

CHAPTER - I

INTRODUCTION

This thesis on "Some Contributions to Acceptance Sampling and Simulation" consists of two parts, (i) Curtailed Sampling Plans by Attributes and (ii) An Investigation into Expected Value and Quantile Methods of Probability Plotting by Simulation.

1.1 Introduction to Part I: Curtailed Sampling Plans by Attributes.

1.1.1 Sampling inspection is of two kinds: namely, lot-by-lot sampling inspection and continuous sampling inspection. In the former, items are formed into lots, a sample is drawn from the lot, and the lot is either accepted or rejected on the basis of the quality of the sample. This is most appropriate for acceptance inspection. In continuous sampling inspection, current inspection results are used to determine whether sampling inspection or screening inspection (i.e. 100% inspection) is to be used for the next articles to be inspected. Sampling plans are further classified depending

on whether the quality characteristics are measured and expressed in actual values, i.e., variables inspection, or whether articles are classified only as defectives or non-defectives, i.e. attributes inspection. We have considered in this thesis, lot-by-lot inspection by attributes.

1.1.2 Any lot-by-lot sampling plan has as its primary purpose the acceptance of good lots and the rejection of bad lots. It is important to define what is meant by a good lot. Naturally, the consumer would like all of his accepted lots to be free of defectives. On the other hand, the manufacturer will usually consider this to be an unreasonable request since some defectives are bound to appear in the manufacturing process. If the manufacturer screens the lots, he may eliminate all the defectives, but at the prohibitive cost of screening. This cost will naturally be reflected in his price to the consumer. Ordinarily, the consumer can tolerate some defectives in his lot, provided the number is not too large. Consequently, the manufacturer and the consumer get together and agree on what constitutes good quality. If lots are submitted at this quality or better, the lot should be accepted, otherwise, rejected. Again this is an imposing task and can be accomplished only with the heavy cost of screening. It is at this point that sampling inspection, with its corresponding advantage of reduced inspection costs,

can be instituted. This advantage should not be under-rated. Few manufacturers or consumers, whoever has to bear the cost of inspection, can stay in business very long if all lots are screened. Strictly speaking, perfect quality is rarely achieved, even with screening [4] .

1.1.3 Having fixed the quality of a good lot, the manufacturer tries to run his production such that, by and large, this quality of the lot is maintained. Since the inspection is assumed to be by attributes, the quality of the lot is characterized by its fraction defective (p), which is nothing but proportion of defectives. We assume that the fraction defective remains constant over the entire production run. But p would change as the time goes on because of some wear and tear in the machinery or some other reasons and the results of the inspection obtained during the execution of the sampling plans may be used to assess the fraction defective and to be sure that the quality remains the same.

1.1.4 There are various lot-by-lot acceptance sampling plans where inspection is by attributes, which implies that articles are classified as defectives or nondefectives during the inspection. The plans are (i) single sampling, (ii) double sampling, (iii) multiple sampling etc. A single sampling procedure can be characterized by the following: one random sample of n items is drawn from a lot of N items; the lot is

accepted if the number of defectives in the sample does not exceed c . Here c is referred to as an acceptance number. A double sampling plan is characterized by the following: A sample of n_1 items is drawn from a lot; the lot is accepted if there are no more than c_1 defectives and is rejected if there are more than c_2 defectives. If there are between c_1+1 and c_2 defective items, a second sample of size n_2 is drawn; the lot is accepted if there are no more than c_2 defectives in the combined sample of n_1+n_2 ; the lot is rejected otherwise. The multiple sampling plan is a straight extension of double sampling plan.

1.1.5 Furthermore, having decided to administer a particular sampling plan, one can have one of the two alternatives during the inspection, namely, (i) to complete the inspection procedure as per the statement of the plan goes or (ii) to stop the inspection in advance when one is certainly knowing, during the inspection procedure, that the lot is of the rejectable type or of the acceptable type. The latter situation is known as curtailment of the inspection. For instance, if inspection had no other purpose than to determine which inspection lots to accept and which to reject, it would be feasible to stop inspection as soon as the rejection number is reached or as soon as it is known that the acceptance number will not be exceeded. Curtailment of inspection is possible in the above situations.

1.1.6 Reference to curtailment of the inspection appears in Sampling Inspection Tables Single and Double Sampling by Dodge and Romig as early as in 1944 [19a] . However they have not encouraged curtailment of inspection under single sampling plan making a remark that the data obtained would be biased and ill-suited for easy computation of process average quality. One more objection for not encouraging the curtailment of the inspection by them is that the inspector will have tendency to minimise his work. In case of double sampling plans they do not encourage the curtailment of the inspection for the first sample, but do not mind if the curtailment of the inspection takes place when second sample is being inspected, for they assume that the inspection of the second sample is to be done by a person other than the usual sampling inspector (perhaps a person of superior grade). We may remark, "will not the data in this case too be anticipated to be biased and ill-suited for easy computation of process average quality?"

1.1.7 It appears that estimation of fraction defective under curtailed sampling plan was introduced by Girshick, Mosteller, and Savage (1946) [28] . Their main problem was to find an unique unbiased estimate of the fraction defective under binomial trials. As an application to the theory developed by them, they have considered a curtailed single sampling

plan which takes into consideration the curtailment of the inspection of a lot both at the rejection and the acceptance stages. They have considered the problem of estimation when one lot is inspected and they have considered one trivial case where estimation is based on two lots. One of the problems raised by them, but not solved, is related with the estimation of p when one is faced with the results of several lots. We have here considered this problem of estimation of the fraction defective when there is curtailment of the inspection during the execution of single sampling plan by attributes and the results of the inspection of several lots (any number of lots will do) are on hand.

1.1.8 One of the characteristics of the curtailed sampling plans, as the name itself suggests, is to have a reduction in the average sample number (ASN). Statistical Research Group, Columbia University (1948) [56] have worked out ASN for the curtailed sampling plan. Later on Burr (1957) [5] and Patil (1963) [48] have worked out ASN for a sampling plan subject to curtailment of inspection in the context of their work, for instance, Patil has introduced ASN of curtailed single sampling plan while giving a different outlook to the curtailed sampling plan in terms of inverse (negative) binomial sampling plan. However, it appears from Craig's remark (1968) [14] that curtailed sampling plan has not yet

been given that importance as it ought to have been received, for he remarks that, textbooks usually advise the uses of single and double attributes acceptance sampling plans to complete the inspection of the sample for single sampling or of the first sample for double sampling, even though enough defectives to reject the lot have already been found. Craig [14] has obtained ASN of curtailed sampling plans - single as well as double. We have discussed in this thesis reduction in ASN and its relation with the asymptotic variance of the maximum likelihood estimate of the fraction defective.

1.1.9 The above discussion gives an idea of the areas which have been discussed in this part of the thesis and perhaps the areas still open for further research. The work that has been covered in the first part of the thesis can be summarized as follows:

In Chapter II we have introduced Curtailed Sampling Plans by Attributes, giving scope for curtailment of inspection, the statements of the Plans considered, the definition of random variables associated with these plans, probability functions of the random variables, etc. Two situations, Situations A and B associated with reporting of the inspection results are described. Situation A takes into consideration

that the complete^{te} information of the inspection results is reported. Situation B takes into consideration censored information of Type I, as defined by Gupta [32] . In this Situation four cases arising from four different modes of reporting the results of sampling inspection of Curtailed Sampling Plans are considered. The estimates of the fraction defective by the method of moments are obtained under the two Situations described above.

In Chapter III we have obtained the maximum likelihood estimates of the fraction defective under both the Situations described in Chapter II. We have obtained the asymptotic variances of these estimates and compared their efficiencies. A Numerical example is provided for illustration.

Circumstances arise in practice when an inspector is prone to classify a defective as a nondefective. This leads to misclassified data. In Chapter IV, maximum likelihood estimates of the fraction defective and the probability of misclassification are obtained, when data from curtailed sampling plans are subject to misclassification. The asymptotic variances and covariance of these estimates are derived. A numerical example is provided.

Miscellaneous aspects of Curtailed Sampling Plans are discussed in Chapter V. They are: relation between ASN and the

asymptotic variances of the maximum likelihood estimates, bias, sufficiency, relation with minimum variance bound estimator, bivariate approach of probability functions etc. Relation between ASN and $V(\hat{p})$ is illustrated by a numerical example.

Chapter VI digresses from the main topic of the thesis. In this chapter, the solution of the maximum likelihood equations for estimation of the parameters of a singly truncated Normal distribution is reduced to a single transcendental equation in a simple statistic and a function of the parameters and a Table for obtaining the solutions of the equations is given for a feasible range of statistic values and is illustrated by a numerical example.

1.2 Introduction to Part II: An Investigation into Expected Value and Quantile Methods of Probability Plotting by Simulation.

1.2.1 This part of the thesis is devoted to Simulation Studies of two methods of probability plotting on ordinary graph paper, which are compared by using Monte Carlo Method.

The Monte Carlo technique consists of a new use of an old procedure. The old procedure is "unrestricted random sampling" (selecting items from a population in such a way

that each item in the population has an equal probability of being selected). This 'new' twist consists of using random sampling to play a game with nature or a man-made system in which an experiment is simulated. In essence, the Monte Carlo technique consists of simulating an experiment to determine some probabilistic property of a population of objects or events by the use of random sampling applied to the components of the objects or events.

Just as the discovery of the laws of gravity is attributed in legend to Newton's observation of a falling apple, so the discovery of the Monte Carlo technique goes back to a legendary mathematician observing the perambulation of a drunkard. Each of the drunkard's steps was supposed to have an equal probability of going in any direction. The mathematician wondered how many steps the drunkard would have to take, on the average, to cover a specified distance away from his starting point. This was called the problem of a "random walk". An application of random sampling called "Stochastic Sampling" was developed to solve this problem, but the method was found to have wide practical applications, and was subsequently given the more colourful name, the Monte Carlo technique [7] .

But it is worth noting that the Monte Carlo Method is

not at all novel to statisticians. For more than fifty years, when statisticians have been confronted with a difficult problem in distribution theory, they have resorted to what they have sometimes called "Model Sampling". The process consists of setting up some sort of urn model or system, or drawings from a table of random numbers, whereby the statistic, the distribution of which is sought, can be observed over and over again and the distribution estimated empirically. The theoretical distribution in question is usually a multiple integral over a region in many dimensions, so, in such cases, "Model Sampling" is clearly a Monte Carlo Method of numerical quadrature. In fact, the distribution of "Student's t " was first determined in this way. Many other examples can be found in the pages of Biometrika and the other Statistical Journals (Symposium on Monte Carlo Methods, held at the University of Florida) [43] .

The use of Monte Carlo Method in building up model may be explained as follows:

Consider a new device which contains two parts that eventually fail. These might be a vacuum tube and a condenser. From past tests, say, we know the 'life curve' of each of the parts. What we want to know, however, is the life curve of the device which contains both of these parts. Putting it

in another way, if $f(t)$ represents the life curve of one of the parts, and $g(t)$ represents the life curve of the other, then the life curve of the device is a function of these two life curves, say, $h[f(t), g(t)]$ or simply $h(t)$. The problem is to know $h(t)$ explicitly when $f(t)$, and $g(t)$ are known. Now in some cases $h(t)$ can be derived by mathematical analysis, for example, when $f(t)$ and $g(t)$ are normal probability density functions. In other instances, however, it need not be possible or practicable to evaluate $h(t)$ by mathematical analysis. For instance if $f(t)$ is Weibull & $g(t)$ is lognormal, a solution may not be possible. If this is the case, one can bring Monte Carlo method to solve the problem. The procedure to obtain the solution by this method will be of the type: one draws hundreds of thousands of observations from $f(t)$ using random numbers and match them with random observations drawn from $g(t)$. For such randomly paired observations from $f(t)$ and $g(t)$, one can determine $\min[f(t), g(t)] = m(t)$, say. One can build up $h(t)$ or assess approximately the curve of $h(t)$ by considering the histogram, frequency curve etc. from the observed set of $m(t)$ and so on. A part of this example has been explained in [7] as one of the uses of Monte Carlo Method. One can find the use of Monte Carlo Method in evaluating certain expressions appearing in other topics of Operations Research such as

replacement, queueing, inventory etc. in [7] itself or in any other standard book on Operations Research for instance [54] .

1.2.2 In our work we have used Monte Carlo Method for comparing two probability plotting methods on ordinary graph paper. Chapter VII is devoted to this problem. The two probability plotting methods considered here are Method I: Expected Value Method, Method II: Quantile Method.

Method I is as follows: Let y_1, y_2, \dots, y_n be a random sample from a population. Let the i th ordered observation be denoted by $y_{(i)}$ $i = 1, 2, \dots, n$, such that $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$. Let the problem be to test whether the sample comes from a Normal population. Then let $X_{(i)}$ be the expected value of the i th ordered observation from $N(0,1)$. Plot $[X_{(i)}, y_{(i)}]$ on ordinary graph paper, taking $X_{(i)}$ along x-axis and $y_{(i)}$ along y-axis. If the points fall almost along a straight line one can draw the conclusion that the sample belongs to a normal population.

Method II is as follows: In this method $X_{(i)}$ of the Method I is replaced by x_{p_i} where x_{p_i} is obtained such that

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

where the problem is still the same to test whether the given sample belongs to a Normal population. In this method, to determine x_{p_i} one has to fix p_i . Three alternatives for p_i are studied; namely $p_i = i/(n+1)$, $p_i = (i - \frac{1}{2})/n$, $p_i = (i - 3/8)/(n + \frac{1}{4})$. We have compared these two methods by drawing actually samples from the hypothetical populations. Both complete as well as censored samples are considered; in case of censored sample, censoring is of type II. The populations considered are Normal and Exponential. Consequences of taking different alternatives of p_i also have been investigated. The basis for the above comparison is the estimate of the parameters of the hypothetical population based on the samples drawn and the mean and variance of these estimates. The estimate of the scale parameter of the population is obtained by the slope of the line fitted to points $[X_{(i)}, y_{(i)}]$ and $[x_{p_i}, y_{(i)}]$ and estimate of the location parameter obtained by the intercept of the line fitted on y-axis. The line is fitted by minimizing the sum of squares of the deviations parallel to (i) y-axis and (ii) x-axis. The former estimates are defined as MVD estimates and the latter MHD estimates. Comparison between MVD and MHD estimates is also done. Tables of $X_{(i)}$ and x_{p_i} are prepared for ready reference for $i=1, 2, \dots, n$, $n=10(1)30$ for both Normal and Exponential populations. These tables are given in Appendix.

1.3 Main Results in the Thesis :

1.3.1 Part I :

(i) The maximum likelihood estimate of the fraction defective (p) under curtailed single sampling plan in Situation A i.e. a situation where complete information of the inspection results is provided is

$$\hat{p} = \frac{\text{Total number of defectives noted}}{\text{Total number of articles inspected}}$$

(ii) \hat{p} given in (i) is

- (a) a ratio of two statistics,
- (b) a biased estimate for p ,
- (c) not a sufficient statistic for p ,
- (d) not a minimum variance bound estimate for p .

(iii) Asymptotic variance of \hat{p} given in (i) is inversely proportional to the average sample number.

(iv) In Situation A iteration is required to estimate p by the method of moments.

(v) In Situation B both the methods, method of maximum likelihood and method of moments, require iteration; but in case of method of maximum likelihood the estimating equation is somewhat simpler.

(vi) Misclassified data of curtailed sampling plans are dealt for obtaining MLE of the fraction defective.

(vii) A table for obtaining the MLE for the parameters of the singly truncated normal distribution is prepared.

1.3.2 Part II :

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

(ii) In quantile Methods, $p_i = \frac{i}{n+1}$ should be discarded.

(iii) In Normal population, $p_i = \frac{i-\frac{1}{2}}{n}$ gives estimates with least variance, while $p_i = \frac{i-3/8}{n+\frac{1}{4}}$ gives the least bias in the estimates.

(iv) In Exponential population, we may generally prefer $p_i = \frac{i-\frac{1}{2}}{n}$.

(v) In Normal population, the Expected Value Method gives results similar to those given by the quantile Method with $p_i = \frac{i-3/8}{n+\frac{1}{4}}$, while in Exponential population, the Expected Value Method gives results similar to those given by the Quantile Method with $p_i = (i-\frac{1}{2})/n$.

1.3.3 A Portion of the material of this thesis has been already published in the following papers :

1. Phatak, A.G., and Bhatt, N.M. (1967),
Estimation of the Fraction Defective in Curtailed
Sampling Plans by Attributes, Technometrics, 9, 219-228.
2. Phatak, A.G.(1968),
Misclassified Data from Curtailed Sampling Plans,
Technometrics, 10, 489-496.
3. Phatak, A.G.(1968),
Censored Sampling in Curtailed Sampling Plans by
Attributes, Technometrics, 10, 854-860.
4. Phatak, A.G. (1964),
A Table for Maximum Likelihood Estimates of the
Parameters of the Singly Truncated Normal Distribution,
Journal of the M.S.University of Baroda (Science Number),
XIII, 1-5.