + 1}}{p a_1^{n_1} a_2^{n_2}} a_1 \int_d^1 z^{n_1} (1-z)^{n_2} dz
 \end{aligned}$$

$$\begin{aligned}
 & \text{by taking } z = \frac{x + a_1}{a_1 + a_2} \text{ and } d = \frac{a_1 + c}{a_1 + a_2} \\
 & = \frac{(a_1 + a_2)(n_1 + 1)}{p(n_1 + n_2 + 2)} \int_d^{\frac{n_1 + 1}{a_1 + 2}(1 - z)^{\frac{n_2}{a_2}}} \frac{d}{B(n_1 + 2, n_2 + 1)} dz = s_1 \\
 & = \frac{(n_1 + a_2)(n_1 + 1)}{p(n_1 + a_2 + 2)} \quad [1 - I_d(n_1 + 2, n_2 + 1) + p]
 \end{aligned}$$

Hence, using relation (3.1.1.4) the selection intensity is

$$I = \frac{\sqrt{(n_1 + 1)(n_1 + a_2 + 3)}}{p\sqrt{(n_2 + 1)}} \quad [(1 - p) - I_d(n_1 + 2, n_2 + 1)] \quad (4.1.3)$$

Using relation (4.1.3) the selection intensities for different values of p , were worked out for different sets of parameters covering their entire range : low values of both the parameters ; high values of one and low value of the other and high values of both the parameters. In addition the selection intensities were also worked out for the two sets of parameters which are close to those obtained by Malhotra (1974) for rate of lay in poultry viz., for $n_1 = 3, n_2 = 1.5$; $n_1 = 4.5, n_2 = 1.5$ and are also given Table 2. The shape of the curves corresponding to the sets of parameters considered in Table 2 are as shown in Fig. 2. The curve is symmetrical, negatively skewed or positively skewed according as $n_1 = n_2$, $n_1 > n_2$ or $n_1 < n_2$. The excess of kurtosis or the peakedness increases with the increase in the value of either one or both the parameters. For the curves with long tail towards low-merit, selection intensity is more for higher values of p (mild selection) than if the distribution were truly normal. For the curves with long tail towards high-merit reverse is the case.

Table 2

R-3264

Selection intensities for 10 different values of p corresponding to
 different sets of values of the parameters n_1 and n_2 of Pearson's
 Type I distribution

Proportion

selected $n_1=1, n_2=1$ $n_1=2, n_2=1$ $n_1=1, n_2=2$ $n_1=48, n_2=1$ $n_1=1, n_2=48$ $n_1=48, n_2=48$ $n_1=9.0, n_2=4.5$,
 $n_2=1.5$ $n_2=1.5$

p	0.1	1.6627	1.5367	1.8235	1.1250	2.1479	1.7434	1.5524	1.4814
0.2	1.4e5e	1.3214	1.4815	1.0393	1.5779	1.3956	1.3218	1.2731	
0.3	1.196e	1.1429	1.2281	0.9475	1.2187	1.1558	1.1372	1.1e45	
0.4	1.01e5	0.9812	1.0157	0.8433	0.9656	0.9642	0.9376	0.9534	
0.5	0.8385	0.8274	0.8274	0.74e2	0.74e2	0.7998	0.82e5	0.81e3	
0.6	0.6737	0.6771	0.6541	0.6437	0.5622	0.6428	0.6719	0.67e1	
0.7	0.5126	0.5263	0.4898	0.5223	0.466e	0.4954	0.5235	0.5276	
0.8	0.3512	0.37e3	0.33e3	0.3945	0.2598	0.3489	0.37e3	0.3784	
0.9	0.1847	0.2e26	0.17e7	0.2387	0.1317	0.1937	0.2e4e	0.2136	

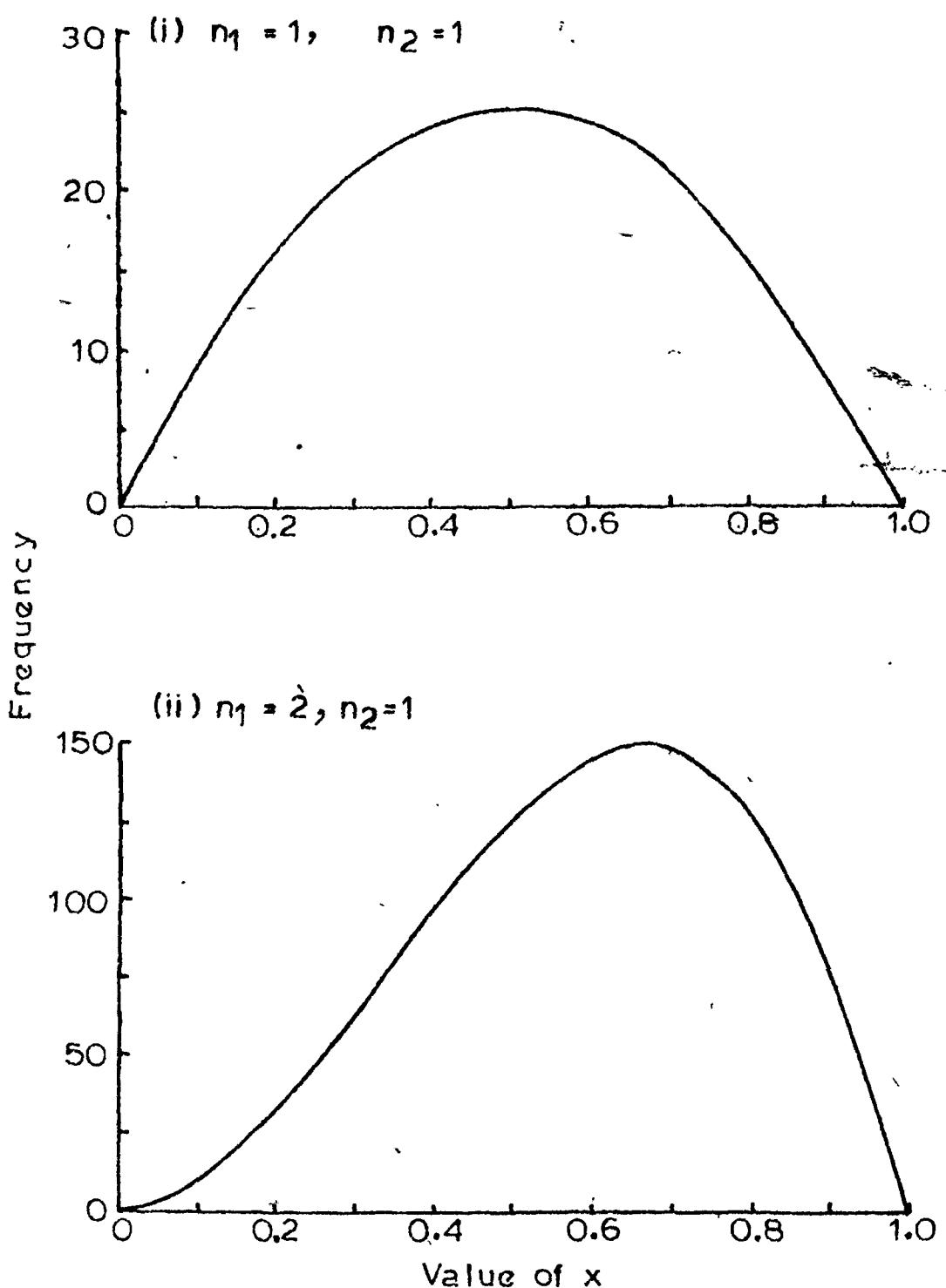
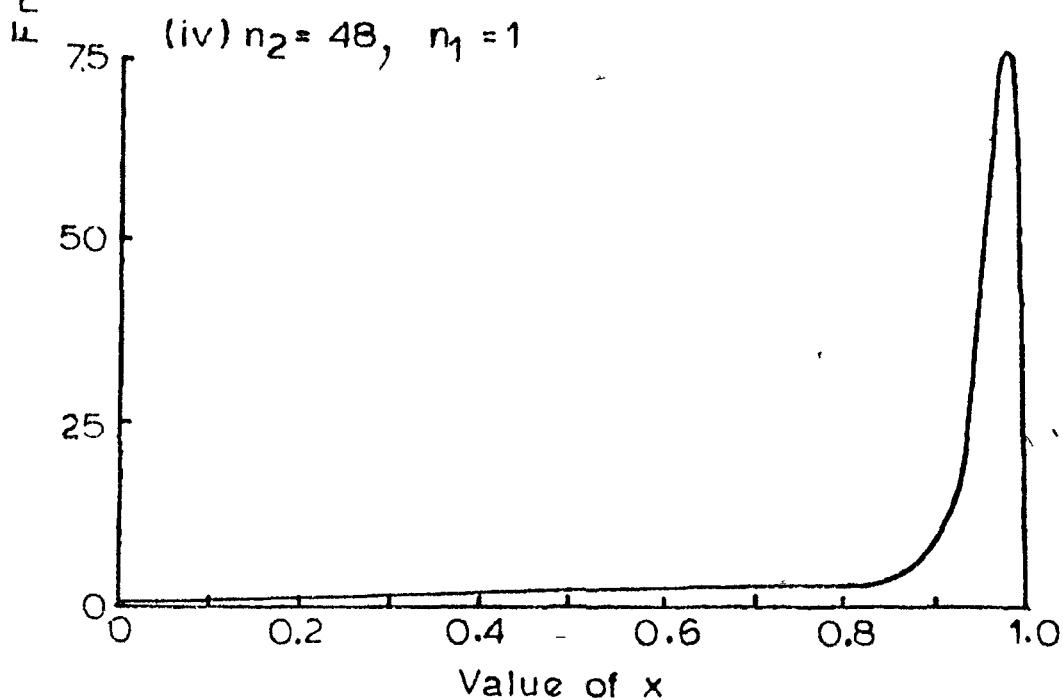
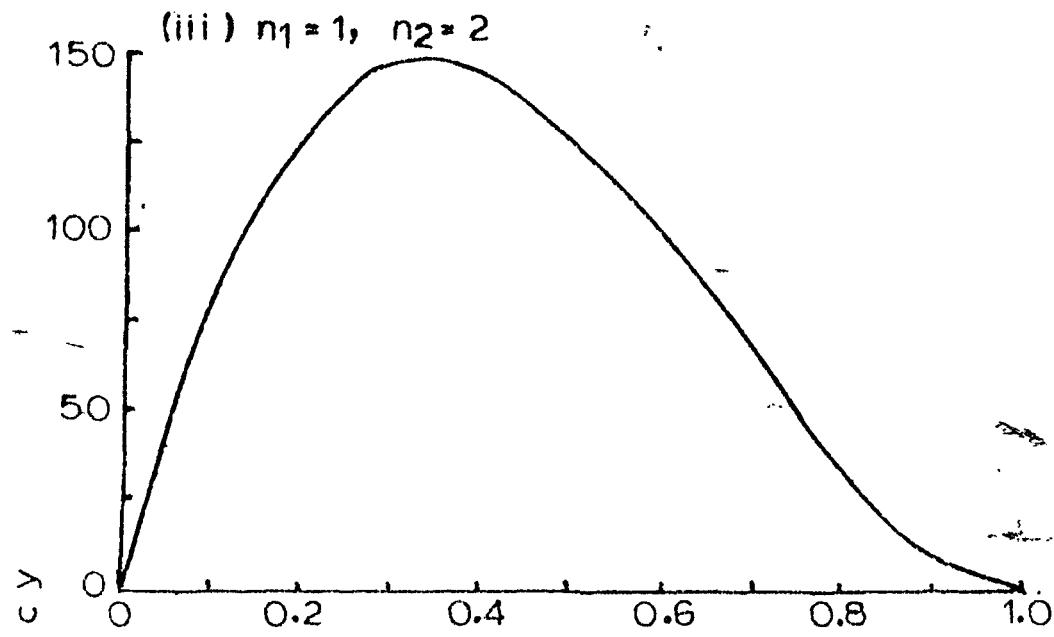
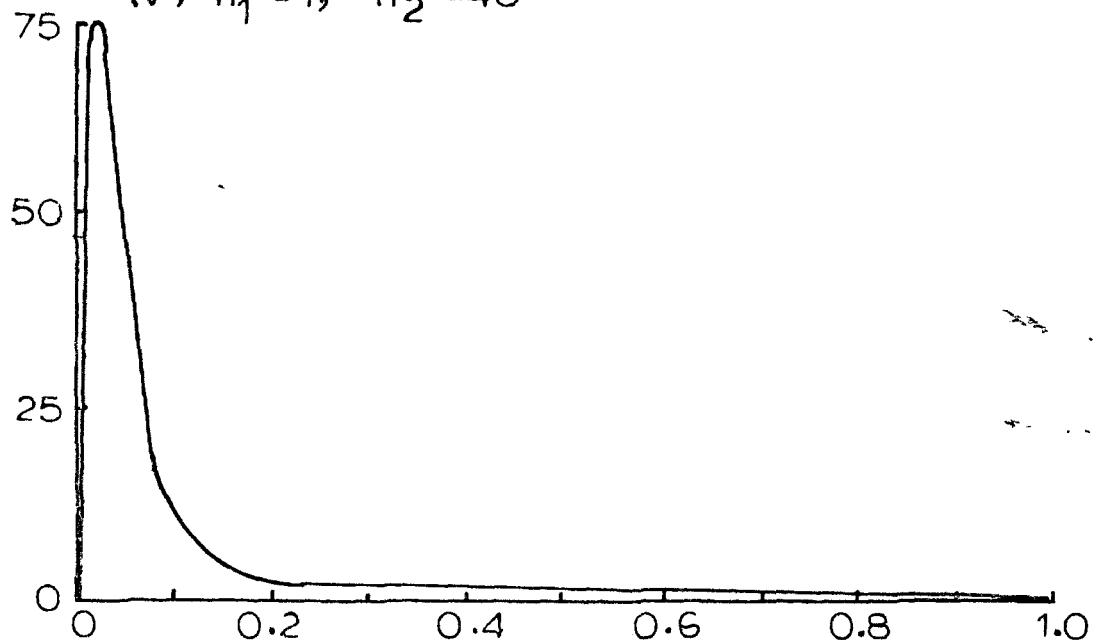


Fig. 2. Pearson's type I distribution for different parametric values



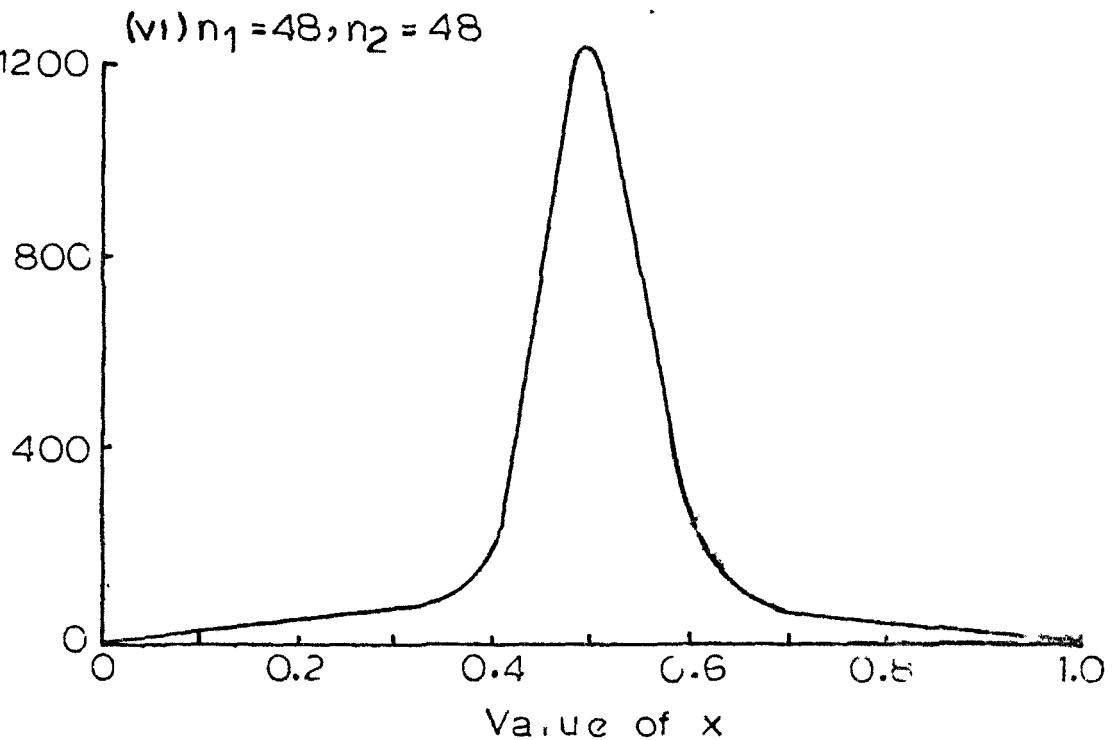
(v) $n_1 = 1, n_2 = 48$

Frequency

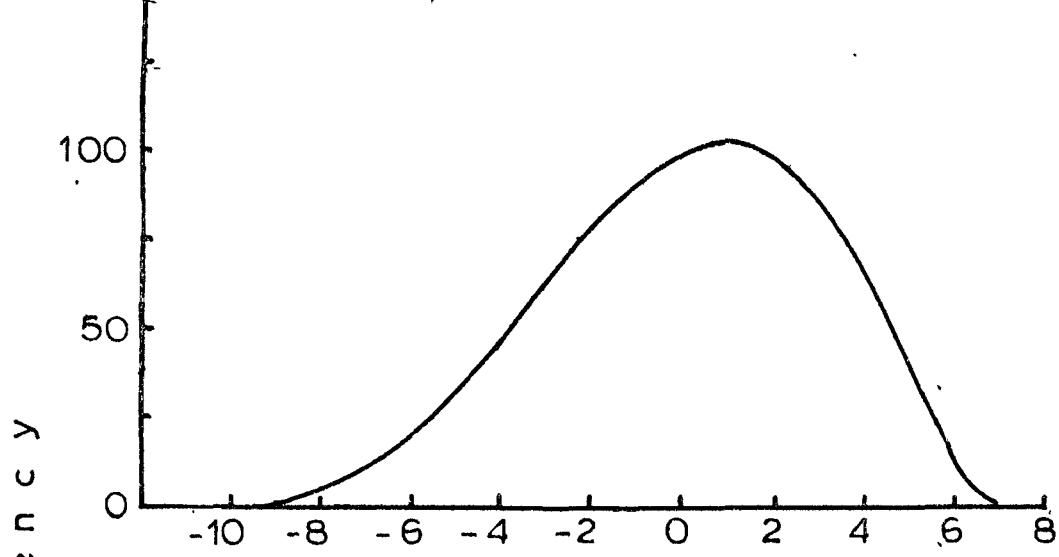


(vi) $n_1 = 48, n_2 = 48$

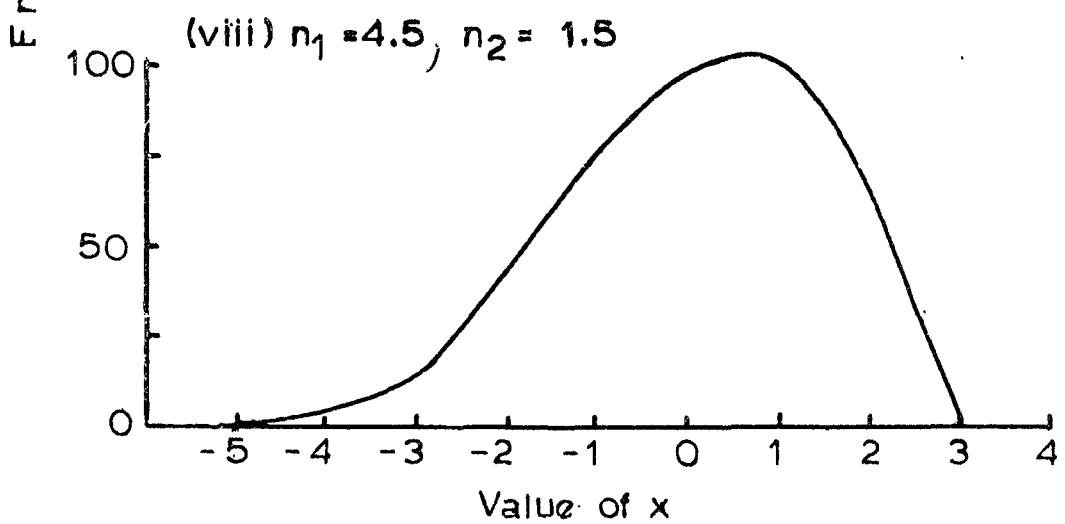
Frequency



(vii) $n_1 = 3.0, n_2 = 1.5$



(viii) $n_1 = 4.5, n_2 = 1.5$



4.1.1 Beta distribution. If we let $x = \frac{x + a_1}{a_1 + a_2}$, $a_1 = m_1 - 1$ and $a_2 = m_2 - 1$ in (4.1.1), the density function of Pearson's Type I distribution, it reduces to beta distribution with density (Mood, Graybill and Boes, 1974)

$$f(x) = \frac{x^{m_1-1} (1-x)^{m_2-1}}{B(m_1, m_2)} \quad 0 < x < 1 \quad (4.1.1.1)$$

The expression for intensity of selection accordingly was obtained directly from (4.1.3) as

$$t = \sqrt{\frac{m_1(m_1 + m_2 + 1)}{m_2}} - \frac{1}{p} \left[(1-p) - I_c(m_1 + 1, m_2) \right] \quad (4.1.1.2)$$

where $I_c(p, q) = B_c(p, q) / B(p, q)$ and $B_c(p, q)$ is the incomplete beta function defined as

$$B_c(p, q) = \int_0^c x^{p-1} (1-x)^{q-1} dx$$

and c is the point of truncation related to p by the relation

$$\begin{aligned} p &= \int_c^1 \frac{x^{m_1-1} (1-x)^{m_2-1}}{B(m_1, m_2)} dx \\ &= 1 - \int_0^c \frac{x^{m_1-1} (1-x)^{m_2-1}}{B(m_1, m_2)} dx \\ &= 1 - I_c(m_1, m_2) \end{aligned} \quad (4.1.1.3)$$

The selection intensities for different values of p for this distribution corresponding to a given set of values of the two parameters m_1 and m_2 can be obtained from that for Pearson's Type I distribution

with parameters $n_1 = m_1 - 1$ and $n_2 = m_2 - 1$ (Table 2).

4.2 Pearson's Type III (Transition Type) Distribution

If the character x has the following density function

$$f(x) = k_o \left(1 + \frac{x}{a}\right)^n e^{-\frac{(n+1)x}{a}} \quad (4.2.1)$$

where $k_o = \frac{(n+1)}{a e^{(n+1)} F(n+1)}$ and $-a \leq x < \infty$

then x is said to follow Pearson's Type III distribution.

The mean and variance of the distribution are given by

$$\bar{x}_o = \frac{a}{n+1}$$

$$\sigma_x^2 = \frac{a^2}{(n+1)}$$

The proportion of selected individuals, p , corresponding to the truncation point, c , is connected by the relation

$$p = k_o \int_c^{\infty} \left(1 + \frac{x}{a}\right)^n e^{-\frac{(n+1)x}{a}} dx$$

$$= \frac{k_o e^{n+1}}{a^n} \int_{(a+c)}^{\infty} z^n e^{-\frac{(n+1)z}{a}} dz,$$

$$\text{where } (x+a) = z, dx = dz$$

$$= \frac{a k_o e^{n+1}}{(n+1)^{n+1}} \int_0^{\infty} y^n e^{-y} dy$$

$$\text{on substituting } (n+1)y = a, y = \frac{a}{n+1}, b = \frac{(n+1)(a+c)}{a}$$

$$= \frac{1}{\Gamma(n+1)} \int_b^{\infty} y^n e^{-y} dy$$

$$\begin{aligned}
 &= 1 - \int_0^b \frac{y^n e^{-y}}{\Gamma(n+1)} dy \\
 &= 1 - I(u, n) \quad (4.2.2)
 \end{aligned}$$

where $u = \frac{b}{\sqrt{n+1}}$ and $I(u, p) = \int_0^{u\sqrt{p+1}} \frac{e^{-x} x^p}{\Gamma(p+1)} dx$

which has been extensively tabulated by Pearson (1946) for different values of u at intervals of 0.1 and $p = -1(0.05) 0(0.1) 5(0.2) 50$.

The mean of the selected individuals is

$$\begin{aligned}
 \bar{x}_s &= \frac{k_e}{p} \int_c^\infty x \left(1 + \frac{x}{a}\right)^n e^{-(n+1)x/a} dx \\
 &\approx \frac{k_e}{p} \int_0^\infty (x+a) \left(1 + \frac{x}{a}\right)^n e^{-(n+1)x/a} dx \\
 &\quad - \frac{k_e a}{p} \int_c^\infty \left(1 + \frac{x}{a}\right)^n e^{-(n+1)x/a} dx \\
 &= \frac{k_e a^2 e^{n+1}}{p} \int_{(a+c)}^\infty z^{n+1} e^{-(n+1)z/a} dz - a
 \end{aligned}$$

where $(x+a)/a = z$, $dx = a dz$

$$\begin{aligned}
 &\approx \frac{a}{p} \int_b^\infty \frac{y^{n+1} e^{-y}}{\Gamma(n+2)} dy - a \\
 &\text{on putting } (n+1)z = ay, \quad \frac{(n+1)(a+z)}{a} = b \quad \text{and}
 \end{aligned}$$

$$\frac{(n+1)}{a} dz = dy$$

$$= \frac{a}{p} \left[1 - I\left(\frac{b}{\sqrt{n+2}}, n+1\right) \right] - p$$

Hence, using relation (3.1.1.4) the selection intensity is

$$i = \frac{\sqrt{n+1}}{p} \left[(1-p) - I \left(\frac{b}{\sqrt{n+1}}, n+1 \right) \right] \quad (4.2.3)$$

The intensities of selection were tabulated (Table 3) for different values of p and $n = 4(1)5(5)20$. In addition, intensities were also worked out for $n = 48$, the value of the parameter of this distribution which was obtained by Pearl and Miner (1919) while studying the variation of lactation yield and fat contents of Ayrshire cows. The nature of the curves for three parametric values i.e., $n = 5, 15, 48$ is shown in Fig. 3. Since all the curves are positively skewed i.e., long tail of the curve is toward high-merit, selection intensity is more for heavy culling and less for mild culling.

4.2.1 Gamma distribution: Since this distribution is a special case of Pearson's Type III distribution where $\frac{(x+a)}{a} = y$, $(n+1) = k$. The expression for selection intensity, i , when the character under consideration follows a gamma distribution will therefore be

$$i = \frac{\sqrt{k}}{p} \left[(1-p) - I \left(\frac{\lambda c}{\sqrt{k+1}}, k \right) \right] \quad (4.2.1.1)$$

where λ is the scale parameter of the distribution and p , the proportion of individuals saved is related to the cut-off point, c , by

$$p = \int_{c}^{\infty} \frac{\lambda^k e^{-\lambda y} y^{k-1}}{\Gamma(k)} dy$$

$$= \int_{\lambda c}^{\infty} \frac{e^{-z} z^{k-1}}{\Gamma(k)} dz, \text{ where } \lambda y = z$$

Table 3

Intensities of selection for different values of p and n for Pearson's Type III distribution.

Properties selected,	$n=4$	$n=5$	$n=10$	$n=15$	$n=20$	$n=40$
$p = .1$	2.054	2.029	1.964	1.932	1.920	1.866
$p = .2$	1.542	1.529	1.504	1.490	1.481	1.454
$p = .3$	1.216	1.215	1.203	1.198	1.194	1.185
$p = .4$	0.978	0.978	0.977	0.977	0.976	0.976
$p = .5$	0.776	0.779	0.787	0.791	0.792	0.795
$p = .6$	0.602	0.607	0.619	0.624	0.627	0.633
$p = .7$	0.445	0.451	0.465	0.471	0.474	0.483
$p = .8$	0.298	0.303	0.316	0.328	0.326	0.336
$p = .9$	0.155	0.158	0.168	0.174	0.176	0.184

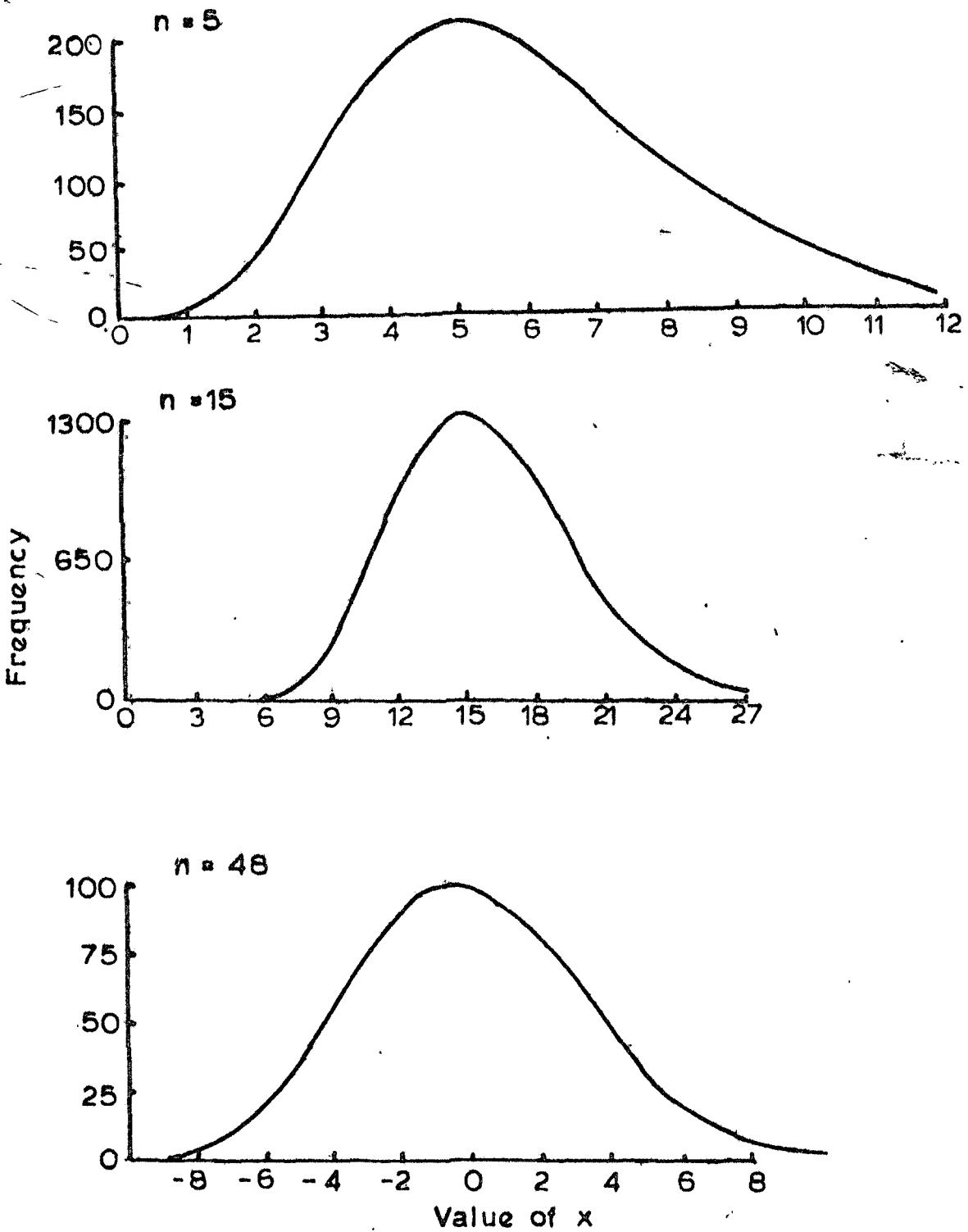


Fig.3. Pearson's type III distribution for different parametric values

$$\begin{aligned}
 &= 1 - \int_0^{AC} \frac{e^{-z} z^{k-1}}{\Gamma(k)} dz \\
 &= 1 - I\left(\frac{\lambda c}{k}, k-1\right) \quad (4.2.1.2)
 \end{aligned}$$

and the function $I(u, p)$ is as defined earlier.

The selection intensities for different values of p , proportion retained, for this distribution which is a particular case of Pearson's Type III distribution can be obtained from Table 3 by taking $n = k = 1$.

4.2.2 Chi-square distribution. Pearson's Type III distribution reduces to chi-square distribution when $\frac{(x+a)}{a} = z$ and $(n+1) = m/2$, where m is the number of degrees of freedom. The expression for selection intensity, I , for χ^2 -distribution from (4.2.3) reduces to

$$I = \frac{\sqrt{m}}{\sqrt{2p}} \left[(1-p) + I\left(\frac{c/2}{\sqrt{\frac{m}{2}+1}}, \frac{m}{2}\right) \right] \quad (4.2.2.1)$$

where p , the proportion of parents selected and c , the point of truncation are related by

$$p = \int_c^{\infty} \frac{e^{-z/2} z^{m/2-1}}{\Gamma(m/2)} dz$$

$$= \int_{c/2}^{\infty} \frac{e^{-y} y^{m/2-1}}{\Gamma(m/2)} dy, \text{ on putting } (z/2) = y$$

$$= 1 - \int_0^{c/2} \frac{e^{-y} y^{m/2-1}}{\Gamma(m/2)} dy$$

$$= 1 - I \left(\frac{c/2}{(m/2)} + \frac{m}{2} - 1 \right) \quad (4.2.2.2)$$

and the function $I(u, p)$ is as defined earlier.

The values of selection intensity for different proportions of individuals retained for this distribution which is a particular case of Pearson's Type III distribution can be obtained from Table 3 by taking $m = 2(n+1)$.

4.3 Exponential distribution

If the density function of character x has the following form

$$f(x) = \theta e^{-\theta x}, \quad x \geq 0 \text{ and } \theta > 0 \\ = 0, \text{ otherwise} \quad (4.3.1)$$

the character is said to follow an exponential distribution with parameter θ .

The mean and variance of this distribution are (Mood, Graybill and Boes, 1974)

$$\bar{x}_0 = 1/\theta$$

$$\sigma_x^2 = 1/\theta^2$$

The proportion of the group chosen, p , corresponding to the truncation point c is

$$p = \int_c^{\infty} \theta e^{-\theta x} dx \\ = e^{-\theta c}$$

$$\text{Thus } c = -(1/\theta) \log_e p \quad (4.3.2)$$

The mean of the selected individuals is

$$\bar{x}_s = \frac{1}{p} \int_c^{\infty} x \theta e^{-\theta x} dx$$

The difference between the mean of the selected parents and the population average is

$$\begin{aligned}
 \text{P.S.D.} &= \bar{x}_s - \bar{x}_e = \frac{1}{p} \left[\int_c^{\infty} e^{-\theta x} x dx - \int_c^{\infty} e^{-\theta x} dx \right] \\
 &= \frac{1}{p} \left[\int_c^{\infty} e^{-\theta x} (\theta x - 1) dx \right] \\
 &= \frac{c e^{-\theta c}}{p}
 \end{aligned}$$

Using relation (4.3.2) we have

$$\bar{x}_s - \bar{x}_e = -(1/e) \log_e p$$

Then, the intensity of selection is

$$\begin{aligned}
 i &= - \left[(1/e) \log_e p \right] / (1/e) \\
 &= - \log_e p
 \end{aligned} \tag{4.3.3}$$

It is a remarkable fact that the intensity of selection for the exponential distribution is independent of the parameter θ .

The intensity of selection can be obtained alternatively from the corresponding expression for the Pearson's Type III distribution noting that when $n = 0$ and $(1/\alpha) = \theta$, Pearson's Type III distribution reduces to exponential distribution.

The selection intensities were computed for different values of $p = 0.001 (0.009) 0.01 (0.01) 0.10 (0.05) 0.95$ (Table 4). The curve is as shown in Fig. 4. The selection intensity is more for intense selection and less for higher values of p than if the distribution were truly normal.

Digitized by srujanika@gmail.com

Intensities of selection for different values of p for experimental distribution

<i>Properties</i>	<i>Intensit y of preferenc e selected.</i>	<i>Intensit y of selection.</i>	<i>P</i>	<i>P</i>
0.001	1.284	0.98	0.98	0.25
0.01	1.650	0.605	0.605	0.29
0.02	3.912	0.46	0.46	0.22
0.03	3.567	0.45	0.45	0.23
0.04	3.219	0.55	0.55	0.05
0.05	2.996	0.55	0.55	0.05
0.06	2.813	0.60	0.60	0.06
0.07	2.659	0.65	0.65	0.07
0.08	2.526	0.70	0.70	0.08
0.09	2.406	0.75	0.75	0.09
0.10	2.303	0.80	0.80	0.10
0.11	2.223	0.85	0.85	0.11
0.12	2.163	0.897	0.897	0.12
0.13	2.116	0.94	0.94	0.13
0.14	2.074	0.98	0.98	0.14
0.15	2.036	1.00	1.00	0.15
0.16	2.000	1.00	1.00	0.16
0.17	1.966	1.00	1.00	0.17
0.18	1.934	1.00	1.00	0.18
0.19	1.904	1.00	1.00	0.19
0.20	1.876	1.00	1.00	0.20

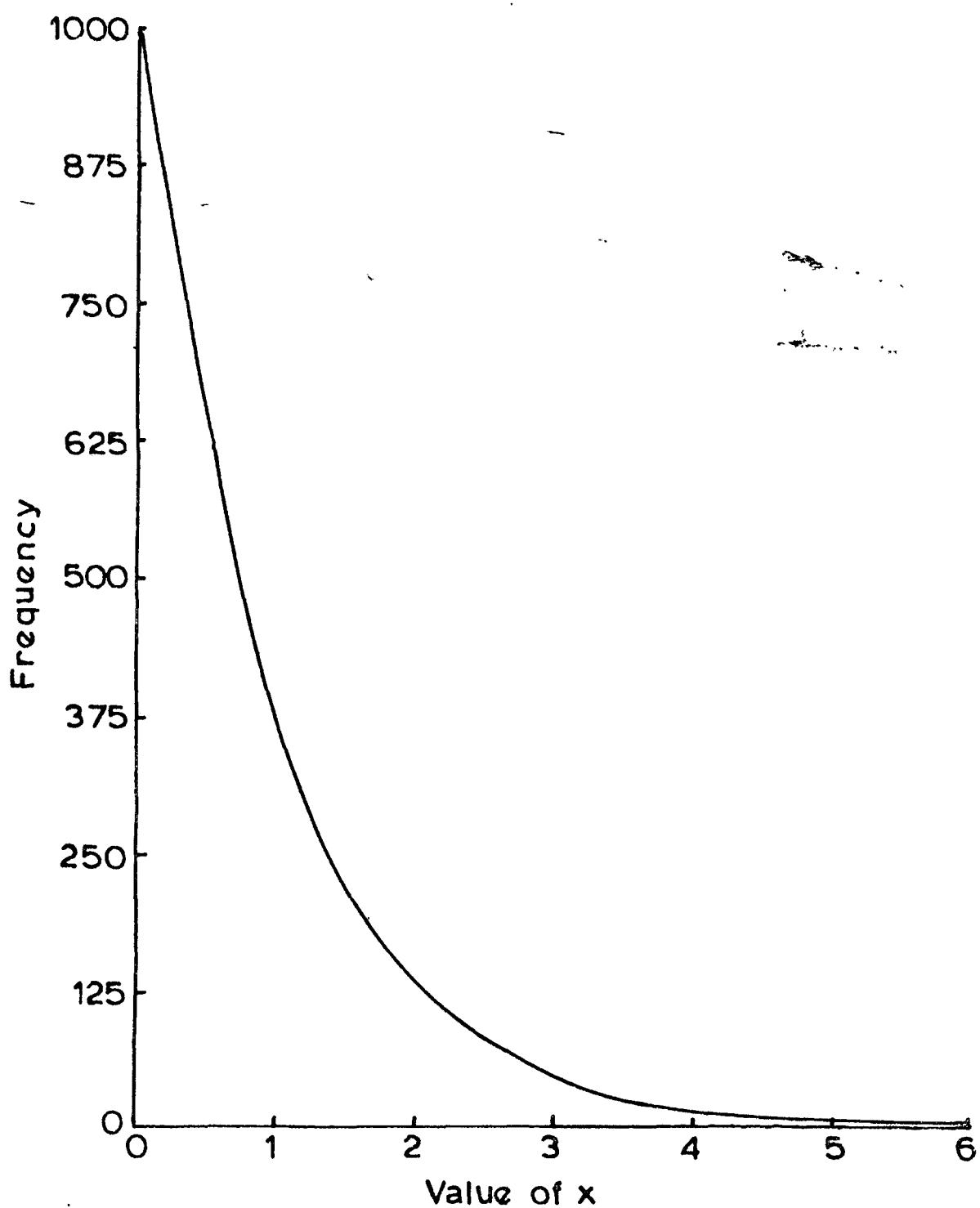


Fig 4 Exponential distribution for $\Theta=1$

4.4 Log-normal Distribution

When the logarithms of observations are distributed normally the character x is said to follow a lognormal distribution. The density function when $\log x$ is normally distributed with mean zero and unit variance is given by

$$f(x) = \frac{1}{\sqrt{2\pi}x} \exp \left[-\frac{1}{2} (\log x)^2 \right], \quad 0 < x < \infty \quad (4.4.1)$$

The mean of this distribution, \bar{x}_e , is

$$\bar{x}_e = \int_0^\infty x \frac{1}{\sqrt{2\pi}x} \exp \left[-\frac{1}{2} (\log x)^2 \right] dx$$

On substituting $\log x = w$,

$$= \int_{-\infty}^{\infty} \frac{e^{w/2}}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} (w+1)^2 \right] dw$$

$$= e^{-1/2}$$

The variance of this distribution, σ_x^2 , similarly, can be shown as

$$\sigma_x^2 = e(e-1)$$

The proportion of individuals saved, p , is related to the truncation point, e , by

$$p = \int_e^\infty \frac{1}{\sqrt{2\pi}x} \exp \left[-\frac{1}{2} (\log x)^2 \right] dx$$

On taking $\log x = w$

$$= \int_{\log e}^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} w^2 \right) dw \quad (4.4.2)$$

The average of the selected parents is

$$\begin{aligned}
 \bar{x}_s &= \frac{1}{p} \int_c^{\infty} \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} (\log x)^2 \right] dx \\
 &= \frac{1}{p} \int_{\log c}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(1-\frac{w}{2})^2} dw, \text{ on substituting } \log x = w \\
 &= \frac{e^{1/2}}{p \log c \sqrt{2\pi}} \int_{\log c}^{\infty} \exp \left[-\frac{1}{2} (w+1)^2 \right] dw \\
 &= \frac{e^{1/2}}{p \log c \sqrt{2\pi}} \int_{\log c-1}^{\infty} \exp \left(-\frac{1}{2} t^2 \right) dt, \text{ by taking } (w+1) = t \\
 &= \frac{e^{1/2}}{p} \left[\frac{1}{2} \cdot \int_{-\infty}^{\log c-1} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} t^2 \right) dt \right]
 \end{aligned}$$

where $\phi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ is the cumulative normal distribution

function tabulated by Sheppard (1903) and which are reproduced in Fisher and Yates (1938).

The intensity of selection, i.e. is therefore

$$\frac{\bar{x}_s - \bar{x}_0}{\bar{x}_x} = \frac{e^{1/2}}{p \sqrt{e-1}} \left[\frac{1}{2} - p \cdot \phi(\log c + 1) \right] \quad (4.4.3)$$

Table 5 gives selection intensities for different values of

$p = 0.01 (0.04) 0.05 (0.08) 0.95$ while the nature of the curve is depicted in Fig. 5. Since the curve is positively skewed, the selection intensity is more for lower values of p and vice versa than if the distribution were truly normal.

Table 5
Probabilities of selection for different values
of p for logarithmic distribution

p	Probability of selection					
0.01	0.28	0.50	0.59	0.64	0.65	0.66
0.05	3.26	6.55	6.55	6.44	6.38	6.31
0.10	2.21	2.21	6.66	6.65	6.59	6.15
0.15	1.89	6.65	6.65	6.31	6.21	6.25
0.20	1.28	6.76	6.76	6.26	6.36	6.36
0.25	1.14	6.75	6.75	6.21	6.16	6.12
0.30	0.97	6.84	6.84	6.16	6.08	6.08
0.35	0.83	6.85	6.85	6.12	6.05	6.05
0.40	0.71	6.90	6.90	6.08	6.03	6.03
0.45	0.61	6.95	6.95	6.04	6.00	6.00

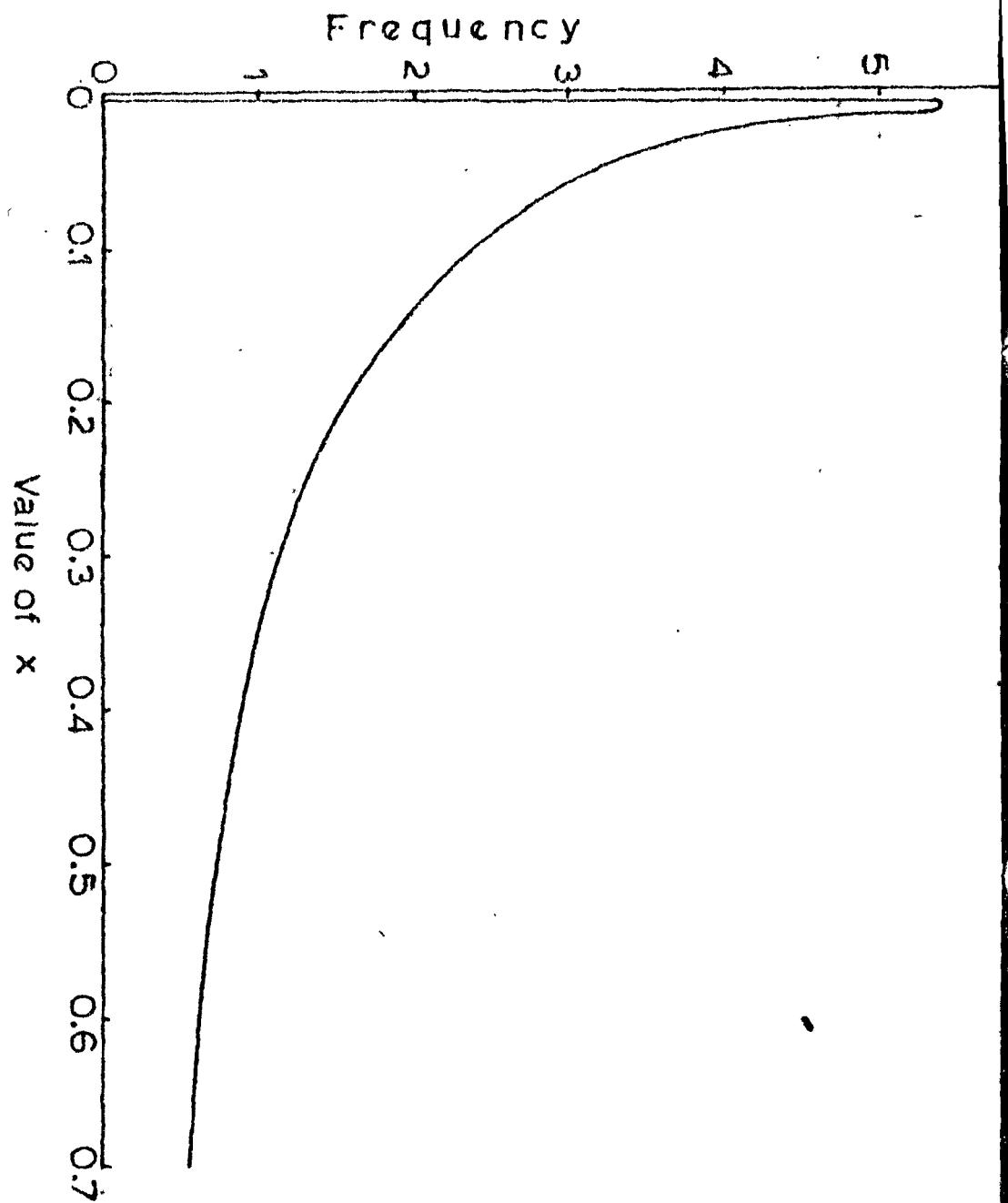


Fig. 5. Lognormal distribution

CHAPTER - V

INTENSITY OF SELECTION IN NON - NORMAL SMALL POPULATIONS

This chapter deals with the derivation of corresponding expressions for selection intensity, i.e., for different non-normal continuous distributions as applicable to small populations. The methodology adopted for their derivation is as outlined earlier in section 3.2.1.

The general frequency functions of Pearson Type I and Type III are rather difficult to handle and as such their derivatives namely beta and gamma distributions respectively were considered.

5.1 Beta Distribution

Suppose that we have n observations each distributed according to the beta probability density function.

$$f(x) = \frac{x^{m_1-1} (1-x)^{m_2-1}}{B(m_1, m_2)}, \quad 0 < x < 1, \quad m_1, m_2 > 0 \quad (5.1.1)$$

The mean and variance of the population are

$$\bar{x}_0 = \frac{m_1}{(m_1 + m_2)}$$

$$\sigma_x^2 = \frac{m_1 m_2}{(m_1 + m_2)^2 (m_1 + m_2 + 1)} \quad (5.1.2)$$

The cumulative distribution function of a beta distributed random variable is

$$F(x) = \int_0^x \frac{x^{m_1-1} (1-x)^{m_2-1}}{B(m_1, m_2)} dx$$

On repeated integration by parts, $F(x)$ can be expressed as

$$F(x) = \sum_{i=0}^{m_2-1} \left(m_1 + m_2 - 1 \right)_c \frac{x^{m_1+m_2-1-i}}{i!} (1-x)^i$$

The expected value of the r -th largest statistic using relation (3.2.1.8) is

$$\begin{aligned} \mu_{r,n} &= \frac{n!}{(n-r)! (r-1)!} \int_0^1 x \left[\sum_{i=0}^{m_2-1} \frac{m_2-1}{m_1+m_2-1} \left(m_1 + m_2 - 1 \right)_c \frac{x^{m_1+m_2-1-i}}{i!} (1-x)^i \right]^{n-r} \left[\sum_{i=0}^{m_2-1} \frac{m_2-1}{m_1+m_2-1} \left(m_1 + m_2 - 1 \right)_c \frac{x^{m_1+m_2-1-i}}{i!} (1-x)^i \right]^{r-1} \frac{x^{m_1-1} (1-x)^{m_2-1}}{B(m_1, m_2)} dx \\ &= \frac{n!}{(n-r)! (r-1)! B(m_1, m_2)} \int_0^1 x^{m_1} (1-x)^{m_2-1} \left[\sum_{i=0}^{m_2-1} \frac{m_2-1}{m_1+m_2-1} \left(m_1 + m_2 - 1 \right)_c \frac{x^{m_1+m_2-1-i}}{i!} (1-x)^i \right]^{n-r} \left[\sum_{p=0}^{r-1} \frac{r-1}{p} \left(m_1 + m_2 - 1 \right)_c \frac{x^{m_1+m_2-1-p}}{p!} (1-x)^{p+1} \right]^{r-p} (-1)^p \left(\sum_{i=0}^{m_2-1} \frac{m_2-1}{m_1+m_2-1} \left(m_1 + m_2 - 1 \right)_c \frac{x^{m_1+m_2-1-i}}{i!} (1-x)^i \right)^{r-p} dx \\ &= \frac{n!}{(n-r)! (r-1)! B(m_1, m_2)} \int_0^1 x^{m_1} (1-x)^{m_2-1} \end{aligned}$$

$$\left[\sum_{p=0}^{r-1} x^{r-1} e_p (-1)^p \sum_{l=0}^{m_2-1} {}_{m_1+m_2-1} c_l \right] x^{\frac{m_1+m_2-1-p}{2}} (1-x)^{\frac{l}{2}} \int^{n+r+p}_{-\frac{1}{2}} dx$$

Now using multinomial expansion (Mood, Graybill and Boes , 1974 , p. 531)
the expression in the parenthesis can be written as

$$\begin{aligned} & \sum_{l=0}^{m_2-1} {}_{m_1+m_2-1} c_l x^{\frac{m_1+m_2-1-p}{2}} (1-x)^{\frac{l}{2}} \int^{n+r+p} \\ & = (n+r+p)! \sum_{\substack{m_2-1 \\ m_1+m_2-1=p \\ \sum n_i = n+r+p}} \prod_{i=0}^{m_2-1} {}_{n_i} c_i x^{\frac{m_1+m_2-1-p}{2}} (1-x)^{\frac{l}{2}} \int^{n+r+p} \\ & = (n+r+p)! \prod_{i=0}^{m_2-1} \frac{n_i!}{\prod_{j=0}^{m_2-1} n_j!} \end{aligned}$$

where the summation extends over all possible combinations of non-negative integers $n_0, n_1, n_2, \dots, n_{m_2-1}$ such that $n_0 + n_1 + \dots + n_{m_2-1} = n+r+p$, $\mu_{x,n}$ can be written as

$$\begin{aligned} \mu_{x,n} &= \frac{n!}{(n+r+p)!} \int_0^1 x^{m_1} (1-x)^{m_2-1} \\ &\quad \left[\sum_{p=0}^{r-1} x^{r-1} e_p (-1)^p (n+r+p)! \prod_{\substack{i=0 \\ \sum n_i = n+r+p}}^{m_2-1} \frac{{}_{n_i} c_i}{\prod_{j=0}^{m_2-1} n_j!} \right. \\ &\quad \left. x^{\frac{m_1+m_2-1-p}{2}} (1-x)^{\frac{l}{2}} \int^{n+r+p} dx \right] \end{aligned}$$

Substituting the values of m_{π^+} , m_{π^-} and a^* in relation (3.2.1.2).

$$(5.1.3) \quad \frac{\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{m_1^n + m_2^n - 1}}{\sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{m_1^n - 1} \cdot \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{m_2^n - 1}}$$

$$\frac{d^{(l-x)}(x^2w^2 + 1)^{-\frac{1}{2}}}{dx} = \frac{(l-x)(l+x)(l+2x)(l+3x)\dots(l+(m-1)x)}{(x^2w^2 - 1)^{\frac{1}{2}}} \cdot \frac{d^{(l-x)}(x^2w^2 + 1)^{-\frac{1}{2}}}{dx}$$

$$\frac{\int_0^x \frac{t^{m-1} (1-t)^{n-1}}{(m+n-1)!} dt}{\int_0^1 \frac{t^{m-1} (1-t)^{n-1}}{(m+n-1)!} dt} = \frac{x^m (1-x)^n}{B(m+n, m+n)} = \frac{x^m (1-x)^n}{(m+n)!}$$

$$\frac{1}{1-x} \cdot \frac{(x-1)}{(1-1-x)(1-x)} \cdot \frac{x}{1+x} = \frac{1}{1-x}$$

$$z_{ux} + l_{ux}) \frac{u}{x} : (1-x) : (x-u)$$

$$P_{r,n} = \frac{(m_1+m_2) \sqrt{(m_1+m_2+1)}}{\sqrt{m_1 m_2}} \left[\frac{n!}{(n-r)! (r-1)! B(m_1, m_2)} \sum_{p=0}^{r-1} \sum_{i=0}^{r-1-p} c_i \frac{(-1)^p}{\prod_{i=0}^{m_2-1} n_i!} \right]$$

$$(n-r+p) : \sum_{\sum n_i = n-r+p} (-m_1-m_2-1)_c \sqrt{n_0} \dots (-m_1-m_2-1)_c \sqrt{n_{m_2-1}}$$

$$\frac{B \sqrt{-m_1 + (m_1+m_2-1)} \sum_{i=0}^{m_2-1} n_i - \sum_{i=0}^{m_2-1} (n_i+1)}{\prod_{i=0}^{m_2-1} n_i!} = \frac{m_1}{m_1+m_2}$$

(5.1.4)

The expected value of the r -th standardized largest deviate from a sample of size n i.e. $\frac{1}{\sqrt{n}}$ has been tabulated for three sets of parametric values which are likely to be encountered in practical situations viz., $m_1 = 2$, $m_2 = 2$; $m_1 = 3$, $m_2 = 2$; $m_1 = 2$, $m_2 = 3$ (Table 6 (a), (b), (c)).

5.2 Gamma Distribution

Suppose we have n independent random observations x_1, x_2, \dots, x_n each distributed according to the gamma probability density function.

$$f(x) = \frac{e^{-x} x^{k-1}}{Q(k)}, \quad 0 < x < \infty \quad (5.2.1)$$

The mean and variance of the distribution are

$$\bar{x}_d = k$$

$$\sigma_x^2 = k$$

The cumulative distribution function $F(x)$ is:

Table 6 (a)

Expected value of the r -th largest standardized deviate for beta distribution with parametric values $m_1 = 2$ and $m_2 = 2$ upto samples of size $n = 10$

Table 6 (b)

Expected value of the r -th largest standardized deviate $\hat{x}_{r,n}$
 for beta distribution with parametric values $m_1 = 3$ and $m_2 = 2$
 upto samples of size $n = 10$

Table 6 (c)

Expected value of the r-th s largest standardized deviate $i_{x,n}$ for beta distribution with parametric values $m_1 = 2$ and $m_2 = 3$ upto sampling of size n=10

$$F(x) = \int_0^x \frac{e^{-t} t^{k-1}}{\Gamma(k)} dt$$

To evaluate $F(x)$, consider Taylor's formula with integral remainder,

$$f(z) = f(a) + zf'(a) + \frac{z^2}{2!} f''(a) + \dots + \frac{z^{k-1}}{(k-1)!} f^{k-1}(a) + \frac{z^k}{(k-1)!} \int f^k(tx)(1-t)^{k-1} dt$$

On putting $f(z) = e^z$, we get

$$e^z = \sum_{i=0}^{k-1} \frac{z^i}{i!} + \frac{z^k}{(k-1)!} \int_0^1 f^k(tx)(1-t)^{k-1} dt$$

$$1 = \sum_{i=0}^{k-1} \frac{e^{-z} z^i}{i!} + \frac{z^k}{(k-1)!} \int_0^1 e^{-z(1-t)} (1-t)^{k-1} dt$$

$$= \sum_{i=0}^{k-1} \frac{e^{-z} z^i}{i!} + \frac{z^k}{(k-1)!} \int_0^z e^{-x} \frac{x^{k-1}}{z^k} dx, \text{ on substituting } z(1-t)=x$$

$$\int_0^z \frac{e^{-x} x^{k-1}}{\Gamma(k)} dx = 1 - \sum_{i=0}^{k-1} \frac{e^{-z} z^i}{i!} \quad (5.2.3)$$

Thus, an incomplete gamma function can be expressed as a partial sum of probabilities in Poisson distribution.

Using relation (5.2.3), $F(x)$ can be written as

$$F(x) = 1 - \sum_{i=0}^{k-1} \frac{e^{-x} z^i}{i!}$$

From relation (3.2.1.8) the expected value of the x -th largest statistic is,

$$\mu_{x,n} = \frac{n!}{(n-x)! (x-1)!} \int_0^\infty x \left[1 - \sum_{i=0}^{k-1} \frac{e^{-x} z^i}{i!} \right]^{n-x} \left[\sum_{i=0}^{k-1} \frac{e^{-x} z^i}{i!} \right]^{x-1} \frac{e^{-x} x^{k-1}}{\Gamma(k)} dx$$

$$= \frac{n!}{(n-r)! (r-1)!} \int_0^{\infty} \sum_{p=0}^{\infty} \frac{n-r}{p} \frac{n-r}{p} c_p (-1)^p \left[\sum_{l=0}^{k-1} \frac{x^l}{l!} \right]^p \left[\sum_{l=0}^{k-1} \frac{e^{-x} x^l}{l!} \right]^{r-1} \frac{e^{-rx} x^k}{k!} dx$$

$$\text{since } \left[\sum_{l=0}^{k-1} \frac{x^l}{l!} \right]^{r-1} = \sum_{p=0}^{\infty} \frac{n-r}{p} \frac{n-r}{p} c_p (-1)^p \left[\sum_{l=0}^{k-1} \frac{e^{-x} x^l}{l!} \right]^p$$

$$= \frac{n!}{(n-r)! (r-1)!} \int_0^{\infty} \sum_{p=0}^{\infty} \frac{n-r}{p} \frac{n-r}{p} c_p \frac{(-1)^p}{p!} e^{-(r+p)x} \sum_{l=0}^{k-1} \frac{x^l}{l!} \sum_{m=0}^{r+p-1} x^m dx$$

$$= \frac{n!}{(n-r)! (r-1)!} \int_0^{\infty} \sum_{p=0}^{\infty} \frac{n-r}{p} \frac{n-r}{p} c_p \frac{(-1)^p}{p!} e^{-(r+p)x} \sum_{m=0}^{(k-1)(r+p-1)} a_m (k, r+p-1) x^m dx$$

$$a_m (k, r+p-1) x^m \sum x^k dx$$

where $a_m (k, r+p-1)$ is the coefficient of x^m in the expansion

of $\left[\sum_{l=0}^{k-1} \frac{x^l}{l!} \right]^{r+p-1}$. m varying from 0 to $(k-1)(r+p-1)$

$$= \frac{n!}{(n-r)! (r-1)!} \sum_{p=0}^{\infty} \frac{n-r}{p} \frac{n-r}{p} c_p \frac{(-1)^p}{p!} \sum_{m=0}^{(k-1)(r+p-1)} \frac{(m+k+1)}{(r+p)^{m+k+1}} a_m (k, r+p-1)$$

(5.2.4)

Substituting the values of \bar{x}_0 , σ_x^2 and $\mu_{x,n}$ in (3.2.1.2) the expression for $t_{x,n}$ is

$$t_{x,n} = \frac{1}{\sqrt{k}} \int_{-\infty}^{\infty} \frac{n!}{(n-r)! (r-1)!} \sum_{p=0}^{\infty} \frac{n-r}{p} \frac{n-r}{p} c_p \frac{(-1)^p}{p!} \sum_{m=0}^{(k-1)(r+p-1)} \frac{(m+k+1)}{(r+p)^{m+k+1}}$$

$$a_m (k, r+p-1) \frac{x^m}{m+k+1} dx \quad (5.2.5)$$

The expected values of the r -th largest standardized deviate $i_{r,n}$ was computed (Table 7) corresponding to the value of $k = 5$ using the table of expected values of the r -th ordered statistic compiled by Gupta (1960) upto samples of size $n = 10$.

5.3 Exponential Distribution

If the observations on character x follow an exponential distribution, then the probability function is

$$f(x) = \theta e^{-\theta x}, \theta > 0, x \geq 0 \quad (5.3.1)$$

The mean and variance of the population are

$$\bar{x}_0 = 1/\theta$$

$$\sigma_x^2 = 1/\theta^2 \quad (5.3.2)$$

The cumulative distribution function $F(x)$ is given by

$$F(x) = \int_0^x \theta e^{-\theta x} dx \\ = 1 - e^{-\theta x}$$

From relation (3.2.1.8) the expected value of the r -th largest statistic is

$$\mu_{r,n} = \frac{n!}{(n-r)! (r-1)!} \int_0^\infty x \left[1 - e^{-\theta x} \right]^{n-r} (e^{-\theta x})^{r-1} (\theta e^{-\theta x}) dx$$

Integrating by parts,

$$= \frac{n! (\theta x) (1 - e^{-\theta x})^{n-r}}{(n-r)! (r-1)!} \left[\frac{(e^{-\theta x})^r}{(-\theta x)} \right]_0^\infty + \frac{n! \theta}{(n-r)! (r-1)!} \int_0^\infty \frac{(e^{-\theta x})^r}{(r\theta)} dx \\ \left[(1 - e^{-\theta x})^{n-r} + x(n-r) (1 - e^{-\theta x})^{n-r-1} \theta (e^{-\theta x}) \right] dx$$

Table 7

Expected values of the r -th largest standardized deviate $i_{r,n}$ for standard gamma distribution with parameter $k = 5$ upto sample of size $n = 10$

$$= \frac{n!}{(n-r)! r!} \int_0^\infty (e^{-\Theta x})^r (1-e^{-\Theta x})^{n-r} dx + \frac{n! (n-r)}{(n-r)! r!} \int_0^\infty e^r (e^{-\Theta x})^{r+1} \\ \times (1-e^{-\Theta x})^{n-r-1} dx$$

On substituting $e^{-\Theta x} = u$, in the first term

$$= \frac{n!}{(n-r)! r!} \int_0^\infty \frac{u^{r+1} (1-u)^{n-r}}{\Theta} du + \frac{n! (n-r)}{r! (n-r-1)!} \int_0^\infty (e^{-\Theta x})^{r+1} (1-e^{-\Theta x})^{n-r-1} x dx \\ = \frac{1}{r\Theta} + \frac{n!}{(n-r-1)!} \frac{1}{r+1} \int_0^\infty (e^{-\Theta x})^{r+1} (1-e^{-\Theta x})^{n-r-1} x dx$$

Proceeding in this manner we obtain

$$\mu_{x,n} = \frac{1}{\Theta r} + \frac{1}{\Theta (r+1)} + \frac{1}{\Theta (r+2)} + \dots + \frac{1}{\Theta n} \\ = \frac{1}{\Theta} \sum_{l=r}^n \frac{1}{l} \quad (5.3.3)$$

Substituting the values of $\mu_{x,n}$, \bar{x}_e and σ_x in relation (3.2.1.2)

we obtain

$$t_{r,n} = L \left[\frac{1}{\Theta} \sum_{l=r}^n \frac{1}{l} - \frac{1}{\Theta} \bar{J} \right] / (1/\Theta) \\ = \sum_{l=r}^n \frac{1}{l} - 1 \quad (5.3.4)$$

The r -th largest deviate in standard deviation units is thus independent of the parameter Θ .

The values of $t_{r,n}$ for samples of size upto $n = 10$ were worked out (Table 8) using the tables of Gupta (1960) on ordered statistics for gamma distribution with parameter $k = 1$.

Table 8

Expected value of the r -th largest standardized deviate $i_{r,n}$ for exponential distribution upto samples of size $n = 10$

5.4 Log-normal Distribution

The density function for the observations from lognormal population when $\log x$ is normally distributed with zero mean and unit variance is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} (\log x)^2 \right] \quad (5.4.1)$$

The mean and variance of the distribution are

$$\begin{aligned}\bar{x}_d &= e^{1/2} \\ \sigma_x^2 &= d(e+1)\end{aligned} \quad (5.4.2)$$

The cumulative distribution function $F(x)$ is

$$\begin{aligned}F(x) &= \int_0^x \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{(\log x)^2}{2} \right] dx \\ &= \int_{-\infty}^{\log x} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{t^2}{2} \right) dt\end{aligned}$$

The expected value of the x -th largest statistic is,

$$\mu_{x,n} = \frac{n!}{(n-x)! (x-1)!} \int x (F(x))^{n-x} (1-F(x))^{x-1} f(x) dx$$

An exact solution of this integral does not exist. This, therefore, has to be numerically integrated. Accordingly, a computer programme was drawn up and run on the computer. However, because of the truncation of exponential - logarithmic series, the values obtained were in serious error. Hence, the values of $\mu_{x,n}$ have not been tabulated.

CHAPTER - VI

EFFECT OF NON-NORMALITY ON RESPONSE TO SELECTION IN LARGE POPULATIONS

6.1 Theoretical Procedure

For a given basis of selection I^* , and a proportion p , of the population saved, we denote the predicted response by ΔG when the basis of selection is assumed to follow normal distribution and by ΔG^* when it follows the actual distribution. Symbolically,

$$\Delta G = I R_{GI} \sigma_G \quad (6.1.1)$$

$$\Delta G^* = I^* R_{GI} \sigma_G \quad (6.1.2)$$

where I and I^* denote respectively the intensity of selection corresponding to the proportion p saved from a normal population and that from the actual population.

The proportionate discrepancy, D in the predicted response to selection on the assumption of normality relative to that based on actual distribution will therefore be given by

$$\begin{aligned} D &= \frac{\Delta G - \Delta G^*}{\Delta G} \\ &= \frac{\frac{I}{I^*} - 1}{\frac{I}{I^*}} - 1, \text{ using (6.1.1) and (6.1.2)} \\ &= R(*) - 1, \text{ say} \end{aligned} \quad (6.1.3)$$

When $R(*) > 1$, the progress is overestimated; when it is less than 1, the progress is underestimated and when $R(*) = 1$, the progress remains unaffected.

For simplicity and uniformity of discussion of results for large and small populations, instead of general Type I and Type III distributions their

derivatives viz., beta, gamma, exponential, and lognormal were considered. This would not affect the general conclusions drawn.

6.2 Discrepancy in Response to Selection

Table 9 gives the percentage discrepancy in response to selection relative to different non-normal distributions. The parametric values considered for beta and gamma distributions cover the entire range of values including those obtained by research workers in the course of their studies (Malhotra, 1974; Pearl and Miner; 1919).

6.2.1 Normal vs. beta distribution. Since beta distribution is characterized by two parameters viz., m_1 and m_2 , the nature of the curve depends on the magnitude and relationship between the two parameters. The curve is symmetrical, negatively skewed, or positively skewed according as (i) $m_1 = m_2$ (ii) $m_1 > m_2$ and (iii) $m_1 < m_2$. The excess of kurtosis or the peakedness increases with the increase in the value of the either one or both the parameters. Thus, we consider the p behaviour of response to selection for the three situations separately.

(Case (i))

The response to selection is overestimated to the same extent for extremely heavy and extremely low cullings. For values of p ranging from 0.2 to 0.8, the progress is underestimated for low values of the parameters and overestimated for high values. For low values of both m_1 and m_2 the underestimation reaches its maximum of 5 percent at $p = 0.5$ and decreases as the value of p deviates from 0.5 on either side. On the other hand, for

Table 9
Percentage discrepancy in response to selection relative to different
non-normal large populations

Prop- ort- ion sele- cted, p	Percentage discrepancy relative to											
	Beta				Gamma				Exponential ^a Lognormal ^b			
	$m_1=2$	$m_1=48$	$m_1=3$	$m_1=49$	$m_1=2$	$m_1=3$	$m_1=2$	$m_1=5.5$	$k=5$	$k=12$	$k=16$	$k=49$
	$m_2=2$	$m_2=48$	$m_2=3$	$m_2=2$	$m_2=3$	$m_2=49$	$m_2=2.5$	$m_2=2.5$				
0.1	5.55	0.66	14.20	48.09	-3.76	-18.29	18.47	-14.54	-9.18	-5.95	-23.80	-20.59
0.2	-0.37	0.30	5.93	34.68	-5.52	-11.29	9.95	-9.85	-6.05	-5.76	-13.00	9.36
0.3	9.310	0.27	1.40	22.31	-5.63	-4.90	4.93	-4.71	-3.28	-2.16	-3.74	19.48
0.4	-4.43	0.16	-1.56	14.52	-4.92	0.03	1.31	-1.29	-1.17	-1.00	5.82	36.04
0.5	-4.85	-0.25	-3.57	7.79	-3.57	7.79	-1.53	2.88	0.93	0.40	15.13	53.43
0.6	-4.43	0.16	-4.92	0.03	-1.56	14.52	-3.91	6.96	3.19	1.74	26.01	69.45
0.7	-3.10	0.27	-5.63	-4.90	1.40	22.31	-5.86	11.63	5.55	2.84	39.13	91.63
0.8	-0.37	0.30	-5.52	-11.29	5.93	34.68	-7.52	17.63	8.59	4.17	56.92	118.72
0.9	5.55	0.66	-3.76	-18.29	14.20	48.09	-8.71	25.93	12.18	5.81	85.71	143.74

high values of both the parameters, the overestimation increases as the value of p deviates from half on either side. Further, the overestimation never exceeds one percent, implying thereby that for high and equal values of the parameters the normal distribution is quite robust. However, for low values of the parameters, the normal approximation to the beta distribution holds either for moderately heavy culling ($p = 0.2$) or for moderately low culling ($p = 0.8$).

Case (ii)

The response is overestimated for intense selection and underestimated for mild selection. The overestimation for low values of p is proportionately more than underestimation for high values of p . Further, the discrepancy in response to selection increases with the increase in the peakedness. For the highly peaked curves i.e., when $m_1 = 49$; $m_2 = 2$, the progress is overestimated by as much as 50 percent for heavy culling and underestimated by as much as 20 percent for low culling. However, for situations which are more likely to be encountered in practice (i.e., low values of both the parameters) the discrepancy in using normal distribution instead of actual distribution is negligible for intermediate values of p .

Case (iii)

Since the only difference in this situation and the previous one is in respect of direction of skewness which is on the right instead of being to the left, the results with respect to overestimation and underestimation of progress are exactly the reverse as in case (ii).

6.2.2 Normal vs. gamma distribution. Gamma distribution is positively skewed i.e., long tail towards high merit, and the degree of skewness and kurtosis decrease with the increase in the value of the parameter, k (Table 10).

Table 10

Coefficients of skewness and kurtosis
for gamma distribution for different
values of k

k	γ_1	γ_2
1	2.000	6.000
5	0.894	1.200
6	0.816	1.000
11	0.603	0.545
16	0.500	0.375
21	0.436	0.286
49	0.286	0.122

The percentage discrepancy, will therefore decrease with the increase in the value of k as can be seen from Table 9. The response is underestimated for heavy culling and overestimated for mild culling. The underestimation does not exceed 24 percent even for extreme culling and for extreme departures from symmetry and flatness (i.e., $p = 0.1$ and $k = 1$). On the other hand, the overestimation of progress in such a situation under mild culling (i.e., $p = 0.9$ and $k = 1$) can be as much as 86 percent of the expected under normal form. For the form of distribution which was obtained by Pearl and Miner (1919) i.e., when $k = 49$, the maximum underestimation under intense selection is of the order of 6 percent whereas the overestimation does not exceed 6 percent even under very mild selection. Thus, for large values of k ,

and moderate culling ($0.2 < p < 0.6$), the approximation of the gamma to the normal distribution in the estimation of response to selection is quite satisfactory,

6.2.3 Normal vs. exponential distribution. In this case the progress is underestimated by about 25 percent for intense selection ($p = 0.1$) and overestimated by as much as 88 percent for mild selection ($p = 0.9$). Except when p is in the neighbourhood of 0.5 and 0.4, it is not advisable to use the normal approximation as the discrepancy for other value of p is too serious to ignore.

6.2.4 Normal vs. lognormal distribution. The response to selection in using normal distribution when actually it is lognormal is always overestimated for all values of $p > 0.2$ and the overestimation increases with the decrease in the rigour of selection. For very mild selection i.e., when $p = 0.9$ the overestimation is as high as 144 percent of that for log-normal. Hence, the use of normal approximation in the case of log-normal distribution is not warranted for any value of p .

CHAPTER - VII

EFFECT OF NON - NORMALITY ON RESPONSE TO SELECTION IN SMALL POPULATIONS

7.1 Theoretical Procedure

Denoting the predicted response by ΔG , corresponding to the selection of r outstanding individuals from a sample of size n , when the basis of selection, I_r , is assumed to follow normal distribution and by ΔG^* when it follows the actual distribution, we have

$$\Delta G = i_{r/n} R_{GI} \sigma_G \quad (7.1.1)$$

$$\Delta G^* = i_{r/n}^* R_{GI} \sigma_G \quad (7.1.2)$$

when $i_{r/n}$ and $i_{r/n}^*$ denote respectively the intensity of selection corresponding to normal and actual distributions. These are to be evaluated as

$$i_{r/n} = (i_{1,n} + i_{2,n} + i_{3,n} + \dots + i_{r,n}) / r$$

$$i_{r/n}^* = (i_{1,n}^* + i_{2,n}^* + i_{3,n}^* + \dots + i_{r,n}^*) / r$$

where $i_{r,n}$ and $i_{r,n}^*$ denotes respectively the r -th largest standardized deviate from the normal and actual distributions.

The proportionate discrepancy, L , in the predicted response to selection on the assumption of normality relative to that based on actual distribution will therefore be given by

$$L = (\Delta G - \Delta G^*) / \Delta G^*$$

$$= \frac{\frac{r}{x/n}}{\frac{*}{x/n}} - 1 \\ = R_{x/n} (*) - 1, \text{ say} \quad (7.1.3)$$

Thus, when $R_{x/n} (*) > 1$, the progress is overestimated ; when it is less than 1, the progress is underestimated and when $R_{x/n} (*) = 1$, the progress remains unaffected.

The different distributions considered are beta, gamma and exponential. The lognormal distribution could not be included for this study for the reason mentioned in section 5.4.

7.2 Discrepancy in Response to Selection

Table 11 gives the percentage discrepancy in response to selection relative to two non-normal small populations of two sizes viz., $n = 5, 10$. In addition, the percentage discrepancy corresponding to large population size ($n = \infty$) is also in the table to see whether a sample of size $n = 10$ could be treated as large. The parametric values considered for beta distribution are those which correspond to the forms more likely to be encountered in practice. The comparison of gamma distribution with normal form has been made for only one value of value k (≈ 5) as the evaluation of $i_{x,n}^*$ for higher values of k involves very heavy and cumbersome calculations.

7.2.1 Normal vs. beta distribution. The behaviour of response to selection for beta distribution in the three cases i.e., when (i) $m_1 = m_2$, (ii) $m_1 > m_2$ and (iii) $m_1 < m_2$ is discussed below.

POST GRADUATE SCHOOL
INDIAN AGRICULTURAL RESEARCH INSTITUTE
NEW DELHI-110012.

APPLICATION FORM FOR THE GRANT OF RAILWAY CONCESSION TO
THE STUDENTS DURING NOTIFIED VACATIONS ONLY.

- NOTE: 1. To reach P.G.School Office four days in advance of the date on which the concession forms are required by the student.
2. To be collected within ten days from P.G.School Office from the day of submission.

1. Name of the student	Sex
2. Roll No.	3. Division
4. Whether belongs to Special Community or not	
5. Particulars of Fellowship held at present if any	
6. Date of Birth	Age
7. Home Town	as shown in the admission application form).
8. Railway Station nearest to home town (As shown in the admission application form).	
9. Period of vacation from	to
10. Purpose of journey	
11. Details of journey proposed	
12.a) Outward journey from Rly Stn.	to
b) Return journey from Rly Stn.	to
13. Date on which railway concession last availed of	
14. Number of railway concession availed of during the Academic year so far (given dates)	
15. Certified here I am not proceeding on study tour during the period of vacation/leave.	
16. Certified that I am not seeking relief from PGS during the period of vacation/leave.	
17. Leave Address:	
18. Certified that the particulars given above are correct.	

Signature of the student

Recommendations of the Chairman of Advisory Committee of the student

Contd...2/-..

Forwarded and recommended/not recommended. The information furnished by the student in his application have been verified and found correct.

Signature of the Chairman of the Advisory Committee of the student with name and Designation

Recommendation of the Professor

Certified that Sh./Smt./Kumari _____
Roll No. _____ M.Sc./Ph.D. student in the Division of _____
has been permitted to proceed to him/her home town during vacation as requested in his/her above application.
Also certified that he/she is not proceeding on study tour during the period of vacation and is also not being relieved from P.G. School during the vacation period.

Signature of the Professor
Division _____

FOR USE OF THE P.G. SCHOOL OFFICE

1. Date of receipt with Dy. No. _____ Date _____
2. Particulars checked and Railway concession form No. _____ prepared on _____ checked and found correct.

Signature
Dealing Asstt.

SUPERINTENDENT
POST GRADUATE SCHOOL, IARI

3. Concession for both/outward/inward journey approved and signed.

Signature of Asstt. Admn. Officer
Post Graduate School, IARI

4. Concession voucher handed over to the student on _____
5. Received the concession vouchers, No. _____

Signature of the student

Table 11

Percentage discrepancy in response to selection relative to different non-normal small populations of size 5 and 10.

Proportions selected p	Sample size, n = 5						Sample size, n = 10						Sample size, n = ∞						Beta								
	Beta			Gamma			Exponential			Beta			Gamma			Exponential			Beta			Gamma			Exponential		
	$m_1=2$	$m_1=3$	$m_1=2$	$m_1=2$	$m_1=3$	$m_1=2$	$m_1=2$	$m_1=3$	$m_1=3$	$m_1=2$	$m_1=3$	$m_1=2$	$m_1=3$	$m_1=2$	$m_1=3$	$m_1=2$	$m_1=3$	$m_1=2$	$m_1=3$	$m_1=2$	$m_1=3$	$m_1=2$	$m_1=3$	$m_1=2$	$m_1=3$		
0.1										5.89	11.27	-3.69	-12.56	-20.22	5.55	14.20	-3.76	-14.54	-23.8								
0.2	-0.26	4.68	-4.27	-7.26	-9.35	-0.30	5.25	-4.81	-8.11	-11.15	-0.37	5.93	-5.52	-9.23	-43.00												
0.3										-2.50	1.58	-4.79	-4.54	-2.80	-3.10	1.40	-5.63	-4.71	-5.74								
0.4	-2.97	-0.73	-3.49	-0.60	5.44	-3.62	-1.09	-4.14	-0.78	5.59	-4.43	-1.56	-4.92	-1.29	5.82												
0.5										-3.96	-2.90	-2.90	2.77	14.45	-4.85	-3.57	-3.57	2.88	15.13								
0.6	-2.97	-3.49	-0.73	6.14	22.89	-3.62	-4.14	-1.09	6.45	24.19	-4.43	-4.92	-1.56	6.96	26.01												
0.7										-2.50	-4.79	-1.38	10.78	35.89	-3.10	-5.63	1.40	11.63	39.13								
0.8	-0.26	-4.27	4.68	14.57	45.50	-0.50	-4.81	5.25	15.85	50.53	-0.37	-5.52	5.95	17.63	56.92												
0.9										5.89	-3.69	11.27	23.02	71.00	5.55	-3.76	14.20	25.93	85.71								

* r , the number of top-ranking individuals selected is obtained by the relation $r = p'n$

Case (I)

The progress is underestimated for all values of p and the discrepancy increases as the sample size increases. The percentage underestimation of progress for all values of p is less than 3 percent for population of size $n = 5$, 4 percent when $n = 10$ and less than 6 percent when the population is large. Thus, normal distribution provides a satisfactory approximation for symmetrical beta distribution.

Case (II)

The progress is overestimated for heavy culling i.e., when a few top-ranking individuals are selected, the percentage overestimation ranges from 10% to 20%. As n increases, the percentage overestimated and the maximum underestimation of $p = 0.6$ to $p = 0.8$ never exceed 10%.

Case (III)

Since this situation is the reverse from the earlier one only in that the long-tail is on the left instead of being on the right, the results with respect to discrepancy are exactly the reverse as in case (II).

7.2.2 Normal vs. gamma distribution. The discrepancy increases with the increase in the sample size and as also with the increase in the deviation of p from 0.5 on either side. The progress is underestimated for low values of p and overestimated for high values of p . Further a comparison of two sets of results corresponding to $n = 10$ and $n = \infty$ shows that a

sample of size 10 can be treated as having come from a large population without any serious error.

7.2.3 Normal vs. exponential distribution. Except for all values of p , in the neighbourhood of 0.3 to 0.4, the percentage underestimation for low values of p and overestimation for high values of p , are too serious to ignore. Further, the discrepancy increases as the sample size increases. However, a sample of size $n = 10$ can be treated as large for the purposes of computing the intensity of selection.

CHAPTER - VIII

SUMMARY

Very rarely, realised selection responses agree with those expected. One of the causes could be the assumption of normality of criterion of selection when in effect it follows a non-normal distribution. That quite a few quantitative traits may depart from the normal form has been borne out by empirical studies made by various investigators. It is the purpose of this thesis to investigate the effect of non-normality on response to selection when the criterion of selection is assumed to be normally distributed. For fixed values of accuracy of selection and genetic standard deviation this problem, however, amounts to studying the relative magnitude of intensity of selection, corresponding to a given proportion of individuals saved, on the assumption of normality vis-a-vis actual distribution.

From the empirical studies, it is seen that Pearsonian Types I and III provided adequate representation to many types of data in the field of animal husbandry and poultry. In a few cases exponential and lognormal distributions also fitted well. Accordingly, the expressions of selection intensity have been derived for Pearson's Type I distribution including its derivative namely beta distribution, Type III along with its derived distributions namely gamma, chi-square and exponential and for lognormal distribution as applicable to both large and small populations. Using these expressions the values of intensities of selection for different proportions of population retained were tabulated for different sample sizes.

It has been observed that beta and gamma distributions for parametric values characterizing common situations can be approximated to normal distribution for moderately heavy and low cullings without any serious error. However, the use of normal approximation for exponential and lognormal distributions is not warranted as the discrepancy in response to selection for almost all values of p , the proportion of individuals saved, is too serious to ignore.

REF E R E N C E S

- Burrows, P.M. (1972). Expected selection differentials for directional selection. *Biometrics* 28, 1091-1100.
- Clayton, G.A., Knight, G.R., Morris, J.A. and Robertson, A. (1957) An experimental check on quantitative genetical theory III Correlated responses. *J.Genet.*, 55 , 171-180.
- Crow, J.F. and Kimura, M. (1970) An Introduction to Population Genetics Theory. 591.pp. Harper and Row, New York.
- David, H.A. (1970) Order Statistics. Wiley, New York.
- Dickerson, G.E. (1951) Effectiveness of selection for economic characters in swine. *J.Anim.Sci.*,10, 12.
- Dickerson, G.E. (1955) Genetic slippage in response to selection for multiple objectives, *Cold.Spr.Harb.Symp.Quant.Biol.*, 20, 213-224.
- Elderton, W.P. and Johnson, N.L (1969) System of Frequency Curves . Cambridge University Press.
- Falconer,D.S. (1953) Selection for large and small size in mice. *J.Genet.*, 51, 470-501.
- Falconer,D.S. (1965) Introduction to Quantitative Genetics. 365, Oliver and Boyd, Edinburgh.
- Federer,W.T.(1951) Evaluation of variance components from a group experiments with multiple classifications. Iowa Agricultural Experiment Station Research Bulletin No. 380.
- Fisher, R.A. and Yates, F.(1938) Statistical Tables for Biological Agricultural and Medical Research. Oliver and Boyd., London.
- Godwin, H.J. (1949) Some low moments of order Statistics. *Ann. Math.Stat.*, 20, 279-285.
- Gowen, J.W.(1949) Study of variation of lactation yield of Jersey cattle. *Genetics*, 34, Jour. of Agric. Res., 16.
- Gupta, S.S. (1960) Order Statistics from gamma distribution, *Technometrics* 2, 243-262.
- Harter, L.H. (1961) Expected values of normal order statistics. *Biometrika* 48, 151-165.

- Hastings, Q.Jr., Mosteller, F., Tukey, J. W. and Winsor, C. P. (1947) Low moments for small samples: a comparative study of order statistics, Ann. Math. Statist., Vol 18, 413-426.
- Henderson, C. R. (1963) Selection index and expected genetic advance, 141-163 in Statistical Genetics and Plant Breeding, NAS-NRC Publication 982.
- Hill, W.G. (1969) On the theory of artificial selection in finite population, Genet. Res., Camb. 13, 143-163.
- Jain, J.P. and Copalan, R. (1976) Statistical Research in genetic engineering of livestock improvement. Second Conference of Agricultural Research Statisticians held at I.A.R.S., New Delhi.
- Jain, J.P. and Narain, P. (1974) The use of population generation matrix in dairy herds. Jr. Ind. Soc. Agric. Stat., 26, 71-92.
- Jardine, R. (1958) Animal Breeding and the estimation of genetic value, Heredity, 12, 499-511.
- Kapteyn, J. C. (1903) Skew Frequency Curves in Biology and Statistics (Groningen Noordhoof).
- Kempthorne, O. (1957) An Introduction to Genetic Statistics. John Wiley and Sons, Inc., New York.
- Lerner, I.M. (1950) Population Genetics and Animal Improvement. Cambridge University Press, New York.
- Lerner, I.M. (1958) The Genetic Basis of Selection. John Wiley and Sons, Inc., New York.
- Le Roy, H.L. (1958) Die Abstammungsbewertung, Z. Tiers. ZuchtBiol., 71, 328-378.
- Lush, J.L. (1945) Animal Breeding Plans. Ames: Iowa State College Press. 3rd edn.
- Lush, J.L. (1947) Family merit and individual merit as bases for selection, Pt. I, Pt. II. Amer. Nat., 81, 241-261; 362-379.
- Lush, J.L. (1948) The Genetics of Population, (Mimeo graphed notes, Dept. of Animal Husbandry, Iowa State College).
- Mahajan, Y.P. and Parkash, O. (1959) Study of frequency distributions of the characters relating to milk yield and the consequences of non-normality on standard tests of significance. Jr. Ind. Soc. Agric. Stat., 21, 163-179.
- Mather, K. (1974) ?
- Mood, M. A., Grabill, A.F. and Boes, C.D. (1974) Introduction to the Theory of Statistics. 3rd edn.

- Osborne, R. (1957a) The use of sire and dam family averages in increasing the efficiency of selective breeding under a hierarchical mating system. *Heredity* 11, 93-116.
- Osborne, R. (1957b) Correction for regression on a secondary trait as a method of increasing the efficiency of selective breeding. *Aust. J. Biol. Sci.*, 10, 865-866.
- Pearl and Miner (1919) Study of variation of lactation yield and fatcontents Ayrshire cows. *Jour. of Agric. Res.*, 16.
- Pearson, K. (1896) Contribution to mathematical theory of evolution. *Phil. Trans. Roy. Soc.*, 186, Pt. I.
- Pearson, K. (1931) Tables for Statisticians and Biometrists, Pt. II, Cambridge University Press, Cambridge, England.
- Pearson, K. (1934) Tables of the Incomplete Beta Function. 1st edn, London: Cambridge University Press.
- Pearson, K. (1948) Tables of the Incomplete Gamma Function, London Cambridge University.
- Pearson, E.S. and Hartley, H.O. (1954) Biometrika Tables for Statisticians.
- Quesenberry, C.P., Mitaker, T.B. and Dickong, J.W. (1978) On testing normality of using several samples: An analysis of peanut aflatoxin data. *Biometrics* 32, 733-739.
- Sheppard, W.E. (1902) New Tables of the Probability Integral. *Biometrika* 2, 174 - 190.
- Skjervold, H. and Odegard, A.K. (1959) Estimation of breeding value on the basis of the individual's own phenotype and ancestor's merit. *Acta Agric. Scandinavica* 9, 341-354.
- Tippett, L.H.C. (1925) On the extreme individuals and the range of samples from a normal population. *Biometrika* 17, 364-387.
- Tocher, J.F. (1928) Investigation of milk yield of dairy cows, *Biometrika* 20 B.
- Young, S.S.L. (1961) The use of sire's and dam's records in animal selection. *Heredity* 16, 91-102.

