APPROXIMATE METHODS FOR DETERMINATION OF OPTIMUM STRATIFICATION BOUNDARY POINTS FOR BETA, GAMMA, AND PARETO DISTRIBUTIONS

BY

RAZAQ BELLO

A thesis submitted to the Postgraduate School, Ahmadu Bello University, Zaria, in partial fulfilment of the requirements for the depree of Doctor of Philosophy in Statistics.

Department of Mathematics,
Faculty of Science,
Ahmadu Bello University,
Zaria, Nigeria.

August, 19 86.

CERTIFICATION

This Thesis entitled

"APPROXIMATE METHODS OF OPTIMUM STRATIFICATION BOUNDARY POINTS FOR BETA, GAMMA, AND PARETO DISTRIBUTIONS"

by RAZAQ BELLO meets the regulations governing the award of the degree of DOCTOR OF PHILOSOPHY of Ahmadu Bello University, and is approved for its contribution to knowledge and literary presentation.

(Name and Signature) Date 21.11.1986
Chairman, Supervisory Committee

(Name and Signature) Date 21.11.1986
Member, Supervisory Committee

(Name and Signature) Date 20.12.1986
Member, Supervisory Committee

(Name and Signature) Date 20.12.1986
Head of Department

(Name and Signature) Date
Dean, Postgraduate School.

D E C L A R A T I O N

I, Razaq Bello, hereby declare that this thesis is not a published or an unpublished work of another author. No part of this thesis has been submitted for a degree in any other University. Where the work of others has been used the source is acknowledged in the text .

ACKNOWLEDGEMENT

I wish to express my sincere thanks to Dr. A.R. Kokan, Reader and Head of Statistics Section, Department of Mathematics Ahmadu Bello University, Zaria for suggesting the problem and for his active guidance, constant encouragement, helpful and valuable remarks at all stages of this work.

I also wish to record my sincere thanks to Professor I.B. Mohammed, Head of the Department of Mathematics, Ahmadu Bello University, Zaria for the opportunity he gave me to complete this work, and for his other supports.

My profound gratitude goes to Dr. E.T. Jolayemi for very useful suggestions, helpful comments and stimulating discussions.

The moral support of the following persons is very well appreciated: Mrs. S.A. Bello, Dr. H.B. Lawal, Dr. M.O. Adewale, Dr. V.V.G. Rao, Mr. J.O. Abolarinwa and Dr. J.O. Oladipo and also S.P. Olaleye.

Finally, I must also record my thanks to Mr. S.I. Babafemi for the pains he took to type this thesis so neatly.

\ B S T R A C T

Stratification is the division of a heterogeneous population into relatively homogeneous sub-populations. Using Stratification we can usually improve the efficiency of the estimate of mean of the population. Stratification may be inevitable in terms of administrative convenience.

Stratification of Hospitals may be by bed capacity, of schools by enrolment, of business concern by volume of sales, of industrial plants by employment and of voluntary organisation by membership. te finite nopulations encounte in practice are all discontinuous. If the populations are sufficiently large, they can very often be approximated by some standard continuous distributions. Thus tables of strata boundaries for these standard distributions can profitably be utilized for the stratification of an. actual population. Under Tchunrow-Neyman allocation, Dalenius (1950) derived equations for the strata boundaries $y_h$which minimize the variance of th estimated population mean. But these equations are troublesome to solve.

A number of approximate methods that could quickly solve the Dalenius equations have been proposed by several authors.

These approximate methods are:

(i) Cum /f Method

(ii) Durbin's Method

(i i i) Ekman's Method

(iv) Equal Range Method

(v) Equal Sized Strata Method

(vi) Equalization of Strata Total Method.

The author has developed computational procedures for the determination of strata boundaries for Beta and Gamma Distributions with various values of the parameters using the above methods except (iii).

The methods are compared using the variance of the stratified sample mean. The best boundaries are those that give the least variance of the stratified sample mean. All the methods construct strata boundaries quickly fox-

Beta Distribution. For Gamma Distribution, only the cum gam method constructs strata boundaries quickly. For ParetO distribution with parameter 0 all the methods consider lead to a very substantial reduction in variance of the estimate of the mean of population except the equal ran method, and all the method construct the strata boundaries quickly.

We conclude that the cum gamma of f method is the best method, of constructing strata boundaries among those considered in this study.

A critical examination of the behavior of the variance as a function of the parameters p and a of the Beta distribution and the number of strata k has shown that for Beta distribution the variance decreases by about 10%, for a unit increase in a when 1 _< p .< 4 and k are fixed; while the variance increases and then decreases when p _> 5. We have the sane conclusion when p and q are interchanged.

For the Gamma distribution with parameters beta and p it is observed that for fixed k the variances increase as one of the numbers v or 8 is increased while the other remains constant.

TABLE OF CONTENT

	PAGE
CHAPTER ONE : INTRODUCTION	1
1.1 FACTORS AFFECTING REDUCTION IN VARIANCE OF THE ESTIMATOR OF THE POPULATION MEAN IN STRATIFIED RANDOM SAMPLING	1
1.1.1 CHOICE OF STRATIFICATION VARIABLE	1
1.1.2 CHOICE OF NUMBER OF STRATA . ..	2
1.1.3 DETERMINATION OF STRATA BOUNDARIES	2
1.1.4 ALLOCATION OF THE SAMPLE	3
1.2 SOME USES OF BETA AND GAMMA DISTRIBUTIONS	4
CHAPTER TWO: REVIEW OF THE LITERATURE ON THE DETERMINATION OF OPTIMUM POINTS OF STRATIFICATION	6
2.1 PROPORTIONAL ALLOCATION	8
2.2 EQUAL ALLOCATION	9
2.3 OPTIMUM ALLOCATION	9
2.4 EQUALIZATION OF STRATUM TOTAL METHOD	10
2.5 EQUALIZATION OF CUMMULATIVE OF SQUARE ROOT OF THE FREQUENCIES METHOD (CUM $\sqrt{f}$ METHOD)	11
2.6 DURBIN'S METHOD	12
2.7 EKMAN'S METHOD	12
2.8 EQUAL RANGES METHOD	13
2.9 EQUAL SIZED STRATA METHOD	13
CHAPTER THREE: EQUATIONS FOR THE STRATIFICATION POINTS FOR THE BETA DISTRIBUTIONS	17
3.1 EQUALIZATION OF STRATUM TOTAL METHOD ..	20
3.1.1 COMPUTATIONAL PROCEDURE	21

TABLE OF CONTENT PAGE

3.2 CUM√f METHOD 21

3.2.1 COMPUTATIONAL PROCEDURE 22

3.3 DURBIN'S METHOD 23

3.3.1 COMPUTATIONAL PROCEDURE 24

3.4. EQUAL RANGES METHOD 24

3.4.1 COMPUTATIONAL PROCEDURE 24

3.5. EQUAL SIZED STRATA METHOD 25

3.5.1 COMPUTATIONAL PROCEDURE 25

3.6 BETA DISTRIBUTION OF THE SECOND KIND 26

3.6.1 CUM √f METHOD 28

3.6.2 EQUALIZATION OF STRATUM TOTAL METHOD 29

3.6.3 EQUAL SIZED STRATA METHOD . .. 30

3.6.4 DURBIN'S METHOD 30

CHAPTER FOUR: EQUATIONS FOR THE STRATIFICATION POINTS OF GAMMA DISTRIBUTIONS 32

4.1 CUM √f METHOD 35

4.1.1 COMPUTATIONAL PROCEDURE 36

4.2 EQUALIZATION OF STRATUM TOTAL METHOD 36

4.2.1 COMPUTATIONAL PROCEDURE 37

4.3 EQUAL RANGES METHOD 38

4.3.1 COMPUTATIONAL PROCEDURE 38

4.4. EQUAL SIZED STRATA METHOD 39

4.4.1 COMPUTATIONAL PROCEDURE 39

4.5 DURBIN'S METHOD 40

4.5.1 COMPUTATIONAL PROCEDURE 41

TABLE OF CONTENTS PAGE

CHAPTER FIVE: EQUATIONS FOR THE STRATIFICATION POINTS OF PARETO DISTRIBUTIONS 42

5.1. CUM $\sqrt{f}$ METHOD 45

5.2. EQUAL SIZED STRATA METHOD 47

5.3. EQUALIZATION OF STRATUM TOTAL METHOD 48

5.4. DURBIN'S METHOD 49

5.5. EQUAL RANGE METHOD 49

CHAPTER SIX: EFFICIENCIES OF THE METHODS 50

6.1. TABLES SHOWING nx (VARIANCE OF ESTIMATES OF POPULATION MEAN) FOR BETA DISTRIBUTION WITH PARAMETERS p AND q UNDER VARIOUS METHODS. ..

6.2. TABLES SHOWING nx (VARIANCE OF ESTIMATES OF POPULATION MEAN) FOR GAMMA DISTRIBUTIONS WITH PARAMETERS p AND β UNDER VARIOUS METHODS 73

6.3. TABLES SHOWING STRATIFICATION BOUNDARY POINTS AND nx (VARIANCE OF ESTIMATE OF POPULATION MEAN FOR PARETO DISTRIBUTION WITH PARAMETER $\theta = 3$ 85

CHAPTER SEVEN: SUMMARY AND CONCLUSIONS 88

REFERENCES 94

CHAPTER ONE

INTRODUCTION

One principal reason for stratification in the design of sample surveys is to reduce the variance of the sample estimates. Another reason is the administrative convinience. Strata may be formed so that different sampling methods and procedures may be employed within each stratum especially when the sampling problems differ markedly in different parts of the population. A heterogeneous population may be divided into sub-populations each of which is internally homogeneous. The internal homogeneity of the groups might lead to some improvement in the efficiency of the estimate compared with a simple random sample drawn from the whole population.

1.1 FACTORS AFFECTING REDUCTION IN VARIANCE OF THE ESTIMATE OF POPULATION MEAN IN STRATIFIED RANDOM SAMPLING

The factors that influence the reduction of variance include:

(i) choice of the stratification variable

(ii) choice of number of strata

(iii) determination of strata boundaries

(iv) allocation of the sample.

1.1.1 CHOICE OF THE STRATIFICATION VARIABLE

Concerning the choice of stratification variable, we may observe that generally more gains acrue from the use of coarser division of several variables than from the

finer division of one variable. Stratifying variables unrelated to each other (but related to the survey variables) should be preferred. But sometimes the stratification variable is the estimation variable itself at a previous survey or pilot survey.

1.1.2 CHOICE OF NUMBER OF STRATA

Regarding the choice of the number K of Strata, Kish (1965) suggested that between three and ten Strata will suffice for any single variable. Cochran (1977) indicated that little reduction in variance is to be expected beyond k = 6. Hess et al (1966) reported that little purpose is served by taking more than six Strata.

1.1.3 DETERMINATION OF STRATA BOUNDARIES

We next consider the determination of stratum boundaries. When the stratification variable and the estimation variable are the same and its probability density function is known (from previous surveys) Dalenius(1957) has given a rule for the determination of strata boundaries for optimum allocation. Dalenius and Gurney (1958) gave the rule for proprotional allocation. Cochran (1963) and Sethi (1963) gave the general equation to be satisfied by stratification boundary points for equal allocation.

Unfortunately the strata optimum boundaries satisfying the equations are ill-adapted to practical computations. Another disadvantage is that these systems of equations are not easily solved. The thrust of much of the work in the

area of optimum stratification has been towards the development of relatively simple rules for the determination of approximately optimum stratum boundaries.

1.1.4. ALLOCATION OF THE SAMPLE

In stratified sampling the values of the sample sizes n_h (sample size in h^{th} stratum) in the respective strata are chosen by the sampler. The sample sizes may be selected to minimize the variance of estimate of the population $V(\bar{y}_{st})$ for specified cost of taking the sample or to minimize the cost for specified $V(\bar{y}_{st})$. When the cost function is in particular linear the two rules lead to the same solution. The sample size in h^{th} stratum should be proportional to $p_h \sigma_h / \sqrt{C_h}$ (p_h is proportion of the units in the h^{th} stratum and σ_h in the standard deviation of the h^{th} stratum and C_h is the cost of per unit in the h^{th} stratum).

If the cost per unit is the same in all strata, this allocation is sometimes called Neyman allocation or optimum allocation. But if the sample is allocated such that n_h is proportional to size of the stratum, this is called proportional allocation. If equal sample sizes are taken in each stratum, this is referred to as equal allocation.- Cochran (1977) has suggested the following rules of conduct.:

(i) One should not resort to optimum allocation unless there are substantial differences among the factors $\sigma_h / \sqrt{C_h}$

(ii) If several values of $\sigma_h / \sqrt{C_h}$ are roughly equal one can use proportional allocation.

1.2. SOME USES OF BETA AND GAMMA DISTRIBUTIONS

In social sciences in which we have proportions, the proportion may be regarded as a variable and this follows Beta Distribution, - (Jolayemi, 1982). Most industrial activities feature practical constraints, thus a given activity cannot be completed in less than a minimum time or longer than a maximum time. The completion time is commonly assumed to be a Beta variable (Bury, 1975).

Gamma models feature a wide variety of shpaes. Therefore the model is capable of representing many diverse phenomena. For example, the life span of components follows the Gamma model. Similarly in Medicine the life expectancy in respect of ailment follows the Gamma Model. The positive nature of Gamma variable makes the Gamma Distribution an attractive model in applied Scientific work. In Inventory and Queeing analysis Gamma model is widely used. (Johnson and Kotz, 1970), (Johnson and Elnadt-Johnson ,1980). Pareto Model is widely used as models for income distribution.

The main purpose of this study is to compare precisions arising out of the application of various rules as approximations to the solutions of sets of Dalenius equations. Strata are constructed for beta and gamma distributions using these approximate rules. The rules are recast in forms suitable for computation.

Chapter two reviews the methods that are used for the determination of stratification boundary points. Chapter three deals with the actual determination of boundary points and evaluation of relative efficiencies taking Beta distribution as the target population, while Chapter four deals with the Gamma distribution, and chapter five deals with Pareto distribution.

In Chapter six we present tables showing efficiencies of the methods. Summary and conclusions are presented in Chapter seven.

CHAPTER TWO

REVIEW OF THE LITERATURE ON THE DETERMINATION OF OPTIMUM POINTS OF STRATIFICATION

One of the major factors that affect reduction of the variance is the choice of the boundary points of the strata. This is why the problem of the determination of optimum stratification points is very important.

The problem has been studied theoretically by Dalenius (1950, 1953, 1957). We give a brief account of the development of the theory. Though the actual distributions encountered in practice are discontinuous, if the population sizes are sufficiently large, they can be approximated by some standard distributions.

Let us suppose that the population under study can be represented by the probability density function

$f(y)$, $a \leq y \leq b$

and it is divided into K strata by points

y_h, $h = 1,2,3,\ldots\ k-1$, where

$a = y_o < y_1 < y_2 < y_3 \ldots\ < y_{k-1} < y_k$.

y_o and y_k being the natural limits of the variable Y.

Let P_h, μ_h and σ_h^2 stand for the proportion of units the mean value of the units and the variance of the units in the h^{th} stratum.

Then

$$p_h = \int_{y_{h-1}}^{y_h} f(y)dy \qquad 2.0.1$$

$$\mu_h = \frac{1}{p_h} \int_{y_{h-1}}^{y_h} yf(y)dy \qquad 2.0.2$$

$$\sigma_h^2 = \frac{1}{p_h} \int_{y_{h-1}}^{y_h} y^2 f(y)dy - \mu_h^2 \qquad 2.0.3$$

Further,

$$\mu = \int_a^b yf(y)dy, \quad \sigma^2 = \int_a^b y^2 f(y)dy - \mu^2 \qquad 2.0.4$$

are the mean and the variance respectively of the population under study. Suppose n_h units are selected at random from the h^{th} stratum. Then an unbiased estimator of the mean μ of the population is

$$\bar{y}_{st} = \sum_{h=1}^{k} p_h \bar{y}_h \qquad 2.0.5$$

where $\bar{y}_h$ is the sample mean of n_h units in the h^{th} stratum. The variance of the estimator $\bar{y}_{st}$ is

$$V(\bar{y}_{st}) = \sum_{h=1}^{k} \frac{p_h^2 \sigma_h^2}{n_h} \qquad 2.0.6$$

When strata sizes are large compared to sample size.

Clearly the variance $V(\bar{y}_{st})$ is a function of the stratum boundaries $(y_1, y_2, y_3, \ldots, y_{k-1})$

Next we consider different types of allocations. Three cases can arise.

CASE 1

PROPORTIONAL ALLOCATION

If the units are allocated to the strata in proportion to p_h's i.e. $n_h = np_h$ where $n = \sum_{h=1}^{k} n_h$ is the total sample size, the variance (2.0.6) becomes

$$V_p(\bar{y}_{st}) = \frac{1}{n}\sum_{h=1}^{k} p_h \sigma_h^2 \qquad 2.1.1$$

The optimum points of stratification of strata in the case of proportional allocation can be obtained by minimizing (2.1.1) with respect to y_h $(h = 1,2,3,\ldots,k-1)$ and the stratification points are given by Dalenius and Gurney (1958) by the set of equations.

$$y_h = \frac{\mu_h + \mu_{h+1}}{2}, \quad h = 1,2,3,\ldots,k-1. \qquad 2.1.2$$

Dalenius and Hodges (1958) applied the above rule to the following distributions.

$$\text{(i)} \quad f(y) = e^{-y}, \quad 0 < y < \infty \qquad 2.1.3$$

$$\text{(ii)} \quad f(y) = ye^{-y}, \quad 0 < y < \infty \qquad 2.1.4$$

$$\text{(iii)} \quad f(y) = 2(1-y), \quad 0 < y < 1 \qquad 2.1.5$$

$$\text{(iv)} \quad f(y) = \sqrt{\frac{2}{\pi}}\, e^{-y^2/2} \quad y \geq 0 \qquad 2.1.6$$

It may be noted that distribution (2.1.3) and (2.1.4) are particular cases of Gamma distribution with parameters $\beta = 1$, $P = 1$ and $\beta = 1$, $P = 2$ respectively, while (2.1.5) is a particular case of Beta distribution of the first kind with parameters $p = 1$, $q = 2$. Moreover (2.1.6) is the right half of the standard normal distribution. The number of strata considered by them was between 2 and 5. The possible reason for this limitation is the computational difficulty.

CASE 2

EQUAL ALLOCATION

For equal allocation i.e. when $n_h = \frac{n}{k}$ the variance of the estimator $\bar{y}_{st}$ is

$$V_{eq}(\bar{y}_{st}) = \frac{k}{n} \sum_{h=1}^{k} p_h^2 \sigma_h^2 \qquad 2.2.1$$

where $n = \sum_{h=1}^{k} n_h$

The optimum points of stratification may be obtained by minimizing (2.2.1) with respect to $y_h (h = 1,2,3,\ldots,k-1)$. The points are given by Cochran (1963) and also by Sethi (1963) by the set of equations.

$$P_h[\sigma_h^2 + (y_h - \mu_h)^2] = P_{h+1}[\sigma_{h+1}^2 + (y_h - \mu_{h+1})^2]$$

$$h = 1,2,3,\ldots,k-1 \qquad 2.2.2$$

We note that when $p_h = 1/k$ i.e when we have constructed equal sized strata (2.1.1) and (2.2.1) are equal and each is equal to

$$V(\bar{y}_{st}) = \frac{1}{nk} \sum_{h=1}^{k} \sigma_h^2 \qquad 2.2.3$$

CASE 3

OPTIMUM ALLOCATION

In the case of optimum allocation

$$n_h = \frac{n\, p_h \sigma_h}{\sum_{h=1}^{k} p_h \sigma_h} \qquad 2.3.1$$

The variance of $\bar{y}_{st}$ is given by

$$V_{opt}(\bar{y}_{st}) = \frac{1}{n} \left(\sum_{h=1}^{k} p_h \sigma_h \right)^2 \qquad 2.3.2$$

This allocation namely 2.3.1 is called Tchuprow - Neyman allocation after Tchuprow (1923) and Neyman (1934) who discovered the rule independently. The optimum points of stratification can as in the previous cases, be obtained by minimizing 2.3.2 with respect to $y_h (h = 1,2,3,\ldots,k-1)$. These stratification points are given by Dalenius (1957) by the set of equations.

$$\frac{\sigma_h^2 + (y_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^2 + (y_h - \mu_{h+1})^2}{\sigma_{h+1}}$$

$h = 1,2,3,\ldots,k-1.$

We remarked that the variance of the estimator depends on the method of allocation and the choice of boundary points of stratification. A theoretical comparison between simple random and stratified random sampling with proportional allocation and optimum allocation shows (Cochran (1977)) that

$$V_{opt}(\bar{y}_{st}) \leq V_p(\bar{y}_{st}) \leq V_{ran}(\bar{y}) \qquad 2.3.4$$

when finite population correction is negligible.
We therefore restrict our consideration to optimum allocation only.

We will now review methods developed for solving 2.3.3

2.4. EQUALIZATION OF STRATUM TOTAL METHOD

Mahalanobis (1952) proposed the equalization of stratum aggregate of the stratification variable i.e. $p_h\mu_h$ = constant, in each stratum. The set of equations

$$\int_{-\infty}^{Y_h} yf(y)dy = \frac{h}{k} \int_{-\infty}^{\infty} yf(y)dy \qquad 2.4.1$$

$$h = 1,2,3,\ldots,k$$

where f is the probability density function of the stratification variable and k is the number of strata are based on the method proposed by Mahalanobis. The roots $Y_1, Y_2, Y_3, \ldots, Y_{k-1}$ of equations 2.4.1 are approximately the optimum stratification points.

2.5. EQUALIZATION OF CUMMULATIVE OF THE SQUARE ROOT OF THE FREQUENCIES METHOD. (CUM $\sqrt{f}$ METHOD)

Dalenius and Hodges (1958, 1959) proposed constructing equal intervals on the cummulative of the square root of the frequencies of the stratification variable. The set of equations

$$\int_{-\infty}^{Y_h} \sqrt{f(y)}dy = \frac{h}{k} \int_{-\infty}^{\infty} \sqrt{f(y)}dy \qquad 2.5.1$$

$$h = 1,2,3,\ldots,k$$

where f(y) as before is the probability density function of the stratification variable and k the number of strata are based on Dalenius rule 2.3.3. The roots $Y_1, Y_2, Y_3, \ldots, Y_{k-1}$ of equations 2.5.1 are approximately the optimum stratification points.

2.6. EQUALIZATION OF CUMMULATIVE OF [f(y) + r(y)] METHOD (DURBIN'S METHOD)

Durbin (1959) proposed that the optimum stratum boundaries may be obtained by taking equal intervals on the cummulative of [f(y) + r(y)] where

$$r(y) = \frac{F(y_k) - F(y_o)}{y_k - y_o} \qquad 2.6.1$$

$$F(y) = \int_{-\infty}^{y} f(t)dt \qquad 2.6.2$$

The set of equations

$$\int_{-\infty}^{y_h} [\, r(y) + f(y)\,]dy = \frac{h}{k}\int_{-\infty}^{\infty} [\, r(y) + f(y)\,]dy \qquad 2.6.3$$

$h = 1,2,3,\ldots,k$ are based on Durbin's Method.

The roots $y_1, y_2, y_3, \ldots, y_{k-1}$ of the equations 2.6.3 are approximately optimum stratification points.

2.7. EKMAN'S METHOD

Ekman (1957) proposed that

$$P_h(y_h - y_{h-1}) = \text{constant}$$

$$h = 1,2,3,\ldots,k \qquad 2.7.1$$

This rule gives approximately the optimum stratification points. Dalenius and Gurney (1957) suggested making $P_h\sigma_h$ = constant in each stratum. But this is exactly the same as the Ekman's Method in the case of a uniform distribution.

2.8. EQUAL RANGES METHOD

If the probability density function has a finite range say $a \leq y \leq b$ where $a<b$ are finite the range $[b-a]$ is divided into equal intervals. Thus to take equal intervals on the range amounts to determining $y_1, y_2, \ldots, y_{k-1}$ such that

$$y_h = a + \frac{h}{k}(b-a) \qquad 2.8.1$$

$$h = 1,2,3,\ldots,k.$$

But when the range of $f(y)$ is infinite the density function can be truncated so that a finite range which practically includes the whole of the population is obtained. Thus we may exclude at the extremes a small proportion of the frequency.

Let $[a,b]$ be the finite interval which includes almost the whole of the population. The boundary points can be obtained by using 2.8.1.

2.9. EQUAL SIZED STRATA METHOD

In this method the stratum weights are made equal in all the strata i.e. P_h = constant for $h = 1,2,3,\ldots,k$. The set of equations

$$\int_{-\infty}^{y_h} f(y)\,dy = \frac{h}{k} \qquad 2.9.1$$

$h = 1,2,3,\ldots,k$ are based on the method of equal sized strata. The roots $y_1, y_2, \ldots, y_{k-1}$ of the equations 2.9.1 are the boundary points.

Kish (1965) argued that the equal ranges method and equal sized strata method can be dismissed on intuitive grounds. He argued that if the range of the distribution is divided into equal intervals so that the ranges $(y_h - y_{h-1})$ of the strata are equal for all h, the relative sizes of p_h of strata becomes too unequal, those on the extreme tail becoming too small, on the other hand, if the area of the population is divided into equal relative sizes i.e. p_h = constant in all strata, then the ranges on the extremes become too great.

Despite these assertions it is necessary to examine their performances relative to other methods because of their grreat convenience.

Kokan (1963) applied the equal Ranges Method and equal sized strata method to the following distributions

(i) $f(y) = e^{-y}$, $y > 0$

(ii) $f(y) = \sqrt{\frac{2}{\pi}}\, e^{-y^2/2}$, $y \leq 0$

(iii) $f(y) = 2(1-y)$, $0 < y < 1$

(iv) $f(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$, $-\infty < y < \infty$

The number of strata considered was between 2 and 5. Kokan reported that Equal Ranges Method is inferior to Equal sized strata for optimum allocation for all the distributions.

Raj (1964) applied the equalization of stratum total method to the following distributions

(i) $f(y) = \sqrt{\frac{2}{\pi}}\, e^{-y^2/2}$, $y \geq 0$

(ii) $f(y) = e^{-y}$, $y \geq 0$

(iii) $f(y) = ye^{-y}$, $y \geq 0$

(iv) $f(y) = 2(1-y)$, $0 < y < 1$.

Also Dalenius and Hodges (1959) applied the Cum $\sqrt{f}$ method to the same set of above distributions (i) - (iv).

Dallas, et al (1975) applied the Cum $\sqrt{f}$ method and cum $\sqrt[3]{f}$ method to the standard normal distribution. They reported that cum $\sqrt{f}$ method performs better than cum $\sqrt[3]{f}$ method. The number of strata considered was between 2 and 10.

Sethi (1963) developed an iterative method for constructing stratum boundaries. The procedure produces a near optimum points of stratification when compared with the cum $\sqrt{f}$ method. It was pointed out that the procedure is considerably more difficult to use.

Cochran (1961), Kokan (1965) and Hess, et al (1966) have examined some of the approximate methods on obtaining optimum stratification through fairly extensive empirical studies. Hess, et al (1966) observed in their study of the empirical population that beyond two strata, Ekman's method excelled as compared with other methods using proportional and optimum allocations of allocation, and Sethi's method, the cum $\sqrt{f}$ method, and equalization of stratum totals method followed in that order of performance.

Kokan (1965) applied four approximate methods to distribution of the number of animals in 214 villages of Uttar Pradash and 183 villages of Bombay, India. He reported that the "Ekman's method"is the least efficient and "Cum $\sqrt{f}$ method" is most efficient in most of the cases. The number of strata he considered was between 2 and 4.

CHAPTER THREE

EQUATIONS FOR THE STRATIFICATION POINTS FOR BETA DISTRIBUTIONS

In this Chapter we consider the determination of the stratification points and the variance of the estimator for optimum allocation for Beta distribution with parameters p and q and we obtain:

(i) equations for the determination of approximately optimum stratification points using various methods reviewed in Chapter Two.

(ii) equations for proportion p_h, mean μ_h and variance σ_h^2 of the h^{th} stratum

$h = 1,2,3,\ldots,k.$

(iii) computational procedure for (i) and (ii)

(iv) $n \times V_{opt}(\bar{y}_{st}) = (\sum p_h \sigma_h)^2$.

Suppose f(y) is the probability density function of a random variable y whose mean is under study. The mean μ and the variance σ^2 are respectively

$$\mu = \int_{-\infty}^{\infty} y f(y)\,dy \qquad 3.0.1$$

$$\sigma^2 = \int_{-\infty}^{\infty} (y-\mu)^2 f(y)\,dy \qquad 3.0.2$$

Now consider the probability density function f(y) of Y of Beta distribution of the first kind

$$f(y) = \frac{y^{p-1}(1-y)^{q-1}}{B(p,q)}, \quad 0 \leq y \leq 1, \quad p > 0,\ q > 0 \qquad 3.0.3$$

where $B(p,q) = \int_0^1 y^{p-1}(1-y)^{q-1}dy$

From 3.0.1, 3.0 3.0.3

$$\mu = \int_0^1 yf(y)dy = \frac{p}{p+q} \qquad 3.0.4$$

$$\sigma^2 = \int_0^1 (y-\mu)^2 f(y)dy = \frac{pq}{(p+q)^2(p+q+1)} \qquad 3.0.5$$

Let $B_y(p,q) = \int_0^y t^{p-1}(1-t)^{q-1}dt \qquad 3.0.6$

$0<y<1,$

$$I_y(p,q) = \frac{B_y(p,q)}{B(p,q)} \qquad 3.0.7$$

$$= \frac{1}{B(p,q)} \int_0^y t^{p-1}(1-t)^{q-1}dt \qquad 3.0.8$$

This is called the incomplete Beta function. By using equations 2.0.1 and 3.0.8

$$P_h = \int_{y_{h-1}}^{y_h} f(y)dy$$

$$= \frac{1}{B(p,q)} \int_{y_{h-1}}^{y_h} t^{p-1}(1-t)^{q-1}dt \qquad 3.0.9$$

$$= I_{y_h}(p,q) - I_{y_{h-1}}(p,q) \qquad 3.0.10$$

$$= \Delta I_{y_{h-1}}(p,q) \qquad 3.0.11$$

By using equations 2.0.2 and 3.0.8

$$P_h\mu_h = \frac{1}{B(p,q)} \int_{y_{h-1}}^{y_h} y^p(1-y)^{q-1}dy \qquad 3.0.12$$

$$= \frac{B(p+1,q)}{B(p,q)} \int_{y_{h-1}}^{y_h} \frac{y^p(1-y)^{q-1}}{B(p+1,q)}dy \qquad 3.0.13$$

But $B(p,q) = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}$, $\Gamma(p) = (p-1)\Gamma(p-1)$ 3.0.14

$p > 1$

Hence

$$P_h\mu_h = \frac{p}{p+q}\{I_{y_h}(p+1,q) - I_{y_{h-1}}(p+1,q)\} \quad 3.0.15$$

$$= \frac{p}{p+q}\{\Delta I_{y_{h-1}}(p+1,q)\} \quad 3.0.16$$

By using equations 2.0.3 and 3.0.8

$$P_h(\sigma_h^2 + \mu_h^2) = \int_{y_{h-1}}^{y_h} y^2 f(y)\,dy \quad 3.0.17$$

$$= \int_{y_{h-1}}^{y_h} \frac{y^{p+1}(1-y)^{q-1}}{B(p,q)}\,dy \quad 3.0.18$$

$$= \frac{B(p+2,q)}{B(p,q)}\,\frac{1}{B(p+2,q)}\int_{y_{h-1}}^{y_h} y^{p+1}(1-y)^{q-1}\,dy \quad 3.0.19$$

$$= \frac{B(p+2,q)}{B(p,q)}\{I_{y_h}(p+2,q) - I_{y_{h-1}}(p+2,q)\} \quad 3.0.20$$

$$= \frac{p(p+1)}{(p+q)(p+q+1)}\{I_{y_h}(p+2,q) - I_{y_{h-1}}(p+2,q)\} \quad 3.0.21$$

$$= \frac{p(p+1)}{(p+q)(p+q+1)}\{\Delta I_{y_{h-1}}(p+2,q)\} \quad 3.0.22$$

P_h's are determined from equations 3.0.10. When P_h's are known, μ_h's are further calculated from equations 3.0.15. With values of p_h and μ_h thus known, values of σ_h^2's are now obtained from 3.0.21.

We note that p_h, μ_h, σ_h^2 are all functions of y_h, and y_h. We next consider the determination of y_h by each method.

3.1. EQUALIZATION OF STRATUM TOTAL METHOD

By making $p_h \mu_h$ constant in each stratum we have

$p_h\mu_h = c$, c is a constant.

$$\sum_{h=1}^{k} p_h\mu_h = \mu \qquad 3.1.1$$

$$\therefore \mu = \frac{p}{p+q} = kc \qquad 3.1.2$$

$$c = \frac{1}{k}\frac{p}{(p+q)} \qquad 3.1.3$$

$$\sum_{i=1}^{h} p_i\mu_i = hc = \frac{hp}{k(p+q)} \qquad 3.1.4$$

$$\text{But } \sum_{i=1}^{h} p_i\mu_i = \int_0^{y_h} \frac{y^p(1-y)^{q-1}}{B(p,q)}dy \qquad 3.1.5$$

$$= \frac{B(p+1,q)}{B(p,q)} \int_0^{y_h} \frac{y^p(1-y)^{q-1}}{B(p+1,q)}dy \qquad 3.1.6$$

$$= \frac{p}{p+q}\{I_{y_h}(p+1,q) - I_0(p+1,q)\} \qquad 3.1.7$$

$I_0(p+1,q) = 0$

$\therefore$ From 3.1.5 and 3.1.7

$$\frac{p}{p+q}\{I_{y_h}(p+1,q)\} = \frac{hp}{k(p+q)} \qquad 3.1.8$$

$$I_{y_h}(p+1,q) = \frac{h}{k} \qquad 3.1.9$$

$h = 1,2,3,\ldots,k.$

$y_1, y_2, y_3, \ldots, y_k$ satisfying 3.1.9 are the approximately optimum stratification boundary points based on Equalization of stratum total method. These values are used in 3.0.10, 3.0.15, 3.o.21 to obtain p_h, μ_h, σ_h^2 respectively.

3.1.1 COMPUTATIONAL PROCEDURE

(i) $h/_K$ is computed for $h = 1,2,3,\ldots,k$

(ii) Incomplete Beta function with parameters $p+1$, q is obtained by a computer programme algorithm such that

$$|I_{y_h}(p+1,q) - \frac{h}{k}| \leq \sigma \qquad 3.1.10$$

where σ is a positive number not greater than 0.01. The process is repeated for all values of h. The set of values of y_h satisfying 3.1.10 are then plugged into 3.0.10, 3.0.15, 3.0.21 and then $(\sum p_h \sigma_h)^2$ is computed for the over-all variance multiplied by n.

3.2 CUM $\sqrt{f}$ METHOD

$$\sqrt{f(y)} = \frac{y^{\frac{p-1}{2}}(1-y)^{\frac{q-1}{2}}}{\sqrt{B(p,q)}} \qquad 3.2.1$$

$$= \frac{y^{\frac{p+1}{2}-1}(1-y)^{\frac{q+1}{2}-1}}{\sqrt{B(p,q)}} \qquad 3.2.2$$

$$= \frac{B(\frac{p+1}{2}, \frac{q+1}{2})}{\sqrt{B(p,q)}} \cdot \frac{y^{\frac{p+1}{2}-1}(1-y)^{\frac{q+1}{2}-1}}{B(\frac{p+1}{2}, \frac{q+1}{2})} \qquad 3.2.3$$

Let $z(y_h) = \int_0^{y_h} \sqrt{f(y)}\,dy \qquad 3.2.4$

We note in 3.2.3 that

$$\int_0^y \frac{y^{\frac{p+1}{2}-1}(1-y)^{\frac{q+1}{2}-1}dy}{B(\frac{p+1}{2}, \frac{q+1}{2})} = I_y(\frac{p+1}{2}, \frac{q+1}{2}) \qquad 3.2.5$$

$$\therefore \quad Z(y_h) = \frac{B(\frac{p+1}{2}, \frac{q+1}{2})}{\sqrt{B(p,q)}} \cdot I_{y_h}(\frac{p+1}{2}, \frac{q+1}{2}) \qquad 3.2.6$$

$$I_{y_k}(\frac{p+1}{2}, \frac{q+1}{2}) = I_1(\frac{p+1}{2}, \frac{q+1}{2}) = 1 \qquad 3.2.7$$

$$\therefore \quad I_{y_k}(\frac{p+1}{2}, \frac{q+1}{2}) = 1 \qquad 3.2.8$$

$$\therefore \quad Z(y_k) = \frac{B(\frac{p+1}{2}, \frac{q+1}{2})}{\sqrt{B(p,q)}} \qquad 3.2.9$$

From 2.5.1

$$Z(y_h) = \frac{h}{k} Z(y_k)$$

From 3.2.6 and 3.2.9

$$I_{y_h}(\frac{p+1}{2}, \frac{q+1}{2}) = \frac{h}{k}, \quad h = 1,2,3,\ldots,k \qquad 3.2.10$$

$y_1, y_2, y_3, \ldots, y_k$ satisfying 3.2.10 are the approximately optimum stratification boundary points based on Cum $\sqrt{f}$ method. And these values are used in 3.0.10, 3.0.15, 3.0.21 to obtain p_h, μ_h and σ_h^2 respectively.

3.2.1 COMPUTATIONAL PROCEDURE

(i) $\frac{h}{k}$ is computed for $h = 1,2,3,\ldots,k$

(ii) Incomplete Beta function with parameters $\frac{p+1}{2}, \frac{q+1}{2}$ is obtained by a computer programme algorithm such that

$$\left| I_{y_h}(\frac{p+1}{2}, \frac{q+1}{2}) - \frac{h}{k} \right| \leq \sigma \qquad 3.2.11$$

where σ is a positive number which is not greater than 0.01. The process is repeated for all values of h. The set of values of y_h satisfying 3.2.11 are then plugged into equations 3.0.10, 3.0.15

and 3.0.21 and then $(\sum p_h \sigma_h)^2$ is computed for the overall variance multiplied by n.

3.3 DURBIN'S METHOD

$$\text{Define } r(y) = \frac{F(y_k) - F(y_o)}{y_k - y_o} \qquad 3.3.1$$

$$F(y_k) = 1,\ F(y_o) = 0$$

And for Beta distribution of the first kind

$$y_o = 0,\ y_k = 1$$

$$r(y) = 1 \qquad 0 < y < 1 \qquad 3.3.2$$

$$\text{Let } Q(y_h) = \text{cum } (r(y) + f(y))$$

$$= \int_0^{y_h} (1 + f(y)\,dy$$

$$= \int_0^{y_h} 1.dy + \int_0^{y_h} f(y)\,dy$$

$$= \int_0^{y_h} 1.dy + \frac{1}{B(p,q)} \int_0^{y_h} y^{p-1}(1-y)^{q-1}dy$$

$$= y_h + I_{y_h}(p,q) \qquad 3.3.3$$

$$Q(y_k) = y_k + I_{y_k}(p,q)$$

$$= 1 + 1$$

$$= 2$$

From 2.6.3

$$y_h + I_{y_h}(p,q) = \frac{h}{k}.2$$

$$= \frac{2h}{k}$$

$$y_h + I_{y_h}(p,q) = \frac{2h}{k},\ h = 1,2,3,\ldots,k \qquad 3.3.4$$

$y_1, y_2, y_3, \ldots, y_k$ 3.3.4 are the approximately optimum stratification boundary points based on Durbin's Method.

3.3.1 COMPUTATIONAL PROCEDURE

(i) $\frac{2h}{k}$ is computed for h = 1,2,3,..,k

(ii) Incomplete Beta function with parameter p and q is obtained by a computer algorithm such that

$$|I_{y_h}(p,q) + y_h - \frac{2h}{k}| \leq \sigma \qquad 3.3.5$$

where σ is a positive number not greater than 0.01. The set of values of y_h satisfying 3.3.5 are then put in equations 3.0.10, 3.0.15, 3.0.21, and then $(\sum p_h \sigma_h)^2$ computed for the over-all variance multiplied by n.

3.4. EQUAL RANGES METHOD

From 2.8.1

$$y_h = y_o + (y_k - y_o)\frac{h}{k}, \quad h = 1,2,3,\ldots,k \qquad 3.4.1$$

$y_o = 0$, $y_k = 1$ for Beta distribution of the first kind, Hence

$$y_h = \frac{h}{k}, \quad h = 1,2,\ldots\ldots k \qquad 3.4.2$$

$y_1, y_2, y_3, \ldots, y_k$ satisfying 3.4.2 are the approximately optimum stratification points based on the Equal Ranges Method.

3.4.1 COMPUTATIONAL PROCEDURE

(i) $\frac{h}{k}$ is computed for h = 1,2,3,...k

(ii) y_h is set equal to $h/_k$,

$$y_h = h/_k \qquad 3.4.1$$

The set of values satisfying 3.4.1 are put in 3.0.10,

3.0.15, 3.0.21 and then $(\sum p_h \sigma_h)^2$ is completed for the overall variance multiplied by n.

3.5. EQUALIZED STRATA METHOD

P_h are made equal in each stratum

i.e P_h = constant = c

$$\sum_{h=1}^{k} p_h = kc$$

$$1 = kc$$

$$c = 1/k$$

$$\therefore \quad P_h = 1/k$$

$$I_{y_h}(p,q) - I_{y_{h-1}}(p,q) = \frac{1}{k} \qquad 3.5.1$$

$$h = 1,2,3,\ldots,k$$

Hence

$$I_{y_h}(p,q) = \frac{h}{k} \qquad 3.5.2$$

$y_1, y_2, y_3, \ldots, y_k$ satisfying 3.5.2 are the approximately optimum stratification boundary points based on equal sized strata method.

3.5.2. COMPUTATIONAL PROCEDURE

(i) h/k is computed for h = 1,2,3...,k.

(ii) Incomplete Beta function with parameters p and q is obtained by a computer algorithm such that:

$$\left| I_{y_h}(p,q) - \frac{h}{k} \right| \leq \sigma \qquad 3.5.3.$$

where σ is positive number not greater than 0.01. The set of values of y_h satisfying 3.5.3 are used in 3.0.10, 3.0.15, 3.0.21 and then $(\sum p_h \sigma_h)^2$ is computed.

3.6. BETA DISTRIBUTION OF THE SECOND KIND

Determination of the stratification points and the variance of the estimator for the mean of the population for Beta Distribution of the second kind with parameters p and q.

The probability density function of a random variable Y whose mean is under study is given by

$$f(y) = \frac{y^{p-1}}{B(p,q)(1+y)^{p+q}} \quad 0<y<\infty \qquad 3.6.1$$

$$p>0, q>0$$

$$\text{Let } J_y(p,q) = \frac{1}{B(p,q)} \int_0^Y \frac{t^{p-1}\,dt}{(1+t)^{p+q}} \qquad 3.6.2$$

In 3.6.2. make the substitutions

$$t = \frac{u}{1-u} \text{ or } u = \frac{t}{1+t} \text{ and } dt = \frac{du}{(1-u)^2} \qquad 3.6.3$$

After some simplifications we will have

$$J_y(p,q) = \int_0^{Y/1+Y} \frac{u^{p-1}(1-u)^{q-1}}{B(p,q)}du \qquad 3.6.4$$

Thus

$$J_y(p,q) = I_{Y/1+Y}(p,q) \quad , \quad 0<y<\infty \qquad 3.6.5$$

Clearly the Beta distribution of the second kind is transformed to Beta distribution of the first kind as shown above.

$I_{Y/1+Y}(p,q)$ is the incomplete Beta function of the first kind, and $\frac{Y}{1+Y}$ can be replaced by z where $0<z<1$.

Now

$$p_h = \int_{y_{h-1}}^{y_h} \frac{t^{p-1}\,dt}{B(p,q)\,(1+t)^{p+q}} \qquad 3.6.6$$

$$= J_{y_h}(p,q) - J_{y_{h-1}}(p,q)$$

$$= I_{\frac{y_h}{1+y_h}}(p,q) - I_{\frac{y_{h-1}}{1+y_{h-1}}}(p,q) \qquad 3.6.7$$

$$p_h\mu_h = \int_{y_{h-1}}^{y_h} \frac{t^{p}\,dt}{B(p,q)\,(1+t)^{p+q}} \qquad 3.6.8$$

In 3.6.8 make the substitutions

$$t = \frac{u}{1-u} \text{ or } u = \frac{t}{1+t} \text{ and } dt = \frac{du}{(1-u)^2} \qquad 3.6.9$$

After some simplifications we will have

$$p_h\mu_h = \frac{B(p+1,q-1)}{B(p,q)} \cdot \left[J_{y_h}(p+1,q-1) - J_{y_{h-1}}(p+1,q-1) \right]$$

$$= \frac{p}{q-1} \cdot \left[I_{\frac{y_h}{1+y_h}}(p+1,q-1) - I_{\frac{y_{h-1}}{1+y_{h-1}}}(p+1,q-1) \right]$$

$$q > 1 \qquad 3.6.10$$

Next

$$p_h(\sigma_h^2 + \mu_h^2) = \int_{y_{h-1}}^{y_h} \frac{t^{p+1}\,dt}{B(p,q)\,(1+t)^{p+q}} \qquad 3.6.11$$

In 3.6.11 make the substitutions

$$t = \frac{u}{1-u} \text{ or } u = \frac{t}{1+t} \text{ and } dt = \frac{du}{(1-u)^2} \qquad 3.6.12$$

After some simplifications we will have

$$p_h(\sigma_h^2 - \mu_h^2) = \frac{B(p+2,q-2)}{B(p,q)} \cdot \left[J_{y_h}(p+2,q-2) - J_{y_{h-1}}(p+2,q-2) \right]$$

$$= \frac{p(p+1)}{(q-1)(q-2)} \cdot \left[I_{\frac{y_h}{1+y_h}}(p+2,q-2) - I_{\frac{y_{h-1}}{1+y_{h-1}}}(p+2,q-2) \right]$$

3.6.13

3.6.1 CUM $\sqrt{f}$ METHOD

$$\sqrt{f} = \frac{y^{\frac{p-1}{2}}}{\sqrt{B(p,q)}\ (1-y)^{\frac{p+q}{2}}} \qquad 0<y<\infty \qquad 3.6.14$$

Let $Q(y_h) = \int_0^{y_h} \sqrt{f(t)}\,dt$

$$= \int_0^{y_h} \frac{t^{\frac{p-1}{2}}\,dt}{\sqrt{B(p,q)}\,.\,(1+t)^{\frac{p+q}{2}}}$$

By putting $t = \frac{z}{1-z} \Rightarrow dt = \frac{dz}{(1-z)^2}$

$$Q(y_h) = \frac{1}{\sqrt{B(p,q)}} \int_0^{\frac{y_h}{1+y_h}} z^{\frac{p-1}{2}}(1-z)^{\frac{q-3}{2}}\,dz$$

$$= \frac{1}{\sqrt{B(p,q)}} \int_0^{\frac{y_h}{1+y_h}} z^{\frac{p+1}{2}-1}(1-z)^{\frac{q-1}{2}-1}\,dz$$

$$= \frac{B(\frac{p+1}{2},\ \frac{q-1}{2})}{\sqrt{B(p,q)}} \cdot \int_0^{\frac{y_h}{1+y_h}} \frac{z^{\frac{p+1}{2}-1}(1-z)^{\frac{q-1}{2}-1}}{B(\frac{p+1}{2},\ \frac{q-1}{2})}\,dz$$

$$Q(y_h) = \frac{B(\frac{p+1}{2}, \frac{q-1}{2})}{\sqrt{B(p,q)}} \cdot I_{\frac{y_h}{1+y_h}}(\frac{p+1}{2}, \frac{q-1}{2}) \qquad 3.6.15$$

$q > 1$

$$\therefore \quad Q(\infty) = \frac{B(\frac{p+1}{2}, \frac{q-1}{2})}{\sqrt{B(p,q)}} I_1(\frac{p+1}{2}, \frac{q-1}{2}) \qquad 3.6.16$$

$$= \frac{B(\frac{p+1}{2}, \frac{q-1}{2})}{\sqrt{B(p,q)}}, \quad q > 1 \qquad 3.6.17$$

From 2.5.1, 3.6.8, and 3.6.9

$$I_{\frac{y_h}{1+y_h}}(\frac{p+1}{2}, \frac{q-1}{2}) = \frac{h}{k}, \quad h = 1,2,3\ldots,k \qquad 3.6.13$$

$y_1, y_2, \ldots, y_k$ satisfying 3.6.13 are approximately optimum stratification boundary points based on cum $\sqrt{f}$ method. We remark that the computational procedure is similar to 3.2.1 except that $q > 2$.

3.6.2. EQUALIZATION OF STRATUM TOTAL METHOD

By making $p_h\mu_h$ constant in each stratum, we have $p_h\mu_h = c$, where c is a constant.

$$p_h\mu_h = \frac{B(p+1,q-1)}{B(p,q)}\{ I_{\frac{y_h}{1+y_h}}(p+1,q-1) - I_{\frac{y_{h-1}}{1+y_{h-1}}}(p+1,q-1)\} \qquad 3.6.19$$

$q > 1$

$$\sum_{h=1}^{k} p_h\mu_h = \mu = \frac{B(p+1,q-1)}{B(p,q)}$$

$$c = \frac{1}{k} \cdot \frac{B(p+1, q-1)}{B(p,q)}$$

$$I_{\frac{y_h}{1+y_h}}(p+1,\ q-1) = \frac{h}{k}\ ,\ h = 1,2,\ldots,k \qquad 3.5.20$$

$$q > 1$$

$y_1, y_2, y_3, \ldots y_{k-1}$ are obtained and they are approximately optimum boundaries based on the equalization of stratum total method. Similarly q must be greater than 2.

3.6.3. EQUAL SIZED STRATA METHOD

We make p_h constant in each stratum and from 3.6.5 we have the set of equations

$$I_{\frac{y_h}{1+y_h}}(p,q) = \frac{h}{k}\ ,\ h = 1,2,\ldots,k \qquad 3.6.21$$

$y_1, y_2, \ldots, y_k$ satisfying 3.6.16 are approximately optimum stratification boundary points based on equal sized strata method. In this case also q must be greater than 2.

3.6.4 DURBIN'S METHOD

Define $r(y) = \dfrac{F(y_k) - F(y_o)}{y_k - y_o}$

Since the range of Beta of the second kind is infinite we truncate the density function.

Let $F(y_k) = 1 - \alpha,\ \ 0 < \alpha < .001$

where

$$F(y_k) = \int_0^{y_k} \frac{y^{p-1}\,dy}{B(p,q)(1+y)^{p+q}}$$

$$= J_{y_k}(p,q)$$

$$= I_{\frac{y_k}{1+y_k}}(p,q)$$

$$\text{Hence } r(y) = I_{y_k/1+y_k}(p,q) \cdot \frac{1}{y_k} \qquad 3.6.22$$

$$\text{Let cum } (r(y) + f(y))) = S(y)$$

$$\therefore S(y_h) = \int_0^{y_h} [r(y) + f(y)]dy$$

$$= \frac{I_{y_k}}{1+y_k}(p;q) \cdot \frac{y_h}{y_k} + \frac{I_{y_h}}{1+y_h}(p,q) \qquad 3.6.23$$

$$S(y_k) = 2\, \frac{I_{y_k}}{1+y_k}(p,q) \qquad 3.6.24$$

From 2.6.3, 3.6.17, 3.6.18, we have

$$\frac{y_h}{y_h} \cdot \frac{I_{y_k}}{1+y_k}(p,q) + \frac{I_{y_h}}{1+y_k}(p,q) = \frac{2h}{k} \cdot \frac{I_{y_k}}{1+y_k}(p,q), \qquad 3.6.25$$

$$h = 1,2,3,\ldots,k$$

The set $y_1, y_2, y_3, \ldots, y_k$ satisfying 3.6.25 are approximately optimum boundary points based on Durbin's Method.

CHAPTER FOUR

EQUATIONS FOR THE STRATIFICATION POINTS OF GAMMA DISTRIBUTION

In this Chapter we discuss the determination of the stratification boundary points and the variance of the estimator of the mean of the population taking Gamma model with parameters p and β as the target population. We obtain:

(i) equations for the approximate determination of optimum stratification points using various methods reviewed in Chapter Two.

(ii) equations for proportion p_h mean μ_h and variance σ_h^2 of the h^{th} stratum

$h = 1,2,3,\ldots,k$

(iii) computational procedure for (i) and (ii), and

(iv) $n\, V_{opt}(\bar{y}_{st}) = (\sum p_h\, \sigma_h)^2$

Suppose the probability density function of a random variable Y is f(y) where

$$f(y) = \frac{e^{-y/\beta}\, y^{p-1}}{\Gamma(p)\ \beta^p}, \quad y > 0, \quad \beta > 0\ p > 0 \qquad 4.0.1$$

β is the scale parameter and p is the shape parameter.

$$\text{Let } G_y(p,\beta) = \int_0^y \frac{t^{p-1}\ e^{-t/\beta}}{\Gamma(p)\ \beta^p}\, dt, \quad y > 0 \qquad 4.0.2$$

This is called the Incomplete Gamma function, with parameters p and β.

By putting $t/_{\beta} = u \Longrightarrow dt = \beta du$

$$G_y(p,\beta) = \int_0^{y/_{\beta}} \frac{u^{p-1} e^{-u}}{\Gamma(p)} du \qquad 4.0.3$$

The mean and variance of y are respectively

$$\mu = \int_0^{\infty} yf(y)dy = p\beta \qquad 4.0.4$$

$$\sigma^2 = \int_0^{\infty} y^2 f(y)dy - \mu^2 = p(p+1)\beta^2 - p^2\beta^2$$

$$= p\beta^2. \qquad 4.0.5$$

From 2.0.1

$$p_h = \int_{y_{h-1}}^{y_h} f(y)dy = \int_{y_{h-1}}^{y_h} \frac{e^{-t/_{\beta}} t^{p-1}}{\Gamma(p)\beta^p} dt \qquad 4.0.6$$

In 4.0.6 make the substitutions

$$t/_{\beta} = z, \; dt = \beta dz \qquad 4.0.7$$

we have

$$p_h = \int_{y_{h-1/\beta}}^{y_{h/\beta}} \frac{e^{-z}(\beta z)^{p-1}}{\Gamma(p)\ \beta^p} \beta dz \qquad 4.0.8$$

$$= \frac{1}{\Gamma(p)} \int_{y_{h-1/\beta}}^{y_{h/\beta}} e^{-z}\ z^{p-1}\ dz \qquad 4.0.9$$

$$= G_{y_h}(p,\beta) - G_{y_{h-1}}(p,\beta) \qquad 4.0.10$$

From 2.0.2

$$p_h\mu_h = \frac{1}{\Gamma(p)\beta^p} \int_{y_{h-1/\beta}}^{y_{h/\beta}} (\beta z)^p e^{-z} \beta dz \qquad 4.0.11$$

$$= \frac{\Gamma(p+1)}{\Gamma(p)\beta^p} \int_{y_{h-1}}^{y_h} \frac{y^p e^{-y/\beta}}{\Gamma(p+1)} dy$$

$$= \frac{\Gamma(p+1)}{\Gamma(p)\beta^p} \int_{y_{h-1}/\beta}^{y_{h/\beta}} (\beta z)^p e^{-z} \beta dz$$

$$= \frac{\beta\,\Gamma(p+1)}{\Gamma(p)} \int_{y_{h-1}/\beta}^{y_{h/\beta}} \frac{z^p e^{-z}}{\Gamma(p+1)} dz$$

$$= \beta p \{ G_y(p+1,\beta) - G_{y_{h-1}}(p+1,\beta) \} \qquad 4.0.12$$

From 2.0.3

$$p_h(\sigma_h^2 + \mu_h^2) = \frac{1}{\Gamma(p)\beta^p} \int_{y_{h-1}}^{y_h} y^{p+1} e^{-y/\beta} dy \qquad 4.0.13$$

$$= \frac{\beta^{p+2}}{\Gamma(p)\beta^p} \int_{y_{h-1}/\beta}^{y_{h/\beta}} z^{p+1} e^{-z} dz$$

$$= \frac{\beta^2 \Gamma(p+2)}{\Gamma(p)} \int_{y_{h-1}/\beta}^{y_{h/\beta}} \frac{z^{p+1} e^{-z}}{\Gamma(p+1)} dz$$

$$= p(p+1)\beta^2 \{ G_{y_h}(p+2,\beta) - G_{y_{h-1}}(p+2,\beta) \} \qquad 4.0.14$$

When $\beta = 2$ we have chi-square distribution with $2p$ degrees of freedom. To determine p_h, μ_h σ_h^2, we use equations 4.0.2, 4.0.10, 4.0.11, 4.0.14.

4.1. CUM $\sqrt{f}$ METHOD

$$\sqrt{f}(y) = \frac{e^{-y/2\beta}\, y^{\frac{p-1}{2}}}{\sqrt{\Gamma(p)\,\beta^{p}}} \qquad 4.1.1$$

Let $T(y_h) = \int_0^{y_h} \sqrt{f(y)}\, dy$ $\qquad 4.1.2$

$$= \frac{1}{\sqrt{\Gamma(p)\,\beta^{p}}} \int_0^{y_h} y^{\frac{p-1}{2}} . e^{-y/2\beta} dy$$

put $y/2\beta = z, dy = 2\beta\, dz$

$$T(y_h) = \frac{1}{\sqrt{\Gamma(p)\,\beta^{p}}} \cdot \int_0^{y_{h/2\beta}} (2\beta z)^{\frac{p-1}{2}} . e^{-z} . 2\beta dz \qquad 4.1.3$$

$$= \frac{2^{\frac{p+1}{2}}\, \beta^{\frac{p+1}{2}}}{\sqrt{\Gamma(p)\,\beta^{p}}} \int_0^{y_h/2\beta} z^{\frac{p-1}{2}} . e^{-z} dx$$

$$= \frac{(2\beta)^{\frac{p+1}{2}}}{\sqrt{\Gamma(p)\,\beta^{p}}} \cdot \Gamma\frac{(p+1)}{2} \int \frac{z^{\frac{p+1}{2}-1} e^{-z}}{\Gamma\frac{(p+1)}{2}}\, dz$$

$$= \frac{(2\beta)^{\frac{p+1}{2}}\ \Gamma\frac{(p+1)}{2}}{\sqrt{\Gamma(p).\beta^{p}}} \cdot G_{y_h/2\beta}\left(\frac{p+1}{2}, 1\right) \qquad 4.1.4$$

$$T(\infty) = \frac{(2\beta)^{\frac{p+1}{2}}\ \Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\Gamma(p)\,\beta^{p}}} \cdot 1 \qquad 4.1.5$$

since $G_{\infty}\left(\frac{p+1}{2}, 1\right) = 1$

From 2.5.1, 4.1.4, 4.1.5

$$G_{y_h/2\beta}\left(\frac{p+1}{2}, 1\right) = \frac{h}{k}, \quad h = 1,2,3,\ldots,k \qquad 4.1.6$$

$y_1, y_2, y_3, \ldots, y_k$ satisfying 4.1.6 are approximately optimum stratification points based on the cum $\sqrt{f}$ method. These values are used in 4.0.10, 4.0.12, and 4.0.14.

4.1.1. COMPUTATIONAL PROCEDURE

(i) $\frac{h}{k}$ is computed for $h = 1,2,3,\ldots,k$

(ii) $\Gamma\left(\frac{p+1}{2}\right)$ is obtained.

(iii) Incomplete Gamma function with parameters $\frac{p+1}{2}$, 1 is obtained by a computer algorithm such that $\left|G_{t_h}\left(\frac{p+1}{2}, 1\right) - \frac{h}{k}\right| \leq \sigma$ 4.1.1

where σ is a positive number which is not greater than 0.02. The process is repeated for various values of h. The set of values $y_h = 2\beta t_h$ where t_h's satisfy 4.1.1 are used in 4.0.11, 4.0.12, 4.0.14 to determine p_h, μ_h and σ_h.

(iv) $n \times \left(\sum p_h \sigma_h\right)^2$ is computed.

4.2. EQUALIZATION OF STRATUM TOTAL METHOD

By making $p_h u_h$ constant we have

$$p_h u_h = c \qquad 4.2.1$$

$$\sum_{h=1}^{k} p_h u_h = kc$$

$$\mu = kc$$

But $\mu = \beta p$

$$\therefore \quad c = \frac{1}{k}\beta p \qquad 4.2.2$$

$$\text{Let } H(y_h) = \int_0^{y_h} yf(y)dy \qquad 4.2.3$$

$$= \frac{1}{\Gamma(p)\beta^p} \int_0^{y_h} y^p e^{-y/\beta}\, dy$$

$$= \beta p \int_0^{y_{h/\beta}} \frac{z^p e^{-z}}{\Gamma(p+1)}\, dz$$

$$= \beta p \{ G_{y_h}(p+1,\beta) \} \qquad 4.2.4$$

$$H(\infty) = \beta p \times 1$$

Using 2.4.1 and 4.2.4

$$P\beta i \; C_{y_h}(p+1,\beta) = \frac{h}{k}\beta p$$

$$\Longrightarrow \quad G_{y_h}(p+1,\beta) = \frac{h}{k}, \quad h = 1,2,3,\ldots,k \qquad 4.2.5$$

$y_1, y_2, y_3, \ldots, y_k$ satisfying 4.2.5 are approximately optimum stratification boundary points based on the equalization of stratum total method. These values are used in 4.0.1, 4.0.12, 4.0.14.

4.2.1 COMPUTATIONAL PROCEDURE

(i) h/k is computed for $h = 1,2,3,\ldots,k$

(ii) $\Gamma(p+1)$ is obtained

(iii) Incomplete Gamma function with parameters $p+1, \beta$ is obtained by a computer algorithm such that

$$\left| G_{y_h}(p+1,\beta) - \frac{h}{k} \right| < \sigma \qquad 4.2.1$$

where σ is positive and is not greater than 0.001. The process is repeated for various values of h. The set of values of y_h are used in 4.0.11, 4.0.12, 4.0.14.

4.3. EQUAL RANGES METHOD

The range of f(y) is infinite, therefore the density function is truncated, and the value of y_k is such that

$$G_{y_k}(p,\beta) = 1-\alpha, \quad 0 < \alpha \leq 0.01 \qquad 4.3.1$$

$$y_h = \frac{h}{k} y_k \ , \ h = 1,2,3,\ldots,k \qquad 4.3.2$$

The set of values y_h satisfying 4.3.1 are approximately stratification boundary points based on the Equal ranges method.

4.3.1 COMPUTATIONAL PROCEDURE

(i) $\Gamma(p)$ is obtained

(ii) Incomplete Gamma function with parameters p and β is obtained by a computer algorithm such that $|1 - G_y(p,\beta)| < \sigma$ 4.3.1 where σ is positive not greater than 0.01. The value of y which satisfy 4.3.1 is the upper boundary of Y and is denoted by y_k, k is the number of strata.

(iii) $y_h = \frac{h}{k} y_k$, $h = 1,2,3,\ldots,k$ 4.3.2

(iv) The set of values of y satisfying 4.3.2 are used in 4.0.10, 4.0.12, 4.0.14

(v) $n \times (\sum_{h=1}^{k} p_h \sigma_h)^2$ is computed for $k = 2,3,\ldots,10$.

4.4. EQUAL SIZED STRATA METHOD

p_h is constant in each stratum.

Note that $\sum_{h=1}^{k} p_h = 1$ 4.4.1

$\therefore kc = 1 \Rightarrow c = \frac{1}{k}$.

From 4.0.10, we have

$$P_h = G_{y_h}(p,\beta) - G_{y_{h-1}}(p,\beta) = \frac{1}{k} \qquad 4.4.2$$

$$\sum_{i=1}^{h} p_i = \int_0^{y_h} f(y)dy$$

$$= G_{y_h}(p,\beta)$$

$$\therefore \quad G_{y_h}(p,\beta) = \frac{h}{k} \ , \ h = 1,2,3,\ldots,k. \qquad 4.4.3$$

Comparing 4.2.5 and 4.1.3 we note that the equations are similar except that the parameters are not the same. The set of values $\{y_h\}$ satisfying 4.4.3 are approximately optimum stratification boundary points based on equal sized strata method.

4.4.1 COMPUTATIONAL PROCEDURE

The procedure is the same as that of equalization of stratum total method except that step (ii) is replaced by $\Gamma(p)$ and step (iii) parameters p and β are used instead of parameters p+1 and β.

4.5. DURBIN'S METHOD

$$r(y) = \frac{F(y_k) - F(y_0)}{y_k - y_0} \qquad 4.5.1$$

$$F(y) = \int_0^y \frac{e^{-t/\beta}\, t^{p-1}}{\Gamma(p)\,\beta^p}\, dt = G_y(p,\beta) \qquad 4.5.2$$

Since the interval is infinite, the density function has to be truncated so that a finite range which practically includes the whole of the population.

Let $F(y_k) = 1-\alpha$ where α is the portion that is deleted.

$$\therefore \quad G_{y_k}(p,\beta) = 1-\alpha \qquad 0 < \alpha \leq 0.01 \qquad 4.5.3$$

$$\text{Let } Q(y_h) = \int_0^{y_h} [r(y) + f(y)]\,dy \qquad 4.5.4$$

$$= \int_0^{y_h} \left[\frac{G_{y_k}(p,\beta)}{y_k} + \frac{y^{p-1}\, e^{-y/\beta}}{\Gamma(p)\,\beta^p} \right] dt \qquad 4.5.5$$

$$= \frac{G_{y_k}(p,\beta)}{y_k} \cdot y_h + G_{y_h}(p,\beta) \qquad 4.5.6$$

$$\therefore \quad Q(y_k) = G_{y_k}(p,\beta) + G_{y_k}(p,\beta)$$

$$= 2\, G_{y_k}(p,\beta) \qquad 4.5.7$$

By using 2.6.3, 4.5.6, 4.5.7

$$G_{y_k}(p,\beta)\, \frac{y_h}{y_k} + G_{y_h}(p,\beta) = \frac{2h}{k} G_{y_k}(p,\beta) \qquad 4.5.8$$

$$h = 1,2,3,\ldots,k.$$

$y_1, y_2, y_3 \ldots, y_k$ satisfying 4.5.8 are approximately optimum stratification boundary points based on Durbin's Method.

4.5.1 COMPUTATIONAL PROCEDURE

(i) $\Gamma(p)$ is obtained

(ii) Incomplete Gamma function with parameters p and β is obtained by a computer algorithm such that

$$| 1 - G_y(p,\beta) | < \sigma_1, \quad 0 < \sigma_1 \leq 0.01 \qquad 4.5.1$$

The values of y which satisfies 4.5.1 is the upper boundary of finite range desired and that value is denoted by y_k. k is the number of strata.

(iii) $G_y(p,\beta)/_{1-\sigma_2}$, $0 < \sigma_2 < \sigma_1$ is computed

(iv) $\frac{2h}{k}$ is computed for various values of h

(v) The set of values of y which satisfy the inequality

$$|G_{y_h}(p,\beta)/_{1-\sigma_2} + y_h - \frac{2h}{k}| < \sigma \text{ where}$$

$0 < \sigma \leq 0.1$ are used.

(vi) $n \times (\sum p_h \sigma_h)^2$ is computed.

CHAPTER FIVE

EQUATIONS FOR THE STRATIFICATION POINTS OF PARETO DISTRIBUTION

In this Chapter we discuss the determination of the stratification boundary points and the variance of the estimator of the mean of the population taking Pareto model with parameter θ as the target population. We obtain

(i) equations for the approximate determination of optimum stratification points using various methods reviewed in Chapter Two,

(ii) Equations for proportion p_h, mean μ_h and variance σ_h^2 of the h^{th} stratum.

Suppose the probability density function of a random Y is f(y) where

$$f(y) = \theta y_o^{\theta} y^{-\theta-1}, \quad y>y_o, \quad \theta > 2 \qquad 5.0.1$$

y_o is called the subsistence wage when used as model for income distribution and θ is the shape parameter.

$$\int_{y_o}^{\infty} f(y)dy = \theta y_o^{\theta} \int_{y_o}^{\infty} y^{-\theta-1} dy \qquad 5.0.2$$

$$= \theta y_o^{\theta} \left[\frac{y^{-\theta}}{-\theta} \right]_{y_o}^{\infty}$$

$$= - y_o^{\theta} \left[0 - y_o^{-\theta} \right]$$

$$= - y_o^{\theta} (- y_o^{-\theta}) \qquad 5.0.3$$

$$= 1.$$

The mean μ and variance σ^2 of y are respectively

$$\mu = \int_{y_o}^{\infty} y f(y) dy \qquad 5.0.4$$

$$= \theta y_o^{\theta} \int_{y_o}^{\infty} y^{-\theta} \, dy$$

$$= \theta y_o^{\theta} \left[\frac{y^{-\theta+1}}{-\theta+1} \right]_{y_o}^{\infty}$$

$$= \theta y_o^{\theta} \left[\frac{y^{-\theta+1}}{1-\theta} \right]^{\infty}$$

$$= \frac{\theta y_o^{\theta}}{1-\theta} \left[-y_o^{-\theta+1} \right]^{y_\theta}$$

$$= - \frac{\theta y_o^{\theta} \; y_o^{-\theta+1}}{1 - \theta}$$

$$= \frac{-\theta y_o}{1-\theta} = \frac{\theta y_o}{\theta-1} \quad , \; \theta > 1 \qquad 5.0.5$$

$$\sigma^2 = \int_{y_o}^{\infty} y^2 f(y) \, dy - \mu^2 \qquad 5.0.6$$

$$= \theta y_o^{\theta} \int_{y_o}^{\infty} y^{-\theta+-} \, dy - \mu^2 \qquad 5.0.7$$

$$= \theta y_o^{\theta} \left[\frac{y^{-\theta+2}}{-\theta+2} \right]_{y_o}^{\infty} - \mu^2 \qquad 5.0.8$$

$$= \theta y_o^{\theta} \left[\frac{y^{-\theta+2}}{2-\theta} \right]_{y_o}^{\infty} - \mu^2 \qquad 5.0.9$$

$$= \frac{\theta y_o^{\theta}}{2-\theta} \left[- \; y_o^{-\theta+2} \right] - \mu^2 \qquad 5.0.11$$

$$= \frac{\theta y_o^{\theta} \; x \; y_o^{-\theta+2}}{\theta - 2} - \mu^2$$

$$= \frac{\theta y_o^2}{\theta-2} - \left(\frac{\theta y_o}{\theta-1} \right)^2 \quad , \; \theta > 2 \qquad 5.0.12$$

From 2.0.1

$$p_h = \theta y_o^{\theta} \int_{y_{h-1}}^{y_h} y^{-\theta-1}\,dy \qquad 5.0.12$$

$$= \theta y_o^{\theta} \left[\frac{y^{-\theta}}{-\theta}\right]_{y_{h-1}}^{y_h}$$

$$= -\, y_o^{\theta}\left[y_h^{-\theta} - y_{h-1}^{-\theta}\right]$$

$$= y_o^{\theta}\left[y_{h-1}^{-\theta} - y_h^{-\theta}\right]$$

$$= y_o^{\theta}\left[\frac{1}{y_{h-1}^{\theta}} - \frac{1}{y_h^{\theta}}\right] \qquad 5.0.13$$

$$= \left(\frac{y_o}{y_{h-1}}\right)^{\theta} - \left(\frac{y_o}{y_h}\right)^{\theta}$$

From 2.0.2

$$p_h u_h = \theta y_o^{\theta} \int_{y_{h-1}}^{y_h} y^{-\theta}\,dy \qquad 5.0.14$$

$$= \theta y_o^{\theta}\left[\frac{y^{-\theta+1}}{-\theta+1}\right]_{y_{h-1}}^{y_h}$$

$$= \frac{\theta y_o^{\theta}}{1-\theta}\left[y_h^{-\theta+1} - y_{h-1}^{-\theta+1}\right]$$

$$= \frac{\theta y_o^{\theta}}{1-\theta}\left[\frac{1}{y_h^{\theta-1}} - \frac{1}{y_{h-1}^{\theta-1}}\right]$$

$$= \frac{\theta y_o^{\theta}}{\theta-1}\left[\frac{1}{y_{h-1}^{\theta-1}} - \frac{1}{y_h^{\theta-1}}\right]$$

$$= \left(\frac{\theta}{\theta-1}\right)\left[\left(\frac{y_o}{y_{h-1}}\right)^{\theta} y_{h-1} - \left(\frac{y_o}{y_h}\right)^{\theta} y_h\right] \qquad 5.0.15$$

From 2.0.3.

$$p_h(\sigma_h^2 + \mu_h^2) = \theta y_o^{\theta} \int_{y_{h-1}}^{y_h} y^{-\theta+1} dy \qquad 5.0.16$$

$$= \theta y_o^{\theta} \left[\frac{y^{-\theta+2}}{-\theta+2} \right]_{y_{h-1}}^{y_h}$$

$$= \frac{\theta y_o^{\theta}}{2-\theta} \left[y_h^{-\theta+2} - y_{h-1}^{-\theta+2} \right]$$

$$= \frac{\theta y_o^{\theta}}{2-\theta} \left[\frac{1}{y_h^{\theta-2}} - \frac{1}{y_{h-1}^{\theta-2}} \right]$$

$$= \left(\frac{\theta}{2-\theta}\right) \left[\left(\frac{y_o}{y_h}\right)^{\theta} y_h^2 - \left(\frac{y_o}{y_{h-1}}\right)^{\theta} y_{h-1}^2 \right] \qquad 5.0.17$$

5.1. Cum $f^{1/2}$ Method

$$f^{1/2} = (\theta y_o^{\theta})^{1/2} \, y^{\frac{-\theta-1}{2}}, \; y > y_o \qquad 5.1.1$$

$$\int_{y_o}^{\infty} f^{1/2} dy = (\theta y_o^{\theta})^{1/2} \int_{y_o}^{\infty} y^{\frac{-\theta-1}{2}} dy \qquad 5.1.2$$

$$= (\theta y_o^{\theta})^{1/2} \frac{\left[y^{\frac{-\theta-1}{2}+1} \right]_{y_o}^{\infty}}{\frac{-\theta-1}{2}+1} \qquad 5.1.3$$

$$= \frac{2(\theta y_o^{\theta})^{1/2}}{(1-\theta)} \cdot \left[y^{\frac{(1-\theta)}{2}} \right]_{y_o}^{\infty} \qquad 5.1.4$$

$$= \frac{2(\theta y_o^{\theta})^{1/2}}{1-\theta} \left[- y_o^{\frac{1-\theta}{2}} \right]$$

$$= 2\left[\frac{\theta y_o^{\theta}}{\theta-1}\right]^{1/2} \cdot y_o^{\frac{1-\theta}{2}}$$

$$= \frac{2}{\theta-1} \cdot \theta^{1/2}\, y_o^{\frac{\theta}{2}} \cdot y_o^{\frac{1}{2}} \cdot y_o^{\frac{-\theta}{2}}$$

$$= \frac{2}{\theta-1} \cdot (\theta y_o)^{1/2} \qquad 5.1.5$$

From 2.5.1 and 5.1.5

$$\int_{y_o}^{y_h} f^{1/2}\, dy = \frac{h}{k} \cdot \frac{2}{\theta-1} \cdot (\theta y_o)^{1/2} \qquad 5.1.6$$

$$\frac{2(\theta y_o^{\theta})^{1/2}}{1-\theta}\left[y^{\frac{1-\theta}{2}}\right]_{y_o}^{y_h} = \frac{h}{k} \cdot \frac{2}{\theta-1} \cdot (\theta y_o)^{1/2} \qquad 5.1.7$$

$$\frac{(\theta y_o^{\theta})^{1/2}}{1-\theta}\left[y_h^{\frac{1-\theta}{2}} - y_o^{\frac{1-\theta}{2}}\right] = \frac{h}{k} \frac{\theta^{1/2}\, y_o^{1/2}}{\theta - 1} \qquad 5.1.8$$

$$\frac{\theta^{1/2}\, y_o^{\theta/2}}{1 - \theta}\left[\frac{1}{y_h^{\frac{\theta-1}{2}}} - \frac{1}{y_o^{\frac{\theta-1}{2}}}\right] = \frac{h}{k} \frac{\theta^{1/2}\, y_o^{1/2}}{\theta - 1} \qquad 5.1.9$$

$$\frac{y_o^{\theta/2}}{\theta-1}\left[\frac{1}{y_o^{\frac{\theta-1}{2}}} - \frac{1}{y_h^{\frac{\theta-1}{2}}}\right] = \frac{h}{k} \frac{y_o^{1/2}}{\theta-1} \qquad 5.1.10$$

$$\frac{1}{y_o(\frac{\theta-1}{2})} - \frac{1}{y_h(\frac{\theta-1}{2})} = \frac{h}{k}\, y_o^{\frac{1-\theta}{2}} \qquad 5.1.11$$

$$\frac{1}{y_o^{\frac{\theta-1}{2}}} - \frac{h}{k}\, y_o^{\frac{1-\theta}{2}} = \frac{1}{y_h^{\frac{\theta-1}{2}}} \qquad 5.1.12$$

$$y_o^{\frac{1-\theta}{2}} - \frac{h}{k}\, y_o^{\frac{1-\theta}{2}} = y_h^{\frac{1-\theta}{2}} \qquad 5.1.13$$

$$(1 - \frac{h}{k})\, y_o^{\frac{1-\theta}{2}} = y_h^{\frac{1-\theta}{2}} \qquad 5.1.14$$

$$y_h = y_o(1 - \frac{h}{k})^{\frac{2}{1-\theta}} \qquad 5.1.15$$

$$y_h = \frac{y_o}{(1 - \frac{h}{k})^{2/(\theta-1)}} \quad h = 1, 2, 3, \ldots, k \qquad 5.1.16$$

$y_1, y_2, \ldots, y_k$ satisfying 5.1.16 are approximately optimum points based on the Cum $\sqrt{f}$ method. These values are used in 5.0.13, 5.0.15, 5.0.17.

5.2. EQUAL SIZED STRATA METHOD

p_h is constant in each stratum. Note that

$$\sum_{h=1}^{k} p_h = 1$$

Then we have

$$\int_{y_o}^{y_h} f(y)\,dy = \frac{h}{k} \qquad 5.2.1$$

$$\theta y_o^{\theta}\left[\frac{y^{-\theta}}{-\theta}\right]_{y_o}^{y_h} = \frac{h}{k} \qquad 5.2.2$$

$$-y_o^{\theta}\left[y_h^{-\theta} - y_o^{-\theta}\right] = \frac{h}{k} \qquad 5.2.3.$$

$$\left[y_h^{-\theta} - y_o^{-\theta}\right] = -\frac{h}{k} y_o^{-\theta} \qquad 5.2.4$$

$$y_h^{-\theta} = y_o^{-\theta} - \frac{h}{k} y_o^{-\theta} \qquad 5.2.5$$

$$= y_o^{-\theta}\left(1 - \frac{h}{k}\right) \qquad 5.2.6$$

$$\frac{1}{y_h^{\theta}} = \frac{1 - \frac{h}{k}}{y_o^{\theta}} \qquad 5.2.7$$

$$y_o^{\theta} = \left(1 - \frac{h}{k}\right) y_h^{\theta} \qquad 5.2.8$$

$$\frac{y_o^{\theta}}{\left(1 - \frac{h}{k}\right)} = y_h^{\theta} \qquad 5.2.9$$

$$y_h = \frac{y_o}{\left(1 - \frac{h}{k}\right)^{1/\theta}}, h = 1,2,3,\ldots,k \qquad 5.2.10$$

y_1, y_2, y_3 satisfying 5.2.10 are approximately optimum stratification boundary points.

5.3. EQUALIZATION OF STRATUM TOTAL METHOD

By making $p_h \mu_h$ constant in each stratum we have

$$p_h \mu_h = c \quad \Rightarrow \quad \sum_{h=1}^{k} p_h \mu_h = \mu = kc$$

$$= \frac{\Theta y_o}{\Theta - 1}$$

$$\Rightarrow C = \frac{1}{k} \frac{\Theta y_o}{\Theta - 1} \qquad 5.3.1$$

$$\therefore \int_{y_o}^{y_h} y f(y)\, dy = \frac{h}{k} \cdot \frac{\Theta y_o}{\Theta - 1} \qquad 5.3.2$$

$$\Theta y_o^{\Theta} \int_{y_o}^{y_h} y^{-\Theta}\, dy = \frac{h}{k} \frac{\Theta y_o}{\Theta - 1} \qquad 5.3.3$$

$$\Theta y_o^{\Theta} \left[\frac{y^{-\Theta+1}}{1-\Theta} \right]_{y_o}^{y_h} = \frac{h}{k} \frac{\Theta y_o}{\Theta - 1} \qquad 5.3.4$$

$$\frac{y_o^{\Theta}}{1-\Theta} \left[y_h^{-\Theta+1} - y_o^{-\Theta+1} \right] = \frac{h}{k} \frac{y_o}{\Theta - 1} \qquad 5.3.5$$

$$\frac{y_o^{\Theta}}{\Theta - 1} \left[y_o^{-\Theta+1} - y_h^{-\Theta+1} \right] = \frac{h}{k} \frac{y_o}{\Theta - 1} \qquad 5.3.$$

$$y_o^{-\Theta+1} - y_h^{-\Theta+1} = \frac{h}{k} y_o^{1-\Theta} \qquad 5.3.7$$

$$y_o^{-\Theta+1} - \frac{h}{k} y_o^{1-\Theta} = y_h^{-\Theta+1} \qquad 5.3.8$$

$$\frac{1}{y_o^{\Theta-1}} - \frac{h}{k} y_o^{\Theta-1} = \frac{1}{y_h^{\Theta-1}} \qquad 5.3.9$$

$$\left(1 - \frac{h}{k}\right) \frac{1}{y_o^{\Theta-1}} = \frac{1}{y_h^{\Theta-1}} \qquad 5.3.10$$

$$y_h^{\Theta-1} = \frac{y_o^{\Theta-1}}{\left(1-\frac{h}{k}\right)} \qquad 5.3.11$$

$$y_h = \frac{y_o}{\left(1-\frac{h}{k}\right)^{1/(\Theta-1)}}, \quad h = 1,2,3,\ldots,k. \qquad 5.3.12$$

$y_1, y_2, y_3, \ldots, y_k$ satisfying 5.3.12 are approximately optimum stratification points based on the equalization of stratum total method.

5.4. DURBINS METHOD

From 2.6.3.

$$\int_{y_o}^{y_h} (\gamma(y) + f(y))dy = \frac{h}{k} \int_{y_o}^{\infty} (\gamma(y) + f(y)dy \qquad 5.4.1$$

$$\gamma(y) = \frac{F(\infty) - F(y_o)}{\infty - y_o} = 0, \; y_k = \infty, \text{ for Pareto Distribution.}$$

Hence 5.4.1 becomes

$$\int_{y_o}^{y_h} f(y)dy = \frac{h}{k} \int_{y_o}^{\infty} f(y)dy. \qquad 5.4.2$$

$$= \frac{h}{k} .$$

But 5.4.2 is the same as 5.2.1 i.e Equal sized Strata Method.

5.5. EQUAL RANGE METHOD

The range of f(y) is infinite, therefore the density function is truncated, and the value of y_k is such that

$$\int_{y_o}^{y_k} f(y)dy = 1-\alpha, \; 0 < \alpha \leq .001 \qquad 5.5.1$$

$$y_h = y_o + \frac{h}{k}(y_k - y_o) \qquad 5.5.2$$

TABLE 6.1.1

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 1, q = 2. MV (Minimum Variance)

	CUM √f	DURBIN'S	$P_h\mu_h = c$	EQUAL RANGE	EQUAL SIZED STRATA	EQUAL SIZED	EQUAL RANGE	MV.
K = 2	0.0150	0.0151	0.0182	0.0182	.0158	.0174	.0187	.0152
K = 3	0.0069	0.0069	0.0094	.0087	.0074	.0087	.0087	.0069
K = 4	0.0039	.0040	.0058	.0050	.0043	.0053	.0050	.0039
K = 5	0.0025	.0026	.0040	.0032	.0028	.0035	.0033	.0023

"The last three columns show Kokan's result".

TABLE 6 1.2

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 3, q = 2.

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.0400000	.0400000	.0400000	.0400000	.0400000
2	.01225064	.01394879	.01248585	.0137147	.01232437
3	.00585103	.00726864	.00606586	.00731614	.00596492
4	.00341252	.00453200	.00359115	.00447717	.00352732
5	.00223181	.00312412	.00238044	.00299816	.00233499
6	.00157221	.00229784	.00170415	.00213950	.00166340
7	.00116697	.00177116	.00127261	.00159994	.00124427
8	.00090059	.00140681	.00098673	.00123998	.00096912
9	.00071569	.00115026	.00079131	.00096497	.00077526
10	.00058270	.00095917	.00064480	.00080585	.00063463

TABLE 6.1.3

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 4 , q = 3 .

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.03061224	.03061224	.03061224	.03061224	.03061224
2	.00984511	.01073875	.00998136	.01065823	.00986385
3	.00478809	.00558876	.00496236	.00664504	.00486110
4	.00281581	.00346904	.00299244	.00428067	.00290954
5	.00184893	.00238631	.00201600	.00292855	.00233499
6	.00130717	.00175549	.00144888	.00211167	.00139288
7	.00097196	.00134915	.00108360	.00158799	.00104894
8	.00075124	.00107100	.00085078	.00123471	.00082049
9	.00059802	.00087330	.00067834	.00098610	.00066003
10	.00048705	.00072763	.00055530	.00080498	.00054224
p = 4 q = 3				p = 4 q = 3	

TABLE 6.1.4

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 5 , q = 4 .

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.02469136	.02469136	.02469136	.02469136	.02469136
2	.00815829	.00869997	.00825943	.00867351	.00816270
3	.00400204	.00452031	.00417181	.00632744	.00406989
4	.00236466	.00280714	.00255096	.00418539	.00245280
5	.00155774	.00192862	.00172403	.00288979	.00164797
6	.00110221	.00141612	.00124688	.00209302	.00118870
7	.00082090	.00108726	.00094772	.00157798	.00089714
8	.00063507	.00086338	.00074165	.00122889	.00070378
9	.00050555	.00070408	.00059537	.00098252	.00056743
10	.00041207	.00058670	.00048598	.00080266	.00046765
p = 5 q =4				p = 5 q = 4	

TABLE 6.1.5

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 6 , q = 5 .

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.02066116	.02066116	.02066116	.02066116	.02066116
2	.00694430	.00731077	.00701408	.00730009	.00694628
3	.00342607	.00379661	.00361815	.00616513	.00348771
4	.00202858	.00235903	.00222266	.00409799	.00211438
5	.00131016	.00162080	.00152251	.00284271	.00142656
6	.00094890	.00118675	.00110759	.00206693	.00103115
7	.00070716	.00091260	.00084501	.00156277	.00078071
8	.00054693	.00072410	.00065966	.00121956	.00061258
9	.00043576	.00058956	.00052341	.00097653	.00049535
10	.00035541	.00049058	.00043600	.00079866	.00040852
p = 6 q = 5				p = 6 q = 5	

TABLE 6.1.6

n x (Variance of estimates of population mean) for Beta distribution with parameters p =7 , q =6 .

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.01775148	.01775148	.01775148	.01775148	.01775148
2	.00603767	.00630114	.00609824	.00629781	.00603804
3	.00299195	.00327422	.00314803	.00607042	.00304894
4	.00177517	.00203195	.00197333	.00399219	.00185589
5	.00117229	.00139531	.00135947	.00278467	.00125271
6	.00083152	.00102225	.00099454	.00203548	.00090639
7	.00062005	.00078513	.00077250	.00154479	.00068737
8	.00047960	.00062274	.00060081	.00120867	.00054201
9	.00038216	.00050775	.00048094	.00096959	.00043716
10	.00031176	.00042183	.00039607	.00079405	.00036102
p = 7	q = 6		p = 7	q = 6	

307163

KASHIM IBR. HIM LIBRARY
RECEIVED
11 MAR 1988
ZARIA, NIGERIA
Ahmadu Bello University

TABLE 6.1.7

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 8, q = 7.

METHOD / NUMBER OF STRATA	CUM $\sqrt{f}$	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.0155556	.0155556	.0155556	.0155556	.0155556
2	.00533754	.00553693	.00539999	.00553562	.0533748
3	.00265276	.00287315	.0280954	.00600214*	.00270268
4	.00157666	.00178511	.00176804	.00386508	.00165074
5	.00104218	.00122487	.00123215	.0027208	.00111612
6	.00073863	.00089828	.00092695	.00200230	.00081055
7	.00055122	.00068918	.00070645	.00152611	.00061602
8	.00042679	.00054689	.00054975	.00119741	.00048464
9	.00033998	.00044561	.00044438	.00096242	.00039220
10	.00027736	.00037014	.00036137	.00078928	.00032019
P = 8 q = 7				p = 8 q = 7	

TABLE 6.1.8

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 9, q = 8 .

METHOD / NUMBER OF STRATA	CUM $\sqrt{f}$	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.01384083	.01384083	.01384083	.01384083	.01384083
2	.00178074	.00493540	.00482075	.00493706	.00478087
3	.00238015	.00256406	.00251313	.00594118	.00242795
4	.00141691	.00159049	.00159386	.00372130	.00148306
5	.00093697	.00109166	.00113633	.00265562	.00100855
6	.00066469	.00079922	.00085106	.0019695[illegible]	.00072934
7	.00049572	.00061349	.00065131	.00150760	.00055627
8	.00038389	.00048672	.0051147	.00118618	.00043871
9	.00030601	.00039614	.0004080[illegible]	.00095523	.00035530
10	.00024971	.00033035	.00033491	.00078447	.00029330
p = 9 q = 8					

TABLE 6.1.9

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 10, q = 9.

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.01246537	.01246537	.01246537	.01246537	.01246537
2	.00432860	.00445103	.00435397	.00445480	.00432622
3	.00215904	.00231012	.00230339	.00587606	.00220427
4	.00128612	.00143490	.00148940	.00356715	.00135185
5	.00085120	.00098604	.00105620	.00259540	.00091493
6	.00060387	.00072122	.00080270	.00199802	.00066512
7	.00045047	.00055336	.00061078	.00158948	.00050722
8	.00034895	.00043918	.00047917	.00117506	.00040083
9	.00027799	.00035774	.00038284	.00093805	.00032417
10	.00022690	.00029768	.00031538	.00077965	.00026761
p = 10 q = 9					

TABLE 6.1.10

n x (Variance of estimates of population mean) for Beta distribution with parameters p =11, q =10.

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.01133787	.01133787	.01133787	.01133787	.01133787
2	.00395387	.00405646	.00399237	.00405808	.00395377
3	.00197533	.00210375	.00212618	.00580800	.00201752
4	.00117663	.00130431	.00136799	.00340838	.00123642
5	.00077922	.00089584	.00098168	.00253920	.00084064
6	.00055269	.00065708	.00073808	.00190776	.00061106
7	.00041278	.00050373	.00057519	.00147169	.00046656
8	.00031970	.00039970	.00044801	.00116403	.00036869
9	.00025481	.00032615	.00036046	.00094091	.00029862
10	.00020803	.00027050	.00029664	.00077483	.00024738
p = 11 q = 10					

TABLE 6.1.11

n x (Variance of estimates of population mean) for Beta distribution with parameters p =12, q =11.

METHOD / NUMBER of STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.01039698	.01039698	.01039698	.01039698	.01039698
2	.00363869	.00372381	.00366743	.00372606	.00363856
3	.00181871	.00193561	.00193302	.00572986	.00185819
4	.00108376	.00119932	.00129022	.00321960	.00113850
5	.00071787	.00082201	.00093196	.00248880	.00077591
6	.00050997	.00060311	.00070275	.00187845	.00056389
7	.00038058	.00046206	.00053900	.00145415	.00043087
8	.00029483	.00036691	.00042208	.00115310	.00033968
9	.00023495	.00029854	.00034361	.00093381	.00027581
10	.00019182	.0023882	.00028113	.00077003	.00022793
p = 12 q = 11					

TABLE 6.1.12

n x (Variance of estimates of population mean) for Beta distribution with parameters P =13, q =12.

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.00960000	.00960000	.00960000	.00960000	.00960000
2	.00337002	.00344055	.00339126	.00344415	.00336962
3	.00168776	.00178533	.00182591	.00564366	.00172423
4	.00100550	.00110871	.00119396	.00309419	.00105728
5	.00066597	.00075915	.00087645	.00244628	.00072273
6	.00047285	.00055645	.00067570	.00184969	.00052665
7	.00035322	.00042701	.00051047	.00143681	.00040051
8	.00027353	.00033968	.00039624	.00114229	.00031517
9	.00021812	.00027652	.00032037	.00092676	.00025633
10	.00017788	.00023001	.00026588	.00076527	.00021264

p = 13 q =12

TABLE 6.1.13

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 14, q = 13.

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.00891632	.00891632	.00891632	.00891632	.00891632
2	.00313782	.00319723	.00315385	.00320182	.00313759
3	.00157337	.00166075	.00170703	.00555016	.00160582
4	.00093698	.00102932	.00113378	.00294443	.00098918
5	.00062107	.00070709	.00083476	.00240526	.00067117
6	.00044110	.00051803	.00064378	.00182109	.00049024
7	.00032944	.00039759	.00048947	.00111968	.00037411
8	.00025511	.00031561	.00038148	.00113159	.00029529
9	.00020349	.00025701	.00030737	.00091978	.00023992
10	.00016605	.00021337	.00025670	.00076054	.00019881
p = 14 q = 13					

TABLE 6.1.14

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 15, q = 14.

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.0083242	.0083242	.0083242	.0083242	.0083242
2	.00293555	.00298663	.00294759	.00299130	.00293560
3	.00147268	.00155224	.00162955	.00145045	.00150812
4	.00087710	.00096264	.00107020	.00080179	.00092839
5	.00058198	.00066124	.00079797	.00237117	.00062953
6	.00041334	.00048339	.00060511	.00179239	.00045922
7	.00030839	.00037169	.00046360	.00140281	.00035105
8	.00023912	.00029448	.00035929	.00112105	.00027760
9	.00019069	.00023985	.00029121	.00091288	.00022488
10	.00015552	.00019982	.00024193	.00075585	.00018721

p = 15 q = 14

TABLE 6.1.15

n x (Variance of estimates of population mean) for Beta distribution with parameters p =16, q =15.

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.00780437	.00780437	.00780437	.00780437	.00780437
2	.00275775	.00280313	.00276671	.00280671	.00275763
3	.00138331	.00145577	.00149279	.00138575	.0014143.
4	.00082650	.00090242	.00102085	.00256705	.00087297
5	.00054702	.00061895	.00076118	.00231137	.00059237
6	.00038857	.00045446	.00057966	.00176338	.00043228
7	.00029012	.00034851	.00044191	.00138627	.00033109
8	.00022474	.00027628	.00034806	.00111068	.00026165
9	.00017930	.00022501	.00028129	.00090605	.00021258
10	.00014627	.00018719	.00023269	.00075119	.00017594
P = 16 q = 15					

TABLE 6.1.16

n x (Variance of estimates of population mean) for Beta distribution with parameters p =17, q =15.

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.00754616	.00754616	.00754616	.00754616	.00754616
2	.00266851	.00271201	.00269505	.00285147	.00266845
3	.00134061	.00140711	.00149728	.00509825	.00137293
4	.00079981	.00087315	.00100450	.00268977	.00084493
5	.00052933	.00060046	.00075674	.00229493	.00057463
6	.00037652	.00043863	.00056680	.00175125	.00042884
7	.00028097	.00033549	.00042207	.00137722	.00032057
8	.00021768	.00026742	.00033566	.00110491	.00025284
9	.00017352	.00021762	.00027213	.00090225	.00020642
10	.00014171	.00018097	.00022766	.00074860	.00017076
p = 17 q = 15					

TABLE 6.1.17

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 18, q = 15.

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.00729217	.00729217	.00729217	.00729217	.00719217
2	.00258040	.00262135	.00262801	.00296274	.00258073
3	.00129603	.00136333	.00147352	.00176849	.00132665
4	.00077345	.00084385	.00098970	.00275781	.00081402
5	.00051217	.00057999	.00073813	.00223230	.00055713
6	.00036377	.00042456	.00054177	.00174062	.00040544
7	.00027155	.00032507	.00040263	.00136773	.00030983
8	.00021045	.00025847	.00031838	.00109883	.00023504
9	.00016790	.00021034	.00026450	.00089824	.00019902
10	.00013696	.00017497	.00022711	.00074585	.00016466

p = 18 q = 15

TABLE 6 1.18

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 19, q = 15.

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.00704399	.00704399	.00704399	.00704399	.00704399
2	.00249408	.00253420	.00253098	.00312237	.00249429
3	.00125244	.00131698	.00140671	.00439273	.00128133
4	.00074730	.00081611	.00097580	.00285558	.00079611
5	.00049541	.00056040	.00072699	.00216160	.00053764
6	.00035203	.00041029	.00052209	.00172991	.00039212
7	.00026288	.00031505	.00038550	.00135799	.00029935
8	.00020373	.00024930	.00030851	.00109247	.00023671
9	.00016234	.00020350	.00026222	.00089404	.00019271
10	.00013215	.00016924	.00021899	.00074298	.00015923

p = 19 q = 15

TABLE 6 .1.19

n x (Variance of estimates of population mean) for Beta distribution with parameters p = 20, q = 15.

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
1	.00680272	.00680272	.00680272	.00680272	.00680272
2	.00241006	.00244779	.00246079	.00331263	.00241040
3	.00121150	.00127077	.00138096	.00300161	.00121157
4	.00072299	.00078832	.00098178	.00296649	.00075301
5	.00047899	.00054189	.00072214	.00209153	.00052597
6	.00034021	.00039658	.00050831	.00171706	.00037967
7	.00025411	.00030455	.00037978	.00134827	.00029072
8	.00019066	.00024088	.00030221	.00108590	.00022842
9	.00015705	.00019656	.00025398	.00088967	.00018611
10	.00012807	.00016380	.00021301	.00073998	.00015515
p = 20 q = 15					

TABLE 6.1.20

COMPARISON OF $\frac{v(\bar{y}_{st},\ k+1)}{v(\bar{y}_{st},\ k)}$ FOR VARIOUS METHODS FOR

PARAMETERS p = 3, q = 2.

K	$(K-1)^2/K^2$	CUMMULATIVE $\sqrt{F}$	$p_i u_i$ =CONST	DURBIN'S METHOD	EQUAL RANGE METHOD	EQUAL SIZED STRATA'S METHOD
k=2	0.25	0.31	0.35	0.31	0.34	0.31
3	0.44	0.47	0.52	0.49	0.53	0.48
4	0.56	0.58	0.62	0.60	0.61	0.60
5	0.64	0.63	0.69	0.66	0.67	0.66
6	0.69	0.70	0.74	0.72	0.71	0.71
7	0.73	0.74	0.77	0.75	0.74	0.75
8	0.77	0.77	0.79	0.76	0.77	0.79
9	0.80	0.79	0.82	0.80	0.78	0.80
10	0.81	0.81	0.83	0.81	0.83	0.82

TABLE 6.1.21

n x (Variance of estimates of population mean) for Beta Distribution

q \ p	k = 6 1	2	3		4
k= 2	.01504747	.01485274	.01226085	q = 2	.00967408
3	.01041785	.01224053	.01130703	3	.00905162
4	.00749578	.00985082	.00984442	4	.00909056
5	.00561209	.00797713	.00844867	5	.00815902
6	.00433566	.00654540	.00725008	6	.00725453
7	.00345628	.00544749	.00625157	7	.00643799
8	.00281333	.00459399	.00542646	8	.00572376
9	.00233323	.00392178	.00474404	9	.00510601
10	.00196732	.00338364	.00417656	10	.00457296
25	.00039275	. --	.00109306	20	.00187060
30	.00027983			25	.00134152
				30	.00100437
				35	.00077956
				40	.00062229

TABLE 6.1.22

n x (Variance of estimates of population mean) for Beta Distribution.

p / q	k = 6 p = 15	k = 6 p = 20	k = 2 p = 20	k = 2 p = 25
2	.00021251	.00015133	.00114651	.00078221
3	.00032773	.00021252	.00156815	.00109551
4	.00038547	.00025834	.00188049	.00134278
5	.00042249	.00029211	.00211039	.00153748
6	.00044496	.00031637	.00227307	.00168820
7	.00045693	.00033332	.00238618	.00180395
8	.00046069	.00034432	.00245863	.00189021
9	.00045956	.00035096	.00250134	.00195326
10	.00045457	.00035455	.00252098	.00199628
11	.00044611			.00202385
12	.00043616			.00203960
13	.00042567			.00204397
14	.00042567			.00204397
15	.00040148			.00202959

TABLE 6.1.23

n x (Variance of estimates of population mean) for Beta Distribution.

p / q	k = 3 p = 3	k = 6 p = 3	q=3 / p	k = 2
2	.00585926	.00157544	2	.01224053
3	.00547047	.00148643	3	.01130703
4	.00478486	.00130744	4	.00985162
5	.00412141	.00112989	5	.00845940
6	.00354318	.00097251	6	.00726067
7	.00305823	.00084065	7	.00626071
8	.00265777	.00073117	8	.00543542
9	.00232475	.00063977	9	.00475193
10	.00204774	.00056416	10	.00418497
25	.00053688		25	.00109551
30	.00039559		30	.00080721
35	.00030409		35	.00061937
40	.00024054		40	.00049013

p / q	p = 5	k = 6 10
p=2	.00104377	.00044678
3	.00113098	.00056487
4	.00110446	.00062628
5	.00103337	.00065188
6	.00095041	.00065538
7	.00086625	.00064539
8	.00078766	.00062711
9	.00071590	.00060259
10	.00065123	.00057945
15	.00042222	
20	.00029186	
25	.00021286	
30	.00016142	
35	.00012670	
40	.00010216	

TABLE 6.2.1

n x (Variance of estimates of population mean) for Gamma distribution with parameters $\beta = 2$ $p = 3$

METHOD / NUMBER OF STRATA	CUM √F	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
k = 1	0.119E+02	0.112E+02	0.100E+02	0.110E+02	0.100E+02
2	3.93942581	3.71335154	3.17501473	6.30409618	3.42353796
3	1.92527035	1.87852821	1.53428352	3.15755081	1.72173779
4	1.14438484	1.14575671	0.89995088	1.82950276	1.04012199
5	0.75742531	0.77260455	0.59007433	1.3[illegible]0759	0.69759639
6	0.53118967	0.56556491	0.41587154	0.96908099	0.50503866
7	0.39736339	0.42640533	0.30870754	0.75984520	0.38038613
8	0.30754511	0.33623095	0.23805551	0.58261495	0.29558018
9	0.24498721	0.27375855	*	0.47009270	0.23907832
10	0.20135505	0.22494894	0.15394110	0.38660965	0.19565433

$\beta = 2$ $p = 3$

TABLE 6.2.2

n x (Variance of estimates of population mean) for Gamma distribution with parameters $\beta = 2$ $p = 4$

METHOD / NUMBER OF STRATA	CUM $\sqrt{f}$	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
k = 1	0.158E+02	0.149E+02	0.135E+02	0.155E+02	0.135E+02
2	5.36589590	5.02552688	4.39959081	7.80964144	4.65658141
3	2.65501823	2.55514747	2.13846434	3.90479311	2.34176242
4	1.57027850	1.55515207	1.25804586	2.51819522	1.41361531
5	1.04081988	1.05425311	0.82684426	1.78629173	0.94647044
6	0.73448537	0.76655577	0.58405705	1.32608058	0.68261652
7	0.54923474	0.57827003	0.43451449	1.01724542	0.51387980
8	0.42337040	0.45467440	0.33573443	0.80166754	0.40228251
9	0.3368760	0.36858014	0.26726696	0.64629706	0.32436201
10	0.27503437	0.30574731		0.53116855	0.26470860

$\beta = 2$ $p = 4$

TABLE 6 2.3

n x (Variance of estimates of population mean) for Gamma distribution with parameters $\beta = 2$ $p = 5$

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
k = 1	0.198E+02	0.185E+02		0.194E-02	0.171E+02
2	6.80013400	3.33447671		9.10917202	5.89317455
3	3.36300802	3.22079845		4.73788386	2.96962675
4	2.01370933	1.96442248		3.24013950	1.79605585
5	1.33272961	1.32837571		2.33611714	1.20109182
6	0.93671959	0.96603835		1.73613774	0.86685475
7	0.69765776	0.73335889		1.32797051	0.65259194
8	0.54285501	0.57632453		1.04208013	0.51015021
9	0.43076820	0.46559422		0.83770567	0.41027579
10	0.35170820	0.38453283		0.68699117	0.33649434
$\beta = 2$ $p = 5$					

TABLE 6.2.4

n × (Variance of estimates of population mean) for Gamma distribution with paramcters β = 2 p = 6

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
k = 1	0.238E+02	0.223E+02		0.234E+02	
2	8.22946828	7.53995150		0.102E+02	
3	4.09950658	3.88177035		5.73398880	
4	2.43748166	2.36738331		4.10465081	
5	1.6093375	1.50615308		2.96195235	
6	1.14070064	1.16451625		2.18370604	
7	0.85377765	0.88614619		1.66040353	
8	0.65794308	0.59395678		1.29976393	
9	0.52507909	0.56219136		1.04297593	
10	0.42091769	0.45291955		0.85441717	
	β = 2	p = 6			

TABLE 5.2.5

n x (Variance of estimates of population mean) for Gamma distribution with parameters $\beta = 2$ $p = 7$

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
k = 1	0.277E+02	0.260E+02		0.273E+02	0.242E+02
2	9.66216330	8.94674055		0.113E+02	8.35771184
3	4.81337965	4.53654843		6.91486433	4.21474526
4	2.87363231	2.77675891		5.07389932	2.54951076
5	1.93036032	1.87495917		3.61960543	1.71463401
6	1.34512352	1.36145700		2.64626651	1.23017637
7	1.00786299	1.03105391		1.62881868	0.93093348
8	0.77897551	0.81182595		1.18997910	0.72778811
9	0.61720296	0.55521680			0.58525277
10	0.50548203	0.54153934			0.48014788
	$\beta = 2$	$p = 7$			

TABLE 6.2.6

n x (Variance of estimates of population mean) for Gamma distribution with parameters $\beta = 2$ $p = 8$

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
k = 1	0.317E+02	0.298E+02		0.278E+02	0.278E+02
2	0.110E+02	0.102E+02		9.37598638	9.59803201
3	5.5328692	6.21814554		6.71274783	4.84113998
4	3.31650318	3.18825175		4.35522985	2.92924541
5	2.18885055	2.15246880		2.94720098	1.96893262
6	1.55421854	1.52148026		2.11633928	1.41070394
7	1.15461980	1.18951185		1.58959751	1.06703887
8	0.89774137	0.93499254		1.23553846	0.83413552
9	0.71675958	0.75290814		0.98667546	0.67182948
10	0.58111759	0.52202019		0.80542465	0.55260364
$\beta = 2$ $p = 8$					

TABLE 6.2.7

n x (Variance of estimates of population mean) for Gamma distribution with parameters $\beta = 2$ $p = 9$

METHOD / NUMBER OF STRATA	CUM $\sqrt{f}$	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
k = 1	0.357E+02	0.335E+02		0.314E+02	0.314E+02
2	0.125E+02	0.115E+02		0.107E+02	0.108E+02
3	6.25656042	5.88061480		8.08521837	5.48169831
4	3.73044243	3.58418063		5.15172752	3.32782669
5	2.47843759	2.42785072		3.48061520	2.22883210
6	1.75719798	1.75805776		2.50394897	1.60581553
7	1.30984275	1.33435897		1.88297461	1.21320136
8	1.01655769	1.02073733		1.46446163	0.95043600
9	0.80992937	0.84903992		1.16987220	0.76325804
10	0.65784411	0.70086716		0.95515664	0.62613509
$\theta = 2$ $p = 9$					

TABLE 6.2.8

n < (Variance or estimates of population mean) for

Gamma distribution with parameters $\beta = 2$ $p = 10$

METHOD / NUMBER OF STRATA	CUM √f	EQUALIZATION OF STRATUM TOTAL	DURBIN	EQUAL RANGE	EQUAL SIZED STRATA
k = 1	0.397E+02	0.371E+02		0.350E+02	0.350E+02
2	0.139E+02	0.128E+02		0.123E+02	0.121E+02
3	6.96017792	5.53572878		9.51807970	5.12692853
4	4.17865642	3.99511689		5.95516111	3.59844802
5	2.76420680	2.70637902		4.03134052	2.48979423
6	1.95769342	1.95733414		2.90745714	1.79221612
7	1.48436042	1.48503846		2.18875009	1.34971410
8	1.13005710	1.16654805		1.70307137	1.05654840
9	0.90453509	0.94405015		1.36083262	0.85116540
10	0.73430445	0.78057957		1.11125583	0.69936509
	$\beta = 2$	$p = 10$			

TABLE 6.2.9

Comparison of Stratification Boundary Points

Gamma Distribution $\beta = 2$, $p = 0.5$

$\chi^2(1)$

	OPS	2.7				
k = 2	SETHI	2.8				
	BELLO	1.2	3.2			
		1.6	4.0			
k = 3		1.6	4.4			
		0.6	2.4.	8.2		
		1.1	2.7	5.2		
k = 4		1.1	2.8	5.5		
		0.2	1.2	3.0	8.2	
		0.8	1.9	3.6	6.0	
k = 5		0.8	2.0	3.6	6.4	
		0.2	0.8	1.8	3.6	8.2

TABLE 6.2.10

Comparison of Stratification Boundary Points
Gamma Distribution $\beta = 2$, $\rho = 6$

$\chi^2(12)$

	DPS	12.6				
k = 2	Sethi	12.7				
	Bello	12.4	36.8			
		10.0	15.5			
k = 3		10.0	16.0			
		9.8	15.6	36.8		
		8.6	12.6	17.5		
k = 4		8.5	12.7	18.1		
		8.2	12.4	17.6	36.8	
		7.8	11.0	14.5	19.0	
k = 5		7.6	11.0	14.5	19.5	
		7.4	10.8	14.2	19.2	36.8

TABLE 6.2.11

Comparison of Stratification Boundary Points
Gamma Distribution $\beta = 2$, $p = .15$

χ^2 (30)

	OPS	30.6				
k = 2	Sethi	30.7				
	Bello	30.0	59.2			
		26.0	35.0			
k = 3		26.0	36.0			
		25.6	35.0	59.2		
		24.0	30.6	38.0		
k = 4		23.6	30.7	38.7		
		23.2	30.0	37.0	59.2	
		22.0	28.0	33.0	40.0	
k = 5		22.0	26.0	34.0	41.0	
		21.6	27.4	32.0	40.0	59.2

TABLE 6.3. COMPARISON OF STRATIFICATION POINTS AND VARIANCE OF ESTIMATE OF POPULATION MEAN FOR PARETO DISTRIBUTION WITH PARAMETER $\theta = 3$

Number of Strata	BOUNDARY POINTS				n* (VARIANCE OF ESTIMATE OF MEAN OF POPULATION) X 10^5			
	Cum √f	$P_h\mu_h = c$	$P_h = c$	Equal Range	Cum √f	$P_h\mu_h = c$	$P_h = c$	Equal Range
1	1500	1500	1500	1500	15.8	15.8	15.8	15.8
	100000	100000	100000	100000				
2	1500	1500	1500	1500	3.98	5.38	7.1	14.9
	3000	2121.32	1889.88	50750				
	100000	100000	100000	100000				
3	1500	1500	1500	1500	1.76	2.79	4.36	14.1
	2250	1837.11	1717.07	34333.33				
	4500	2598.08	2163.37	67166.67				
	100000	100000	100000	100000				
4	1500	1500	1500	1500	0.994	1.74	3.07	13.3
	2000	1732.05	1650.96	26125				
	3000	2121.32	1889.88	50750				
	6000	3600	2381.10	75375				
	100000	100000	100000	100000				
5	1500	1500	1500	1500	0.636	1.20	2.32	12.6
	1875	1677.05	1615.83	21200				
	2500	1936.49	1778.48	60600				
	7500	3351.10	2564.96	80300				
	100000	100000	100000	100000				

	1500	1500	1500	1500				
	1800	1643.17	1593.99	17916.67				
	2250	1837.12	1717.07	34333.33				
6	3000	2121.32	1889.88	50750	0.442	0.884	1.85	11.9
	4500	2598.08	2163.37	67166.67				
	9000	3674.23	2725.68	83583.33				
	100000	100000	100000	100000				
	1500	1500	1500	1500				
	1750	1620.19	1579.08	15571.43				
	2100	1774.83	1678.08	29642.86				
	2625	1984.31	1807.61	43764.29				
7	3500	2291.29	1989.53	57785.71	0.325	0.681	1.52	11.3
	5250	2306.24	2277.44	71857.14				
	10500	3968.63	2869.40	85928.57				
	100000	100000	100000	100000				
	1500	1500	1500	1500				
	1714.29	1603.57	1562.27	13812.5				
	2000	1732.05	1650.96	26125.0				
	2400	1897.37	1754.31	38437.50				
	3000	2121.32	1889.88	50750				
8	4000	2449.49	2080.08	63062.5				
	6000	3000	2381.10	75375	0.249	0.542	1.29	10.7
	12000	4242.64	3000	87687.5				
	100000	100000	100000	100000				

TABLE 6.3. FOR PARETO DISTRIBUTION CONT'D

	1500	1500	1500	1500				
	1687.5	1590.99	1560.06	12444.44				
	1928.57	1700.84	1631.07	23388.89				
	2250	1837.12	1717.07	34333.33				
	2700	2012.46	1824.66	45277.78				
9	3375	2250	1965.56	56222.22	0.196	0.444	1.11	10.1
	4500	2508.08	2163.37	67166.67				
	6750	3131.98	2476.45	78111.11				
	13500	4500	3120.13	89055.56				
	100000	100000	100000	100000				
	1500	1500	1500	1500				
	1666.67	1581.14	1553.62	11350				
	1875	1677.05	1615.83	21200				
	2142.86	1792.84	1689.37	31050				
	2500	1936.49	1778.45	40900				
	3000	2121.32	1889.88	50750	0.159	0.359	0.971	9.68
	3750	2371.71	2035.81	60600				
	5000	2738.61	2240.70	70350				
	7500	3654.10	2564.96	80300				
	15000	4743.42	3231.65	90150				
	100000	100000	100000	100000				

TABLE 6.3.1

n x (VARIANCE OF ESTIMATE OF POPULATION MEAN) FOR PARETO DISTRIBUTION WITH PARAMETER $\theta = 2.5$

NUMBER OF STRATA	Cum $\sqrt{f}$	$P_h \mu_h \times 10^5 = c$	$P_h = c$	Equal Range
1	36.4	46.4	36.4	36.4
2	8.27	11.5	16.7	31.3
3	3.58	5.77	10.4	27.6
4	2.00	3.50	7.42	24.8
5	1.27	2.36	5.68	22.5
6	0.887	1.70	4.55	20.5
7	0.653	1.29	3.77	18.8
8	0.501	1.01	3.20	17.4
9	0.397	0.821	2.76	16.0
10	0.323	0.68	2.43	14.9

CHAPTER SEVEN

SUMMARY AND CONCLUSIONS

Five of the six methods reviewed in this study are used to determine the stratum boundaries for the Beta distribution with parameters p and q. All these five methods construct the stratum boundaries quickly. The computer time for each of the methods is not the same. The "equal ranges method" takes less computing time compared with other methods. The "equal ranges method" is closely followed by "cum $\sqrt{f}$ method". The "equal sized strata method" and Durbin's method take more time compared with others.

For k strata given by each method the values of $n\,V(\bar{y}_{st}) = \{\sum_{h=1}^{k} p_h \sigma_h\}^2$, were computed for k = 2,3,4,...,10. The best boundaries are those that give the smallest variance $V(\bar{y}_{st})$ for the estimated population mean for a stratified random sample of size n with optimum choices of n_h in the individual strata. The values of $n\,V(\bar{y}_{st})$ for Beta distribution with parameters p and q appear in Tables 6.1.(2) to 6.1.(18). It is clear from the tables that the "cum $\sqrt{f}$ method" ranked first for all the values of p and q considered in this study.

This is closely followed by the "equal sized strata method", but this is the method which Kish (1965) had dismissed on intuitive grounds. As postulated by him the "equal range method" ranked last except when

(i) $p = 3$, $q = 2$, and $k \neq 3$

(ii) $p = 4$, $q = 3$, $k = 2$, $k = 5$.

It is expected that $V(\bar{y}_{st})$ should decrease as k increases. This is generally true. Exceptions occur in "equal ranges methods" when $p \geq 8$, $q \geq 7$ where

$$V(\bar{y}_{st}, k = 3) > V(\bar{y}_{st}, k = 2)$$

That is increasing k does not always lead to a reduction in variance of the estimate of the mean of the population.

As pointed out in Chapter two on the review of literature on the determination of optimum points of stratification

$$f(y) = 2(1-y), \quad 0 < y < 1$$

is a special case of the Beta distribution with $p = 1$, $q = 2$. The results obtained here agreed very well with those obtained by Kokan (1965b). See Table 5.1(1) for the "equal ranges method", but for the "equal sized strata method" $V(\bar{y}_{st})$ obtained in this study is slightly less than that obtained by him. The same table 5.1.(1) shows that the "cum $\sqrt{f}$ method" and Durbin's method agree very closely for $2 \leq k \leq 5$. And for $k = 2$, $V(\bar{y}_{st})$ is less than the result quoted by Kokan (1965b).

It is also note worthy that the Durbin's method ranked third consistently well for all values of p and q considered in this work. Cochran (1961) in his study of eight empricel population had reported that the cum $\sqrt{f}$

method and Ekman's method performed consistently well. This study has also shown that the cum√f method is the best. The equalization of stratum total method is consistent in this study contrary to the findings Cochran (1961) who found it erractic. We recommend that the cum √f method be used whenever stratification points are needed for a population that resembles the Beta distribution.

For very high values of p and q, $p > 100$ $q > 100$ the computer algorithms designed for the methods do not construct the strata boundaries. This is an area in which further research can be done.

Dalenius (1957) had earlier proposed this relationship. For all the methods considered, there is very littlr reduction in $V(\bar{y}_{st},k)$ when $k > 6$. We therefore recommend that k need not be greater than 6.

It is observed that $V(\bar{y}_{st},k)$ is inversely proportional to k^2. Examination of Table 6.1.21 shows that for $k \geq 6$

$$\frac{V(\bar{y}_{st}, k)}{V(\bar{y}_{st}, k-1)}$$

agrees very closely with the proposition that

$$V(\bar{y}_{st},k) \propto \frac{1}{k^2}.$$

A critical examination of Tables 5.1.22, 5.1.23, 5.1.24 will show that for fixed p in the range $1 \leq p \leq 4$ and for fixed k

$$\frac{V(\bar{y}_{st}, q)}{V(\bar{y}_{st},q+1)} = 0.9, \quad q \geq 1.$$

Also in this study we observe that for fixed q in the range $1 \leq q \leq 4$ and for fixed k

$$\frac{V(\bar{y}_{st}, p)}{V(\bar{y}_{st}, p+1)} \doteq 0.9, \quad p \geq 1$$

We therefore propose the following relationship

(i) $V(\bar{y}_{st}, q) = 0.9^s V(\bar{y}_{st}, q+s), q,s = 1,2,3,\ldots,$
$1 \leq p \leq 4$, p and k are fixed

(ii) $V(\bar{y}_{st}, p) = 0.9^s V(\bar{y}_{st}, p+s), p,s = 1,2,3,\ldots,$
$1 \leq q \leq 4$, q and k are fixed.

But for fixed p in the range $5 \leq p \leq 100$ and fixed k, $V(\bar{y}_{st})$ first increases and then decreases as q is increased. Similarly for fixed q in the range $5 \leq q \leq 100$ and for fixed k, $V(\bar{y}_{st})$ first increases and then decreases as p is increased. We also observe that beyond the turning point for the variance the relationships (i) and (ii) still hold as q or p is increased.

Considering the efficiences of the methods we recommend that the cum $\sqrt{f}$ method be used in determining stratification boundary points for population that resembles Beta distribution.

A lot of computer time is required for constructing stratification boundary points for Gamma distribution. The computer time for each of the methods is not the same. The cum $\sqrt{f}$ method takes less computing time than the other methods. The worst in terms of computing time is the Durbin's Method.

Since the Gamma distribution has infinite range, the density function is truncated. To delete the same proportion at the extreme is not possible as this depends on the choice of σ. But within 1% allowance for σ we can compare the efficiences of the methods.

For k strata given by each method the values of

$$n\, V(\bar{y}_{st}) = \left(\sum_{h=1}^{k} P_h \sigma_h\right)^2$$

were computed for $k = 2,3,\ldots,10$. The tables of $n\, V(\bar{y}_{st})$ for various parameters appear in Tables 6.2.1 to 6.2.8. Tables 6.2.9, 6.2.10, 6.2.11 show stratification points for $\chi^2(1)$, $\chi^2(12)$ and $\chi^2(30)$ respectively.

Chi-square distribution is a special case of Gamma distribution. The tables show a very close agreement between Sethi's result, and those of the author for $\chi^2(12)$ and $\chi^2(30)$, but a wide disagreement for $\chi^2(1)$.

The discrepancy for $\chi^2(1)$ may be due to approximation used by Sethi. The rule used by Sethi was that if the optimum points of stratification for the normal distribution for optimum allocation are $p_1, p_2, p_3, \ldots, p_{k-1}$ percent points of the normal distribution then the optimum stratification points for $\chi^2(n)$ for optimum allocation are closely approximated by $p_1, p_2, \ldots, p_{k-1}$ percent points of $\chi^2(n + 1.3)$ distribution. For large values of n both Sethi's result and that of the author agree.

A critical examination of the behavior of the variance as a function of the parameters of the Gamma distribution with parameters β and p show that the variance increases as β increases while p and k are fixed. And also the variance increases as p increases while β and k are fixed. Further study of the behavior of the variance as a function of one of the parameters is advocated.

Only the cum $\sqrt{f}$ method constructs the strata boundaries for Gamma distribution quickly. It is also the most efficient. We recommend that cum $\sqrt{f}$ method be used for determination of stratification points for population that resembles that of Gamma distribution.

We also observed that the variance n $V(\bar{y}_{st})$ is inversely proportional to k^2 for $k \geq 6$.

$$\frac{V(\bar{y}_{st}), k)}{V(\bar{y}_{st}, k-1)}$$

agrees very closely well with the proportion that $V(\bar{y}_{st}, k)$ is inversely proportional to k^2.

For Pareto distribution, all the methods considered construct the boundary points quickly. For a given subsistence wage and appropriate parameter θ, each of the methods leads to a very substantial reduction in variance of the estimate of the mean of population except the equal range method. Even when the number of strata is 2, the reduction in variance of the estimate of the mean of population is more than 75% for the cum $\sqrt{f}$ method. The cum $\sqrt{f}$ method ranked first, equalization of stratum total method, equal sized strata method, and equal range method followed in that order.

R E F E R E N C E S

1. Abramowitz M., Stegun I.A. (1968). Hand book of Mathematical functions. Dover Publications. New York.

2. Bury K.V. (1975). Statistical Models in Applied Science. John Willey Series.

3. Block E., (1958). Numerical Consideration for the Stratifications of variables following Logarithmic Normal Distribution. Skandinavisk Aktuarietidskrift. 1. 185 - 200.

4. Chatterjee S., (1968). Multivariate Stratified Surveys. Journal of American Statistical Association. Vol. 63. 530 - 534.

5. Cochran W.G., (1970). Sampling Techniques. 3rd Edition. John Willey and Sons Inc. New York.

6. Cochran W.G. (1961). Comparison of Methods for determining Stratum boundaries. Bulletin of the International Statistical Institute 38 (2) Tokyo. 345 - 358.

7. Dalenius T., (1950). The problem of Optimum Stratification. Skandinavisk Aktuarietidskrift Vol. 33. 203 - 213.

8. Dalenius T., Gurney M., (1951). The problem of Optimum Stratification II. Skand. Aktuarietidskrift. Vol. 34. 133 - 148.

9. Dalenius T., Hodges J.L., (1957). The Choice of Stratification Points. Skand. Aktuarietidskrift. Vol. 40. 198 - 203.

10. Dalenius T., Hodges J.L. Jr., (1959). Minimum variance stratification. Journal of the American Statistical Association Vol. 54 88 - 101.

11. Dallas W.A., Kish L., Cornell G.R., (1975). Implications of Optimum and Approximately Optimum Stratification. Proceedings of the American Statistical Association.

12. Durbin J., (1959). Review of Sampling in Sweden. Journal of Royal Statistics. A 122. pp. 246 - 248.

13. Ekman G., (1959). An Approximation Useful in Univariate Stratatification. Annuals of Mathematical Statistics Vol. 30. 219 - 229.

14. Elandt-Johnson R.C., Normal L. Johnson., (1980). Survival Models and Data Analysis. John Willey Series.

15. Ghosh P., (1963) Optimum Stratification with two characters. The Annuals of Mathematical Statistics Vol. 34. 866 - 872.

16. Hansen M.H., Hurwitz, W.G., Madow W.G., (1953). Sample Survey Method and Theory. Vol. I & II Willey Series. New York.

17. Hayash C., Maruyawa F., Ishida M.D., (1950). On some Criteria for Stratification. Institute of Mathematical Statistics. 77 - 86.

18. Hess I., Sethi V.K., Balakrishan T.R., (1966). Stratification: A Practical Investigation. Journal of American Statistical Association. Vol. 61. No. 313. 74 - 90.

19. Herlerkar R., (1967). The problem of Optimum Stratification II. Skand. Aktuarietidskrift. 183 - 202.

20. Issi K., Taga Y., (1969). On Optimum Stratification for Multivariate Distribution. Skand. Aktuarietidskrift. 24 - 38.

21. Jarque C.M., (1981). A solution to the problem of optimum stratification in Multivariate Sampling. Journal of the Royal Statistical Society (Series C). Vol. 30 No.2 164 - 169.

22. Jolayemi E.T., (1982). A C_p Method to select a Loglinear Model. Unpublished Ph.D. thesis. The University of Michigan. U.S.A.

23. Kish L., (1965). Survey Sampling. John Willey and Sons Inc. New York.

24. Kish L., Dallas W.A., (1978). Multivariate and Multipurpose Stratification. Journal of American Statistical Association 73. 24 - 34.

25. Kitagawa T., (1955). Some contributions to the design of sample Surveys. Sankhya 14. 317 - 362.

26. Kokan A.R., (1963). Optimum allocation in Multivariate Surveys. Journal of Royal Statistical Society. A 126. 557 - 565.

27 Kokan A.R., (1965a). Contribution to theory and Practive of Sampling Connected with Multivariate Surveys and non-response. Unpublished Ph.D. thesis.

28. Kokan A.R., (1965b). On two Methods of Constructing Strata. Journal of the India Statistical Association. Vol. 3. 9 - 16.

29. Konijn H.S., (1973). Statistical theory of sample design and analysis. American Elsevier Publishing Company. New York.

30. Kpedekpo G.M.K., 1973). Recent Advances on some Aspects of stratified Sample Design. A review of the Literature Metrika. 20 54 - 64.

31. Lanke J., (1975). Some contributions to the theory of Survey Sampling. Unpublished Ph.D. Thesis. Lund.

32. Lawless J.F., (1982). Statistical Models and Methods for Life time Data. John Willey and Sons. New York.

33. Mahalanobis P.C. (1952). Some aspects of design of Sample Surveys. Sankhya 12 1 - 7.

34. Murthy M.N., (1967). Sampling theory and Methods. Statistical Publishing Society. Calcutta.

35. NAG Fortran Library. Mark 9. The Numerical Algorithms Group. U.K.

36. Neyman J., (1934). On the two different aspects of the representative methods: The method of Stratified Sampling and the method of purposive selection. Journal of Royal Statistical Society. 97. 558 - 625.

37. Nilson G., (1967). Optimum Stratification. According to the Method of Least Squares. Skand. Aktuarietidskrift. 128 - 136.

38. Pearson K., (1965). Tables of Incomplete Gamma Function. Cambridge University Press. London.

39/ Raj. D., (1964). On forming strata of equal aggregate size. Journal of American Statistical Association. Vol. 59. 481 - 485.

40. Raj. D., (1968). Sampling Theory. McGraw-Hill Book Company. New York.

41. Sethi V.K., (1963). A note on Optimum Stratification of population for Estimating the population Mean. The Australian Journal of Statistics. Vol. 5. 20 - 33.

42. Sethi V.K., (1960). On the possibility of improving upon the principle of equalization of strata total. Journal of Social Science, Agra University, Institute of Social Sciences Vol. II. 47 - 51.

43. Serfling R.J., (1968). Approximately Optimal Stratification. Journal of the American Statistical Association. Vol. 63. 1298 - 1309.

44. Singh R., Sukhatme B.V., (1969). Optimum Stratification. Annuals of the Institute of Statistical Mathematics Vol. 21. 515 - 528.

45. Sukhatame P.V., Sukhatme B.V., (1970). Sampling theory of Surveys with Applications. 2nd Edition. Food and Agricultural Organization. Rome.

46. Taga Y., (1971). On convergence of Optimum Stratification for empiric distribution function in univariate case. Institute of Statistical Mathematics. Vol. 23., 355 - 363.

47. Tschuprow A.A., (1923). On the mathematical expectation of moments of frequency distributions in the case of correlated observations. Metron, 2. 461 - 493, 646 - 683.

48. Yule, G.U., Kendal M.G., (1937). An Introduction to the theory of Statistics. Griffin and Company Limited.