1970

Acceptable Regions for Approximations in Quality Control

Du Hung-Ching

ACCEPTABLE REGIONS FOR APPROXIMATIONS IN QUALITY CONTROL

BY

HUNG-CHING DU

A thesis submitted
in partial fulfillment of the requirements for the
degree Master of Science, Major in
Mechanical Engineering,
South Dakota State University

1970

ACCEPTABLE REGIONS FOR APPROXIMATIONS

IN QUALITY CONTROL

This thesis is approved as a creditable and independent investigation by a candidate for the degree, Master of Science, and is acceptable as meeting the thesis requirements for this degree, but without implying that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Thesis Adviser — Date

Head, Mechanical Engineering Department — Date

ACKNOWLEDGMENTS

The author is indebted to Dr. Richard P. Covert for his initiation, guidance, counseling, and many clarifying discussions to the work presented here.

The author also wishes to thank Professor John F. Sandfort for his guidance throughout the author's graduate program.

HCD

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	1
	Review of Literature	6
	Statement of Problems	8
II.	EFFECT OF VARIABLES INVOLVED IN HYPERGEOMETRIC, BINOMIAL AND POISSON PROBABILITY DISTRIBUTIONS	10
III.	METHODS, PROCEDURES AND COMPUTER PROGRAMS	18
IV.	RESULTS	24
V.	CONCLUSIONS	40
	LITERATURE CITED	42

LIST OF TABLES

Table		Page
2-1	Probability of Hypergeometric, Binomial and Poisson Distributions for U=100, S=20, c=0, p=0.01 and p=0.10 . . .	11
4-1	Comparison of Hypergeometric, Binomial and Poisson Probability Distributions for c=0 and c ≤ 1	39

LIST OF FIGURES

Figure		Page
2-1.	Probability of Hypergeometric, Binomial and Poisson Distributions vs sample size for U=150, p=0.02 and c=0	13
2-2.	Probability of Hypergeometric, Binomial and Poisson Distributions vs sample size for U=150, p=0.02 and c=1	15
2-3.	Cumulative Probability of Hypergeometric and Binomial Distribution vs Sample Size for U=150, p=0.02 and $c \leq 1$	16
2-4.	Cumulative Probability of Hypergeometric and Poisson Distributions vs Sample Size for U=150, p=0.02 and $c \leq 1$	17
4-1.	Error Lines for Hypergeometric vs Binomial Distributions c=0, p=0.01 to 0.10	26
4-2.	Error Lines for Hypergeometric vs Poisson Distributions c=0, p=0.01 to 0.10	28
4-3.	Error Lines for Hypergeometric vs Binomial Distributions $c \leq 1$, p=0.01 and 0.02	31
4-4.	Error Lines for Hypergeometric vs Binomial Distributions $c \leq 1$, p=0.03, 0.04 and 0.05	32
4-5.	Error Lines for Hypergeometric vs Binomial Distributions $c \leq 1$, p=0.06 and 0.07	33
4-6.	Error Lines for Hypergeometric vs Binomial Distributions $c \leq 1$, p=0.08, 0.09 and 0.10	33
4-7.	Error Lines for Hypergeometric vs Poisson Distributions $c \leq 1$, p=0.01 and 0.02	35

Figure		Page
4-8.	Error Lines for Hypergeometric vs Poisson Distributions $c \leqslant 1$, p=0.03 and 0.04	36
4-9.	Error Lines for Hypergeometric vs Poisson Distributions $c \leqslant 1$, p=0.05, 0.06 and 0.07	37
4-10.	Error Lines for Hypergeometric vs Poisson Distributions $c \leqslant 1$, p=0.08, 0.09 and 0.10	37
	Flow Chart for Computing the Hypergeometric and Binomial Probabilities where 2% are Existing	20

CHAPTER I

INTRODUCTION

The manufacturer's inspection of his own product or the product received from outside vender, serves two purposes:

1. To provide a basis for action with regard to the materials and goods at hand. For instance; to decide whether the particular article or group of articles should be utilized, or whether some alternative disposition should be made, such as: inspected further, sorted, repaired, reworked or scrapped.
2. To provide a basis for action with regard to the future production process. For instance; to decide whether the process should be left alone, or whether action taken to find and eliminate disturbing causes.

Statistical Quality Control achieves these two purposes through sampling inspection. Thus, when parts are received from an outside vender, the inspection department may specify that a random sample of size "S" is to be drawn from a lot size (or universe size) "U" in which it is expected that there will be 'p' fraction defectives. It is desired to find the probability that the sample will contain "c" or less defectives. Theoretically, the probability that the lot is acceptable follows the hypergeometric distribution whenever a sample is drawn from a finite lot. Therefore, if a sample of five is drawn from a lot of 50 with 4% defectives, the probability of finding 1 or less defective can be computed from:

$$PH(c \leq 1)=PH(0)+PH(1) \tag{1-1}$$

where:

$PH(c \leq 1)$ = the probability of finding zero or one defective using the hypergeometric distribution

$PH(0)$ = the probability of exactly zero defective

$PH(1)$ = the probability of exactly one defective

The standard computational notation is

$$PH(c \leq 1)= \frac{{}_{48}C_5}{{}_{50}C_5} + \frac{({}_{48}C_4)({}_{2}C_1)}{{}_{50}C_5} \tag{1-2}$$

where

${}_{48}C_5$ = the number of different possible samples consisting entirely of good articles from a lot of 50 with 4% defectives

$= \frac{48!}{43!5!}$

${}_{50}C_5$ = the number of different possible samples of 5 articles taken from a lot of 50

$= \frac{50!}{45!5!}$

The other terms are similarly found.

Therefore, $PH(c \leq 1)=0.8081 + 0.1836 = 0.9917$ (1-3)

The computation of hypergeometric probability is obviously lengthy and time-consuming. This is particularly true if the sample size and allowed number of defectives are large. For a practical solution (that is, an economical amount of calculation) approximations are frequently used.

The two most important approximations to the hypergeometric distribution, both in the theory of probability and in its applications are the binomial distribution and the Poisson distribution. Each will be discussed in more detail.

1. The binominal (Bernoulli) distribution.

If the probability of occurrence of an event "E" in any single trial is p, where $0 \leq p \leq 1$, and the probability of nonoccurrence of "E" is q, where q=1-p the probability of exact "c" occurrence and N-c nonoccurrences of "E" in "n" independent trials, is given by

$$PB(c)= \frac{n!}{c!(n-c)!}(p)^{c}(q)^{n-c} = {}_{n}c_{c}(p)^{c}(q)^{n-c} \tag{1-4}$$

where: PB(c) = the probability of exactly c defectives and the probability "c" or less occurrences is given by

$$PB(S \leq c) = \sum_{S=0}^{c} {}_{n}c_{c}(p)^{S}(q)^{n-S} \tag{1-5}$$

where: S = 0,1,2,----,c

Since the expression on the right-hand side of equation (1-4) is the (c+1)th term in the binomial expansion of $(p+q)^n$, the number of occurrences "c", is said to be distributed in accordance with binomial probability distribution. It is also called the "point binomial", since a variable so distributed can assume only integer values from 0 to n, and in consequence the probabilities are concentrated at "points. The binomial probability distribution is based on the theory that a sample is drawn from an infinite lot size. It is considered a good approximation to the hypergeometric distribution when the sample size is small and the lot size is large. Using the same example as with the hypergeometric

$$\begin{aligned} PB(S \le 1) &= \sum_{S=0}^{1} {}_5C_S (p)^S (q)^{5-S} \\ &= {}_5C_0 (p)^0 (1-p)^5 + {}_5C_1 (p)^1 (1-p)^4 \\ &= (1)(1)(0.96)^5 + (5)(0.04)(0.96)^4 \\ &= 0.8156 + 0.1699 = 0.9855 \end{aligned} \tag{1-6}$$

Calculations involving the use of the binomial are also burdensome if many terms are involved and if the sample size and the allowed number of defectives are also large.

2. The Poisson Distribution

The Poisson distribution is also called "Poisson's Exponential Binomial Limit". Frequently, it can be used to approximate the binomial probability distribution. The probability of "c" occurrences is

$$PP(c) = \frac{(np)^c}{c!} e^{-np} \tag{1-7}$$

where: n = sample size

p = fraction defectives

PP(c) = the probability of exactly c defectives

The probability of "c" or less occurrences is given by

$$PP(S \leq c) = \sum_{S=0}^{c} \frac{(np)^S}{S!} e^{-np} \tag{1-8}$$

The Poisson probability distribution is considered a good approximation to hypergeometric distribution when the sample size is large and the fraction defective is small. In order to show the relation between hypergeometric probability and Poisson probability, the same example is used again.

$$\begin{aligned} PP(S \leq 1) &= \frac{(np)^0}{0!} e^{-np} + \frac{(np)^1}{1!} e^{-np} \\ &= e^{-(5)(0.04)} + (5)(0.04)\, e^{-(5)(0.04)} \\ &= 0.819 + 0.163 \\ &= 0.982 \end{aligned}$$

From the examples, the errors by using binomial and Poisson approximations to theoretical hypergeometric distribution are shown to be 0.0062 and 0.0097 respectively.

The binomial and Poisson approximations to the hypergeometric distribution are based on the assumption that a finite population can be assumed to be infinite when the effect of each individual member becomes small. Obviously, there is no definite line that can be laid down between finite and assumed infinite populations, since the individual situation will define the acceptable error.

Review of Literature

Comparisons, nomographs and tables have been developed for the hypergeometric distribution, binomial distribution and the Poisson distribution by a number of investigators.

Kane and Rokhsar (1), compared the Poisson and hypergeometric distributions for small lot sizes as follows:

lot size U=50 to 100

sample size S=1,2,4,8,12,16,20,24 and 28

if the absolute value of the differences between the cumulative terms of Poisson and hypergeometric probabilities, (PP-PH) was larger than 0.01 the difference was declared significant. A typical (PP-PH) vs d (number of defects in the sample) chart was plotted for U=50, S=1,2,4,8,12,16 and D=25 (average number of defects in the lot). As a general rule, they concluded that (PP-PH) is less than 0.01 when d is equal to zero, one and S/2.

Duncan (2), made a table comparing the hypergeometric, binomial and Poisson distributions. The comparison was based on p×S=0.5 where p is fraction defectives and S is sample size. Three sets of comparison were made:

$$\left.\begin{array}{llll} (1) & p=0.25 & S=2 & U=8,20,40 \\ (2) & p=0.1 & S=5 & U=20,50,100 \\ (3) & p=0.02 & S=25 & U=100,250,500 \end{array}\right\} \text{c (acceptance number) = } 0,1,2,3,4,5,6$$

His results were as follows:

comparison	smallest error	U value	c value	difference used	largest error	U value	c value	difference used
(1)	0.5%	40	2	PB-PH	12%	8	1	PP-PH
(2)	0.5%	100	0	PB-PH	9%	20	1	PP-PH
(3)	0.8%	500	0	PB-PH	7.5%	100	1	PP-PH

The largest error always occurs at the smallest lot size in each set, while the smallest error always occurs at the largest lot size in each set. In general, the larger the lot size the smaller the error.

Larson (3), developed a nomograph of the cumulative binomial distribution, which can be used to solve both cumulative and point probabilities. The nomograph consists of three variables, p=1%-5%, S=2---1000 and c=0---200. The lot size is assumed to be infinite. The nomograph covers the range of binomial distribution needed for practical applications. It is a geometric approximation based on the duality principle of projective geometry. The accuracy is quite adequate for practical applications, assuming an infinite population.

In the National Bureau of Standards, Applied Mathematics Series 6 (4), two tables of binomial probability distributions have been constructed. One is for individual terms of probability, which gives exactly "c" occurrence in "n" independent trials, when the probability of occurrence in any single trial c is 0.5 or less for $2 \leq n \leq 49$ and $1 \leq c \leq n$. For practical applications, the maximum sample size n=49 is too small.

Lieberman and Owen (5), generated three tables of hypergeometric probability on an IBM 704 for

lot size	sample size	no. of defective items in the lot (k)
2--100	1--50	0--50
1000	500	0----500
100--2000	50--1000	S-1, S, S=U/2

The point probabilities were obtained by taking antilogarithms correct to at least eight decimal places. The cumulative probabilities were calculated by summing the point probabilities. The results were rounded off to six decimal places within the IBM 704 computer and printed. As a check on the accuracy of the table they made calculations on desk calculators of randomly selected values from each set of 200. No discrepancies were found.

Statement of the Problem

A review of the literature indicates that previous studies have provided only limited information in the comparison of the probabilities of hypergeometric distribution, binomial distribution and Poisson distribution, particularly when lot size is over 100, although Duncan showed that the error tends to be small in this range.

This study proposes to develop a technique to indicate the limiting sample size for a given fraction defectives, lot size and acceptance number for a specified level of error, using both binomial and Poisson distributions.

The expected form of the output will be a series of graphs relating lot size and sample size that will indicate the 2% error limit for the approximating distribution, fraction defective and acceptance number.

Further, the computer program will be available and with changes in input can be used with other error limits, as well as other values for the variables involved.

CHAPTER II

EFFECT OF VARIABLES INVOLVED IN HYPERGEOMETRIC, BINOMIAL AND POISSON PROBABILITY DISTRIBUTIONS

The hypergeometric, binomial and Poisson probability distributions were discussed in Chapter I as useful in statistical quality control. It was suggested that the binomial and Poisson distributions could be used under certain circumstances to approximate the hypergeometric distribution. The conditions under which these approximations hold need amplification and further examination. The three conditions are:

(1) the assumption of infinite lot size

(2) the effect of the amount of fraction defectives

(3) the effect of sample size

Each will be discussed further.

(1) The assumption of infinite lot size

The binomial and Poisson probability distributions are based on the theory that the sample is drawn from an infinite population. Therefore, smaller errors are expected as the lot size increases while sample size remains unchanged. It is expected that for some acceptable error, the effect of lot size will cease to be important at a specific value of lot size and beyond that point the lot size can be assumed to be infinite.

(2) The effect of the amount of fraction defectives

If the lot size and the sample size are constant, the larger the fraction defective, the greater the error will be from using one of the approximation method. This can be demonstrated by inserting a series or numerical values. Such values give result as shown in Table 2-1. It will be remembered that both approximations were intended for a large lot size and a small fraction defectives.

TABLE 2-1

PROBABILITY OF HYPERGEOMETRIC, BINOMIAL AND POISSON DISTRIBUTIONS FOR U=100, S=20, c=0, p=0.01 AND p=0.10

Probability	p=0.01	p=0.10
hypergeometric(PH)	0.800	0.0755
binomial(PB)	0.8101	0.1213
Poisson(PP)	0.819	0.135
PB-PH	1.01%	5.58%
PP-PH	1.9%	5.95%

(3) The effect of sample size

When the lot size and the fraction defective are fixed and the sample size is allowed to vary, then the probabilities of an exact event occurring in each of the three distributions will tend to differ as a result of distribution assumed and the assumption of finite or infinite population.

Perhaps of more importance is the direction of the error for the probabilities of finding exactly zero, one and greater defectives from a given lot size. For c=0, the probabilities computed from the binomial and Poisson distributions are increasingly greater than the probability computed from the hypergeometric distribution. It can be noted theoretically that the hypergeometric distribution reaches zero probability at the lot size, while the approximations approach zero asymptotically.

The numerical illustrations in Figure 2-1, 2-2, 2-3 and 2-4 are based on a given lot size of 150 units and a fraction defectives of 0.02. They show the effect of increasing sample size for the hypergeometric, binomial and Poisson distributions and acceptance numbers of 0 and 1. Figures are on the pages following their first mention in the text.

In Figure 2-1, the probability value is computed for exactly zero defectives in the sample. As expected, the approximations of the binomial and Poisson distributions are always greater than the hypergeometric distribution and deviate from theoretical value of hypergeometric probability as sample size increases.

For probabilities of exactly 1 or 2 or more defectives, the values obtained by the approximation are smaller when the sample size is small, and become larger only when the sample size becomes large.

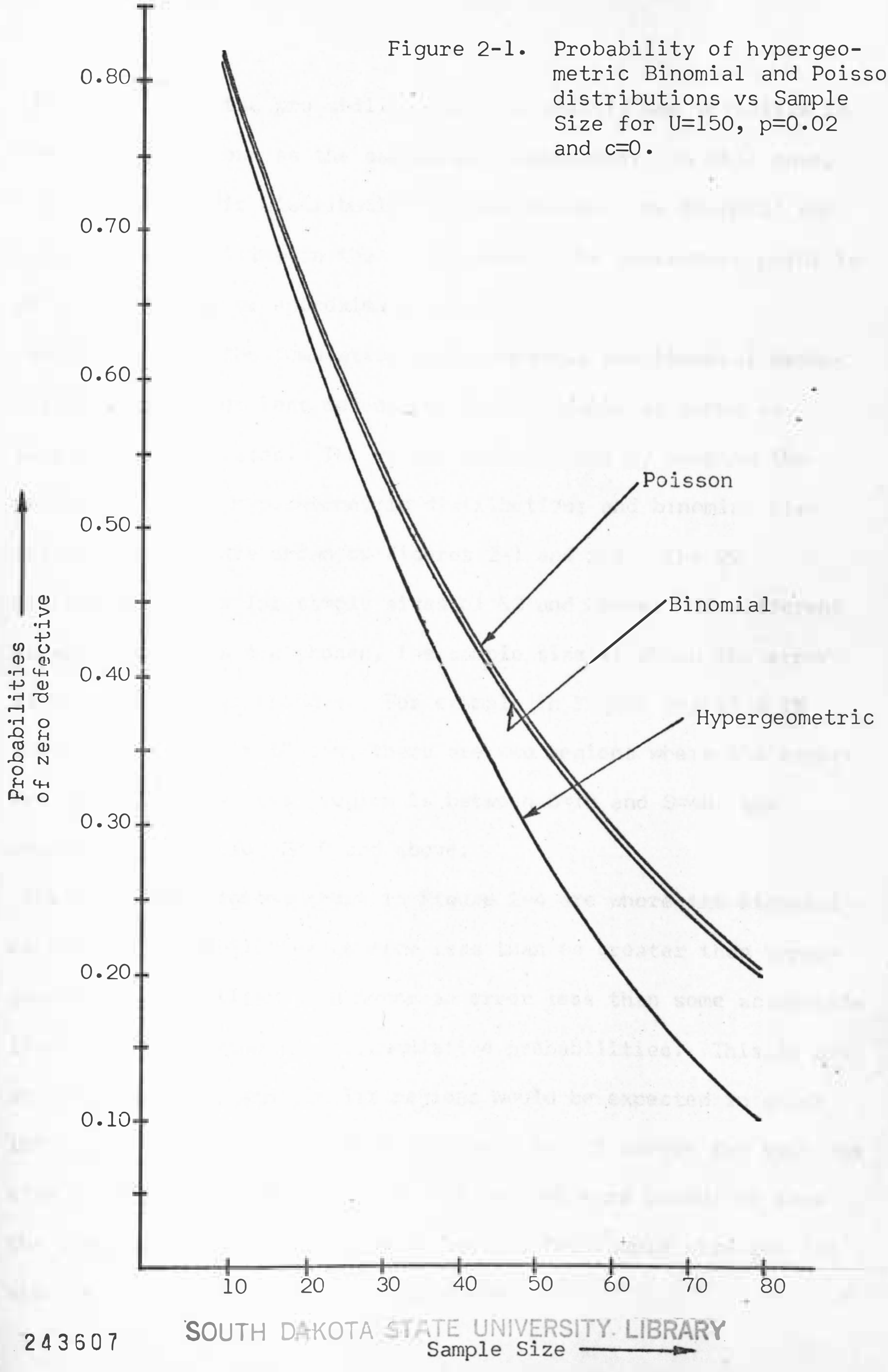


Figure 2-1. Probability of hypergeometric Binomial and Poisson distributions vs Sample Size for U=150, p=0.02 and c=0.

In Figure 2-2, the probability value of exactly one defective in the sample is shown as the sample size increases. In this case, the hypergeometric distribution is greater than the binomial and Poisson probabilities in the range shown. The cross-over point is at a sample size of approximately 90.

In Figure 2-3, the cumulative hypergeometric and binomial probabilities of one or less defectives in the sample is shown as sample size increases. The values are obtained by summing the probabilities of hypergeometric distributions and binomial distributions that are shown on Figures 2-1 and 2-2. The 2% difference occurs for sample sizes of 63 and above. If different acceptable errors are chosen, the sample size at which the error is acceptable also changes. For example in Figure 2-3 if a 1% acceptable error is chosen, there are two regions where the errors are over 1%. The first region is between S=18 and S=48, the second region is for S=60 and above.

The transient regions shown in Figure 2-4 are where the binomial or Poisson probabilities go from less than to greater than hypergeometric probability, and hence an error less than some acceptable limit would be expected for cumulative probabilities. This is not an isolated case, and similar regions would be expected in other lot sizes. However, to plot a different set of curves for each lot size is bulky and not desirable. It is much more useful to show the acceptable and non-acceptable regions for sample size and lot size on a single plot for a given error.

Figure 2-2. Probability of Hypergeometric Binomial and Poisson distributions vs Sample Size for U=150, p=0.02 and c=1.

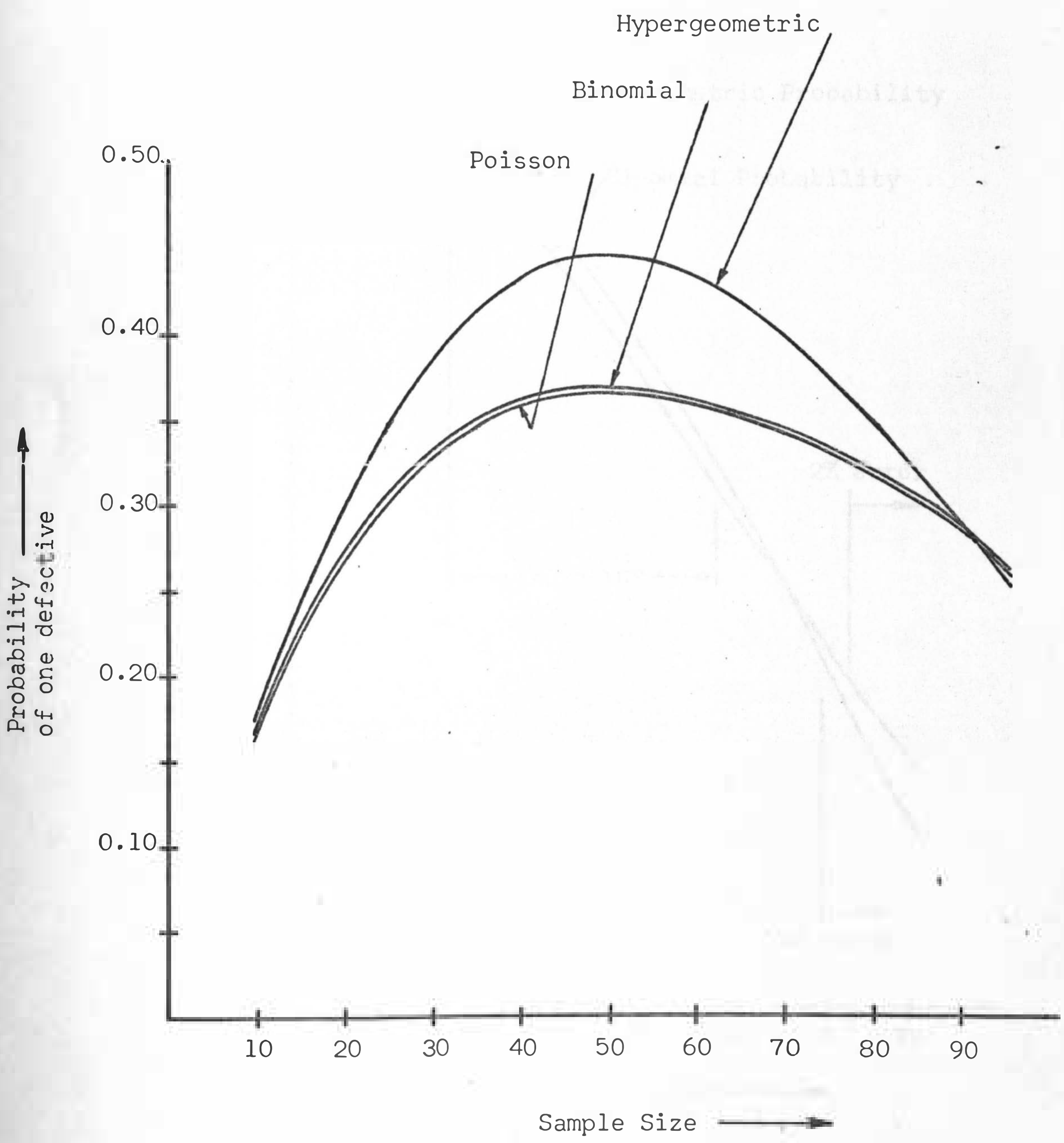


Figure 2-3. Cumulative Probability of Hypergeometric and Binomial distributions vs Sample Size for U=150, p=0.02 and c ≤ 1.

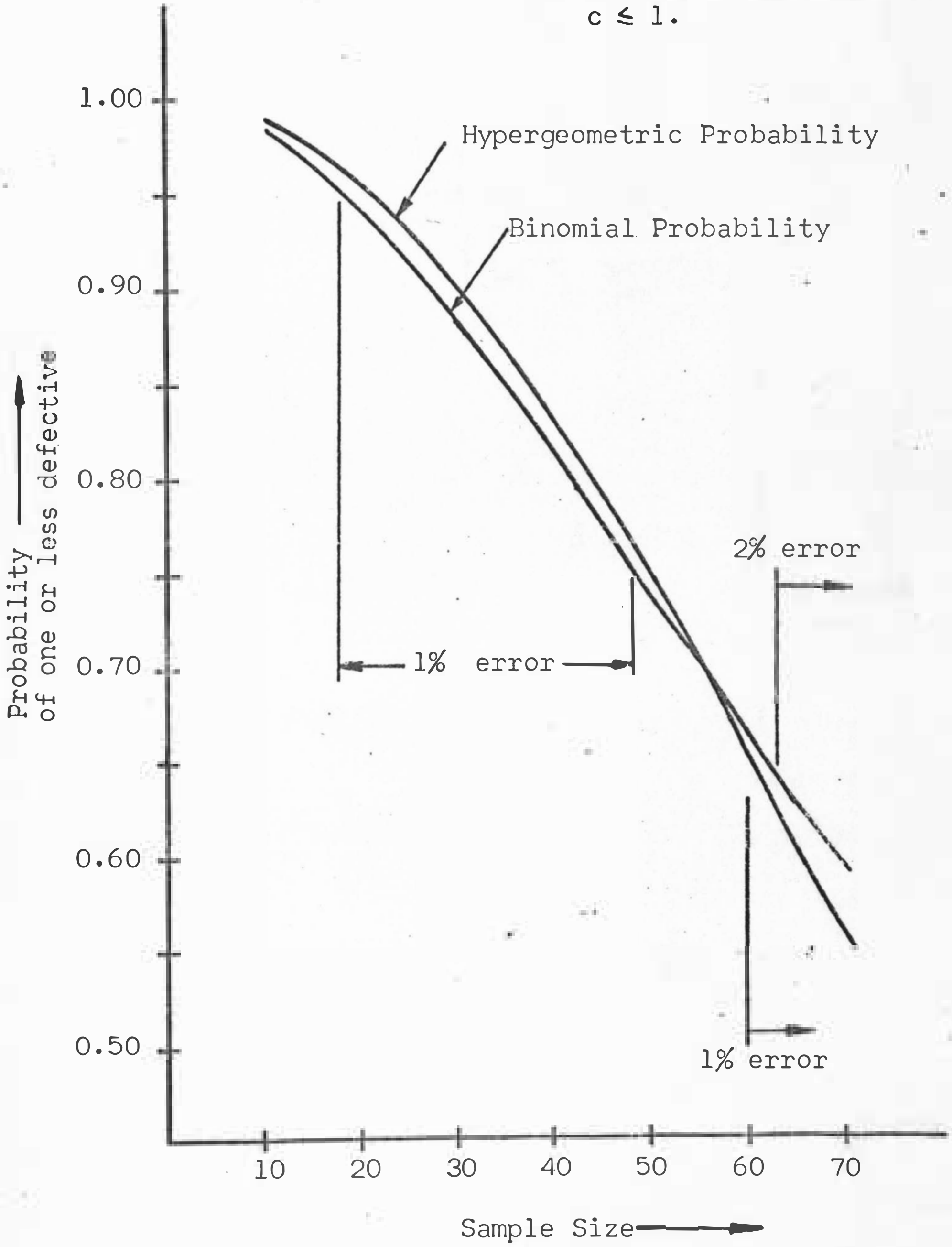
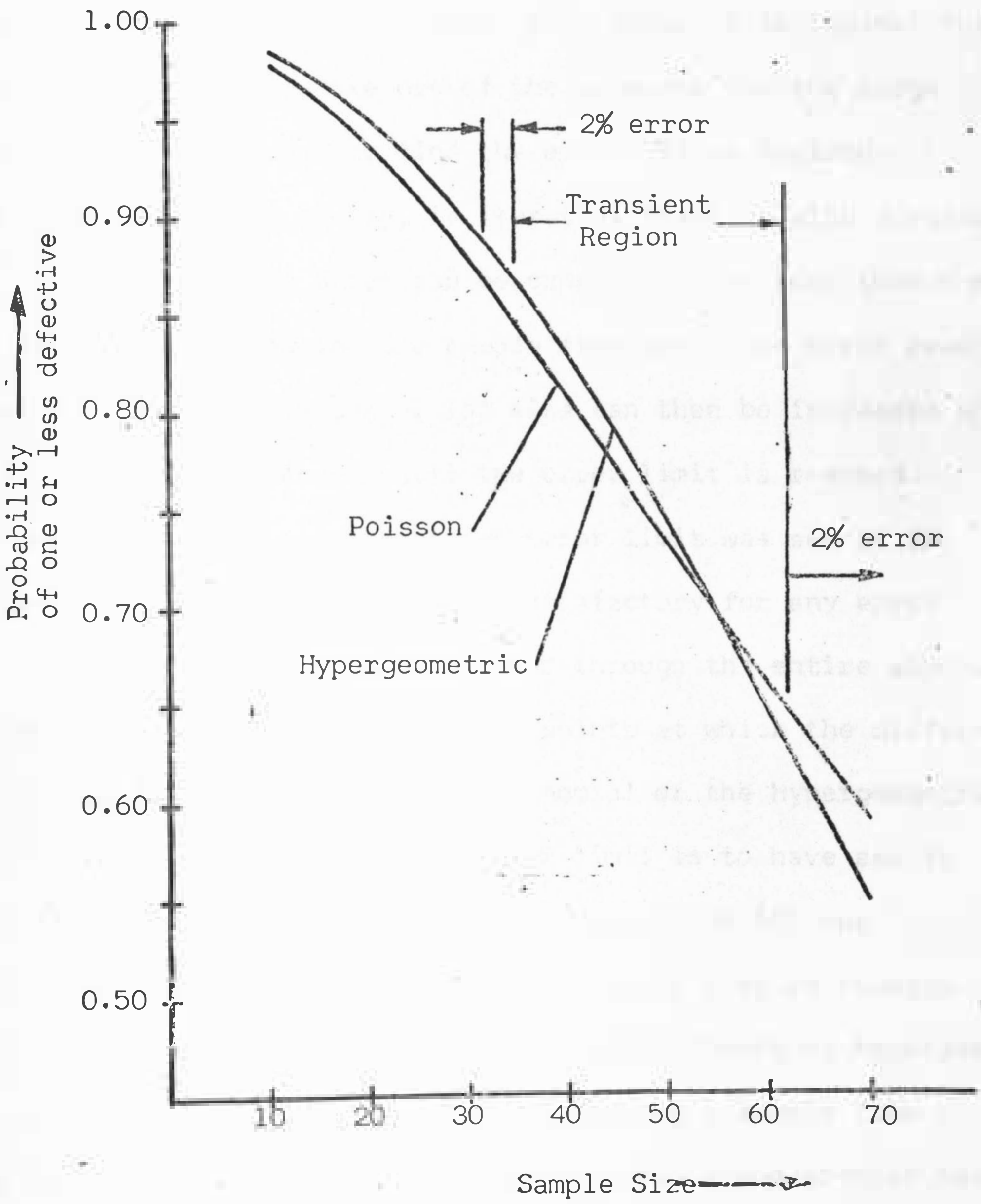


Figure 2-4. Cumulative Probability of Hypergeometric and Poisson distribution vs Sample Size for U=150, p=0.02 and c ≤ 1.



CHAPTER III

METHODS, PROCEDURES AND COMPUTER PROGRAMS

The computation for even a single point, as shown in Chapter I, becomes time-consuming if done by hand. Therefore, it is logical that a procedure be developed to make use of the computer for the large amounts of computation needed to find the error limits desired.

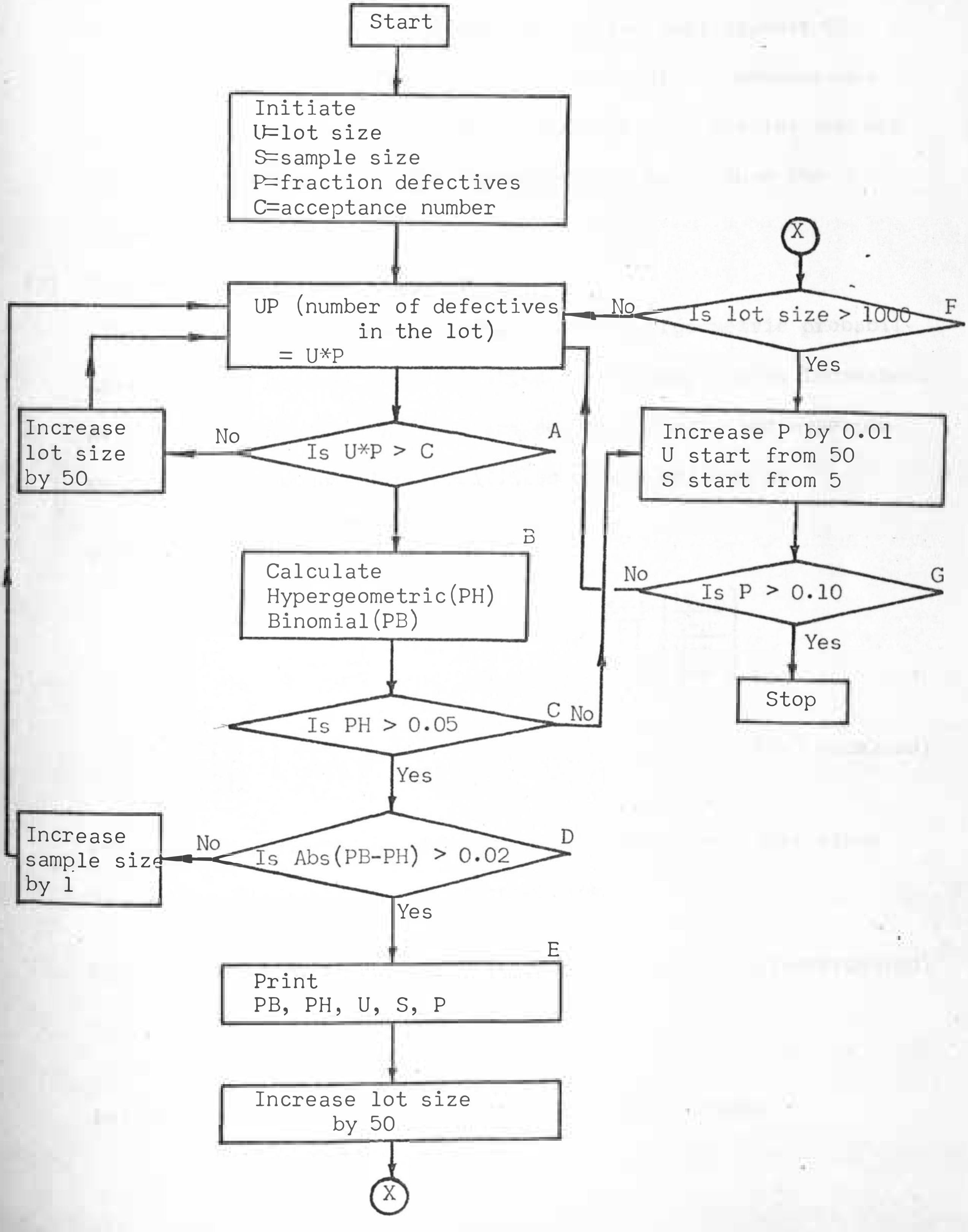
The procedure, of necessity, is iterative starting with a value of sample size in which the error can be expected to be less than the specified limit, and increasing the sample size until the error reaches the desired limit. The value of lot size can then be increased and the sample size again increased until the error limit is reached.

For the purposes of this paper the error limit was set at 2%. The logic of the procedure, however, is satisfactory for any error desired. A 2% difference criterion is used through the entire study.

The procedure used to determine the points at which the difference between the hypergeometric and the binomial or the hypergeometric and the Poisson falls outside of the preset limit is to have sample size(S) increase while lot size(U), fraction defectives(P) and acceptance number (C) remain unchanged. The sample size is changed rather than one of the other variables because the theory of hypergeometric probability distribution is based on drawing a sample from a finite lot and hence, the value of S/U is less than 1 and greater than 0. The binomial and Poisson approximations are based on the theory that sample is drawn from an infinite lot size and therefore, the value of S/U is zero. In fact, when U is very large, and S is very small, the ratio of S/U approaches zero, and the value of binomial probability is close to the theoretical value of hypergeometric probability. When S is increased while U remains unchanged the ratio of S/U is becoming greater and the binomial probability will deviate from the value of the hypergeometric probability. At some point, the absolute difference between the hypergeometric and the binomial probabilities or the hypergeometric and the Poisson probabilities will be greater than 2%. As is mentioned in Chapter II, two regions might be expected in which the errors are over 2% on a curve of lot size vs sample size for given fraction defectives and given acceptance number. The region where 2% error occurs at a larger sample size as well as the boundary in which the error again falls below 2%, was determined by hand computation since the computer program was designed to terminate upon finding a 2% error.

Basically, four computer programs have been written to supply the needed information. The first program, written for the hypergeometric vs the binomial distributions, covers the range of lot size of 50 to 1000 by increments of 50, while the fraction defectives varies from 0.01 to 0.10 by increments of 0.01, and the acceptance number is zero. A flow chart of the program is shown on page 20. A number of points may need additional explanation:

FLOW CHART FOR COMPUTING THE HYPERGEOMETRIC AND BINOMIAL PROBABILITIES WHERE 2% ERRORS ARE EXISTING



(1) Decision block A

Testing of U*P (number of defectives in the lot) against C (acceptance number of defectives) is to eliminate unnecessary computation. Obviously, if U*P is less than C, the lot can not be rejected since there are less defective parts than the acceptable limit.

(2) Computation block B

This program is designed to compute the hypergeometric probabilities and the binomial probabilities as the sample size increases. As indicated in equation (1-2) and equation (1-4), the hypergeometric and the binomial probabilities can be written as

$$PH(0) = \frac{{}_{U-Up}C_S}{{}_{U}C_S}$$

$$= \frac{{}_{UA}C_S}{{}_{U}C_S} = \frac{\frac{UA}{S} \times \frac{UA-1}{S-1} \times --- \frac{UA-S+1}{1}}{\frac{U}{S} \times \frac{U-1}{S-1} \times --- \frac{U-S+1}{1}} = \prod_{n=0}^{S-1} \left[\frac{\frac{UA-n}{S-n}}{\frac{U-n}{S-n}} \right]$$

$$= \prod_{n=0}^{S-1} \left[\frac{UA-n}{U-n} \right] \qquad \text{(1-2 repeated)}$$

PH(0) can be computed by means of an iterative procedure since values are decreased by one for each step.

$$PB(0) = {}_{S}C_0 (p)^0 (1-p)^S \qquad \text{(1-4 repeated)}$$

$$= (1-p)^S$$

PB(0) can be computed by straightforward computation.

(3) Decision block C

Since the binomial distribution approaches the hypergeometric distribution as the lot size increases, at some point the error introduced will never exceed the 2% limit. This will occur when the hypergeometric probabilities become small. To prevent the computer from excessive search the computer does not compute the binomial or the Poisson distribution when the value of the hypergeometric probability is less than 0.05.

(4) Decision block D

Testing the absolute difference between hypergeometric and binomial probabilities against preset 2% limit. Since the criterion was set at 2% the purpose of the program is to find out when the difference between two probabilities will fall outside of the preset criterion as sample size increases.

(5) Output block E

Once the difference between hypergeometric and binomial probabilities begins to fall outside of 2% the answer is reached. The current values of binomial probability, hypergeometric probability, sample size, lot size and fraction defectives are printed out.

(6) Decision blocks F and G

Testing the lot size against 1000 and testing the fractives against 0.10. The upper limits of this study for lot size and fraction defectives are 1000 and 0.10.

The second program is written for the hypergeometric vs the Poisson probabilities. It covers the same range of lot size, fraction defectives and acceptance numbers. The only difference between this program and the first program is that the Poisson probability distribution is computed rather than the binomial probability can be written as

$$PP(0) = \frac{(Sp)^0}{0!} e^{-(S)(p)}$$

$$= e^{-(S)(p)} \qquad \text{(1-7 repeated)}$$

Obviously, PP(0) can be computed by straightforward computation. The programs for acceptance numbers greater than zero are almost identical to the programs for an acceptance number of zero, except that the probabilities of individual terms must be summed, thus the cumulative hypergeometric, binomial and Poisson probabilities can be written as

$$PH = \sum_{n=0}^{c} PH(n) \qquad (3\text{-}1)$$

$$PB = \sum_{n=0}^{c} PB(n) \qquad (3\text{-}2)$$

$$PP = \sum_{n=0}^{c} PP(n) \qquad (3\text{-}3)$$

Because of the limited computer time available for this study, acceptance numbers of zero and one were selected to demonstrate both the procedure and output.

CHAPTER IV

RESULTS

This paper is intended to show a procedure for determining those regions in which the binomial and Poisson distributions could be used in place of the hypergeometric distribution and demonstrate some of the regions. As the computer programs were designed to output the values for which the 2% error was found, the results are best shown graphically. In these graphs, the absciassa is the lot size while the ordinate is the sample size. The graphs will indicate the acceptable regions where the errors introduced by the approximation methods are less than 2% and the non-acceptable regions where the errors introduced by the approximation methods are over 2%.

It is reasonable to expect that the number of defectives in a lot should be a finite integer since a non-integer number of defectives is meaningless within a given lot. Therefore, all the computed values on each figure are discrete points and must satisfy the conditions that lot size times fraction defectives in the lot equals an integer. When a lot size is small, some fraction defectives may have no meaning. The discrete points have been connected with straight lines to better define the regions and for use when the average fraction defective over several lots is known to produce values other than integer values.

Regions were verified on figures by computing the probability for each distribution for an arbitrary point within that region. These results are included in the Table 4-1 that follows all of the graphical presentations. The graphs will be discussed in more detail as follows:

(1) Figure 4-1, error lines for hypergeometric vs binomial distributions, c=0

Lot size vs sample size for various fraction defectives from 0.01 to 0.10 are plotted in Figure 4-1. The investigated area has been divided into three regions. Two of the regions are regions where the approximations hold and the third is the region where the errors introduced by approximations are over 2%. The acceptable regions are labeled A and C, while the non-acceptable region is B.

Region A is that region in which the lot size has become sufficiently large that the lot size may be assumed infinite and therefore, the error is less than 2%, regardless of the sample size. The region is specifically marked for the situation with fraction defectives for 0.02. The point $\bar{A}$(U=700, S=90, p=0.02) has been arbitrarily selected to show the error introduced by using binomial approximation at this point. Region A does not exist for p=0.01 because maximum lot size of 1000 was reached before Region A was found.

Region B is above and to the left of the two percent error line. In this region, the hypergeometric probabilities can not be approximated by binomial probabilities. The point $\bar{B}$(U=300, S=60, p=0.01) has been arbitrarily selected to show the error introduced by using binomial approximation at this point.

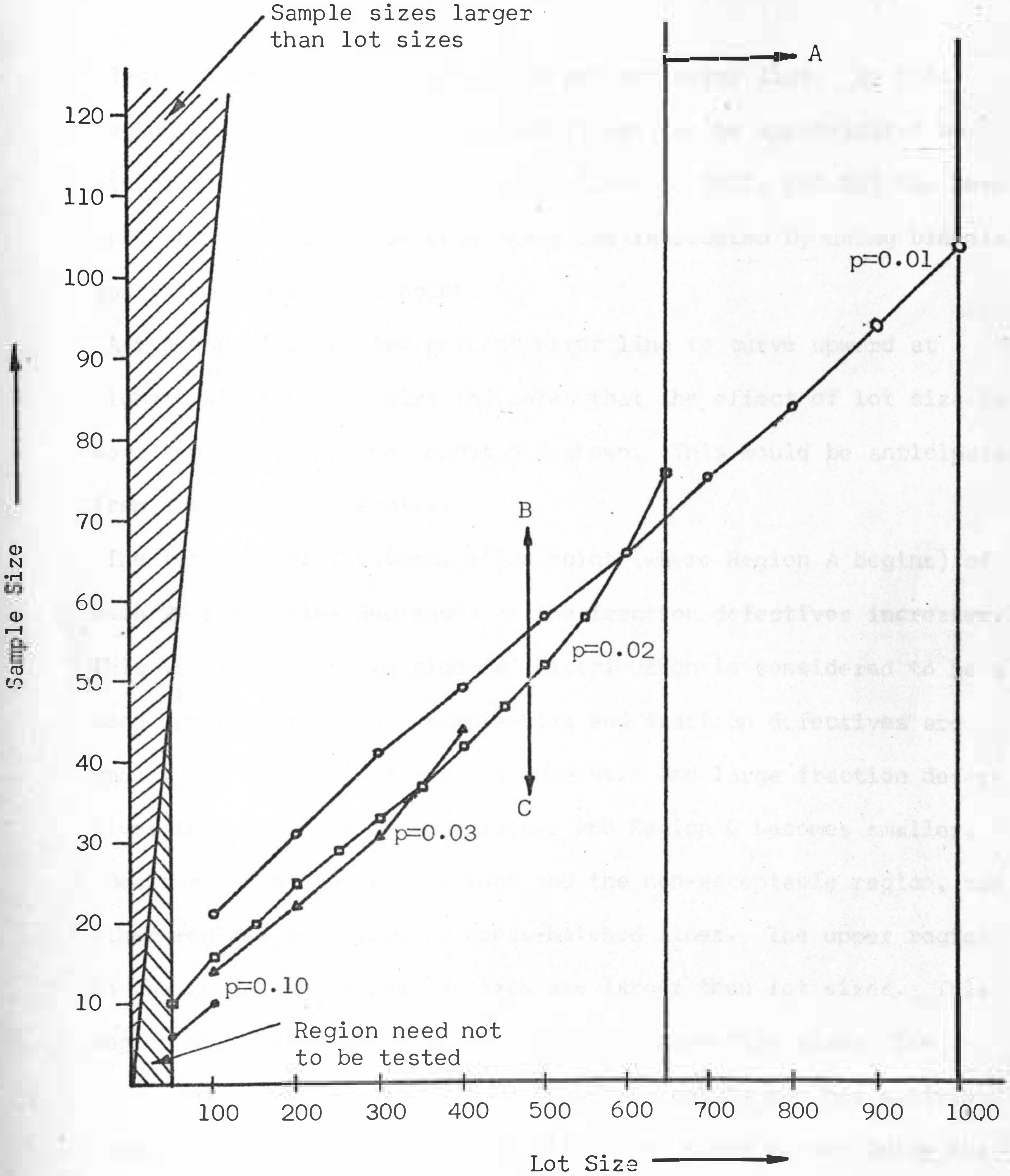


Figure 4-1. Error Lines for Hypergeometric vs Binomial Distribution, c=0, p=0.01 to 0.10.

Region C is on and below the two percent error line. In this region, the hypergeometric probabilities can be approximated by binomial probabilities. The point $\bar{C}$(U=400, S=20, p=0.03) has been arbitrarily selected to show the error introduced by using binomial approximation at this point.

A tendency for the two percent error line to curve upward at higher values of lot size indicates that the effect of lot size is more important for the conditions shown. This would be anticipated from the Region A results.

The lot size of the termination point (where Region A begins) of each 2% error line decreases as the fraction defectives increases. This is expected since binomial distribution is considered to be a good approximation when sample size and fraction defectives are small. In the case of small sample size and large fraction defectives thus Region A becomes larger and Region C becomes smaller.

Besides the acceptable regions and the non-acceptable region, two other regions are shown by cross-hatched lines. The upper region is the region where sample sizes are larger than lot sizes. This region will never exist in any sampling inspection plan. The lower region is the region where discrete points can not satisfy condition of integer defectives within the scope or are below the predetermined limit on lot size.

(2) Figure 4-2, error lines for hypergeometric vs Poisson distributions, c=0

The results for the comparison of the hypergeometric and the Poisson

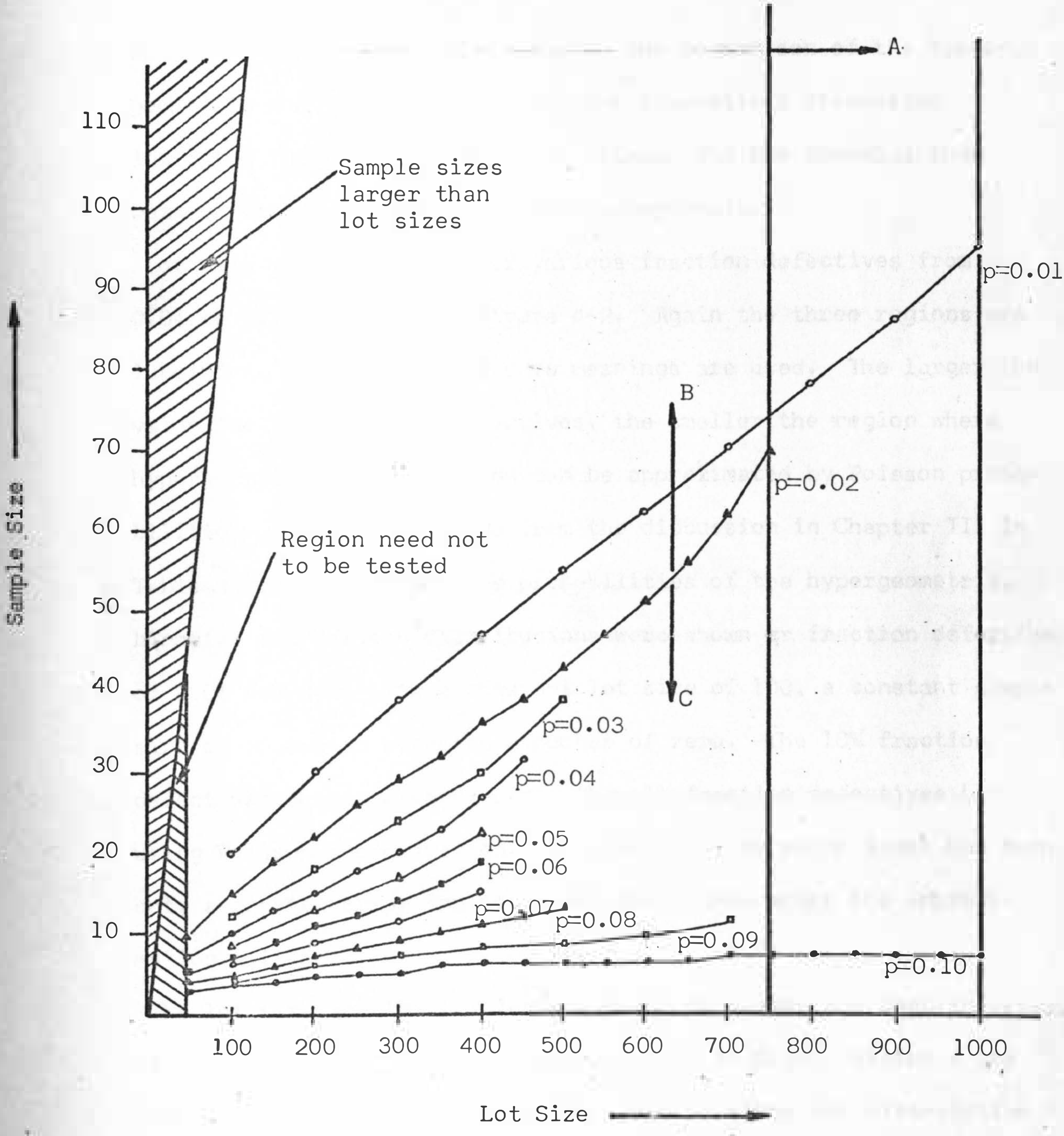


Figure 4-2. Error Lines for Hypergeometric vs Poisson Distribution, c=0, p=0.01 to 0.10.

should be expected to be similar to the comparison of the hypergeometric and the binomial since the theoretical discussion indicated less error between the Poisson and the binomial than between either of them and the hypergeometric.

Lot size vs sample size for various fraction defectives from 0.01 to 0.10 is shown on Figure 4-2. Again the three regions are labeled A, B and C and the same meanings are used. The larger the value for the fraction defectives, the smaller the region where hypergeometric probabilities can be approximated by Poisson probabilities. This is expected from the discussion in Chapter II, in Table 2-1, for examples the probabilities of the hypergeometric, binomial and Poisson distributions were shown or fraction defectives of 0.01 and 0.10 with a constant lot size of 100, a constant sample size of 20 and an acceptance number of zero. The 10% fraction defectives showed greater error than 1% fraction defectives by using Poisson approximation. Therefore, if an error limit has been set, a greater error implies a smaller region where the approximation will hold.

The lot size for the termination point decreases and then increases as fraction defectives increase from 0.01 to 0.10. Region A and Region C become smaller. This is expected since the distribution is considered to be a good approximation when sample size is large and fraction defectives is small. In the case of small sample size and large fraction defectives, thus Region A and Region C become smaller. Two percent error lines become flatter as fraction defectives increase, therefore, Region B and Region C change from trapezoid to almost rectangular shapes.

Again, points are arbitrarily selected in each region to demonstrate the errors introduced by Poisson approximation.

(3) Figure 4-3, 4-4, 4-5 and 4-6, error lines for hypergeometric vs binomial distributions, $c \leq 1$

For the acceptance numbers of one or greater, Figure 2-4, page 17 demonstrated that it is possible to have an error greater than any given percent in two regions on a curve of lot size vs sample size. For the results when the acceptance number of one or greater, it is therefore necessary to define five regions. The first three, A, B and C, will be the same as before. The additional regions D and E will sometimes appear as an area and sometimes as a line. Region D represents a 2% error at the lower sample size region where hypergeometric probabilities are greater than binomial probabilities. Region E is the region in which the cumulative sum of errors is small although the component errors may be large. Region E is actually a part of Region C, but is distinguished in this paper because of the difference in the sign of the error. The normal B region, at higher sample size region, is beyond the end point of Region E where binomial probabilities are greater than hypergeometric probabilities.

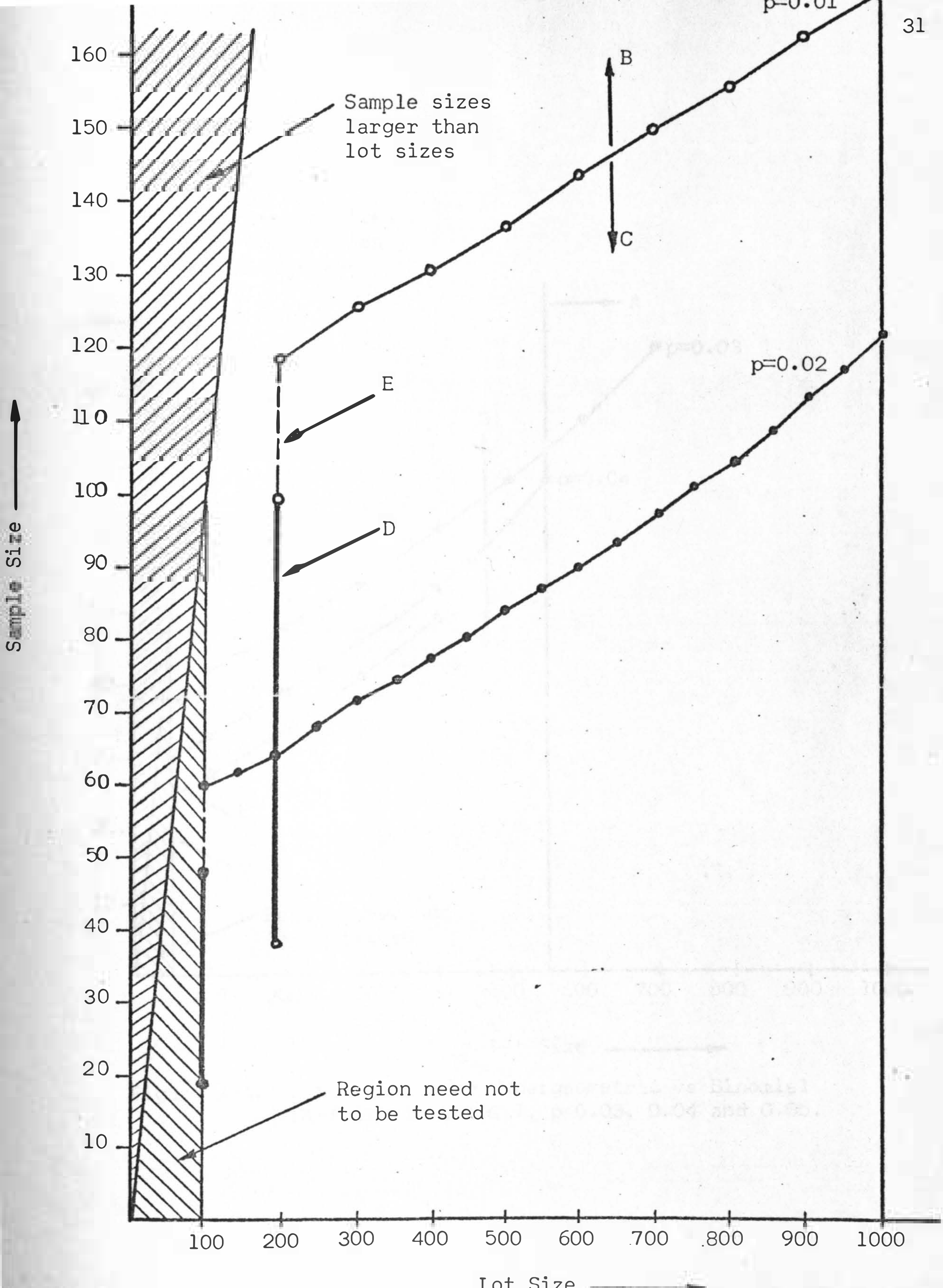


Figure 4-3. Error Lines for Hypergeometric vs Binomial Distributions, $c \leq 1$, $p=0.01$ and 0.02.

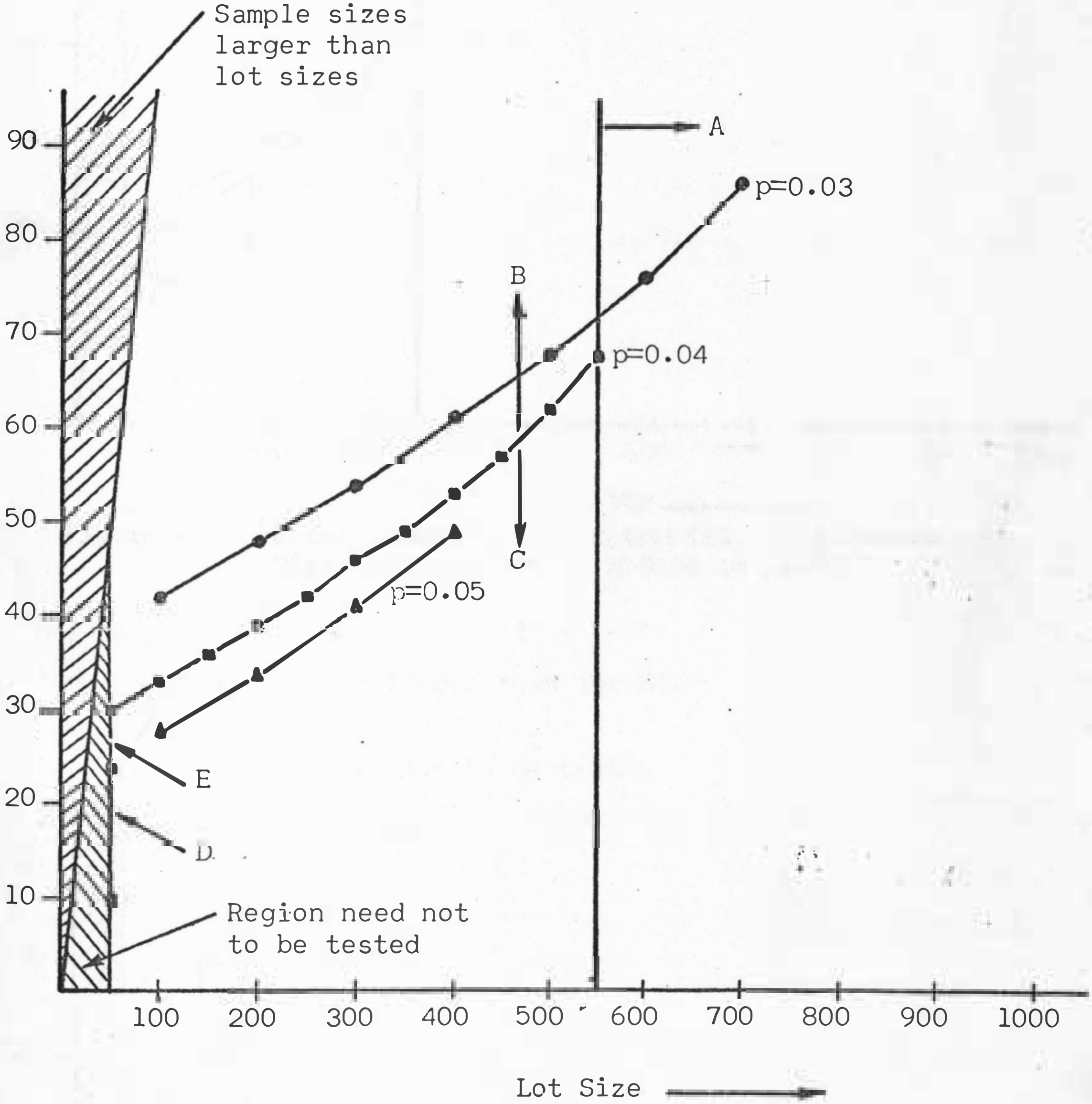


Figure 4-4. Error Lines for Hypergeometric vs Binomial Distribution, $c \leq 1$, $p=0.03$, 0.04 and 0.05.

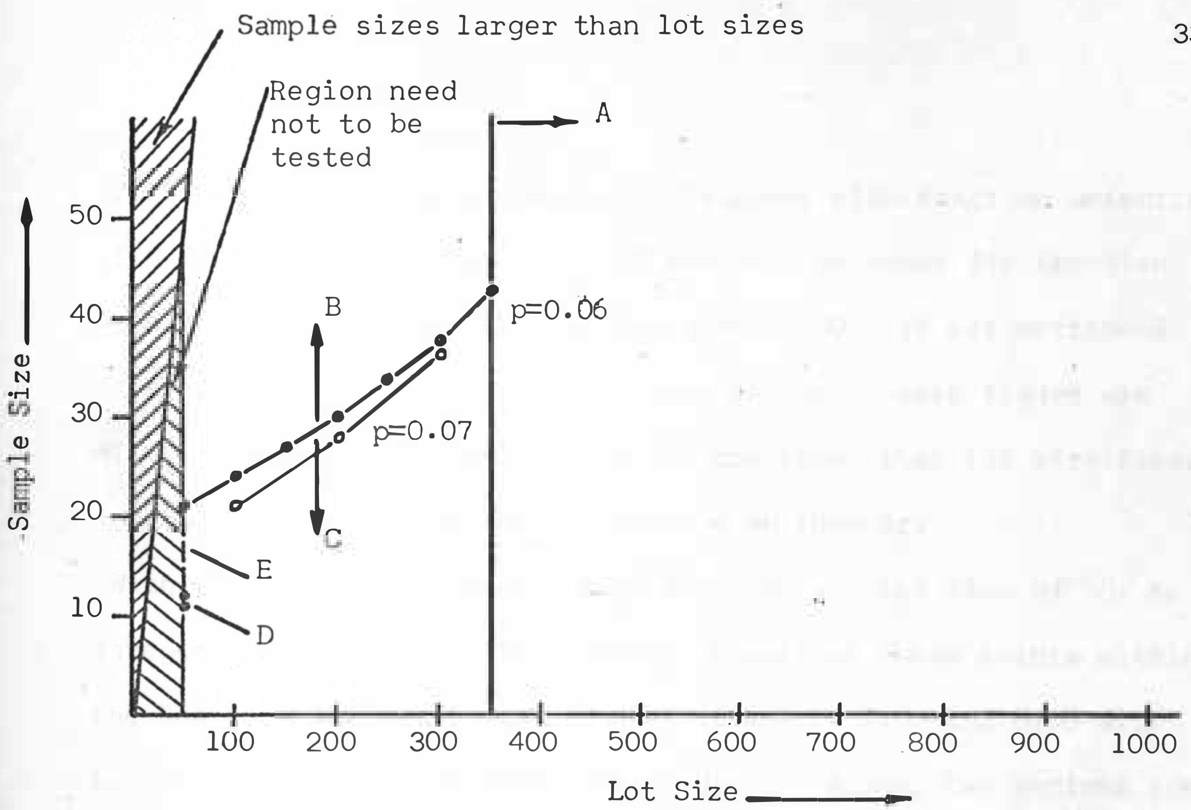


Figure 4-5. Error Lines for Hypergeometric vs Binomial Distribution, c ≤ 1, p=0.06 and 0.07.

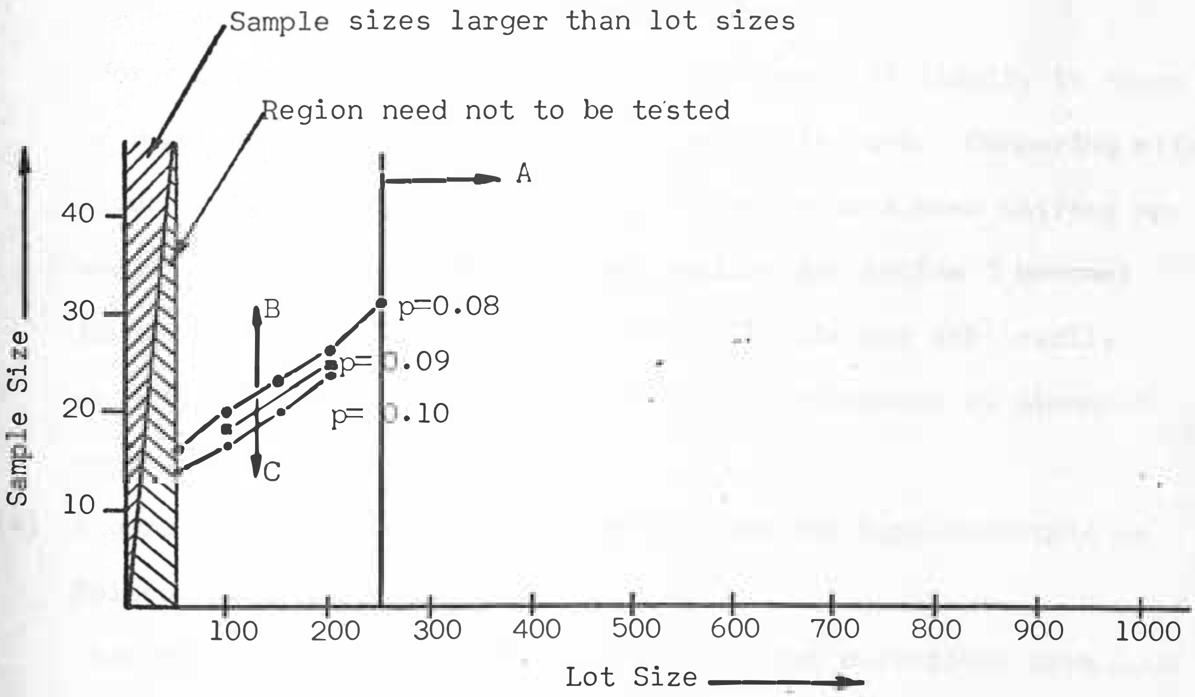


Figure 4-6. Error Lines for Hypergeometric vs Binomial Distribution, c ≤ 1, p=0.08, 0.09 and 0.10.

For $c \leqslant 1$, Region D and Region E disappear with fraction defectives of 0.07 and higher. Region D and Region E do occur for fraction defectives of 0.03 and 0.05 at lot size of 50. It was mentioned early in this chapter that all of the points on each figure are discrete points and must satisfy the condition that lot size times fraction defectives in the lot must be an integer.

With fraction defectives of 0.03 and 0.05 and lot size of 50, no integer number of defectives exist. Therefore those points within the Region D and E for fraction defectives of 0.03 and 0.05 with lot size of 50 are considered meaningless. Again, two regions are shown by cross-hatched lines; one is the region where sample sizes are larger than lot sizes, and the other is the region that does not need to be tested for finding 2% errors.

For $c \leqslant 1$, the trend of 2% error lines showed similarity to those of Hypergeometric vs Binomial distribution for c=0. Comparing with Figure 4-1, error lines in Figure 4-3 to 4-6 have been shifted upward. Therefore, Region B becomes smaller and Region C becomes larger while Region A becomes smaller. Points are arbitrarily selected in Figure 4-4 to show the errors introduced by binomial approximation at those points.

(4) Figure 4-7, 4-8, 4-9 and 4-10, error lines for hypergeometric vs Poisson distributions, $c \leqslant 1$

Lot size vs sample size for various fraction defectives from 0.01 to 0.10 are plotted on Figures 4-7, 4-8, 4-9 and 4-10. The figures

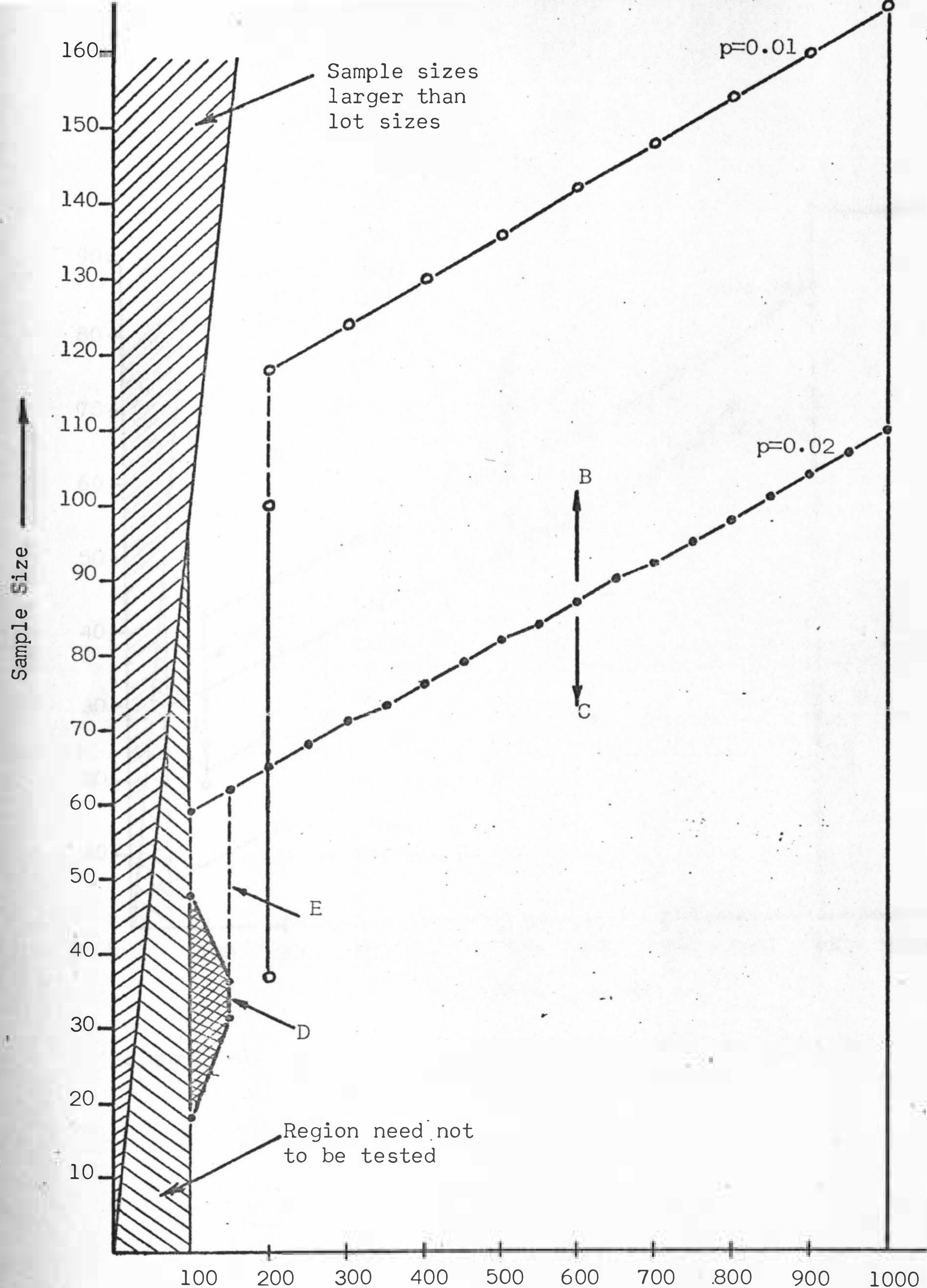


Figure 4-7. Error Lines for Hypergeometric vs Poisson Distribution, $c \leq 1$, p=0.01 and 0.02.

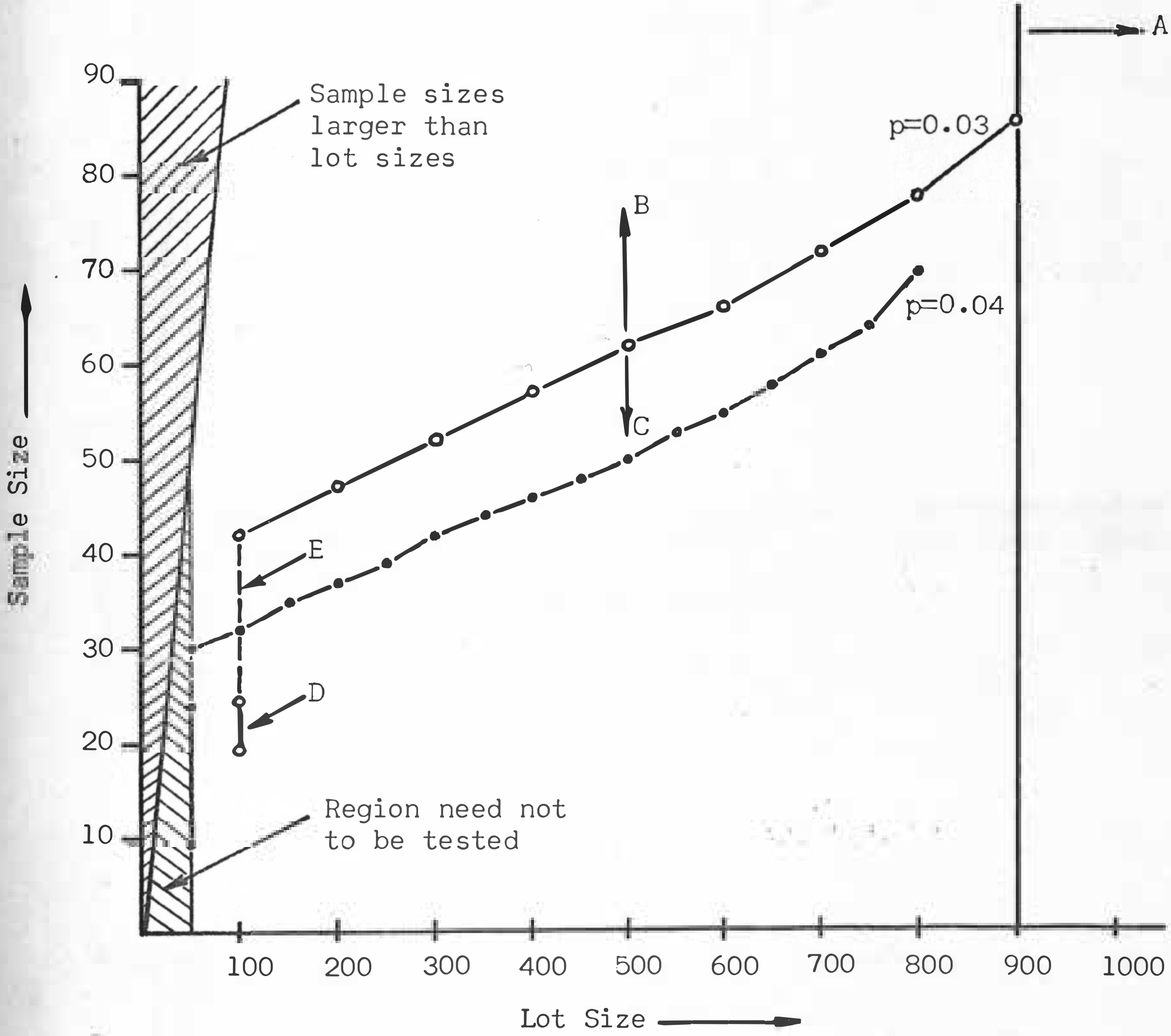


Figure 4-8. Error Lines for Hypergeometric vs Poisson Distribution, $c \leq 1$, p=0.03 and 0.04.

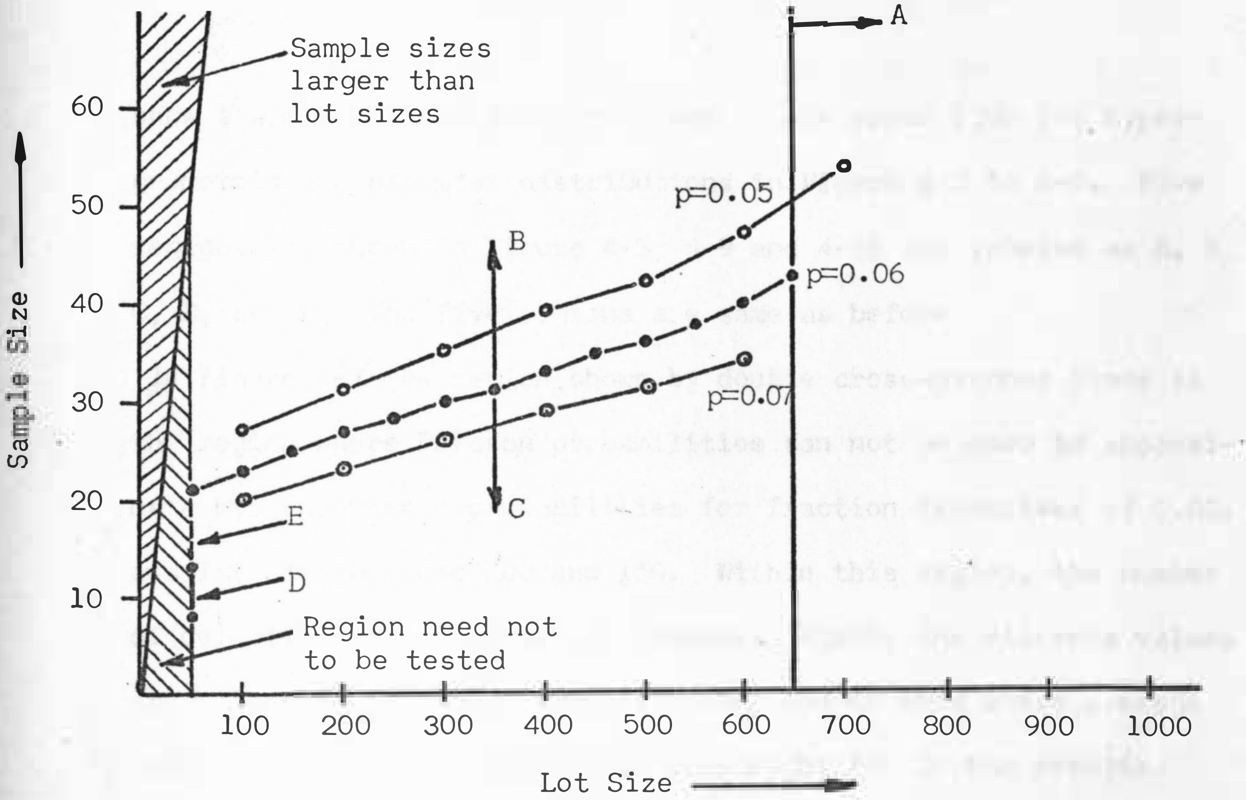


Figure 4-9. Error Lines for Hypergeometric vs Poisson Distribution, c ≤ 1, p=0.05, 0.06 and 0.07.

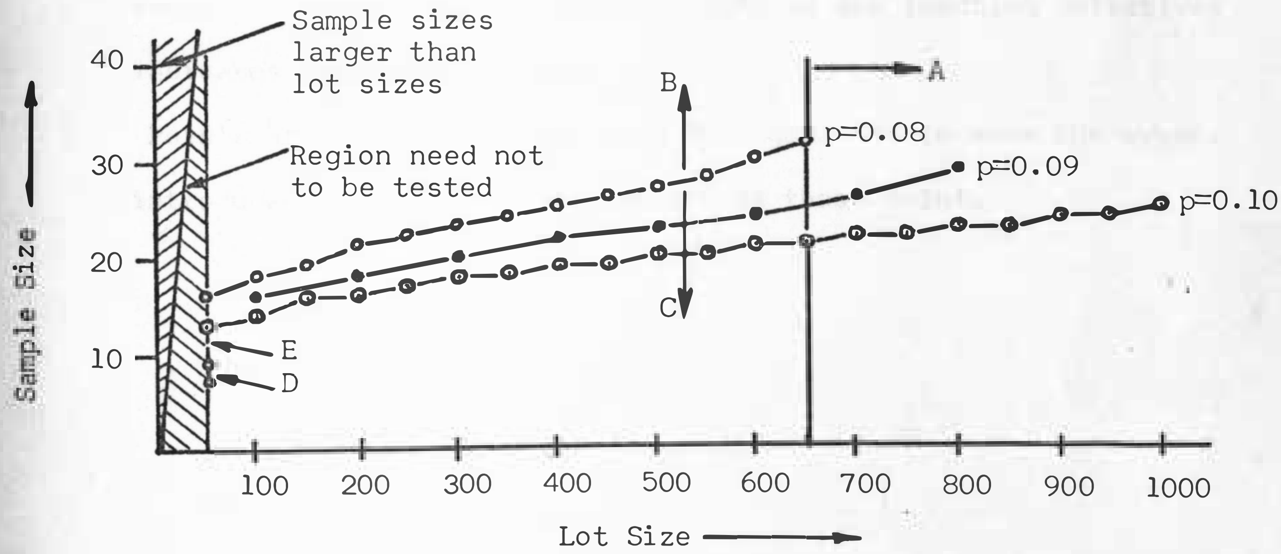


Figure 4-10. Error Lines for Hypergeometric vs Poisson Distribution, c ≤ 1, p=0.08, 0.09 and 0.10.

show a similarity to the comparison of the error line for hypergeometric and binomial distributions in Figure 4-3 to 4-8. Five regions are shown in Figure 4-8, 4-9 and 4-10 and labeled as A, B, C, D, and E. The five regions are same as before.

In Figure 4-7 the region shown by double cross-hatched lines is the region where Poisson probabilities can not be used to approximate hypergeometric probabilities for fraction defectives of 0.02, and lot size between 100 and 150. Within this region, the number of defectives per lot are not integer. Again, the discrete values are connected to aid in identification and to show where average number of defectives in several lots might lie in the results.

The error lines show a similarity to those in Figure 4-2 for c=0 except that the error lines in Figure 4-7 to 4-10 have been shifted upward. Therefore, Region B becomes smaller and Region C becomes larger while the shapes change from trapezoid to almost rectangular because the error lines become flatter as the fraction defectives increases from 0.01 to 0.10.

Points are arbitrarily selected in Figure 4-8 to show the errors introduced by Poisson approximation at those points.

TABLE 4-1

Figure 4-1 c=0	(PH) Hyper-geometric	(PB) Binomial	(PP) Poisson	Difference PH vs PB	Difference PH vs PP
$\bar{A}$(U=700, S=90, p=0.02)	0.1481	0.1614	-----	1.33%	-----
$\bar{B}$(U=300, S=60, p=0.01)	0.5224	0.5445	-----	2.21%	-----
$\bar{C}$(U=400, S=20, p=0.03)	0.5355	0.5445	-----	0.9%	-----
Figure 4-2 c=0					
$\bar{A}$(U=500, S=35, p=0.04)	0.2275	-----	0.2474	-----	1.99%
$\bar{B}$(U=300, S=55, p=0.01)	0.5429	-----	0.577	-----	3.41%
$\bar{C}$(U=300, S=35, p=0.01)	0.689	-----	0.705	-----	1.6%
Figure 4-4 c $\leq$ 1					
$\bar{A}$(U=700, S=70, p=0.04)	0.2093	0.2260	-----	1.57%	-----
$\bar{B}$(U=200, S=70, p=0.04)	0.1636	0.2260	-----	6.24%	-----
$\bar{C}$(U=300, S=30, p=0.04)	0.6629	0.6625	-----	0.04%	-----
$\bar{D}$(U=50, S=20, p=0.04)	0.845	0.8103	-----	3.47%	-----
$\bar{E}$(U=50, S=28, p=0.04)	0.6914	0.6907	-----	0.07%	-----
Figure 4-8 c $\leq$ 1					
$\bar{A}$(U=1000, S=70, p=0.03)	0.3652	-----	0.3805	-----	1.53%
$\bar{B}$(U=600, S=80, p=0.03)	0.2823	-----	0.308	-----	2.57%
$\bar{C}$(U=600, S=40, p=0.03)	0.6595	-----	0.663	-----	0.35%
$\bar{D}$(U=100, S=20, p=0.03)	0.901	-----	0.878	-----	2.3%
$\bar{E}$(U=100, S=40, p=0.03)	0.650	-----	0.663	-----	1.3%

Note: $\underline{\bar{A}}$ refers to the point in Region A., etc.
$\underline{A}$, C, E, are acceptable points
B, D, are non-acceptable points

CHAPTER V

CONCLUSION

Binomial and Poisson probability distributions have been used to approximate the hypergeometric probability distribution in sampling inspection problems because the calculations involved in the hypergeometric probability distribution are lengthy and time-consuming. It is well-known that using these approximations introduces error, but the error limitations have not been well defined. This study provides a computer technique to determine the regions on the sample/lot plane in which the approximations are valid for a given level of accuracy.

The output of the program has been demonstrated for a 2% error limit and acceptance values of 0 and 1 for lots from 50 to 1000 and fraction defectives from 0.01 to 0.10.

This demonstration output is of interest by itself. The regions where errors introduced by binomial and Poisson approximations are greater than 2% have been found to occur in two segments. One, as generally expected, occurs when the sample size is large compared with the lot size. In addition, a region of lower sample size was found for acceptance values of 1, resulting from the non-compensation errors for the segments of the overall probability. The second error area, called Region D in the result, was computed by hand since the computer program was designed to terminate upon finding a 2% error. The alternative was to search all sample sizes for all lot sizes. Logic to recognize a Region D is needed to maximize results and minimize computer time.

Recommendations are as follows:

(1) The computer programs should be modified to find the region D.

(2) Ranges of lot size, fraction defectives and acceptance number should be extended.

The present study was conceived of as a demonstration of the computer technique. Because of limited computer time available, the tests were not carried further. Additional information would improve the usefulness of the output graphs.

(3) The binomial can be approximated by the normal distribution, therefore, the normal distribution should also be studied in those regions of fraction defectives that the normal distribution might be a better approximation to the hypergeometric distribution than the binomial or Poisson distributions.

Note: Decks of computer programs are available from Dr. Richard P. Covert, Mechanical Engineering Department of South Dakota State University, Brookings, South Dakota.

LITERATURE CITED

(1) Kane, G. E. and Rokhsar, A. E., "The effect of hypergeometric probability distribution on the design on sampling plans for small lot sizes", Journal of Industrial Engineering, Nov.-Dec., 1959.

(2) Duncan, A. J., Quality Control and Engineering Statistics, Richard D. Irwin, Inc., 1965.

(3) Larson, H. R., "A nomograph of the cumulative binomial distribution", Journal of Industrial Engineering, April 1967.

(4) National Bureau of Standards, Applied Mathematics Series 6, Table of Binomial Probability Distribution.

(5) Lieberman, G. J. and Owen, D. B., Table of the Hypergeometric Probability Distribution. Standford University Press, California, 1961.