OPTIMIZATION OF PREVENTATIVE SAMPLING AND STRATIFIED SAMPLING

by 1264

THIRUVENGADAM JANAKIRAMAN

B.E. (Hons) (Mechanical), Madras University

Madras, India, 1961

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1970

Approved by

Chung-Lai Hwang

Major Professor

TABLE OF CONTENTS

1. INTRODUCTION

Quality is of utmost importance in any sophisticated system. It is more so in the case of life support systems whose reliability, which is vitally important, depends on the quality of the components that go to make the system. Modern technology has developed to ensure high precision and quality in the manufacture of products. Due to increased specialization of technology of production, it is not economical for any industry, however large, to manufacture all the parts needed for the final product. This has necessitated obtaining certain parts from outside suppliers. Also, even for the parts made within the industry, raw materials are received from outside sources. Since the quality of the final product, to a very great extent, will depend on that of the parts received from the supplier as well as the raw material received, a very strict vigilance on the quality of incoming goods is warranted. Sampling inspection at this stage could be stated as the starting point of a quality control program in the whole project.

An attempt has been made to design an optimal sampling plan with the total cost of inspection as the main consideration. The total cost of inspection includes two elements, namely, the cost of inspection and the cost of undetected defective articles going out of the inspection department. The process average or the probability of defect of the incoming lot is an important aspect in optimization of sampling. In general, this process average is taken as fixed and the sampling scheme is calculated to give the optimal results for that fixed process average (3). The approach used in this report is to consider the

process average as being largely caused by the sampling scheme adopted (3,6).

Hwang, Fan and Tillman [4] have formulated such a sampling problem and solved a five variable problem by the application of the discrete version of Pontryagin's maximum principle. In this report the same numerical example of allocating a resource for sampling inspection among five products has been formulated and solved using the sequential unconstrained minimization technique [SUMT]. Both these methods yield identical optimal sample sizes.

The stratified sampling, which has the advantage of higher precision has been discussed in the next section. A numerical problem with four strata and two variates taken from Cochran [2] has been formulated in the form of SUMT general formulation. The problem is to find the sampling plan with minimum cost where the variances for the two variates are equal to or less than the specified value. Assuming the cost of inspection to be linearly related to the sample size, the above mentioned problem is one of minimizing a linear objective function subjected to nonlinear constraints. The problem is solved using SUMT and the results are presented in section 4 of this report. Assuming a nonlinear cost function, which is more realistic in some cases, the problem becomes one of minimizing a nonlinear objective function subject to nonlinear constraints. The results of such a problem obtained by using SUMT has also been presented.

Finally, parameter estimation has been carried out on the model which is used in preventative sampling where the probability of defect

is assumed to be a function of the fraction of sampling. Marquardt's method [5], one of the most powerful techniques available for the nonlinear parametric estimation is used to estimate the three parameters involved using a set of assumed data points. Even with the initial guesses which are 100% more or less of the correct value, the method estimates the parameters to an accuracy of about $\pm$ 3%. The same data points were used for the parametric estimation by Bard's method [1] for a comparitive study. It was found that Marquardt's method takes less computation time than Bard's method.

1. Bard, Y., "Nonlinear Parameter Estimation and Programming". IBM Publication. (1967).

2. Cochran, W. G., Sampling Techniques, 2nd Ed., Wiley, New York, 1963.

3. Horsnell, G., "Economical Acceptance Sampling Plans:, Journal of Royal Statistical Society, Vol. A-120, pp. 148-175 (1957).

4. Hwang, C. L., L. T. Fan, and F. A. Tillman, "Optimal Preventative Sampling", Kansas State University, Unpublished Lecture Notes.

5. Marquardt, D. W., "Least Square Estimation of Nonlinear Parameters", A computer program in Fortran IV Language, IBM Share Library, Distribution No. 309401 (1966).

6. Whittle, P., "Optimum Preventative Sampling", Journal of Operations Research Society of America, Vol. 2, pp. 197-203 (1954).

2. ACCEPTANCE SAMPLING

2.1 INTRODUCTION

One of the problems faced by industry is the design of a proper acceptance sampling procedure. Since the quality of the final product will depend on the quality of the parts and raw materials received, it is very important to have a proper acceptance sample plan to get the desired quality level of input.

One of the problems in a sampling inspection procedure is to decide the sample size. The problem can be solved in various ways. The approaches are as follows [5]:

(1) By applying one of the standard sampling inspection tables such as the Dodge and Romig Tables [4], the Military Standard 105-D [9], or the Philips SSS Tables [6].

(2) By choosing numerical values for a suitable set of parameters (AOQL, Producer's or Consumer's Risk point, etc), and constructing a corresponding sampling plan,

(3) On the basis of an economic theory which takes into consideration various costs [7,8,10].

Methods (1) and (2) are more commonly used due to the simplicity of calculation. Though these two methods are convenient for use in industry, one can never be certain that they will lead to what must be considered as an optimum sample size. They may in certain cases lead to sample sizes which are not the best from cost point of view [5]. The third method based on the economic theory has much less doubt as to the optimality of the decision. The main draw back with the cost

basis is that it depends on various costs such as cost of accepting a defective sample etc. which cannot be estimated with reasonable degree of precision. A very good approach to estimate the cost of accepting defective material based on the theory of probability has been given by Stacy et al [10].

2.2 PLANS FOR ACCEPTANCE SAMPLING BY ATTRIBUTES

Whenever there is material transactions between the producer and customer or two departments within a single organization, there is always the problem of acceptance of goods. Sampling inspection is preferred to 100 per cent inspection for two reasons. First it saves money since only a fraction of the lot is to be tested and second, in cases which involve destructive testing, it is absolutely necessary. Since acceptance sampling is desirable and often necessary, we ought to make sure that we choose the right plan for our purposes and that we use it properly.

For a sampling plan, we decide the sample size n which is to be choosen without bias from a lot size N. Then the sample is inspected and, if it contains say c or fewer defectives, the lot is accepted, while if there are more than c defectives in the sample, the lot is rejected. Deciding the numbers n and c constitute the sampling plan choosen. This method is known as single sampling.

The desire to give a lot a second chance leads to double sampling. The main advantage of this method is that the average number of articles inspected is much less than the single sampling method. A sample of n_1 pieces is chosen at random. Then if it contains c_1 defectives or fewer, the lot is accepted, or if it contains more than c_2 defectives,

the lot is rejected. If, however, the first sample contains more than c_1 defectives but not more than c_2 defectives, a second sample of n_2 pieces is drawn. Then the total number of defectives in the two samples is compared with c_2. If the total is equal to or less than c_2, then the lot is accepted; otherwise it is rejected.

The development of various standard acceptance sampling plans was and is a group project as in the case of many other industrial and scientific achievements. The early work was largely concentrated in the Bell Telephone Laboratories, beginning about 1923 [3]. The famous Dodge-Romig Tables were due to work done at Bell Telephone Laboratories [4].

Early in 1945, the statistical research group of Columbia University was asked by the United States Navy to prepare a manual of tables, procedures and principles for sampling inspection. Subsequently the attribute tables were adopted for use by both the United States Army and United Stated Navy [5].

In general it is best to use one of the sampling plans already worked out. This is the most convenient, since it saves the industrial worker a great amount of calculation, trouble and time.

The various standard plans provide adequate flexibility and variety of approach to take care of most of the industrial sampling problems. However, if a particular situation warrents, it is possible to design a specific plan to suit to the specific requirement [3].

Many authors have discussed the use of optimization in the decision of sampling size [1,5,7,10]. The basic principle involved is the minimization of the total cost of inspection. The first cost to be

considered is the cost of inspecting the sample. This is calculated by multiplying the sample size by the unit cost of inspection. The other element of the cost arises due to the fact that in any sampling plan there is always a possibility of accepting bad articles. A cost of accepting a defective article is assumed and the cost of accepting bad articles is calculated. The objective is to minimize this total cost subject to certain conditions like total number of articles etc.

The quality of the incoming material, the probability of defect of the lot, is an important factor that has to be considered in the design of the sampling plan. Most of the authors assume a constant probability of defect and, based on this, they have proceeded to the optimization of sampling size. The next chapter presents a different approach for the optimization of sampling size. This approach assumes that the probability of defects is influenced by the sample size choosen for inspection.

REFERENCES

1. Barter, K. E., "Sample Sizes for an Acceptance Number of Zero", Industrial Quality Control, pp. 322-324 (1967).

2. Brenner, M. E., "A cost model for determining the sample size in the simulation of Inventory Systems", Journal of Industrial Engineering, Vol. 17, pp. 141-144 (1966).

3. Burr, I. W., Engineering Statistics and Quality Control, McGraw-Hill, New York, 1953.

4. Dodge, H. F., and H. G. Romig, Sampling Inspection Tables, John Wiley, New York, 1944.

5. Hamaker, H. C., "Some Basic Principles of Sampling Inspection by Attributes", Journal of Royal Society of Statistics, Vol. 7, No. 3, pp. 149-159 (1958).

6. Hamaker, H. C., C. J. J. M. Taudin and F. G. Willemze, "The Practical Application of Sampling Inspection Plans and Tables", Philips Technical Review, Vol 11, pp. 362-372 (1950).

7. Horsnell, G., "Economical Acceptance Sampling Plans", Journal of Royal Statistical Society, Vol. A-120, pp. 148-175 (1957).

8. Martin, C. A., "The Cost of Breakeven Point in Attribute Sampling", Industrial Quality Control, pp. 137-143 (1964).

9. Military Standard 105A, Sampling Procedures and Tables for Inspection by attributes, US Government Printing Office, 1950.

10. Stacy, E. W., L. E. Hunsinger, and J. E. Price, "Determining Economic Sampling Plans", IBM SYST J. No. 3 (1969).

3. PREVENTATIVE SAMPLING

3.1 INTRODUCTION:

The two basic purposes of sampling inspection are: (1) to check the quality level of the goods and (2) to help to maintain a certain quality level in the goods. The former function is quite apparent and the latter needs to be explained. The second purpose, though unfortunately not very commonly recognized, justifies the establishment of sampling rather than the first one. Maintenance of a quality level is achieved since the very act of inspection produces a reaction which improves the quality of the population [24]. In acceptance sampling if the customer does not inspect the products received, the supplier is likely to underestimate the importance of the quality and he may relax his effort to maintain the quality at the best possible level [13]. Thus a routine inspection procedure, more than serving the purpose of discrimination of good lots from the bad, exerts a decided influence on the quality of the lots received and may be desirable for this reason alone [14]. The mere knowledge that a regular inspection is being carried out encourages a sense of responsibility and makes the producer more careful. It is natural that this reaction increases as the sampling is intensified, and thus a more rigorous inspection results in a smaller probability of defect [24]. The above advantage will be more effective if the supplier is fully aware of the inspection done by the customer [16]. Thus an effective feedback information from inspector to producer or supplier is essential for the success of this sampling inspection.

Most of the authors [17,23,24] who have done work in the area of optimization of the sample size, have based their approach on a constant process average. As expressed by Hill in the discussion of a paper by Horsnell [14], the influence of the sampling scheme on the process curve is not taken into consideration when one assumes a constant process curve. In this light, selecting a sample size is mainly a task of deciding what plan will give the vendor an economic incentive to offer the desired quality. As for practical examples, we can think of random sampling of tax returns, or checking the drivers' license of teenagers. Obviously one hundred percent checking in these situations is cost-prohibitive, whereas no inspection will lead to carelessness and dishonesty. Hence random sampling is the only answer. Considering this human reaction mentioned above which results in the prevention of future defects, we may refer to this as the preventative sampling.

Based on the above considerations an attempt has been made to allocate an effort or resource to obtain an optimal preventative sampling. The problem has been formulated in the form of a non-linear programming problem. The sequential unconstrained minimization technique (SUMT), one of the most powerful nonlinear programming, has been used to solve a numerical example of resource allocation among five products. The results obtained have been presented.

3.2 FORMULATION OF THE PROBLEM:

Let us assume that a manufacturing company is producing N types of products, each being different with respect to the value of the

product, probability of being defective, quantity of production of each type and so on. We are to find the optimum preventative sampling procedure subject to some constraints. To decide the criterion for optimality, let us consider the various costs involved.

Let a^n = the quantity of the n^{th} type product.*

θ^n = the sampling fraction of the n^{th} type.

v^n = the value of each of the n^{th} product.

$p^n(\theta^n)$ = the probability of a^n being defective, a function of θ^n.

I^n = the inspection cost of each article at the n^{th} stage.

$p^n(\theta^n)$ is a monotone decreasing function of θ^n.

An essential assumption is that the probability of being defective, p^n, is a function of the fraction of sampling, θ^n. The relation could be assumed to correspond to the one shown in Fig. 3.1 [24]. It can be seen that $p^n(\theta^n)$ decreases very rapidly initially with small increases in θ^n, but thereafter it tends to be constant. This constant value of $p^n(\theta^n)$ which cannot be reduced to zero even for very large θ^n is called the residual "unavoidable defect".

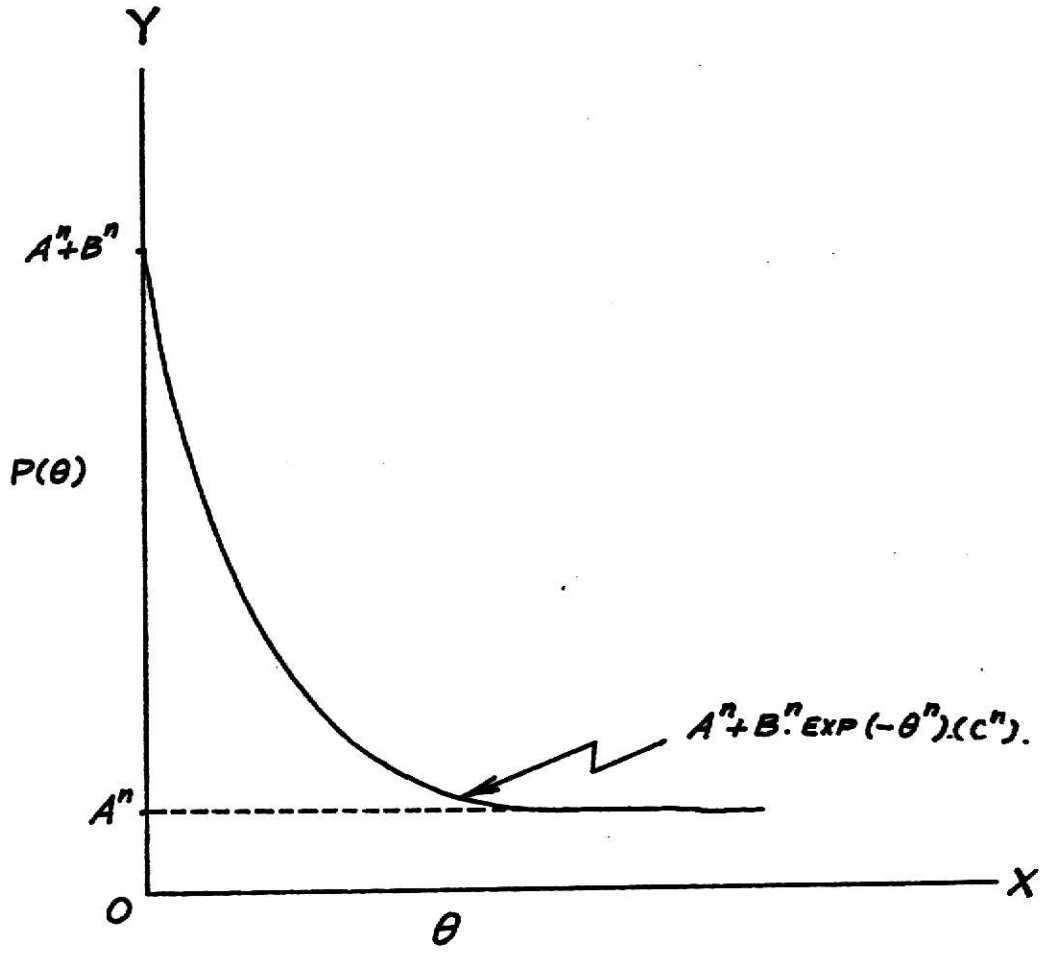
The equation

$$p^n(\theta^n) = A^n + B^n(e)^{-c^n\theta^n} \tag{3.1}$$

may be used as a fairly good representative of the $p^n(\theta^n)$ function where A^n,B^n and C^n are constants [24].

---

*The superscript n indicates the stage number. The exponents are written with parenthesis or brackets such as $(x^n)^2$.



RELATION BETWEEN
THE FRACTION OF SAMPLING, θ
& THE PROBABILITY OF DEFECT, P(θ).

FIG. 3.1 .

In equation (3.1), A^n represents the proportion of articles that are 'unavoidably' defective [13]. When θ^n attains the maximum possible value of unity, $p^n(\theta^n)$ will be nearly equal to A^n since $B^n e^{-C^n\theta^n}$ becomes negligible. B^n measures the proportion of defects added to A^n depending upon lack of inspection. C^n measures the weight or effect of sampling on the "avoidable" defects. Thus the quantity C^n can be looked upon as the "deterrent" effect of sampling on the probability of being defective. It is also regarded as "elasticity" of reaction to sampling. It may be pointed out here that the idea of preventative sampling is based on the existence of the quantity C^n [16]. The larger the value of C^n, the greater is the scope of improvement by the use of preventative sampling.

The total cost of inspection, then, becomes $I^n a^n \theta^n$. The other cost to be considered is the cost of accepting defective articles. $a^n(1-\theta^n)$ represents the quantity of the products not sampled. Therefore $a^n v^n p^n(\theta^n)(1-\theta^n)$ is the value of the undetected defective quantity. This is the cost of accepting defective articles. Thus, the combined cost becomes

$$\begin{aligned}\text{combined cost} &= \text{cost of inspection} + \\ &\qquad \text{cost of accepting defective articles} \\ &= I^n a^n \theta^n + a^n v^n p^n(\theta^n)(1-\theta^n) \qquad (3.2)\end{aligned}$$

The criterion for optimality is the one that makes "the sum of the total expected value of the undetected faulty articles, and the total cost of inspection" as small as possible. The problem is then finding an optimal sampling fraction θ^n to minimize the objective

$$S = \sum_{n=1}^{N} [a^n v^n p^n(\theta^n)(1-\theta^n) + I^n a^n \theta^n] \qquad (3.3)$$

subject to the constraints,

$$\sum_{n=1}^{N} a^n \theta^n = D\alpha. \tag{3.4}$$

and $0 \leq \theta^n \leq 1, \quad n = 1, 2, \ldots, N,$ (3.5)

where $\sum_{n=1}^{N} a^n = D$ is the total quantity of N products and α is the fraction sampling of the total quantity D, and therefore $D\alpha$ represents the sampling capacity restricted by man power and/or equipment.

3.3 THE SEQUENTIAL UNCONSTRAINED MINIMIZATION TECHNIQUE [15]:

A number of constrained minimization techniques have been developed recently for finding the minimum or maximum of a function with several variables subject to certain constraints. A technique which was originally proposed by Carroll [2,3] and further developed by Fiacco - and McCormick [6,7,8,9,10] is introduced here.

This technique, known as the sequential unconstrained minimization technique (SUMT), is considered as one of the simplest and most efficient methods for solving the constrained nonlinear optimization problem. Transformation of a constrained minimization problem into an unconstrained minimization problem enables us to use available techniques without inventing a new technique to solve constrained minimization problems. Many methods for minimizing an unconstrained function are known and newer ones are continually being developed.

The general nonlinear programming problem with nonlinear inequality constraints is to choose x to

$$\left.\begin{array}{l} \text{minimize } f(x) \\ \text{subject to} \\ \qquad g_i(x) \geq 0, \quad i = 1, 2, \ldots, m \end{array}\right\}$$

where x is an n-dimensional column vector $(x_1, x_2, \ldots, x_n)^T$. The superscript T denotes transposition. If the variables are required to be non-negative, such constraints are included in the g_i's. Functions f(x) and $g_i(x)$, i = 1, 2, ..., m, can take linear or nonlinear form.

To solve this problem the following algorithm is presented [6,7,9,10]. Define the function (called the P function)

$$P(x, r_k) = f(x) + r_k \sum_{i=1}^{m} \frac{1}{g_i(x)} \tag{3.7}$$

where r_k is a positive constant. Subscript k indicates the number of time P function has been set up to solve the problem given by equation (3.6). The conditions imposed on the P function are as follows:

(1) r_k, k = 1, 2, ..., is a positive real number and $r_1 > r_2 > \ldots > r_k > \ldots > 0$. This indicates that $\{r_k\}$ is a strictly monotonic decreasing sequence and $r_k \to 0$ as $k \to \infty$.

(2) $R^0 = \{x \mid g_i(x) > 0, i = 1, 2, \ldots, m\}$ is non-empty. This condition indicates that at least one point must exist within the interior of the feasible region.

(3) The functions f(x), $g_1(x)$, ..., $g_m(x)$ are twice continuously differentiable.

(4) The function f(x) is convex.

(5) The functions $g_1(x)$, ..., $g_m(x)$ are concave.

(6) For every finite K, $\{x \mid f(x) \leq K;\ x \in R\}$ is a bounded set, where $R = \{x \mid g_i(x) \geq 0,\ i = 1, 2, \ldots, m\}$.

(7) The function $P(x, r_k) = f(x) + r_k \sum_{i=1}^{m} \frac{1}{g_i(x)}$ is, for each $r > 0$, strictly convex for $x \in R^0$. This also indicates that either $f(x)$ is strictly convex or one of $g_1, \ldots, g_m$ is strictly concave.

Practical experience indicates that the problems given by equation (3.6) can be solved even when these conditions are not met. The three conditions which are absolutely required to obtain any useful results are conditions (1), (2), and (6). Condition (1) guarantees that the sequential minimization of the P function will eventually lead to the solution of minimization of function $f(x)$. Condition (2) eliminates problems with equality constraints. Condition (6) eliminates problems having local minimum at infinite points.

The characteristics of the P function are as follows:

(1) $\lim_{k \to \infty} r_k \sum_{i=1}^{m} \frac{1}{g_i(x)} = 0,$

(2) $\lim_{k \to \infty} f[x(r_k)] = u^*,$

(3) $\lim_{k \to \infty} P[x(r_k), r_k] = u^*,$

(4) $\{f[x(r_k)]\}$ is a monotonically decreasing sequence,

(5) $\{\sum_{i=1}^{m} \frac{1}{g_i(x)}\}$ is a monotonically increasing sequence.

The proofs of these characteristics are presented in detail in Fiacco and McCormick [6,7,8,9,10]. The proofs of (4) and (5) will be presented below since these proofs will help to clarify the intuitive concept of the method.

Let

$$P(x, r_k) = f(x) + r_k \sum_{i=1}^{m} \frac{1}{g_i(x)} = f(x) + S(r_k)\ I(x)$$

where $S(r_k) = r_k$, and $I(x) = \sum_{i=1}^{m} \frac{1}{g_i(x)}$. Also let f^k denote $f[x(r_k)]$ and I^k denote $I[x(r_k)]$. Since each $x(r_k)$ is a global unconstrained minimum in the interior of the compact set, the following relationships must hold.

$$f^k + S(r_k)\ I^k \leq f^{k+1} + S(r_k)\ I^{k+1} \tag{3.8}$$

and

$$f^{k+1} + S(r_{k+1})\ I^{k+1} \leq f^k + S(r_{k+1})\ I^k \tag{3.9}$$

Adding inequality constraints, equations (3.8) and (3.9), one obtains

$$S(r_k)(I^k - I^{k+1}) \leq S(r_{k+1})(I^k - I^{k+1}). \tag{3.10}$$

Shifting the right-hand side to the left-hand side of inequality sign, the inequality, equation (3.10), becomes

$$[S(r_k) - S(r_{k+1})]\ (I^k - I^{k+1}) \leq 0 \tag{3.11}$$

Since $\{S(r_k)\}$ is a monotonically decreasing sequence, $[S(r_k) - S(r_{k+1})] > 0$. This implies that the following relationship exists.

$$I^k - I^{k+1} \leq 0, \tag{3.12}$$

that is,

$$I^k \leq I^{k+1} \quad \text{for all } k = 0, 1, \ldots \tag{3.13}$$

Hence, $\{\sum_{i=1}^{m} \frac{1}{g_i(x)}\}$ is a monotonically increasing sequence.

Using the inequality, equation (3.13), in equation (3.9), one obtains

$$f^{k+1} + S(r_{k+1})I^{k+1} \leq f^k + S(r_{k+1})I^{k+1} \tag{3.14}$$

which implies

$$f^{k+1} \leq f^k \tag{3.15}$$

Hence, $\{f[x(r_k)]\}$ is a monotonically decreasing sequence.

INTUITIVE CONCEPT OF P FUNCTION

The term $r_k \sum_{i=1}^{m} \frac{1}{g_i(x)}$ in P function of equation (3.7) can be considered as a penalty factor attached to the objective function f(x). By adding the penalty term, the minimization of P function will assure a minimum to be in the interior of the inequality constrained region by avoiding crossing the boundaries of feasible region. Since the feasible boundary is defined by one or more of the $g_i(x) = 0$, i=1,...,m, the value of $r_k \sum_{i=1}^{m} \frac{1}{g_i(x)}$ will approach to infinity as the value of x

approaches to one of the boundary lines. Hence the value of x will tend to remain inside the inequality-constrained region.

The motivation behind this formulation of the P function is the transformation of the original constrained problem into a sequence of unconstrained minimization problem. The desirability of this transformation lies in the fact that numerous methods for minimizing an unconstrained function are known and newer methods are continually being developed [4,5,11,12,19,22]. Thus, by this transformation it becomes possible to solve the more formidable constrained problem without inventing new techniques.

COMPUTATIONAL PROCEDURE

The procedure for using SUMT is summarized below [6,7].

(1) Select the initial value of r_0 arbitrarily or use the formula for selecting r_0, which is available in reference [7].

(2) Select a feasible starting point $x^0 = (x_1^0, x_2^0, \ldots, x_n^0)$. If the feasible point can not be easily obtained, select x^0 arbitrarily. The computer program will minimize the following P function and obtain the feasible point.

$$P(x, r_k) = -g_s(x) + r_k \sum_{t \in T} \frac{1}{g_t(x)}$$

where $g_s(x^0) \leq 0$ and $T = \{t \mid g_t(x^0) > 0\}$. Note that the constraint function of $g_s(x) \geq 0$ is violated.

(3) Minimize the P function for the current value of r_k by using the second-order optimum gradient method.

(4) Check if a stopping criterion such as

$$\frac{f[x(r_k)]}{G[x(r_k)]} - 1 < \varepsilon \qquad (11)$$

is satisfied. The solution is the optimal one if the criterion is satisfied; otherwise go to step 5. The dual value, $G[x(r_k)]$, is defined as [5]

$$G[x(r_k)] = f[x(r_k)] - r_k \sum_{i=1}^{m} \frac{1}{g_i[x(r_k)]}$$

(5) Set $k = k+1$ and $r_{k+1} = r_k/C$, where $C > 1$. Repeat the iteration from step 3.

The procedures described above will have to satisfy two stopping criteria before any meaningful optimal solution can be obtained. Stopping criterion used for terminating minimization of P function [Step 3] is in the following forms.

(i) $$\left|\nabla_x P^T(x, r) \left|\frac{\partial^2 P(x,r)}{\partial x_i \partial x_j}\right|^{-1} \nabla_x P(x,r)\right| < \varepsilon'$$

or

(ii) $$\left|\nabla_x P^T(x, r) \left|\frac{\partial^2 P(x,r)}{\partial x_i \partial x_j}\right|^{-1} \nabla_x P(x,r)\right| < \frac{P(x,r_{k-1}) - P(x,r_k)}{5}$$

or

(iii) $$\left|\nabla_x P(x,r)\right| < \varepsilon'$$

The first stopping criterion was used throughout this study with ε' in the range of 10^{-3} to 10^{-5}. The stopping criterion for terminating overall minimization of $f[x(r_k)]$ may take the following form in addition to the form given by equation (11).

$$r_k \sum_{i=1}^{m} \frac{1}{g_i[x(r_k)]} < \varepsilon$$

Equation (11) was used in the numerical examples presented in this work with ε equal to 10^{-5}. The procedure should not be terminated until both criteria are satisfied. If these stopping criteria are not satisfied within a specified time limit, the iterations will be terminated by the given time limit.

A computer program entitled "RAC Computer Program Implementing the Sequential Unconstrained Minimization Technique for Nonlinear Programming" is available. Its SHARE number is 3189 [18]. The program is written in FORTRAN IV and can be used on IBM 360. With minor modifications the program can be run on any sufficiently large computer with a FORTRAN compiler.

3.4. SOLUTION BY THE SEQUENTIAL UNCONSTRAINED MINIMIZATION TECHNIQUE:

In the present problem, we have a non-linear objective function and linear constraints. It is solved by the sequential unconstrained minimization technique.

The problem formulated in section 3.2 is formulated in the general form of the SUMT.

$$\text{Min} \qquad S = \sum_{n=1}^{N} [a^n v^n p^n (\theta^n)(1-\theta^n) + I^n a^n \theta^n]$$

subject to the constraints

$$\left(\sum_{n=1}^{N} a^n \theta^n \right) - D\alpha \geq 0,$$

$$\theta^n \geq 0, \quad n = 1, 2, \ldots, N$$

and $\quad 1 - \theta^n \geq 0, \quad n = 1, 2, \ldots, N.$

Let us consider a numerical example of N = 5. The data are given in Table 3.1. The problem has been solved on the IBM 360/50 computer using SUMT. Table 3.2 gives the results of the optimal values obtained from various starting points. A very wide range of starting points which include both feasible and nonfeasible regions, have been used to study the power of the technique, and the effect on the computational time taken to arrive at the optimal value. As can be seen from the table even the very bad starting points converged to the optimal very quickly. Hence this technique could be used successfully in larger problems where it is not possible to guess a feasible starting point. The results compare very well with the ones obtained by Hwang, Fan and Tillman [16] using a discrete version of Pontryagin's maximum principle. These are tabulated in table 3.3. Table 3.4 shows the detailed computer output from the starting point [0.1, 0.1, 0.1, 0.1, 0.1]. For stopping criterion values of ε' and ε were chosen to be 1×10^{-5}.

3.5 DISCUSSION AND CONCLUDING REMARKS

A study of Table 3.2 shows that the number of iterations and hence the computation time is not affected very much by the starting points. The starting trial point [0.1, 0.1, 0.1, 0.1, 0.1] which is fairly close to the optimum converged to the optimum with 14 iterations. The other starting point [50, 100, 800, 5, 0] which is very far from the optimum value also converged to the optimum in only 16 iterations. Convergence to the optimum even from such

Table 3.1 Data for numerical example

$N = 5$

$A^n = A = 2\%$ unavoidably defective

$B^n = B = 20\%$ avoidably defective

$C^n = C = 25$ (incentive factor)

$\alpha = 10\%$ overall inspection percent

Products n	Value ($/unit) v^n	Inspection Cost ($/unit) I^n	Quantity (units) a^n
1	5	0.05	400
2	10	0.20	250
3	15	0.40	100
4	20	0.80	150
5	25	1.25	100
D			1000

Table 3.2 Optimal Sample Sizes from Different Starting Values

$r_0 = 3000 \qquad c = 4$

Trial No.	Starting Values for θ^n	Optimal Values for θ^n	No. of Iterations, k
(1)	1.00	0.07693	26
	2.00	0.10305	
	3.00	0.11722	
	4.00	0.12492	
	5.00	0.12995	
(2)	0.0	0.07693	12
	0.0	0.10305	
	0.0	0.11722	
	0.0	0.12492	
	0.0	0.12995	
(3)	50	0.07693	16
	100	0.10305	
	800	0.11722	
	5	0.12492	
	0.0	0.12995	
(4)	0.1	0.07693	13
	0.1	0.10305	
	0.1	0.11722	
	0.1	0.12492	
	0.1	0.12995	

Table 3.3 Comparison of Results

Inspection Fraction to Different Articles	Results by Maximum Principle [8]	Results by Nonlinear Programming (SUMT)
θ^1	0.0771	0.0769
θ^2	0.1026	0.1031
θ^3	0.1170	0.1172
θ^4	0.1251	0.1249
θ^5	0.1301	0.1300

Table 3.4 Computer Results of Optimum Preventative Sampling Model

Number of Iterations k	Value of r	Value of θ_1	θ_2	θ_3	θ_4	θ_5	Value of f(θ)	Value of P(θ)
0		0.1	0.1	0.1	0.1	0.1	412.4	0
1	3000	0.0923	0.0826	0.1300	0.1064	0.1310	404.1	379335
2	750	0.0923	0.0826	0.1300	0.1065	0.1300	404.0	95136
3	187.5	0.0920	0.0829	0.1297	0.1069	0.1300	403.5	24087
4	46.8	0.0912	0.0840	0.1286	0.1083	0.1298	401.9	6323
5	11.7	0.0885	0.0874	0.1256	0.1121	0.1294	397.7	1874
6	2.92	0.0836	0.0937	0.1211	0.1181	0.1297	392.6	766
7	0.732	0.0792	0.0992	0.1183	0.1226	0.1299	390	484
8	0.183	0.0774	0.1017	0.1175	0.1245	0.1304	389	413
9	0.045	0.0769	0.1025	0.1173	0.1251	0.1306	389	395
10	0.011	0.0768	0.1028	0.1173	0.1253	0.1306	389	390
11	0.0028	0.0769	0.1025	0.1173	0.1251	0.1306	389	389
12	0.00007	0.0768	0.1028	0.1173	0.1254	0.1307	388.7	388.8
13	0.000014	0.0768	0.1029	0.1174	0.1254	0.1307	388.7	388.75
14	0.000005	0.0768	0.1029	0.1174	0.1254	0.1307	388.69	388.70

a bad starting point shows the power of the technique. In cases where the number of products is very large and where the feasible region is difficult to guess, this technique could be very successfully employed. It was observed that the computer time for this five dimensional problem was only 15 seconds on the average. The results obtained by SUMT are identical to the ones obtained by the application of maximum principle. The minimum cost of inspection is $388.69. From Table 3.4 we can see that this optimum is reached in 14 iterations.

REFERENCES

1. Bracken, N., and G. P. McCormick, Selected applications of nonlinear programming, John Wiley, New York, 1968.

2. Carroll, C W., "An Operations Research Approach to the Economic Optimization of a Kraft Pulping Process," Ph.D. Dissertation, Institute of Paper Chemistry, Appleton, Wisc., 1959.

3. Carroll, C. W., "The Created Response Surface Technique for Optimizing Nonlinear Restrained Systems," Operations Research, 9, 169-184 (1961).

4. Davidon, W. C., "Variable Matric Method for Minimization," Research and Development Report ANL-5990 (Revised), Argonne National Laboratory, U. S. Atomic Energy Commission, 1959.

5. Fan, L. T., C. L. Hwang, and F. A. Tillman, "A Sequential Simplex Pattern Search Solution to Production Planning Problems," AIIE Trans. Vol. 1, No. 3 (1969).

6. Fiacco, A. V., and G. P. McCormick, "The Sequential Unconstrained Minimization Technique for Nonlinear Programming: A Primal-Dual Method," Management Sci., 10, 360-366 (1964).

7. Fiacco, A. V., and G. P. McCormick, "Computational Algorithm for the Sequential Unconstrained Minimization Technique for Nonlinear Programming," Management Sci., 10, 601-617 (1964).

8. Fiacco, A. V., and G. P. McCormick, "SUMT without parameters," Systems Research Memorandum No. 121, Technical Institute, Northwestern University, Evanston, Illinois, 1965.

9. Fiacco, A. V., and G. P. McCormick, "Extension of SUMT for Nonlinear Programming: Equality Constraints and Extrapolation," Management Sci., 12 (11): 816-829 (1966).

10. Fiacco, A. V., and G. P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, Wiley, New York, 1968.

11. Fletcher, R. and M. J. D. Powell, "A Rapidly Convergent Descent Method for Minimization," Computer J., 6, 163-168 (1963).

12. Fletcher, R. and C. M. Reeves, "Function Minimization by Conjugate Gradients," Computer J., 7, 149-154 (1964).

13. Hamacker, H. C., "Some Basic Principles of Sampling Inspection by Attributes", Journal of Royal Statistical Society, Vol VII, No. 3, 149-159 (1958).

14. Horsnell, G., "Economical Acceptance Sampling Plans", Journal of Royal Statistical Society, A 120, 148-168 (1957).

15. Hsu, F. T., L. T. Fan and C. L. Hwang, "Sequential Unconstrained Minimization Technique (SUMT) for Optimal Production Planning", A paper submitted for publication, August 1969.

16. Hwang, C. L., L. T. Fan, and F. A. Tillman, "Optimal Preventative Sampling", Kansas State University, unpublished lecture note.

17. Martin, A. C., "The Cost Breakeven Point in Attribute Sampling", Industrial Quality Control, 137-143 (1964).

18. Taubert, W. H., "A Search Decision Rule for the Aggregate Scheduling Problem," Management Science, 14, B-343 (1968).

19. Powell, M. J. D., "An Efficient Method of Finding the Minimum of a Function of Several Variables without Calculating Derivatives," Computer J., 7, 155-162 (1964).

20. Powell, M. J. D., "A Method for Non-Linear Constraints in Minimization Problems," T.P. 310, Atomic Energy Research Establishment, Harwell, England, 1967.

21. Self, G. D., "Hypergeometrically Computed LTDD Sampling Planner", Journal of Industrial Engineering", 608-611 (1968).

22. Spang, H. A. III, "A Review of Minimization Techniques for Nonlinear Functions," SIAM, Rev., 4, 343-365 (1962).

23. Stacy, E. W., L. E. Hunsinger and J. E. Price, "Determining Economic Sampling Plans", IBM SYST Journal, No. 3, 220-244 (1969).

24. Whittle, P., "Optimum Preventative Sampling", Journal of Operations Research Society of America, Vol. 2, 197-203 (1954).

4. STRATIFIED SAMPLING

4.1 INTRODUCTION

In sampling practice we find that if the population is very heterogeneous and considerations of cost limit the size of the sample, it is impossible to get a sufficiently precise estimate by taking a simple random sample from the entire population [10]. Populations encountered in practice are generally very heterogeneous. To cite an example, suppose we wish to estimate the average percent defective of products in a certain factory. It is known that the majority of the products are good or fair quality, and there are only a few bad quality products. It is also known that these very bad quality products account for a substantial amount of total sales. In other words, the sales distribution of products is highly skewed. When simple random sampling procedure is applied to such a distribution, there is a chance that either too many of the very bad products will be included in the sample or none at all. As a result, the sample may not adequately represent the population. In such cases, stratification is used to increase the precision of estimation.

4.2 DESCRIPTION

In stratified sampling, the population of N units is first divided into subpopulations of N_1, N_2, ..., N_L units, respectively. The subpopulations are nonoverlapping, and together they comprise the whole population, so that,

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

The subpopulations are called STRATA. To obtain the full benefit from stratification, the values of the N_h must be known, where the subscript 'h' denotes the h-th stratum. The process of breaking down the populations into strata, selecting simple random samples from each stratum, and combining these into a single sample to estimate population parameters is called stratified random sampling [2]. The sample sizes within the strata are denoted by n_1, n_2, ..., n_L respectively.

Stratification is a common technique. From the strict statistical point of view, the primary cause for stratified sampling is to increase efficiency [3]. This increase in efficiency is achieved because of the reduction in standard deviations of the strata over the standard deviation of the total unstratified population. We may list the advantages as follows [2,3,10,11]:

(1) If data of known precision are wanted for certain subdivisions of the population, it is advisable to treat each subdivision as a "population" in its own right. For example we might want to know not only the overall company's failure rate of a product, but also the failure of the product in each factory of the nation-wide company. When this is the case, stratified sampling offers a means of regulating the accuracy with which we estimate these failure rates. The accuracy of a sample estimate depends upon the sample size. If we draw an unrestricted random sample from the total population, we cannot control the sample size within each stratum. If we draw a stratified sample, we can control the sample size in each stratum and thus control the accuracy of the estimate for that stratum.

(2) Administrative convenience may dictate the use of stratification. In many situations it is less difficult to sample strata than to treat the whole population as a single unit. Suppose we wish to estimate the probability of defects of a product manufactured by a company, in many locations all over the nation. In such a case, drawing an unrestricted sample from the total population of products, scattered across the country would present a difficult problem. Thus the convenience of drawing a stratified sample often makes it worthwhile.

(3) Sampling problems may differ in different parts of the population. With human populations, people living in institutions (e.g., hotels, hospitals, prisons) are often placed in a different stratum from people living in ordinary homes because a different approach to the sampling is appropriate for the two situations. In sampling business we may posses a list of the large firms, which are placed in a separate stratum.

(4) Stratification may produce a gain in precision in the estimates of characteristics of the whole population. It may be possible to divide a heterogeneous population into subpopulations, each of which is internally homogeneous. This is suggested by the name strata, with its implication of a division into layers. If each stratum is homogeneous, in that the measurements vary little from one unit to another, a precise estimate of any stratum mean can be obtained from a small sample in that stratum. These estimates can be combined into a precise estimate of the whole population.

The theory of stratified sampling deals with the properties of the estimates from a stratified sample and with the best choice of the sample

sizes n_h to obtain maximum precision. In this development, it is taken for granted that the strata have already been constructed.

4.3 PROPERTIES OF THE ESTIMATES

Let the suffix h denote the stratum and i the unit within the stratum. The following symbols all refer to stratum h.

N_h = total number of units, in a stratum

n_h = number of units in the sample of the stratum

y_{hi} = value obtained for the i^{th} unit,

$W_h = \frac{N_h}{N}$, stratum weight,

$f_h = \frac{n_h}{N_h}$, sampling fraction in the stratum,

$$\bar{Y}_h = \frac{\sum_{i=1}^{N_h} y_{hi}}{N_h}, \text{ true mean,}$$

$$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}, \text{ sample mean,}$$

$$S_h^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2}{N_h - 1}, \text{ true variance,}$$

s_h^2 = estimate of S_h^2 .

For the population mean per unit, the estimate used in stratified sampling is $\bar{y}_{st}$, where [1]

$$\bar{y}_{st} = \frac{\sum_{h=1}^{L} N_h \bar{y}_h}{N} \tag{4.1}$$

where

$$N = N_1 + N_2 + \ldots + N_L$$

The sample mean, $\bar{y}$, is given by the equation, [2],

$$\bar{y} = \frac{\sum_{h=1}^{L} n_h \bar{y}_h}{n} \tag{4.2}$$

From equations (4.1) and (4.2) it can be seen that $\bar{y}$ coincides with $\bar{y}_{st}$ provided that in every stratum .

$$\frac{n_h}{n} = \frac{N_h}{N} \quad \text{or} \quad \frac{n_h}{N_h} = \frac{n}{N} \quad \text{or} \quad f_h = f$$

The above condition implies that the sampling fraction is the same in all strata. This stratification is called as stratification with proportional allocation of the n_k.

It can be shown that* with stratified random sampling, an unbiased estimate of the variance of $\bar{y}_{st}$ is, [2]

$$V(\bar{y}_{st}) = s^2(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{L} N_h(N_h - n_h) \frac{s_h^2}{n_h} \tag{4.3}$$

A more convenient form would be [2]

$$s^2(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 s_h^2}{n_h} - \sum_{h=1}^{L} \frac{W_h s_h^2}{N} \tag{4.4}$$

*See Appendix I for the proof.

4.4 OPTIMAL ALLOCATION

In stratified sampling, the values of the sample sizes, n_h, in the respective strata are chosen by the sampler. They may be selected to minimize the variance $V(\bar{y}_{st})$ for a specified cost of taking the sample, or to minimize the cost for a specified value of $V(\bar{y}_{st})$ [8]. The simplest cost function is of the form [2,10]

$$\text{cost} = C = c_o + \sum c_h n_h. \tag{4.5}$$

The above equation implies that within any stratum, the cost is proportional to the size of the sample, but the cost per unit, c_h, may vary from stratum to stratum. The term, c_o, represents an overhead cost. This is justified when the major item of the cost is that of taking the measurement on each unit. Certain types of sampling may have costs that are not linearly related to the number of units in the sample in the various strata [2]. To start with let us consider the linear cost function given by equation (4.5).

It was proved* that in stratified sampling with the above cost function, the variance of the estimated mean $\bar{y}_{st}$ is a minimum when n_h is proportional to $N_h S_n/\sqrt{c_h}$ such as, [1]

$$n_h = n \cdot \frac{N_h S_h/\sqrt{c_h}}{\sum (N_h S_h/\sqrt{c_h})} \tag{4.6}$$

From the above equation, it can be seen that

(1) n_h will be larger if N_h, the stratum size is larger

*See Appendix II for the proof.

(2) n_h will be larger if S_h, the standard deviation of that stratum is larger.

(3) n_h will be larger if c_h, the unit cost of inspection in that stratum is smaller.

Equation (4.6) gives us the value of n_h in terms of n, but it does not indicate the actual value of n, the total sample size. The answer depends on whether the sample is chosen to meet a specified total cost C or to give a specified variance V for $\bar{y}_{st}$. If the cost is fixed, the optimum values of n_h is substituted in the cost function, equation (4.5), and solved for n. This gives,

$$n = \frac{(C - c_o) \sum(N_h S_h / \sqrt{c_h})}{\sum(N_h S_h \sqrt{c_h})} . \tag{4.7}$$

If the cost per unit is the same in all strata, the cost becomes $C = c_o + cn$, and optimum allocation for fixed cost reduces to optimum allocation for fixed sample size.

In stratified random sampling, $V(\bar{y}_{st})$ is minimized for a fixed total size of sample n if,

$$n_h = n \cdot \frac{W_h S_h}{\sum W_h S_h} = n \cdot \frac{N_h S_h}{\sum N_h S_h} . \tag{4.8}$$

This allocation is sometimes called the Neyman allocation, after Neyman whose proof gave the result prominence [2]. Hess et al [6] concluded that Neyman allocation gave the highest precision, when compared with other types of allocation. It should be pointed out that if intelligently used, stratification nearly always results in a better

estimate as compared to the estimate obtained by the simple random sample [2]. It is not to be taken for granted that any stratified random sample gives a smaller variance than a simple random variable. If the values of the sample sizes allocated to different strata are far from the optimum, stratified sampling may have a higher variance. This shows the importance of finding suitable techniques for optimum allocation of sample size in order to utilize this useful tool of stratification. Optimal allocation is very easy when we are dealing with only one variate. When two or more variates are involved, optimization could be achieved through the use of nonlinear programming [8].

The use of linear and nonlinear programming in approaching this type of problem has been discussed by several authors [2,6,8]. The use of linear programming in sample surveys to determine allocations when several characters are under study was first suggested by Dalenius [4] and Nordbotten [9] illustrated this approach by a numerical example. For problems with two strata and several variates, a graphical solution has been proposed by Dalenius [5]. A general mathematical approach useful for problems with a small number of strata and variates is given by Yates [12]. Bracken and McCormick [1] have formulated such a problem as a nonlinear programming problem.

4.5 EXAMPLE PROBLEM

Let us consider an example problem formulated by Cochran [2], with four strata and two variates. The index h denotes the stratum and j the variate, where h = 1, ..., L and j = 1, ..., K. From equation (4.1) we see that the estimate of the population mean of the j^{th} variate is

$$E(\bar{y}_j) = \frac{\sum_{h=1}^{L} N_h \bar{y}_{jh}}{N} = \sum_{h=1}^{L} W_h \bar{y}_{jh} \tag{4.9}$$

where $\bar{y}_{jh}$ is the h^{th} stratum mean of the j^{th} variate. From equation (4.4) the variance of the estimate $\bar{y}_j$ is

$$V(\bar{y}_j) = \sum_{h=1}^{L} \frac{W_h^2 s_{jh}^2}{n_h} - \sum_{h=1}^{L} \frac{W_h s_{jh}^2}{N_h} \tag{4.10}$$

where s_{jh}^2 is the known sampling variance for the j^{th} variate in the h^{th} stratum.

The optimal sampling problem is to minimize sampling cost subject to the constraints that the variance of the estimate of the population mean must be equal to or less than a specified value for all the K variates. Thus the constraints could be written as follows [8]:

$$\sum_{h=1}^{L} \frac{W_h^2 s_{jh}^2}{n_h} - \sum_{h=1}^{L} \frac{W_h s_{jh}^2}{N_h} \leq V_j, \quad j = 1, \ldots, K \tag{4.11}$$

where V_j is the upper limit on the variance of the estimate of the mean of the j^{th} variate. In the above constraint, everything except the stratum sample sizes $n_h (h = 1, \ldots, L)$ are known. The other obvious constraints are that the sample size n_h must be nonnegative, and it must be equal to or less than the total number of units in the stratum. Thus the upper and lower limits may be specified as

$$0 \leq n_h \leq N_h, \quad h = 1, \ldots, L \tag{4.12}$$

Since the minimization of the total cost is being considered, the total cost becomes the objective function. This could be a function

as stated in equation (4.5). The nonlinear programming model for obtaining optimal sample sizes with respect to this cost function is as follows:

Choose $n_h (h = 1, \ldots, L)$ to minimize the linear cost function given by equation (4.5) subject to the nonlinear constraints, equation (4.11), and to the nonnegativity restrictions and upper bounds given by equation (4.2).

The example that is being considered here is with four strata and two variates [1] and the problem is to find the sampling plan with minimum cost where the variances of the estimates of the population mean for the two variates are equal to or less than the specified values. The total cost equation is of the form

$1 + \sum_{h=1}^{4} n_h$, $(c_0 = 1, c_1=c_2=c_3=c_4=1)$. Table 4.1 gives the data of the problem, including the population sizes in the strata, known variances of the two variates in the four strata and unit cost in sampling in the four strata. The upper limits on the variances of estimates of the population means of the two variates are

$$V_1 \leq .04, \qquad V_2 \leq .01.$$

The nonlinear programming problem is as follows. Choose n_1, n_2, n_3, and n_4 to minimize the linear criterion function

$$(1) + (1)(n_1) + (1)(n_2) + (1)(n_3) + (1)(n_4)$$

subject to

Table 4.1 Data for Stratified Sampling Problem [1]

Stratum h	Stratum Population N_h	Stratum Weight W_h	Variances of j^{th} Variate s^2_{j1}	s^2_{j2}	Cost per Unit of Sample c_h
1	400,000	0.4	25	1	1
2	300,000	0.3	25	4	1
3	200,000	0.2	25	16	1
4	100,000	0.1	25	64	1

$$\frac{(.4^2)(25)}{n_1} + \frac{(.3^2)(25)}{n_2} + \frac{(.2^2)(25)}{n_3} + \frac{(.1^2)(25)}{n_4}$$

$$- \left[\frac{(.4)(25)}{400,000} + \frac{(.3)(25)}{300,000} + \frac{(.2)(25)}{200,000} + \frac{(.1)(25)}{100,000}\right] \leq .04,$$

$$\frac{(.4^2)(1)}{n_1} + \frac{(.3^2)(4)}{n_2} + \frac{(.2^2)(16)}{n_3} + \frac{(.1^2)(64)}{n_4}$$

$$- \left[\frac{(.4)(1)}{400,000} + \frac{(.3)(4)}{300,000} + \frac{(.2)(16)}{200,000} + \frac{(.1)(64)}{100,000}\right] \leq .01,$$

$$0 \leq n_1 \leq 400,000,$$

$$0 \leq n_2 \leq 300,000,$$

$$0 \leq n_3 \leq 200,000,$$

$$0 \leq n_4 \leq 100,000.$$

The above nonlinear programming problem is solved by the sequential unconstrained minimization technique (SUMT) and optimal sample sizes and costs are given in Table 4.2.

The total cost is $1 + \sum_{h=1}^{4} c_h n_h = 728$. The results are identical to the ones obtained by Bracken and McCormick [1].

Table 4.3 gives the total number of iterations taken to converge to the optimum from various initial guesses. Starting points were

Table 4.2 Results of the Problem

Stratum h	Optimal Sample Size n_h	Cost $c_h n_h$
1	193	193
2	180	180
3	185	185
4	169	169

choosen from both the feasible and nonfeasible regions. A study of the Table 4.3 indicates the power of this technique. It could be observed that even when the initial points were much away from the optimum, the program did not take too many iterations to converge to the optimum. The program was written in FORTRAN and run on IBM 360/50 computer in FORTRAN G LEVEL. The average execution time for this four dimensional problem was 40 seconds.

Certain types of sampling may have costs that are not linearly related to the number of units in the sample in the various strata. A more general cost function might be [1]

$$C = c_o' + \sum_{h=1}^{L} c_h' n_h^p \,. \tag{4.13}$$

Though linear cost function may be taken as a good approximation in many cases, equation (4.13) with $p < 1$ may be more realistic in some cases. For this more general cost function, the following nonlinear programming problem would arise:

Choosen $n_h (h = 1, \ldots, L)$ to minimize the nonlinear criterion function given by equation (4.13) subject to constraints of equations (4.11) and (4.12). The numerical problem given by Cochran [2] has been modified to have a nonlinear cost function as given by equation (4.13) with a value of $p = 0.5$. Thus this becomes a problem with nonlinear objective function with non-linear constraints. This problem has also been solved by SUMT and Table 4.4 gives the optimal sample sizes.

Tables 4.5 gives the detailed output of the computer program for the linear cost function. Initial value of r is selected as 3000.

Table 4.3 Results of the Problem of Linear Cost Function with Various Starting Points

$r_0 = 3000 \qquad c = 4$

No.	Starting Values for Sample Sizes	Optimal Values for Sample Sizes	No. of Iterations, k
(1)	100	193	20
	100	180	
	100	185	
	100	169	
(2)	40	193	20
	100	180	
	800	185	
	5	169	
(3)	0.1	193	20
	0.1	180	
	0.1	185	
	0.1	169	
(4)	0.0	193	20
	0.0	180	
	0.0	185	
	0.0	169	

Table 4.4 Results of the Problem with Nonlinear Cost Function with Various Starting Points

$r_0 = 3000 \quad c = 4$

No.	Starting Values of the Sample Sizes	Optimal Values of the Sample Sizes	No. of Iterations, k
	200	84315	
	200	96063	
(1)	200	120134	12
	200	97402	
	0	84298	
	0	96116	
(2)	0	120132	12
	0	97402	
	100	84289	
	100	96094	
(3)	100	120137	12
	100	97402	
	300	84286	
	300	96087	
(4)	300	120125	12
	300	97402	

Table 4.5 Computer Results of Optimum Stratified Sampling (Linear Cost Function)

Number of Iterations k	Value of r	Values of n_1	n_2	n_3	n_4	Value of f(n)	Value of P(n)
0		100	100	100	100	Infeasible	
1	3000	3655	4021	4750	4584	17011	405591
2	750	1910	2096	2472	2385	8864	110082
3	187.5	1038	1133	1333	1285	4791	32132
4	46.68	603	652	763	734	2754	10608
5	11.72	388	413	478	458	1737	4009
6	2.93	282	293	334	318	1230	2102
7	0.73	233	225	261	247	977	1321
8	0.183	211	206	224	209	851	1000
9	0.045	201	192	204	189	789	857
10	0.011	197	186	195	179	757	790
11	0.0028	195	182	190	174	742	758
12	0.007	194	181	187	171	734	742
13	0.0018	193.8	180.3	186	170	730	734
14	0.000044	193.8	179.9	185.6	169.3	728	730
15	0.000011	193.6	179.7	185.3	169	727.6	728.6
16	0.000002	193.4	179.6	185.1	168.8	727.1	727.6
17	0.0000006	193.4	179.5	185.0	168.7	726.9	727.1
18	0.0000001	193.4	179.5	185.0	168.7	726.8	726.9
19	0.00000004	193.4	179.5	185.0	168.7	726.7	726.8

Values of ε' and ε are choosen to be 1×10^{-5}. Table 4.6 gives the detailed output of the computer program for the nonlinear cost function. r_0 was taken as 3000 while ε' and ε are selected as 1×10^{-5}.

4.6 DISCUSSION AND CONCLUDING REMARKS

As seen from Table 4.3, a wide range of starting points yield the same optimal values. The average computation time was observed to be 30 seconds. In the second problem we have a nonlinear objective function with nonlinear constraints. Even in this case various starting points were used to study the reaction. Again it was found that convergence to the optimum is independent of the starting points. The average computational time was observed to be 40 seconds.

From Table 4.5 we can see the minimum cost of inspection with the linear cost function to be \$726.7. This optimum is attained in 19 iterations. The starting point choosen was infeasible and hence the program locates a starting point which is feasible. In this case it is the point [3655, 4021, 4750, 4584] with an objective function of 17011. By observing the rate of reduction of this objective function, we can see that the starting value of 17011 reduces to a near optimal value of 730 in 13 iteration. After 19 iterations this converges to the optimal value of 726.7. The same value is obtained by all the four starting points selected.

Table 4.5 gives similar results for the nonlinear cost function, the minimum cost of inspection being \$1259.

Table 4.6 Computer Results of Optimum Stratified Sampling (Nonlinear Cost Function)

Number of Iteration k	Value of n	Value of n_1	Value of n_2	Value of n_3	Value of n_4	Value of f(n)	Value of P(n)
0		100	100	100	100	Infeasible	
1	3000	84286	96017	120128	97402	1250	374216
2	750	84289	96094	120137	97402	1259	94498
3	187	84289	96094	120137	97402	1259	24568
4	46.8	84289	96094	120137	97402	1259	7086
5	11.7	84289	96094	120137	97402	1259	2715
6	2.9	84289	96094	120137	97402	1259	1623
7	0.73	84289	96094	120137	97402	1259	1350
8	0.18	84289	96094	120137	97402	1259	1281
9	0.04	84289	96094	120137	97402	1259	1264
10	0.001	84289	96094	120137	97402	1259	1260
11	0.00028	84289	96094	120137	97402	1259	1259.37
12	0.00007	84289	96094	120137	97402	1259	1259

REFERENCES

1. Bracken, J. and G. P. McCormick, Selected Applications of Nonlinear Programming, John Wiley, New York, 1968.

2. Cochran, W. G., Sampling Techniques, 2nd Ed., John Wiley, New York, 1963.

3. Cyert, R. M. and H. J. Davidson, Statistical Sampling for Accounting Information, Printice Hall, New Jersey, 1962.

4. Dalenius, T., Sampling in Sweeden, Contributions to the Methods and Theories of Sample Survey and Practice, Almqvist and Wiksell, Stockholm, 1957.

5. Dalenius, T., "The multivariate sampling problem", Skand Aktuar Tidskr, Vol 36, 92-102 (1957).

6. Hess, I., ET AL, "Stratification- A Practical Investigation", Journal of American Statistical Association, 74-90 (1966).

7. Jagannathan, R., "The Programming Approach in Multiple Character Studies", Econometrica, Vol. 33, 236-237 (1965).

8. Kokan, A. R., "Optimum Allocation in Multivariate Surveys", Royal Statistical Society Journal, Vol 126 (1963).

9. Norbotten, S., "Allocation in Stratified Sampling Means of Linear Programming", Skand. Aktrarietidskr., Vol 39, 1-6 (1956).

10. Raj. D., Sampling Theory, McGraw-Hill, New York, 1968.

11. Yamane, T., Elementary Sampling Theory, Printice Hall, New Jersey, 1967.

12. Yates, F., Sampling methods for Census and Surveys, Griffin, London, 1960.

5. PARAMETER ESTIMATION OF THE PREVENTATIVE SAMPLING MODEL

5.1 INTRODUCTION

Engineers frequently propose mathematical models or equations which attempt to describe the relationship between physically measured variables. Such equations may be derived from physical laws directly or they may be simply convenient forms for summarizing tabulated data obtained by what is known as 'curve fitting'. These models usually contain parameters or coefficients whose values are unknown. The object of parameter estimation is to find the values of these parameters that cause the proposed equation to give the best possible fit to the data. The closeness of fit depends on the differences between the observed values of certain variables, and the values predicted by the equations. There are various criterions as to what constitutes the "best fit". The simplest and hence the most common is the sum of squares of the differences between the observed values from the data and the predicted values from the model.

We characterize the models we shall deal with by specifying the kinds of variable that appear, and the forms of the equations relating them. Thus a mathematical model may be a differential equation or an algebraic one. Also we speak of linear or nonlinear models which refers to the manner in which the unknown parameters enter the equation.

A mathematical model can explicitly written in the form:

$$y = f(x_i, p_j), \tag{5.1}$$

where y is the measured variable, x_i, i = 1, ..., n are the measured independent variable, and p_j, j = 1, ..., P are parameters to be estimated.

Then, if equation (5.1) can be represented as

$$y = p_1 z_1 + p_2 z_2 + \ldots + p_P z_P, \tag{5.2}$$

where, $z_1, z_2, \ldots, z_P$ are functions of x_i's only, the model is said to be linear with respect to the parameters. Models which cannot be expressed in the form given by equation (5.2) are classified as non-linear model. Here we deal with algebraic nonlinear models.

Ideally speaking, one would like to find values of p_j's of equation (5.1) that will satisfy equations exactly for each experiment. Due to errors in measurement and inaccuracies in the model, we cannot, however, hope for an exact fit. Thus in equation (5.1) there exists a difference between y and $f(x_i, p_j)$. We may write,

$$u = f(x_i, p_j) - y,$$

where u, often referred to as the residual, represents the departure of the predicted values $f(x_i, p_j)$ from the observed values y. The task of parameter estimator is to find values of p_j which minimizes some appropriate function of the u. We shall confine ourselves with the least square method of parametric estimation.

In the least square method, the sum of squares of deviations of pre dicted values from measured values is minimized.

It may be pointed out, here, that the least square is not very valid if several variables are observed at each experiment. It does not make sense to add together sums of squares of say pressures and temperatures. This problem may be overcome by assigning a weight factor to each variable and minimizing the weighted sum of squares. However, in general, when several variables are measured, it is preferable to resort to other methods.

We minimize the objective function ϕ given by

$$\phi = \sum_{u=1}^{N} (y_u - \hat{y}_u)^2$$

where, y_u = measured value of the dependent variable

$\hat{y}_u$ = estimated value of the dependent variable

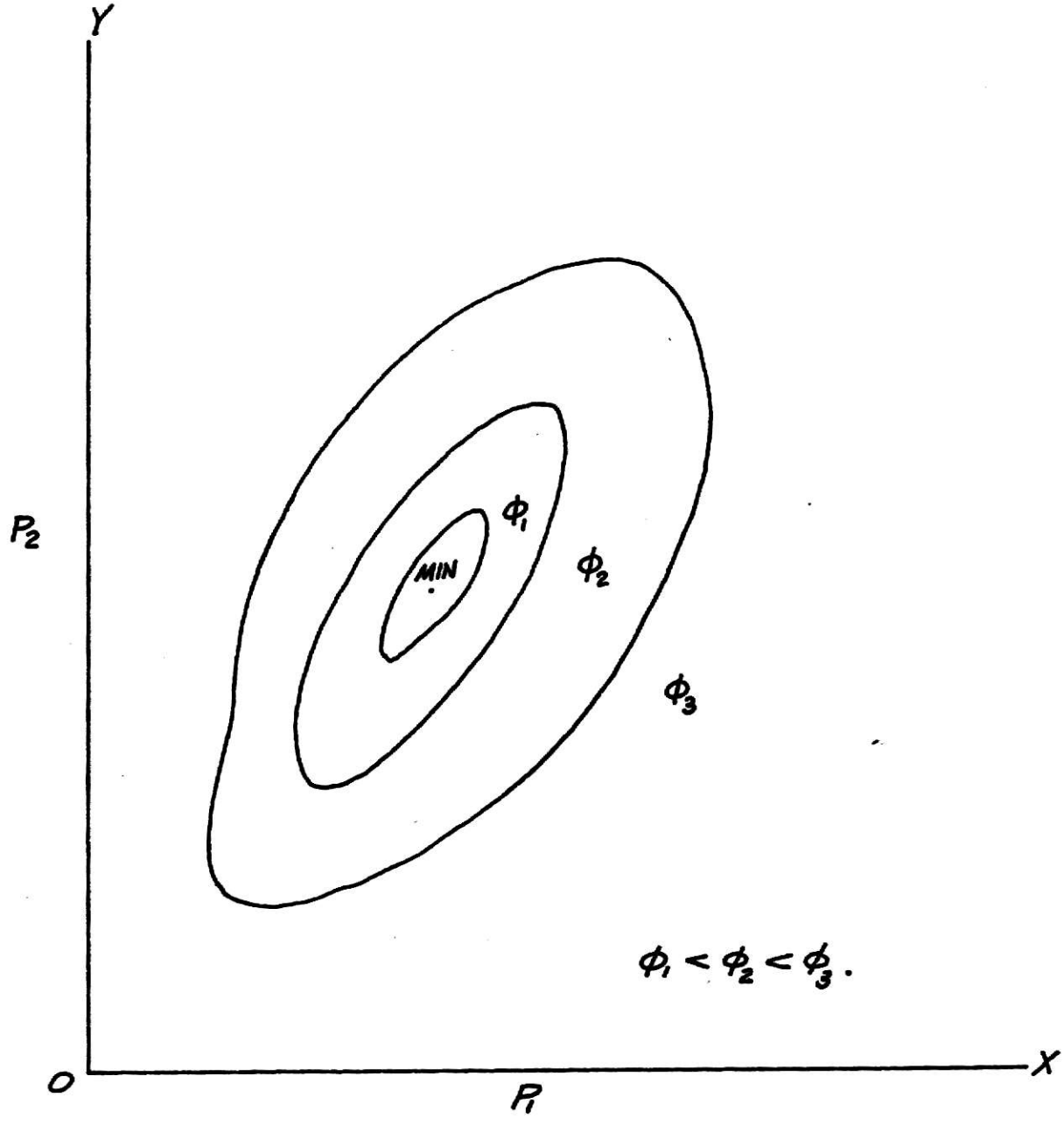
u = 1, ..., N, denotes the number of data points

The contours for the above objective function for a 2-parameter model can be visualized as in Fig. 5.1. We observe that in the case of nonlinear models, these contours are highly distorted except in regions near the minimum, where they may be approximately circular.

5.2 METHODS OF PARAMETER ESTIMATION

Most algorithms for estimating least squares for nonlinear models are centered around two basic approaches. On the one hand, the function may be expanded in terms of the Taylor's series and corrections on several parameters calculated on the assumption of local linearity. On the other, modifications based on the steepest descent methods have been used. The first method is known for its slow convergence if the initial guess is poor, but has the advantage of quick convergence in the vicinity of optimum. The second method has rapid convergence when the starting point is far from the optimum, but converges very slowly in the neighborhood of the minimum.

Marquardt's method is an optimum interpolation between these two methods. The mathematical basis for the Gauss and steepest descent method are given before presenting Marquardt's modification.



CONTOURS OF
TWO - PARAMETER MODEL .

FIG. 5.1 .

(a) Gauss Method.

This is an iterative approach based on the linearization of a function, $f(\bar{x}_u, \bar{p}) = y_u$ by a truncated Taylor series expansion about $\bar{p}^n$. The vector $\bar{p}^n$ represents the parameter vector after the n^{th} iteration: subscript u refers to a specific data point.

An expansion around $(\bar{x}_u, \bar{p}^n)$ gives

$$f(\bar{x}_u, \bar{p}^{n+1}) = f(\bar{x}_u, \bar{p}^n) + \sum_{j=1}^{P} \left(\frac{\partial f_u}{\partial p_j}\right) \delta_{tj}^n \tag{5.3}$$

where,

$$\delta_{tj}^{n\dagger} = p_j^{n+1} - p_j^n \,. \tag{5.4}$$

Then,

$$\phi = \sum_{u=1}^{N} \left[y_u - f(\bar{x}_u, \bar{p}^n) - \sum_{j=1}^{P} \left(\frac{\partial f_u}{\partial p_j}\right) \delta_{tj}^n\right]^2 \tag{5.5}$$

At the minimum,

$$\frac{\partial \phi}{\partial p_j} = 0. \tag{5.6}$$

Differentiating equation (5.5) with respect to p_j we get,

$$\frac{\partial \phi}{\partial p_j} = -\,2 \sum_{u=1}^{N} \left[y_u - f(\bar{x}_u, \bar{p}^n) - \sum_{j=1}^{P} \frac{\partial f_u}{\partial p_j} \delta_{tj}^n\right] \frac{\partial f_u}{\partial p_j} = 0 \tag{5.7}$$

---

†Subscript 't' on δ refers to the direction given by Taylor's series expansion. Subscript ' ' refers to the steepest descent method, non subscripted δ, to be used later, refers to the Marquardt's method.

Equation (5.7) is linear in δ_{tj}^{n}.

By defining $B_{uj} = \frac{\partial f_u}{\partial p_j}$, $u = 1, 2, \ldots, N$; $j = 1, 2, \ldots, P$, and

$$\bar{f}^n = \begin{bmatrix} f(\bar{x}_1, \bar{p}^n) \\ f(\bar{x}_2, \bar{p}^n) \\ \vdots \\ f(\bar{x}_u, \bar{p}^n) \end{bmatrix}, \qquad \delta_t^n = \begin{bmatrix} p_1^{n+1} - p_1^n \\ p_2^{n+1} - p_2^n \\ \vdots \\ p_P^{n+1} - p_P^n \end{bmatrix}$$

the correction vector δ_t^n can be found from equation (5.7) as given below

$$\delta_t^n = [B^T \; B]^{-1} \; B^T \; (\bar{y} - \bar{f}^n) \tag{5.8}$$

$$= A^{-1} \; g \tag{5.8a}$$

where,

$$A = B^T B$$

$$g = B^T(\bar{y} - \bar{f}^n)$$

Assuming initial parameter values, we determine the correction vector δ_t, which is used for getting a better parameter value for the next iteration. Thus we have, $p_j^{n+1} = p_j^n + \delta_{tj}^n$. The iterations are carried on until the correction vector, δ_t, becomes small enough. This method converges rapidly in the vicinity of the minimum. However if the initial guess is far from the minimum, the method may converge slowly, oscillate widely or even diverge.

(b) Steepest Descent Method.

The method of steepest descent simply moves from the current trial value, in the direction of the negative gradient of ϕ. The components and direction of the correction vector at successive iterations is given by

$$\delta_g = - \left(\frac{\partial \phi}{\partial p_1} , \ldots, \frac{\partial \phi}{\partial p_p} \right) \Bigg/ \left(\sum_{j=1}^{P} \left(\frac{\partial \phi}{\partial p_j} \right)^2 \right)^{1/2}$$

Thus, a search is made in the direction of the negative gradient of ϕ. The step size is so chosen as to give the minimum ϕ in the direction of δ_g. In effect this method seeks to calculate corrections such that at each iteration the value of ϕ will decrease most rapidly.

This method is successful for highly nonlinear contours of ϕ. The method converges very rapidly initially. However, in the vicinity of the minimum the convergence is very slow.

5.3 MARQUARDT'S MODIFICATION (2,4,6)

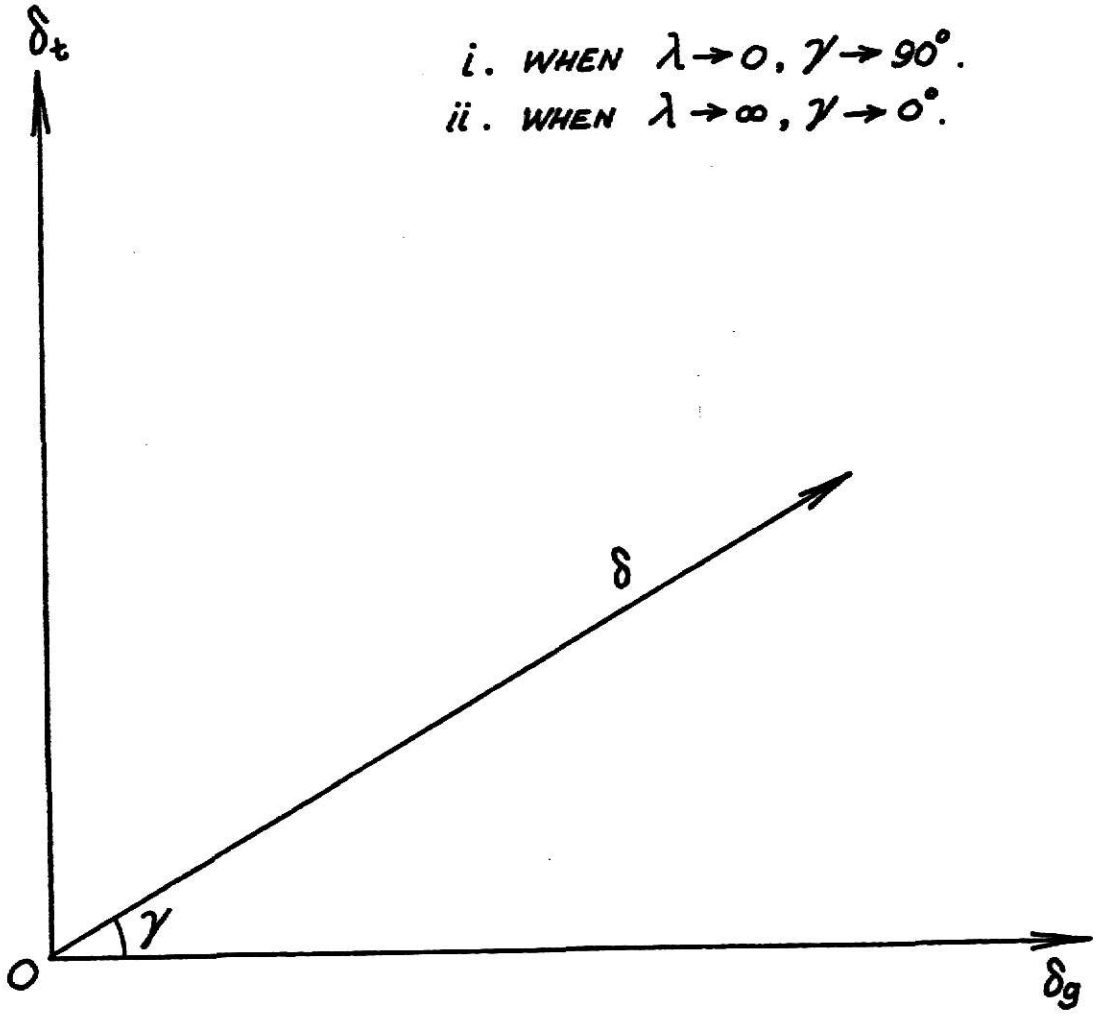
Gauss method has the advantage of quick convergence in the vicinity of optimum and its main draw back is its failure to converge when the initial guess is poor. On the other hand, gradient method has rapid convergence when the starting point is far from the optimum, but once the near optimum value is obtained, the convergence to the optimum is extremely slow. Marquardt's method represents a compromise between the linearization (or the Gauss) method and the steepest descent method, and combines the best features of both, while avoiding their most serious limitations. Marquardt has devised an algorithm in such a manner as to take advantage of the fast convergence of the steepest

descent method in the early stages and that of the Gauss method as the search gets closer to the minimum (4).

The idea of this method can be explained briefly as follows. Suppose we start from a certain point in the parameter space. If the method of steepest descent is applied, a certain vector direction δ_g, where g stands for gradient, is obtained for movement away from the initial point. By the nature of the contours of the objective function, this may be best local direction in which to move to attain smaller values of the objective function, ϕ, but may not be the best overall direction (2). However, the best direction to move should be within 90^o of the negative gradient of ϕ. Otherwise, ϕ will increase successively. The Gauss (or the Taylor series) method leads to another correction vector δ_t, where t stands for the Taylor series, which is given by equation (5.8). Marquardt found that for a number of practical problems he studied, the angle between δ_g and δ_t fell in the range of 80^o and 90^o, and thus these two directions are at right angles to each other as depicted in Fig. 5.2. From these considerations, it would seem reasonable that any improved method will in some sense interpolate between δ_g and δ_t. The vector, δ, in Fig. 5.2 represents the correction vector of Marquardt's method. The algorithm is so constructed that its position will depend on the value of a constant λ. By a proper choice of λ at each stage the best direction δ is obtained.

The theoretical basis for the algorithm is based on the following three theorems (4).

Theorem 1. Let $\lambda \geq 0$ be arbitrary and let δ_0 satisfy the equation



RELATIVE POSITIONS OF CORRECTION VECTORS OF GAUSS AND STEEPEST DESCENT METHODS.

FIG. 5.2.

$$(A + \lambda I)\delta_0 = g \tag{5.9}$$

Then, δ_0 minimizes ϕ on the sphere whose radius $||\delta||$ satisfies

$$||\delta||^2 = ||\delta_0||^2$$

Recall that A and g are defined in equation (5.8a).

Theorem 2. Let $\delta(\lambda)$ be the solution of equation (5.9) for a given value of λ. Then $||\delta(\lambda)||^2$ is a continuous decreasing function of λ, such that as $\lambda \to \infty$, $||\delta(\lambda)||^2 \to 0$.

Theorem 3. Let γ be the angle between δ_0 and δ_g. Then γ is a continuous monotone decreasing function of λ such that as $\lambda \to \infty$, $\gamma \to 0$. Since δ_g is independent of λ, it follows that δ_g rotates towards δ_g as $\lambda \to \infty$.

The broad outline of the appropriate algorithm is as follows. Specifically at the n^{th} iteration the equation

$$(A^n + \lambda^n I)\ \delta^n = g^n \tag{5.10}$$

is constructed and solved for δ^n. The new trial vector

$$p^{n+1} = p^n + \delta^n$$

will lead to a new sum of squares ϕ^{n+1}. Since our objective is to minimize this objective function, λ is chosen so that

$$\phi^{n+1} < \phi^n.$$

The key element which changes the nature of approach of Marquardt's method is the value of the constant, λ. The two important properties of

λ are (1) its effect when it has a low value, and (2) its effect when it attains larger values. To study the effect of λ attaining a low value, let us rewrite the equation (5.8), for the correction vector for the Gauss method,

$$\delta_t^n = (B^T B)^{-1} B^T (\bar{y} - f^n) \tag{5.8}$$

The corresponding relationship in Marquardt's method is

$$\delta^n = (B^T B + \lambda I)^{-1} B^T (\bar{y} - f^n) \tag{5.11}$$

By comparing equation (5.8) with equation (5.11), we find that as λ tends to zero, equation (5.11) approaches the value of equation (5.8). Thus low values of λ make the Marquardt's method similar to that of Gauss. When the starting point is very near the optimum, Gauss method is more efficient and at such cases we have to ensure to use low values for λ.

To study the effect of larger values of λ, we find from Theorem 3, that as $\lambda \to \infty$, $\gamma \to 0$. From Fig. 2, we can see that larger values of λ essentially makes the Marquardt's method similar to that of steepest descent method. Thus by reducing or increasing the value of λ, the basic equation (5.10) is made to be an optimal interpolation between the Gauss and the steepest descent methods. The problem reduces to finding a methodical procedure which makes λ to attain smaller values near the optimal value and to increase λ while the optimum value is far away. The procedure of the selection of λ is given below. Procedure for the selection of λ [4]:

Let $\nu > 1$.

Let λ^{n-1} denote the value of λ from the previous iteration. Initially

let $\lambda^{(0)} = 10^{-2}$, say.

Compute $\phi^{n+1}(\lambda^{n-1})$ and $\phi^{n+1}(\lambda^{n-1}/\nu)$.

(i) If $\phi^{n+1}(\lambda^{n-1}/\nu) \leq \phi^n$, let $\lambda^n = \lambda^{n-1}/\nu$.

(ii) If $\phi^{n+1}(\lambda^{n-1}/\nu) > \phi^n$, and $\phi^{n+1}(\lambda^{n-1}) \leq \phi^n$, let $\lambda^n = \lambda^{n-1}$.

(iii) If $\phi^{n+1}(\lambda^{n-1}/\nu) > \phi^n$, and $\phi^{n+1}(\lambda^{n-1}) > \phi^n$, increase λ by successive multiplication by ν until for some smallest w, $\phi^{n+1}(\lambda^{n-1}\nu^w) \leq \phi^n$. Let $\lambda = \lambda^{n-1}\,\nu^w$.

In practice, $\nu = 10$ has been found to be a good choice.

Stopping Criterion:

The solution is said to be converged when

$$\frac{\delta_j^n}{\tau + p_j^n} < \varepsilon \quad \text{for all } j,$$

for some suitably small $\varepsilon > 0$, say 10^{-5} and τ, say 10^{-3}.

As stated above let us consider a low value of 10^{-2} for λ. If the starting point is near the optimum, either condition (i) or (ii) will be satisfied, and thus λ will be kept at a low value, operating the Marquardt's method as Gauss method. On the contrary, if the starting value is far from the optimum, condition (iii) will be satisfied, and this will result in an increase of λ. Higher value of λ transforms Marquardt's method to have the steepest descent approach which is the better approach when the optimum is far away.

A computer program entitled "Nonlinear Least Squares by D. W. Marquardt" is available. Its SHARE number is 309401 [5]. The program is written in FORTRAN IV. With some minor modifications this was run on an IBM 360/50 computer.

5.4 MATHEMATICAL MODEL:

The relation between the fraction of sampling, θ^n, and the probability of defect $p^n(\theta^n)$ as given by equation (3.1) is taken as the mathematical model for the parametric estimation. Considering only one product equation (3.1) can be written as:

$$p(\theta) = A + Be^{-C\theta}, \tag{5.12}$$

where A, B and C are the parameters to be estimated in the above nonlinear equation. As no experimental data are available, data points were generated using random numbers and the standard deviation of errors. Values of A, B and C were taken from the data of the numerical example solved in Section 3, to generate the data points. 50 data points were generated using varying standard deviations.

Parametric estimation has been done using Marquardt's method. Different initial guesses for the parameters were used to study the effect. The program was written in FORTRAN and was run in IBM 360/50 in FORTRAN IV G LEVEL. It was observed that with even high values of standard deviation and very poor initial guesses, the convergence to the correct values were obtained, which shows the power of the method.

Bard's method of nonlinear parameter estimation was also used for parameter estimation of this model using the same data points. The results compare well with the ones obtained by Marquardt's method. Results obtained by these two methods indicating the number of iterations taken to converge are tabulated in Tables 5.1, 5.2 and 5.3. It was observed that Marquardt's method is better than Bard's method considering the number of iterations and the computer time for execution.

Table 5.1 Computer results of parametric estimation on preventative sampling model by Marquardt's method.

[correct value of the parameters are A = 0.02, B = 0.2 and C = 25.0].

No.	Initial Values	Mean of Error	Std. Dev. of Error	Final Values	No. of Iterations
(1)	A = 0.025 B = 0.25 C = 20.0	0.001	0.002	A = 0.0202 B = 0.206 C = 25.65	4
(2)	"	"	0.003	A = 0.0198 B = 0.209 C = 25.99	9
(3)	"	"	0.004	A = 0.0194 B = 0.212 C = 26.34	13
(4)	A = 0.04 B = 0.1 C = 30.0	"	0.002	A = 0.020 B = 0.205 C = 25.65	18
(5)	"	"	0.003	A = 0.019 B = 0.204 C = 19.44	19
(6)	"	"	0.004	A = 0.018 B = 0.208 C = 19.54	21

Table 5.2 Computer results parametric estimation on preventative sampling model by Bard's method:

[correct values of the parameters are A = 0.02, B = 0.2 and C = 25.0]

No.	Initial Values	Mean of Error	Std. Dev. of Error	Final Values	No. of Iterations
(1)	A = 0.025 B = 0.25 C = 20.0	0.001	0.002	A = 0.0202 B = 0.206 C = 25.64	31
(2)	"	"	0.003	A = 0.0198 B = 0.209 C = 25.98	31
(3)	"	"	0.004	A = 0.0197 B = 0.2123 C = 26.33	31
(4)	A = 0.04 B = 0.1 C = 30.0	"	0.002	A = 0.0202 B = 0.205 C = 25.66	26
(5)	"	"	0.003	A = 0.0198 B = 0.209 C = 26.0	25
(6)	"	"	0.004	A = 0.0194 B = 0.212 C = 26.35	25

Table 5.3 Comparitive results of parametric estimation by Marquardt's and Bard's method.

[correct values of the parameters are A = 0.02, B = 0.2 and C = 25.0]

No.	Final Values of Parameters		No. of Iterations	
	Marquardt's	Bard's	Marquardt's	Bard's
(1)	A = 0.0202 B = 0.206 C = 25.65	A = 0.0202 B = 0.206 C = 25.64	4	31
(2)	A = 0.0198 B = 0.209 C = 25.99	A = 0.0198 B = 0.209 C = 25.98	9	31
(3)	A = 0.0194 B = 0.212 C = 26.34	A = 0.0197 B = 0.2123 C = 26.33	13	31
(4)	A = 0.0200 B = 0.205 C = 25.65	A = 0.0202 B = 0.205 C = 25.66	18	26
(5)	A = 0.019 B = 0.204 C = 19.44	A = 0.0198 B = 0.209 C = 26.0	19	25
(6)	A = 0.019 B = 0.208 C = 19.54	A = 0.0194 B = 0.212 C = 26.35	21	25

A study was made to find the effect of error on the parameter estimation. Using a starting value of 0.04, 0.1, and 30, Bard's method was used to estimate the parameters. Using the equation (5.12), the probability of defect for various sampling fractions were calculated. To this normally distributed error of a particular standard deviation and 0.001 mean was added. This error was generated using the subroutine GAUSS (1). The resulting numerical values were then considered as experimental data and the nominal parameter values described by these data are of course known. Table 5.4 shows the results for various standard deviations, with 10, 20, 30, 40 and 50 data points and with a mean of 0.001.

For each distribution, five sample populations were selected from the same distribution to allow some consideration on the effect of different sample populations. In each run the value of the sample fraction is allowed to vary from 0 to 1 by generating random numbers. Accordingly, information based on the survey of the effect of error and number of data points on this mathematical model for a normal distribution of error can be obtained.

5.5 DISCUSSION AND CONCLUDING REMARKS

The set of data points consist of the independent variable, θ, and the dependent variable, $p(\theta)$. Random numbers are generated to obtain various values of θ whose value varies from 0 to 1. For each value of θ, $p(\theta)$ is calculated using the relation

$$p(\theta) = 0.02 + 0.2\ e^{-(25.0)(\theta)}.$$

The error which is calculated using the mean and the standard deviation is added to this calculated value of $p(\theta)$ for the value of

Table 5.4 Parameter Estimates for the Three Parameter Preventative Sampling Model Based on Generated Data.*

mean of error = 0.001

Run No.	Sample Population	No. of Data Points	Estimated Values P_1	P_2	P_3	$Sx10^2$	No. of evals.
	σ = 0.0002						
1	1	10	0.02	0.16	23.38	$0.17x10^{-5}$	59
2	2	20	0.02	0.20	25.02	$0.14x10^{-4}$	49
3	3	30	0.02	0.19	24.86	$0.2\ x10^{-4}$	51
4	4	40	0.02	0.19	24.95	$0.63x10^{-4}$	51
5	5	50	0.02	0.20	25.00	$0.85x10^{-4}$	53
	σ = 0.002						
6	1	10	0.02	0.04	14.08	$0.16x10^{-3}$	208
7	2	20	0.02	0.20	25.15	$0.14x10^{-2}$	45
8	3	30	0.02	0.17	23.6	$0.2\ x10^{-2}$	57
9	4	40	0.02	0.18	24.41	$0.63x10^{-2}$	57
10	5	50	0.02	0.20	25.6	$0.85x10^{-2}$	51
	σ = 0.01						
11	1	10	0.03	-0.73	22.8	0.26	23
12	2	20	0.017	0.22	25.5	0.035	35
13	3	30	0.018	0.12	19.3	0.05	79
14	4	40	0.015	0.11	20.7	0.15	121
15	5	50	0.017	0.23	28.6	0.21	39
	σ = 0.02						
16	1	10	0.038	-0.80	23.14	0.39	14
17	2	20	0.014	0.24	25.90	0.14	27
18	3	30	0.015	0.09	16.22	0.20	89
19	4	40	0.009	0.04	10.94	0.63	374
20	5	50	0.013	0.28	33.68	0.85	40

*Values of the parameters used to generate data:- $P_1 = 0.02$, $P_2 = 0.20$, $P_3 = 25.0$

the dependent variable. Errors of different standard deviations were introduced in the data points. As can be seen from Tables 5.1 and 5.2, smaller standard deviations converge to better estimations. Different starting points were used to study the effect. From Tables 5.1 and 5.2, it could be concluded that a better initial guess results in a better estimation of the parameters. The average computer time taken by Marquardt's method was 40 seconds. The average computer time taken for each problem by Bard's method was found to be about 80 seconds. From Table 5.3 it could be seen that Marquardt's method is a better method than the Bard's method.

Table 5.4 gives the computer results of four different standard deviations namely 0.0002, 0.002, 0.01, 0.02. The mean of the error was kept at 0.001. For each standard deviation five sets of data points namely 10,20,30,40 and 50 were used. When the standard deviation was 0.0002, or in other words, when the error is smaller, the value of parameters converges to the correct value. When the standard deviation exceeds 0.01 convergence to the correct value was not obtained. This may be taken as the limit for the experimental error that could be allowed while collecting the data.

As to the number of data points, at the standard deviation of 0.0002, the results indicates that 10 data points are not enough to obtain correct estimation of parameters. For this standard deviation, 20 data points gives a good estimation. Increasing the data points beyond 20 points does not result in a better estimation. When the standard deviation is 0.02, which represents larger error, the program does not converge to the correct value whatever be the number of data points.

NOMENCLATURE:

$$A^{(pxp)} = B^T B$$

$$A^* = (a_{jj}^*) = \frac{a_{jj'}}{\sqrt{a_{jj'}} \cdot \sqrt{a_{j'j}}}$$

$$B_{uj} = \frac{\partial f_u}{\partial p_j} , \; u = 1, \ldots, N; \quad j = 1, \ldots, p.$$

$$g^* = g_j^t = \left(\frac{g_i}{\sqrt{a_{jj}}}\right) .$$

$$g^{(px1)} = B^T (\bar{y} - f^n)$$

N = no. of data points

p_j = parameter to be estimated

p = no. of parameters.

x_i = Measurable independent variables

y = Measured dependent variable

y_u = Measured value of the dependent variable

$\hat{y}_u$ = estimated value of the dependent variable.

γ = angle between δ_0 and δ_g ve, the directions between the steepest descent direction and the Marquardt's direction

δ_t = correction vector applied to parameters in the Gauss method using Taylor's expansion

δ_g = correction vector used in steepest descent method

δ or δ_0 = correction vector used in Marquardt

δ_{tj}^n = correction vector in Gauss method, for the j^{th} component used in n^{th} iteration of the process

$\delta_j^* = \delta_j \cdot \sqrt{a_{jj}}$

s = objective function - term of the squares of the difference between the value of labor and the value derived from the model

s^n = value of the objective function at the n^{th} iteration

ν = a constant greater than 1 used in the calculation of λ, (normally = 10)

τ = constant used in the calculation of stopping criterion (choosen as 10^{-3})

ε = constant used in the calculation of stopping criterion (choosen as 10^{-5})

REFERENCES

1. Bard, Y., "Nonlinear Parameter Estimation and Programming", IBM Publication (1967).

2. Draper, N.R., and H. Smith, Applied Regression Analysis, John Wiley, New York, 1966.

3. Marquardt, D. M., "Solution of Nonlinear Chemical Engineering Models", Chem. Engr. Prog., Vol. 55, pp 65-70 (1959).

4. Marquardt, D. W., "An Algorithm for Least Square Estimation of Nonlinear Parameters", Journal of Society of Industrial and Applied Mathematics, Vol. 2, pp 431-441 (1963).

5. Marquardt, D. W., "Least Square Estimation of Nonlinear Parameters", A computer program in Fortran IV language, IBM Share Library, Distribution No. 309401 (1966).

6. Stemetz, Ph.D., Thesis, "Nonlinear Parameter Estimation and Sensitivity Analysis for Chemical Process Systems", Kansas State University, 1969.

6. RECOMMENDATION FOR FURTHER STUDY

1. Including the Cost of Replacement in the Optimum Preventative Sampling

The total cost equation in the preventative sampling, takes into consideration the two costs, namely, (1) the cost of inspection and (2) the expected cost of the undetected defective quantity going out. In calculating the second cost it is equated to $a^n v^n p^n(\theta^n)\ (1-\theta^n)$ in equation (3-_). This approach assumes that a defective article could be replaced by a good one at a later stage without any additional cost. In some cases, however, this may involve some additional cost, namely the cost of replacement. This can be equated to $a^n r^n p^n(\theta^n)(1-\theta^n)$, where r^n represents the cost of replacement for each unit of the n^{th} type. The total cost, in such a case, can be written as

$$S = \sum_{n=1}^{N} [I^n a^n \theta^n + a^n r^n p^n(\theta^n)(1-\theta^n) +$$

$$a^n v^n p^n(\theta^n(1-\theta^n)].$$

The optimal policy based on this cost function can be obtained with out difficulties.

2. Additional constraint on Each Product on the Preventative Sampling Problem

The constraint that is considered in the formulation of the preventative sampling, is one which an upper limit for the total number of products inspected is given. In this case, it is assumed that all the products need the same equipment for inspection. But in some cases, each product may be tested in a different equipment. Therefore, each

and in that equipment may has the upper limit of the testing capacity, and this becomes the constraint for each product. In the numerical example, instead of a single constraint $\sum_{n=1}^{N} a^n\theta^n \leq 100$, we will have $a^1\theta^1 \leq a^2\theta^2 \leq k_2$ etc where k_1, k_2, ..., k_N will represent the capacities of testing different products.

3. Statistical Analysis on Parameter Estimation

In the Preventating Sampling Model, further analysis on the effect of varying the standard deviation on the final estimate of the parameters could be carried out. Normal experimental errors vary from 5% to 10% in practical cases. It is possible to find the effect of such range of errors on the accuracy of the final results. Such an analysis may show the relationship between the percentage of error in expt. data and the percentage of error in the estimation of parameters. Additional statistical test such as χ^2 can be done on this parameter estimation.

ACKNOWLEDGEMENTS

The author wishes to express his sincere gratitude to his major professor, Dr. C. L. Hwang, for his guidance and assistance during the preparation of this report.

He also wishes to express his appreciation to Dr. S. Konz for his valuable suggestions in this work.

This work was partly supported by NASA Grant No. NGR-17-001-034.

APPENDIX I

The variance of the y_i in a finite population is usually defined as

$$\sigma^2 = \frac{\sum_{1}^{N} (y_i - \bar{Y})^2}{N}$$

As a matter of notation, results are presented in terms of a slightly different expression, in which the divisor (N-1) is used instead of N. We have

$$S^2 = \frac{\sum_{1}^{N} (y_i - \bar{Y})^2}{N-1}$$

We now consider the variance of $\bar{y}$. By this, we mean $E(\bar{y} - \bar{Y})^2$ taken over all ${}_N C_n$ samples.

For an unbiased sampling, by definition

$$E(\bar{y}) = \bar{Y}.$$

$$E[(y_1 - \bar{Y})^2 + (y_2 - \bar{Y})^2 + \ldots + (y_n - \bar{Y})^2]$$

$$= \frac{n}{N} [(y_1 - \bar{Y})^2 + (y_2 - \bar{Y})^2 + \ldots + (y_N - \bar{Y})^2] \qquad (1)$$

It follows that

$$E[(y_1 - \bar{Y}) \cdot (y_2 - \bar{Y}) + (y_1 - \bar{Y})(y_3 - \bar{Y}) + \ldots + (y_{n-1} - \bar{Y})(y_n - \bar{Y})] \qquad (2)$$

$$= \frac{n(n-1)}{N(N-1)} [(y_1 - \bar{Y})(y_2 - \bar{Y}) + (y_1 - \bar{Y})(y_2 - \bar{Y})$$

$$+ \ldots + (y_{N-1} - \bar{Y})(y_N - \bar{Y})] \qquad (3)$$

Squaring (3), we get,

$$n^2E(\bar{y} - \bar{Y})^2 = \frac{n}{N}\{(y_1 - \bar{Y})^2 + \ldots + (y_N - \bar{Y})^2 + \frac{2(n-1)}{N-1}[(y_1 - \bar{Y})(y_2 - \bar{Y}) + \ldots$$

$$+ (y_{N-1} - \bar{Y})(y_N - \bar{Y})]\}$$

$$n^2E(\bar{y} - \bar{Y})^2 = \frac{n}{N}\{(1 - \frac{n-1}{N-1})[(y_1 - \bar{Y})^2 + \ldots + (y_N - \bar{Y})^2] + \frac{n-1}{N-1}[(y_1 - \bar{Y}) +$$

$$\ldots + (y_N - \bar{Y})]^2\}$$

The second term inside the curly bracket vanishes, since the sum of the y_i equals $N\bar{Y}$. Division by n^2 gives,

$$V(\bar{y}) = E(\bar{y} - \bar{Y})^2 = \frac{N-n}{nN(N-1)} \cdot \sum_{i=1}^{N} (y_i - \bar{Y})^2$$

$$= \frac{N-n}{n \cdot N} \cdot S^2 = \frac{S^2}{n}(1-f). \qquad (4)$$

Now we have,

$$(\bar{y}_{st} - \bar{Y}) = \frac{\sum N_h \cdot \bar{y}_h}{N} - \frac{\sum N_h \cdot \bar{Y}_h}{N}$$

$$= \frac{\sum N_h (\bar{y}_h - \bar{Y}_h)}{N} .$$

$$(\bar{y}_{st} - \bar{Y})^2 = \frac{\sum N_h^2 (\bar{y}_h - \bar{Y}_h)^2}{N^2}$$

$$+ \frac{2\sum N_h \cdot N_j (\bar{y}_h - \bar{Y}_h)(\bar{y}_j - Y_j)}{N^2} .$$

$(\bar{y}_j - \bar{Y}_j) = 0$ since $\bar{y}_j$ is unbiased.

$$V(\bar{y}_{st}) = \frac{\sum N_h^2 \, E(\bar{y}_h - \bar{Y}_h)^2}{N^2}$$

$$= \sum W_h^2 \cdot V(\bar{y}_h) \, .$$

Substituting in equation (4), we get,

$$V(\bar{y}_{st}) = \frac{\sum W_h^2 \cdot S_h^2}{n_h} \cdot (1-f_n).$$

$$= \sum_{h=1}^{L} \frac{W_h^2 \cdot S_h^2}{n_h} - \sum_{n=1}^{L} \frac{W_h \, S_h^2}{N} \, .$$

APPENDIX II

The problem is to minimize

$$V(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 S_h^2}{n_h} - \sum_{n=1}^{L} \frac{W_h^2 S_h^2}{N_h}$$

subject to the restriction

$$c_1 n_1 + c_2 n_2 + \ldots + c_L n_L = C - c_o$$

Using the calculus method of Lagrange multipliers, we select the n_h and the multiplier λ to minimize

$$V(\bar{y}_{st}) + \lambda(\sum c_h n_h - C - c_o)$$

$$= \sum \frac{W_h^2 S_h^2}{n_h} - \sum \frac{W_h^2 S_h^2}{N_h} + \lambda(c_1 n_1 + c_2 n_2 + \ldots + c_L n_L - C + c_o).$$

Differentiating with respect to n_h gives the equation,

$$- \frac{W_h^2 S_h^2}{n_h^2} + \lambda c_h = 0 \quad (h = 1, 2, \ldots, L)$$

that is,

$$n_h \cdot \sqrt{\lambda} = \frac{W_h S_h}{\sqrt{c_h}} \tag{1}$$

Summing over all the strata, we get,

$$n\sqrt{\lambda} = \sum \frac{W_h \cdot S_h}{\sqrt{c_h}} \tag{2}$$

Finally, the ratio of (1) and (2) gives

$$\frac{n_h}{n} = \frac{W_h S_h / \sqrt{c_h}}{\sum [W_h S_h / \sqrt{c_h}]}$$

$$= \frac{N_h S_h / \sqrt{c_h}}{\sum (N_h S_h / \sqrt{c_h})} .$$

OPTIMIZATION OF PREVENTATIVE SAMPLING AND STRATIFIED SAMPLING

by

THIRUVENGADAM JANAKIRAMAN

B.E. (Hons) (Mechanical), Madras University

Madras, India, 1961

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1970

This report concerns the optimal allocation of resources in sampling inspection among different products. The sampling inspection is considered to produce a reaction which prevents the occurance of faults in the future, so that it could be termed as the preventative sampling. The criterion for optimality is the minimization of the sum of the expected cost of undetected defective articles and the total cost of inspection. The problem is formulated in the form of a nonlinear programming problem. The sequential unconstrained minimization technique (SUMT), which is considered as one of the simplest and most efficient method for solving the constrained nonlinear optimization problem, is used to obtain an optimal sample plan for five products.

Parameter estimation is carried out on the preventative sampling model which assumes that the probability of defects is affected by the sample size selected. Two methods, namely, Marquardt's method and Bard's method are used for the parameter estimation and the results obtained by the two methods are compared.

Problems of optimal sample size allocations in stratified sampling is formulated. The problems are to find the sampling plan which minimizes the inspection cost subject to the constraints that the variances for the two variates are equal to or less than the specified values. Two numerical problems which have four strata and two variates are solved by using the sequential unconstrained minimization technique (SUMT).