

Determination of Optimal Double Sampling Plan using Genetic Algorithm

Sundram Sampath
Department of Statistics
University of Madras
Chennai, India.
sampath1959@yahoo.com

S.P Deepa
Department of Statistics
SDNB Vaishnav College for Women
Chennai, India
spdeepa_2007@yahoo.co.in

Abstract

Designing double sampling plan requires identification of sample sizes and acceptance numbers. In this paper a genetic algorithm has been designed for the selection of optimal acceptance numbers and sample sizes for the specified producer's risk and consumer's risk. Implementation of the algorithm has been illustrated numerically for different choices of quantities involved in a double sampling plan.

Keywords: *Double sampling plan, Average sample number, Average total inspection and genetic algorithm.*

1. Introduction.

Acceptance Sampling Plans spell out the conditions for acceptance or rejection of the immediate lot inspected. In industries, acceptance sampling plans are necessary for several reasons. The primary two reasons being (i) in the evaluation process of the produced lot, 100% inspection is practically impossible particularly when the quality of the manufactured items can be assessed only on performing tests of destructive nature and (ii) Customers may insist on mandatory sampling procedures, which must be met. Sampling plans can be broadly classified into two kinds, namely, sampling plans for attributes and sampling plans for variables. In this paper, it is aimed to address certain issues related to sampling plans for attributes, more specifically, the design of double sampling plan. Single sampling plan is the basic to all acceptance sampling plans. Implementation of an attributes single sampling plan is very simple. It involves taking a random sample of size n from a lot of size N . The sample may be intended to represent the lot itself (Type A sampling) or the process used to produce the lot (Type B sampling). The number of defectives (or defects) d found in the sample is compared to an acceptance number c . If the number found is less than or equal to c , the lot is accepted. If the number found is greater than

c , the lot is rejected. Analytic procedures are available in Schilling and Neubauer (2008) for determination of single-sampling plans for specified values of

p_0 - Acceptable quality level (AQL)

p_1 - Rejectable quality level (RQL)

α - Producer's risk (PR)

β - Consumer's risk (CR)

Double and multiple sampling plans are designed in order to give a questionable lot an additional chance. Thus, in double sampling if the results of the first sample are not conclusive in leading to acceptance or rejection, a second sample is taken which then leads to a decision on the disposition of the lot. This approach makes sense, not only as a result of experience, but also in terms of the mathematical properties of the procedure. It has been noticed (Schilling and Neubauer (2008)) that the average sample number (ASN) can usually be made to be less for a double-sampling plan than for a single-sampling plan with the same protection. The second section of the paper gives mathematical and statistical details related to the implementation of double sampling plan. It is to be noted that (Guenther (1969)), more than one double sampling plans can be developed which satisfy a given set of values for consumer's and producer's risk. Hence in this work it is desired to develop double sampling plans possessing certain optimal properties, like minimum Average Sample Number (ASN) and Average Total Inspection (ATI). The nonlinear nature of the constraints associated with the consumer's and producer's risk makes it impossible to obtain a solution using classical methods. Hence, in this paper genetic algorithm has been used in the determination of optimal double sampling plans. Details related to the genetic algorithm are presented in the third section and the computational results of the study are reported in the fourth section.

2. Double Sampling Plan

Application of a double sampling plan requires that a first sample of size n_1 be drawn at random from the lot (usually assumed large). The number of defectives d is counted and compared to the first sample acceptance number c_1

If $d_1 \leq c_1$, the lot is accepted

If $d_1 > c_2$, the lot is rejected

If $c_1 < d_1 \leq c_2$, a second sample of size n_2 is taken. The number of defectives d_2 contained in the second sample is determined. The total number of defectives $d_1 + d_2$ is compared to the acceptance number c_2 for the second sample. If $d_1 + d_2 \leq c_2$, the lot is accepted and the lot is rejected if $d_1 + d_2 > c_2$.

Many parameters are used in the designing of acceptance sampling plans which include AQL and Lot Tolerance Percent Defective (LTPD). Guenther (1970) has given a procedure for designing a double sampling plan which takes into consideration, p_0 , p_1 , α and β . Olorunniwo and Salas (1982) developed a

computer program for facilitating the double sampling plan design with any value of α and β but limited to $n_1 = n_2$ or kn_2 where k is an integer.

A double sampling is characterized by the four parameters n_1, n_2, c_1 and c_2 . Guenther's procedure designs a double sampling plan by seeking a solution for these four parameters that satisfy the following set of inequalities.

$$\begin{aligned} P(\text{Accept Lot}) &\geq 1 - \alpha \quad \text{if } \theta = \theta_0 \\ &\leq \beta \quad \text{if } \theta = \theta_1 > \theta_0 \end{aligned} \quad (1)$$

or equivalently,

$$\begin{aligned} P(\text{Reject Lot}) &\leq \alpha \quad \text{if } \theta = \theta_0 \\ &\geq 1 - \beta \quad \text{if } \theta = \theta_1 > \theta_0 \end{aligned} \quad (2)$$

where θ_0 is the value of parameter which will be accepted by the sampling plan at least $100(1 - \alpha)$ percent on the time. It is pertinent to note that the inequalities given in (1) and hence (2) have infinite number of solutions. Hence, the practitioner has to choose a double sampling plan in an optimal manner. Here one can define optimality in different ways. For example, one may be interested in seeking a double sampling plan that minimizes ASN or ATI. Since, expressions appearing in the inequalities assume complicated forms (depending on the underlying distribution) which cannot be solved explicitly, one need to adopt unconventional procedures to secure an optimum double sampling plan. The availability of high-speed computing facilities and the introduction of evolutionary algorithms make it possible to seek solutions for the inequalities given in (1) or (2) which minimizes a meaningful quantity like ASN or ATI. Cheng and Chen (2006) developed a genetic algorithm for optimizing multi objective functions related to an attribute double sampling plan. More specifically, they aimed at the determination of a double sampling plan that minimizes quantities involving consumer's risk and producer's risk. In this paper, genetic algorithm is developed for minimizing average sample number (ASN) and average total inspection (ATI) through which optimum sample sizes and acceptance numbers are determined for a specified α, β and proportion of defective units.

If the lot size is large then the observed number of defective units in the sample is a binomial random variable. Let p_d be the lot fraction defective. In this case the probability of x defectives out of a random sample of n items can be obtained by using binomial distribution defined as

$$p(x) = \frac{n!}{x!(n-x)!} p_d^x (1 - p_d)^{n-x} \quad (3)$$

In other words, the acceptance probability of a lot can be calculated by employing (3) when the specific p_d is known. In double sampling plans, the acceptance probability (denoted as P_a) of a lot is decided from the combined

samples. If P_{a1} and P_{a2} denote the probability of acceptance on the first and second samples, respectively, then P_a is the sum of P_{a1} and P_{a2} . That is,

$$P_a = P_{a1} + P_{a2} \quad (4)$$

The method of calculating P_{a1} is straightforward. The lot will be accepted in the first sampling if the defective items (x_1) are less than or equal to acceptance number (c_1). However, to calculate the probability of acceptance on the second sample, two conditions should meet sequentially. The first is that the lot is neither accepted nor rejected in the first sampling and the other is that the lot is accepted in the second sampling. Thus we get the expressions for P_{a1} and P_{a2} as

$$P_{a1} = P(X_1 \leq c_1) \quad (5)$$

and

$$P_{a2} = \sum_{i=c_1+1}^{c_2} P(X_1 = i) \times P(X_2 \leq (c_2 - i)) \quad (6)$$

respectively. If the acceptance parameters n_1, n_2, c_1 and c_2 are known, the probability of accepting a lot can be given in any required quality level. However, if Acceptable Quality Level (AQL) p_0 , Rejectable Quality Level (RQL) p_1 , producer's risk α and consumer's risk β are specified, then a proper combination of acceptance parameters n_1, n_2, c_1 and c_2 have to be found to fit the predefined p_0, α and p_1, β on OC curve. The relations between $(p_0, \alpha, p_1, \beta)$ and the acceptance parameters n_1, n_2, c_1 and c_2 are

$$(1-\alpha) = \sum_{x=0}^{c_1} \frac{n_1!}{x!(n_1-x)!} p_0^x (1-p_0)^{n_1-x} + \sum_{x=0}^{c_2-i} \sum_{i=c_1+1}^{c_2} \frac{n_1!}{i!(n_1-i)!} p_0^i (1-p_0)^{n_1-i} \times \frac{n_2!}{x!(n_2-x)!} p_0^x (1-p_0)^{n_2-x} \quad (7)$$

and

$$\beta = \sum_{x=0}^{c_1} \frac{n_1!}{x!(n_1-x)!} p_1^x (1-p_1)^{n_1-x} + \sum_{x=0}^{c_2-i} \sum_{i=c_1+1}^{c_2} \frac{n_1!}{i!(n_1-i)!} p_1^i (1-p_1)^{n_1-i} \times \frac{n_2!}{x!(n_2-x)!} p_1^x (1-p_1)^{n_2-x} \quad (8)$$

2.1 Average sample number

In double sampling plan, the size of the sample selected depends on whether or not the second sample is necessary. The probability of drawing a second sample varies with the fraction defective in the incoming lot. With complete inspection of the second sample, the average sample size in double sampling is equal to the size of the first sample times the probability that there will only be one sample plus the size of the combined samples times the probability that a second sample

will be necessary. If the second sample has complete inspection, a general formula for average sample number (ASN) in double sampling is

$$ASN = n_1 + (1 - P_1)n_2 \quad (9)$$

where P_1 is the probability of making a lot disposition decision on the first sample.

2.2 Average total inspection

Acceptance sampling programs usually requires corrective action when the lots are rejected. This generally takes the form of 100 percent inspection of rejected lots, with all defective items either removed for subsequent rework or returned to the supplier or replaced from the stock of known good items. Such sampling programs are called rectifying inspection programs. These programs are used in situations where the manufacturer wishes to know the average level of quality that is likely to result at a given stage of manufacturing operations. Thus, rectifying inspection programs are used either at receiving inspection, in process inspection of semi finished products or at final inspection of finished goods.

Important measure relative to rectifying inspection is the average total inspection (ATI) required by the sampling program. In double sampling plan the average total inspection can be visualized in three possible situations (i) only n_1 items will be inspected if the lot is accepted on the basis of the first sample and its chance is P_{a1} . (ii) $(n_1 + n_2)$ items will be inspected if the lot is accepted on the basis of the second sample and its chance is P_{a2} and (iii) the entire lot of N items will be inspected if the lot is rejected and the chance of this is $1 - P_a$. ATI for double sampling is

$$ATI = n_1 + (1 - P_{a1})n_2 + (N - n_1 - n_2)(1 - P_a) \quad (10)$$

where P_{a1} is the probability of acceptance on the first sample, P_a is the acceptance probability of the lot and the lot size $N = (n_1 + n_2) * 10$, following Naidu, Babu and Rajendra (2006). Since the acceptance parameters (n_1, n_2, c_1, c_2) can take several sets of nonnegative integral values satisfying (1) and (2), getting a closed-form solution is a difficult process. Hence, it is felt that genetic algorithms can be used to determine a double sampling plan such that either ASN or ATI is minimized. The details related to the implementation of genetic algorithm are given in the following section.

3. Genetic algorithm

Genetic Algorithms (GA) are adaptive heuristic search algorithms premised on the evolutionary ideas of natural selection and genetic. The basic concept of GA is designed to simulate processes in natural system necessary for evolution, specifically those that follow the principles first laid down by Charles Darwin of survival of the fittest. As such they represent an intelligent exploitation of a

random search within a defined search space to solve a problem. First pioneered by Holland (1975), Genetic Algorithms have been widely studied, experimented and applied in many fields. GA provides an alternative method to solving problem that consistently outperforms other traditional methods in most of the problems link. Many of the real world problems involved finding optimal parameters that might prove difficult for traditional methods but ideal for GA. A population of individuals is maintained within search space for a GA, each representing a possible solution to a given problem. Each individual is coded as a finite length vector of components. These individuals are linked to chromosomes and the components are analogous to genes. Thus, a chromosome comprises several genes. A fitness score is assigned to each chromosome representing the abilities of an individual to compete. The individual with the optimal or generally near optimal fitness score is sought. The GA aims to use selective breeding of the solution to produce off springs better than parents do by combining information from the chromosomes. This process is repeated until the strings in the new generation are identical or certain termination conditions are met.

3.1 Chromosome formation

In this study, individual chromosomes are formed in such a manner so that the search process considers all possible solutions for the determination of optimal double sampling plans can be represented and evaluated. Each individual chromosome is treated as a collection of twelve genes. The first eight genes is the octal encoding of the sample sizes n_1 and n_2 , each indicating four genes and the remaining four genes are divided into two sets of two genes each representing the octal encoding of the acceptance numbers c_1 and c_2 respectively. For example, if $n_1=140, n_2=158, c_1=2$ and $c_2=5$ then the individual formation of the chromosome is as follows:

$$\overbrace{0 \ 2 \ 1 \ 4}^{n_1} \ \overbrace{0 \ 2 \ 3 \ 6}^{n_2} \ \overbrace{0 \ 2}^{c_1} \ \overbrace{0 \ 5}^{c_2}$$

While defining the initial population we have ensured that the chromosomes which are included in the initial pool satisfy the probabilistic constraints given in (1) and (2). All those chromosomes which fail to satisfy the constraints were eliminated from the pool constituted to form the initial population of two hundred chromosomes.

3.2. Fitness value

The fitness value corresponding to each chromosome is determined by computing the value of ASN (or ATI) corresponding to the contents of the 12 bits defining the chromosome by making use of the expression given in (9) (or (10)). It is to be noted that a chromosome having smaller value for ASN (or ATI) is more fitter. Note that computation of expressions given in (9) and (10) makes use of the expressions given in the initial part of Section 2.

3.3. Selection

The selection operator involves randomly choosing individuals of the population to enter a mating pool. The operator is carefully formulated to ensure that better individuals of the population (with higher fitness) have a greater probability of being selected for mating and less potent individuals of the population have smaller probability of being selected. Having some positive probability for choosing such less potent individuals is important to ensure that the search process is global and does not simply converge to the nearest local optimum. Selection is one of the important aspects of the GA process, and there are several ways for the selection: some of these are tournament selection, ranking selection, and proportional selection. In this work we have used what is known as Roulette wheel selection procedure. Details related to Roulette wheel selection can be obtained from any standard text book on Soft Computing like, Sivanandam and Deepa (2008).

3.4. Crossover and Mutation

Once a pair of individual has been selected, crossover of two chromosomes can take place to produce a new set of off springs. In this study, single point crossover mechanism is applied. A number is randomly drawn from 1 to 11, say, r . On swapping the contents of the bit positions to the right of the r th bit, in the selected parents, we generate two child chromosomes.

The mutation operator takes each individual and randomly changes a portion of its genetic character. When a bit position is considered for mutation, the value placed in the location is added ± 1 with equal chance provided the value is different from 0 and 7. In the case of 0 is chosen the mutated value will be either 0 or 1 with equal probability and in the case of 7 either 6 or 7 is taken as the mutated value. The probability of mutation occurring on any individual is determined by the user specified mutation rate. Mutated individual are put back into the pool after the mutation process. The number of iterations is usually set as the termination criterion for stopping the evolution process while performing GA operation. In the present study, the number of iterations is limited to 1000.

4. Determination of optimal sampling plan

The above explained genetic operations have been used in the process of determining optimal sampling plan for a wide range of p_0 and p_1 , with specified producer's and consumer's risk. The optimal sampling plans are listed in the tables furnished in the Appendix. Table 1 (given in Appendix) gives optimal double sampling plans for different choices of p_0 and p_1 as indicated with $\alpha = 0.05$ and $\beta = 0.10$ such that ASN is minimized. Similarly, Table 2 (given in Appendix) gives optimal double sampling plans such that ATI is minimized for $\alpha = 0.05$ and $\beta = 0.10$.

5. Conclusions

Thus in this paper, an application of genetic algorithm has been considered in the determination of optimal double sampling plan such that either ASN or ATI is minimized subject to the probabilistic constraints considered by Guenther (1969). It is pertinent to note that there exists several sets of solutions satisfying the constraints given in (1) and (2) and traditional methods like calculus, fail to identify the sample sizes and acceptance numbers such that quantities like ASN, ATI etc are minimized. The methodology to be followed in different stages of the implementation of genetic algorithm has been clearly spelt out in Section 3 of this paper and it was found to work very well in the process of determination of optimal double sampling plans. Implementation of genetic algorithm has been carried out using C++ and the coding can be obtained from the authors on request.

References

1. Cheng, T.M., and Chen, Y.L. (2006). A GA Mechanism for Optimizing the Design of attribute-double-sampling-plan, *Automation in Construction*, 16(3), 345-353.
2. Guenther, W.C. (1969). Use of the binomial, hypergeometric and Poisson tables to obtain sampling plans, *Journal of Quality Technology*, 1(2): 105-109.
3. Guenther, W.C. (1970). A procedure for finding double sampling plans for attributes, *Journal of Quality Technology*, 2(4): 219–225.
4. Holland, J. (1975): *Adaptation in Natural and Artificial System*, University of Michigan Press, Ann Arbor, Michigan.
5. Naidu, N.V.R., Babu, K.M., and Rajendra, G. (2006). *Total Quality Management*, New Age International (P) Ltd.
6. Olorunniwo, F.O., and Salas, J.R. (1982). An algorithm for determining doubleattribute sampling plans, *Journal of Quality Technology*, 14(3), 166-171.
7. Schilling, E.G., and Neubauer, D.V. (2008). *Acceptance Sampling in Quality Control*, CRC Press.
8. Sivanandam, S.N., and Deepa, S.N. (2008). *Principles of Soft Computing*, Wiley India Private Limited, New Delhi.

Appendix

Table 1: Double Sampling plans minimizing ASN

p_0	p_1	n_1	n_2	c_1	c_2	Minimum ASN
0.005	0.05	59	86	0	2	80.734062
	0.06	42	119	0	2	64.443680
	0.07	32	130	0	2	51.193451
	0.08	47	19	0	1	50.545547
	0.09	26	50	0	1	31.734432
0.01	0.05	79	147	1	4	106.344124
	0.06	70	129	1	4	89.939507
	0.07	58	68	1	3	65.609756
	0.08	49	146	1	3	61.399845
	0.09	31	34	0	2	39.977722

Table 2: Double Sampling Plans minimizing ATI

p_0	p_1	n_1	n_2	c_1	c_2	Minimum ATI
0.005	0.05	79	161	1	5	90.683800
	0.06	53	71	0	3	73.231926
	0.07	57	118	1	4	62.635845
	0.08	50	132	1	5	53.751930
	0.09	44	189	1	9	47.890190
0.01	0.05	114	184	2	8	140.712784
	0.06	69	137	1	6	96.928795
	0.07	60	164	1	6	80.787010
	0.08	49	153	1	8	62.512157
	0.09	26	96	0	4	54.317287