BAYESIAN ONE-SIDED CREDIBILITY BOUNDS
FOR A PROPORTION
USING A TWO-STAGE SAMPLING PLAN
INVOLVING IMPERFECT AND PERFECT CLASSIFICATION.

by

LINDA R. NEDEN

A thesis
presented to the University of Manitoba
in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

Winnipeg, Manitoba

ISBN 0-315-33598-X

BAYESIAN ONE-SIDED CREDIBILITY BOUNDS FOR A PROPORTION USING A TWO-STAGE SAMPLING PLAN INVOLVING IMPERFECT AND PERFECT CLASSIFICATION

BY

LINDA R. NEDEN

A thesis submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

MASTER OF SCIENCE

ABSTRACT

It is desired to estimate the proportion of nonconforming items in a population, and to place a lower "confidence" bound on this proportion. The sampling plan is subject to the following constraints: items in an initial sample are classified by an imperfect classifier, a small subsample is taken from the group classed as nonconforming, and these items are reclassified by a perfect classifier.

Bayes Theorem is used to obtain a posterior distribution for the proportion of nonconforming items in the population. From this an estimate of the proportion and lower credibility bounds may be determined using numerical methods. Prior distributions for the proportion of nonconforming items in the population and for the probability of misclassifying conforming items are modelled using independent beta distributions.

A computer program is provided that accepts a set of observed values and a set of values for the parameters of the priors. Using this input the expected value and standard deviation of the posterior distribution are calculated, as well as $(1-\alpha)100\%$ lower credibility bounds. Plots of the posterior density function and posterior cumulative distribution function are also generated.

ACKNOWLEDGEMENTS

I am immensely grateful to my major advisor, Dr. John F. Brewster for his guidance and advice in all phases of my graduate program, especially his editorial advice. In addition, the comments and criticisms of the other committee members, Dr. Brian D. MacPherson and Dr. Gerry Losey, are appreciated.

I would like to thank Jeff Babb and the Canadian Grain Commission for providing the practical problem on which this thesis is based.

The financial support I received from the National Scientific and Engineering Research Council, in the form of a Summer Research Grant and a Scholarship is gratefully acknowledged.

My special thanks to K. Chris Kirby of the Computer Services Department for programming advice and suggestions on how to produce the graphical material in this thesis.

The interest of my fellow classmates in the progress of my work is much appreciated.

And finally, without my husband's care and encouragement I would never have got to this point. Thank you Ron.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION AND SUMMARY

In a sample taken from a population, the proportion of items with a certain characteristic can provide an estimate of the proportion of items with this characteristic in the population. In certain situations it may be very expensive to correctly classify items in a sample, but there may be a cheaper classification method that does not classify items correctly all of the time. To economically obtain an estimate of the population proportion in such situations it may be desirable to use some combination of these two methods.

An example from the health research field of the use of two types of classifiers is presented by Deming (1977) in which the problem is to estimate the proportion of persons with a particular psychopathology. A large sample of persons are screened at the first stage by trained interviewers and placed into two groups, those apparently having the psychopathology and those appearing not to have it. At the second stage a small sample from each of these two groups is examined by a psychiatrist who makes the final determination of their psychiatric and medical characteristics. The estimate of the population proportion is formed by weighting the proportion of cases in each of the screening stage groups by the proportion of persons found to have the psychopathology in each of the second stage samples.

Perfect and imperfect classifiers are also used in quality control procedures. Frequently a simple method is used for classifying items in a sample from a batch off a production line. If the batch is rejected then the classification of the sample items may be redone using more thorough methods. In many cases two types of classifiers are available, although they may not be used in the manner being considered in this thesis. For example, a sample of items might be classified by the perfect classifier. This set of items is then reclassified by an inspector and the proportion of correct classifications is used to evaluate the inspector's performance.

In order to place a lower bound on a binomial proportion, where the data is subject to misclassification, a two-stage sampling plan may be used to provide information about the misclassification rate. At the first stage a random sample is classified by a fallible method into two categories. From one of the first stage categories a subsample is taken, which is classified by an exact, but expensive, method.

Research into this type of problem was motivated by the need to solve a specific inspection problem from the grain industry. In this thesis, the specific problem will be described in detail, and a Bayesian solution will be developed. This solution can easily be adapted for use in similar situations arising in other fields.

1.1 THE SPECIFIC PROBLEM

A mixture of more than one variety of wheat can occur in a boxcar shipment. In the problem under consideration the mixture consists of two varieties that look very similar, one conforming to licencing standards and the other not. An estimate of the proportion of nonconforming kernels, p_0, is used to grade the carlot; a poorer grade of wheat has in it a larger percentage of the nonconforming variety. As payment to the shipper is based on the grade of wheat, lowering the grade on a carlot has serious consequences. Therefore it is desirable to be able to provide, with a given degree of confidence, a lower bound on p_0.

1.2 THE SAMPLING PROCEDURE

In order to estimate p_0 in any one boxcar, and to place a lower bound on p_0, a two-stage sampling plan is used. At the first stage a random sample of three hundred kernels is taken. These kernels are visually inspected and classified into two groups, those thought to be conforming and those thought to be nonconforming. The characteristics for correct visual classification are not always present, so some kernels may be misclassified into either group. At the second stage a subsample of ten kernels is take from the group classified as nonconforming and these are analyzed by an expensive laboratory technique which correctly classifies each kernel.

No sample is taken from the group classified as conforming, and thus no estimate is available for the number of nonconforming kernels misclassified into this group. It will therefore be assumed that there is no misclassification of nonconforming kernels. This assumption, as will be explained later, provides a conservative result, that is the carlot is less likely to be assigned to a lower grade, which is to the shippers' advantage.

1.3 THE SOLUTION

A Bayesian approach is used to find a solution. This method allows information which has been gathered from the sample to be combined with any information which may be available (prior to the sample being taken) concerning the proportion of interest and the misclassification rate. This prior "information" may be subjective feelings about the distribution of the unknown parameters or it may be based on more specific information provided by an examination of estimates of parameters in earlier samples. The two relevant factors about which there may be some prior knowledge are: the proportion of the nonconforming variety present in previous boxcar shipments, and the proportion of the conforming variety correctly classified by the inspection process in other samples. If it is felt that the process is relatively stable over time, then it would be sensible to utilize the knowledge from past samples in the solution to the current problem. In this particular solution, knowledge about the population proportion and the misclassification probabilities is modelled by independent beta distributions. The beta distribution is a function of two parameters and changes to these parameters permit the approximation of a wide variety of distributions. In particular, a suitable choice of parameters reduces the beta distribution to a uniform distribution which can be used to represent lack of knowledge (that is, from what is known, "any proportion is equally likely to occur"). These prior distributions are combined with the sample information (the number classified as nonconforming from the three hundred kernel sample and the number classified as nonconforming from the ten kernel subsample) to produce a posterior distribution for p_0, the proportion of nonconforming kernels in the carlot. The mean of this distribution provides a point estimate of p_0 and this posterior distribution can also be used to provide a lower bound for p_0. For example, if the 5th percentile of this distribution is found to be 0.15, then one may be 95% confident that p_0 is 0.15 or larger, in the sense that the posterior probability that p_0 exceeds 0.15 is 0.95. In Bayesian terminology, 0.15 is thus a 95% lower credibility bound for p_0.

1.4 AN EXAMPLE

Assume there is no information from previous shipments to incorporate. An inspector classifies a 300 kernel sample as 255 kernels conforming and 45 kernels nonconforming. A 10 kernel subsample is taken at random from the 45 kernels thought to be nonconforming. If the laboratory analysis found all 10 kernels to be nonconforming, indicating that the inspection process is quite effective, a point estimate for p_0 is the posterior mean at 14.28% and the 95% lower bound is 10.63% (Figure 1.1). Suppose high grade wheat can contain at most 8% of the nonconforming Variety. Then, as the lower bound is greater than 8%, there is sufficient evidence to conclude that this boxcar contains enough nonconforming kernels to be graded as lowgrade wheat.

However if the subsample indicated that the inspector was less accurate, with only 8 kernels in the subsample being found to be nonconforming, then a point estimate for p_0 would be 11.70% and the 95% lower bound would be 7.72% (Figure 1.2). In this case the estimate of p_0 would still be greater than 8%, but the lower bound would be below 8%. This indicates that, due to the uncertainty associated with the sampling and inspection processes, there is not sufficient evidence to conclude that the boxcar contains enough nonconforming kernels to grade the carlot as low grade wheat.

If it is known from earlier samples that the proportion of Variety 1 kernels that were correctly classified was

Figure 1.1: POSTERIOR DENSITY FUNCTION FOR p_0

EXPERIMENTAL EVIDENCE:	300 kernels randomly selected from the boxcar; 45 kernels classified as nonconforming, of which 10 are selected and all 10 are confirmed to be nonconforming.
PRIOR SPECIFICATION:	Independent uniform priors, $\beta(1,1)$, on p_0 and on the probability of misclassifying conforming kernels.
POSTERIOR VALUES:	Posterior mean of p_0 = 0.1428; 95% lower bound for p_0 = 0.1063.

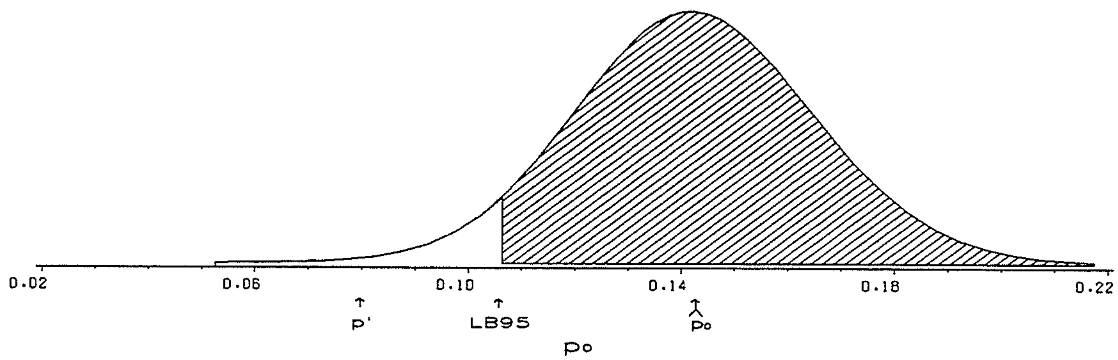


Figure 1.2: POSTERIOR DENSITY FUNCTION FOR p_0

EXPERIMENTAL EVIDENCE:	As above, except that of the 10 kernels in the subsample, 8 are found to be nonconforming.
PRIOR SPECIFICATION:	As above.
POSTERIOR VALUES:	Posterior mean of p_0 = 0.1170. 95% lower bound for p_0 = 0.0772.

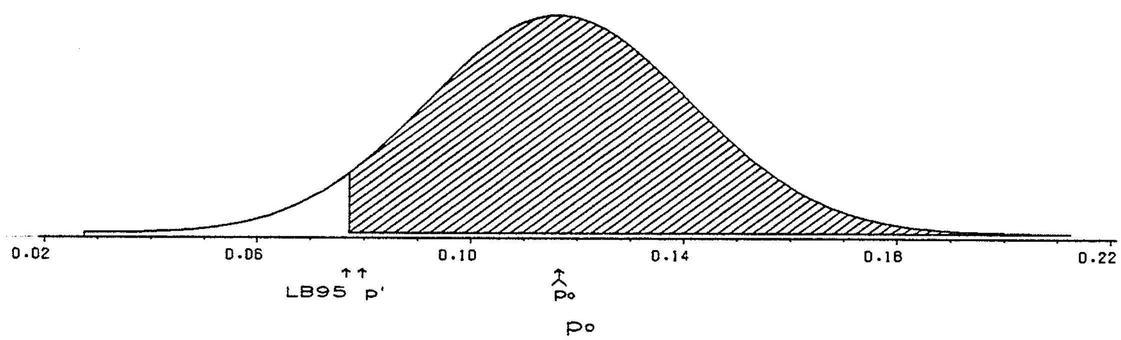
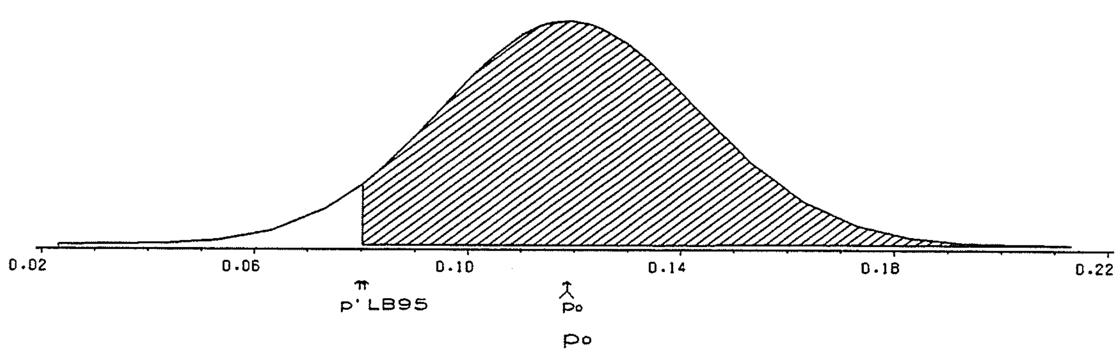


Figure 1.3: POSTERIOR DENSITY FUNCTION FOR p_0

EXPERIMENTAL EVIDENCE:	As in Figure 1.2.
PRIOR SPECIFICATION:	Independent uniform prior, $\beta(1,1)$, on p_0, and $\beta(10,1)$ prior on the probability of misclassifying conforming kernels.
POSTERIOR VALUES:	Posterior mean of p_0 = 0.1194; 95% lower bound for p_0 = 0.0802.



0.91, on average, and if 95% of the time the proportion of Variety 1 kernels that were correctly classified was 0.74 or greater, then inclusion of this information into the analysis would change the result. One prior distribution that would have these properties would be a beta distribution with parameters (10,1). Using this as a prior for the inspection step, and with the same sample data as in the previous example, the point estimate for p_0 would become 11.94% and the 95% lower bound would become 8.02% (Figure 1.3)-- a large enough value to grade the boxcarlot as low grade wheat.

1.5 ORGANIZATION OF THESIS

A detailed explanation of the problem is given in Chapter II, together with the notation used and a discussion of conditions particular to this problem. Chapter III contains an overview of relevant Bayesian methods. In Chapter IV a Bayesian solution to the specific problem is presented. A literature survey of related work is given in Chapter V. Chapter VI contains concluding remarks about the problem, the proposed solution, and possible adaptations and extensions of the solution. A computer program to generate a solution for observed sample values can be found in Appendix A.

Chapter II

THE PROBLEM

2.1 THE GENERAL PROBLEM

It is desired to estimate the proportion of items in a population having a particular characteristic and to place a lower bound (in some statistical sense) on this proportion. The reason for placing a lower bound on this proportion is to see whether the sample provides sufficient evidence for concluding that the proportion exceeds some critical value. The case where there are two methods of classifying items in a sample is being considered. One is an infallible, but very expensive, method, such that only a small, fixed number of items can be classified. The other is a fallible, but less expensive, method, and thus many more items, within practical limitations, can be classified.

2.2 THE SPECIFIC PROBLEM

The specific problem under consideration arises in the grading of boxcarlots of wheat. The boxcar may contain a mixture of two varieties of wheat: Variety 1, that conforms to grading standards and Variety 2, that does not. The shipper is paid less if the carlot contains more than a specified proportion, p', of the nonconforming variety. It is therefore desirable to test the following hypotheses concerning p_0, the proportion of Variety 2 in the carlot:

$$H_0: p_0 \le p'$$

$$H_1: p_0 > p'.$$

The specified value, p', can be quite small (less than 0.1).

Determining a lower bound on the proportion p_0, with a certain degree of "confidence", will permit the testing of these hypotheses. If the lower bound is found to be greater than p', then the null hypothesis will be rejected in favour of the alternative.

In this thesis Bayesian methods will be used to determine this lower bound, that is, for some values of α, $0 < \alpha < 1$, $(1-\alpha)\cdot 100\%$ lower credibility bounds will be found. Hence the associated tests will have a Bayesian interpretation.

The two varieties of wheat mentioned previously are very similar in appearance. Correct visual identification is based on subtle variations in shape and colour and on the relative lengths and positions of parts of the kernels. Due to variations in growing conditions, storage periods and other uncontrollable factors the characteristics for correct visual classification are not always distinguishable or present. Thus, during visual inspection some kernels of wheat may be misclassified. The misclassification can be in either direction: a Variety 1 kernel may be classified as Variety 2, or a Variety 2 kernel may be classified as Variety 1. There is a laboratory technique available that can give an exact identification of kernels of wheat, but this method is very expensive and so only an extremely small number of kernels can be classified.

2.3 NOTATION

To clarify discussion of the problem the following notation is introduced and a diagram of the sampling procedure is presented in Figure 2.1.

p_0 = proportion of non-conforming kernels in the carlot.

p_1 = probability of a conforming kernel being correctly identified at the inspection step.

p_2 = probability of a non-conforming kernel being correctly identified at the inspection step.

n = total number of kernels in the initial sample.

y = number of non-conforming kernels in the initial sample.

n_1 = number of kernels in the initial sample that are classified as conforming at the inspection step.

n_2 = number of kernels in the initial sample that are classified as non-conforming at the inspection step.

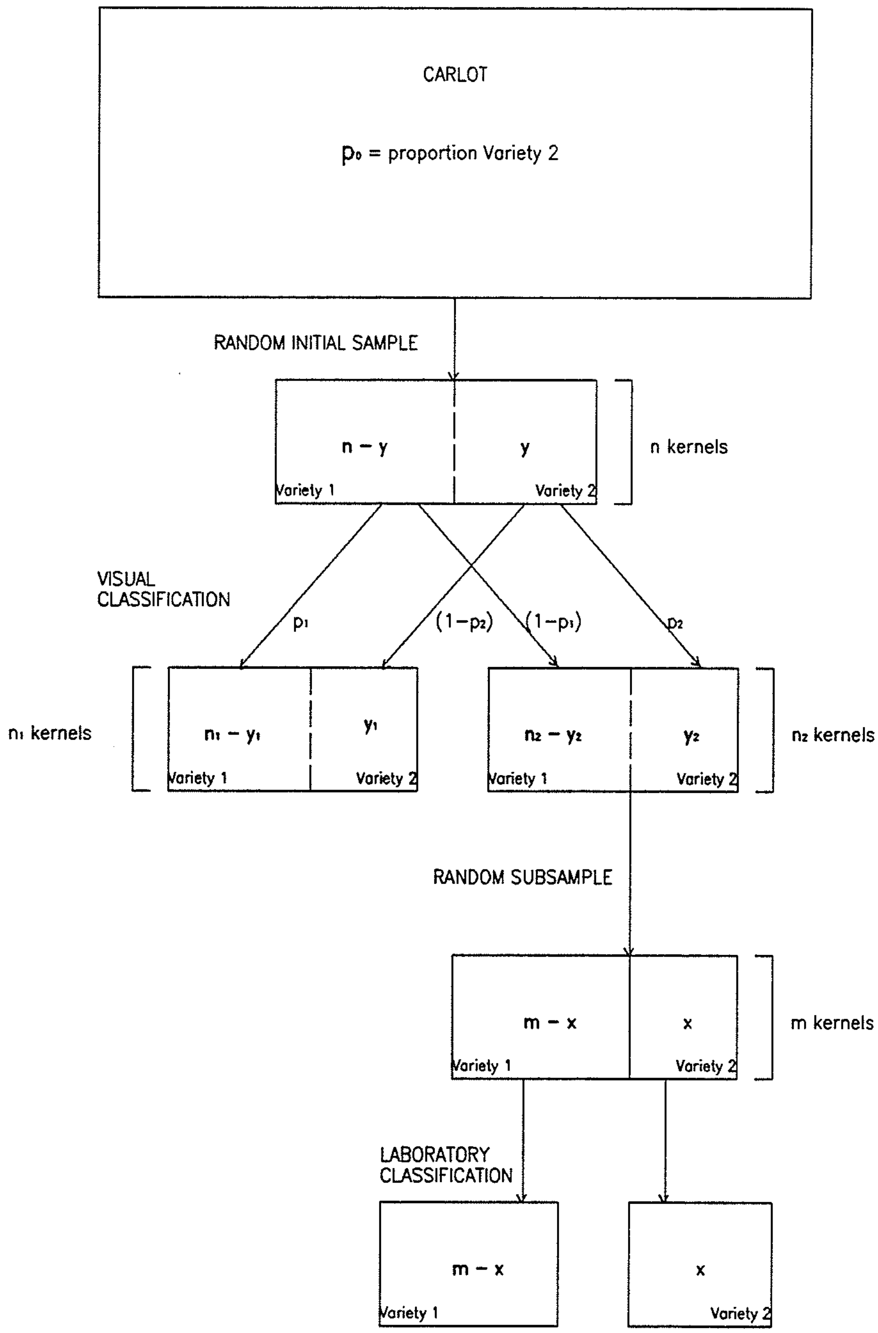
y_1 = number of non-conforming kernels in the initial sample classified as conforming at the inspection step.

y_2 = number of non-conforming kernels in the initial sample classified as non-conforming at the inspection step.

m = total number of kernels in the subsample.

x = number of kernels confirmed to be non-conforming by the laboratory technique.

Figure 2.1: SAMPLING PLAN



2.4 SAMPLING PLAN

The carlot contains a proportion p_0 of Variety 2 kernels. An initial random sample of n kernels is taken from the carlot. It contains an unknown number, y, kernels of Variety 2. These n kernels are classified by visual inspection into 2 groups: n_1 kernels thought to be Variety 1 and n_2 kernels thought to be Variety 2. The n_1 kernels thought to be Variety 1 consist of an unknown number, y_1, kernels of Variety 2 kernels misclassified as Variety 1 and n_1-y_1 kernels of Variety 1. Similarly the n_2 group consists of y_2 kernels of Variety 2 and n_2-y_2 kernels of Variety 1 misclassified as Variety 2.

The probability of a Variety 1 kernel being correctly classified as Variety 1 is p_1 and the probability of a Variety 2 kernel being correctly classified as Variety 2 is p_2. The probability p_1 is not necessarily equal to p_2. It is assumed that these probabilities of correct classification are constant from kernel to kernel, independent of the classifications of any other kernels.

Since no sample is taken from the n_1 group and as only a one-sided bound on p_0 is desired, an assumption that $p_2 = 1$ will be made in deriving the answer. Under this assumption $y_1 = 0$ and therefore all kernels in the n_1 group are considered to be Variety 1. The assumption that $p_2 = 1$ provides a conservative result, as any smaller value of p_2 would cause the estimated value of p_0 and the lower bound for p_0 to be larger. From the hypothesis testing point of view, if H_0 can be rejected with the assumption that $p_2 = 1$, then it would also be rejected for smaller values of p_2.

A random subsample of size m is taken from the n_2 kernels thought to be Variety 2. These m kernels are examined by the laboratory technique and classified as m-x kernels of Variety 1 and x kernels of Variety 2. In $n_2 < m$ no subsample is taken.

The values of n, n_1, n_2, m and x can be observed. For example, in an actual problem an initial sample of n = 300 kernels would be taken from the carlot, these kernels being divided by visual inspection into two groups, $n_1 = 255$ thought to be Variety 1 and $n_2 = 45$ thought to be Variety 2. A subsample of m = 10 kernels would then be taken from the 45 kernels thought to be Variety 2. Upon laboratory analyses, x = 8 kernels may be found to be of Variety 2 and m-x = 2 kernels may be found to be of Variety 1. The question is "How can p_0 be estimated and how can a lower bound for p_0 be determined, in order to test the hypotheses H_0: $p_0 \leq 0.08$ vs H_1: $p_0 > 0.08$?"

2.5 CONDITIONS PARTICULAR TO THIS PROBLEM

There are two methods available for classifying kernels of wheat, visual inspection and laboratory analysis. The imperfect classification method, visual inspection, is relatively inexpensive, however, use of this method alone will lead to a biased estimate of the proportion of Variety 2 wheat in the carlot (Bross 1954). If the misclassification rates were known to be the same for kernels from any boxcar, calibration samples could be classified by both the imperfect and perfect methods, and the average results could be used to correct for bias in the estimates from samples classified only by the imperfect method (Trader 1983, Wooding 1979). In this problem, however, the misclassification rates may not be constant from carlot to carlot and thus information about the misclassification rates needs to be obtained from each carlot.

The perfect classification method, laboratory analysis of the kernels, is very expensive and thus it is impractical to use this method to classify all kernels in the initial sample. Laboratory analysis is therefore used to reclassify a subsample of kernels, previously classified by visual inspection, in order to provide information about the probability of correctly classifying kernels. Because of the nature of the laboratory analysis the subsample size is fixed and no variation in size of subsample, either with the size of the group it is removed from, or with other factors which might affect optimal subsample size, is considered.

As a lower bound is desired in this problem, the whole of the subsample is taken from the n_2 group, those classified as Variety 2 by inspection. As mentioned above, this results in a conservative answer, in view of the assumptions that will be made. If an upper bound was also desired, a second subsample would need to be taken from the n_1 group in order to obtain information about the probability of misclassification in the other direction. In fact it is clear that samples should be taken from both groups (occasionally at least) in order that the effects of both kinds of misclassification can be assessed.

The proportion of interest, p_0, may be quite small. This small proportion, together with the small size of the subsample prohibits the use of asymptotic maximum likelihood techniques based on normal approximations. In order to combine the information available from the two stages of this particular sampling procedure, to incorporate information available prior to sampling and to cope with the limitations mentioned above, Bayesian methods will be used to estimate the proportion p_0, and to provide a lower bound for p_0.

Chapter III

BAYESIAN METHODS

Before developing the Bayesian solution to the sampling problem presented in Chapter II, an overview of relevant Bayesian methods and their purpose and implication will be presented. A simple binomial example will be used to describe these methods. In Bayesian inference, in order to make statements about a parameter p, information from the sample (expressed through the likelihood function) is combined with any prior information that exists about p (expressed through the prior distribution). The effect of the choice of the prior on the answer in the binomial example will be examined as motivation for the choice of the prior in the problem of interest. A basic general reference for Bayesian procedures is Box and Tiao (1973).

3.1 A BINOMIAL EXAMPLE

Consider a binomial random variable Y, so that

$$f(y|p) = C(n,y)\, p^{y} (1-p)^{n-y}, \qquad y = 0,1,\ldots,n.$$

For a fixed observed y, if $f(y|p)$ is regarded as a function of p ($0 \leq p \leq 1$), then it is referred to as the likelihood function (unique up to a multiplicative constant). The likelihood functions for two sets of data ($n = 25$, $y = 5$ and $n = 25$, $y = 15$) are shown in Figures 3.1(a) and 3.1(b).

Figure 3.1: LIKELIHOOD FUNCTIONS

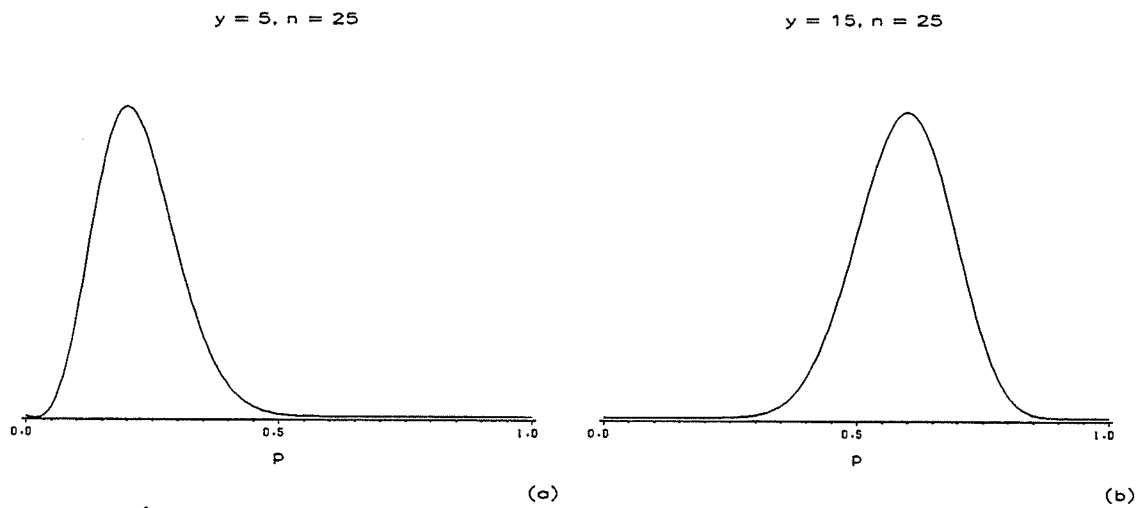


Figure 3.2: PRIOR PROBABILITY DENSITY FUNCTIONS

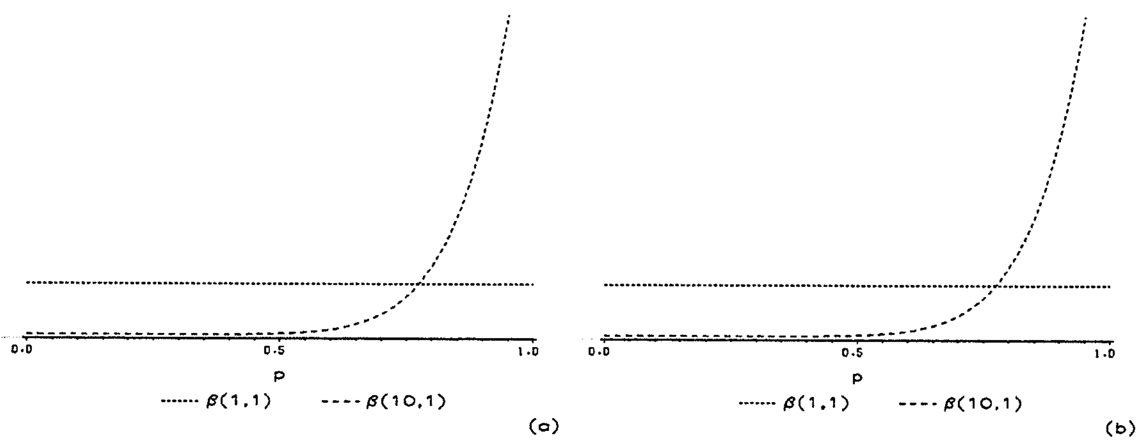
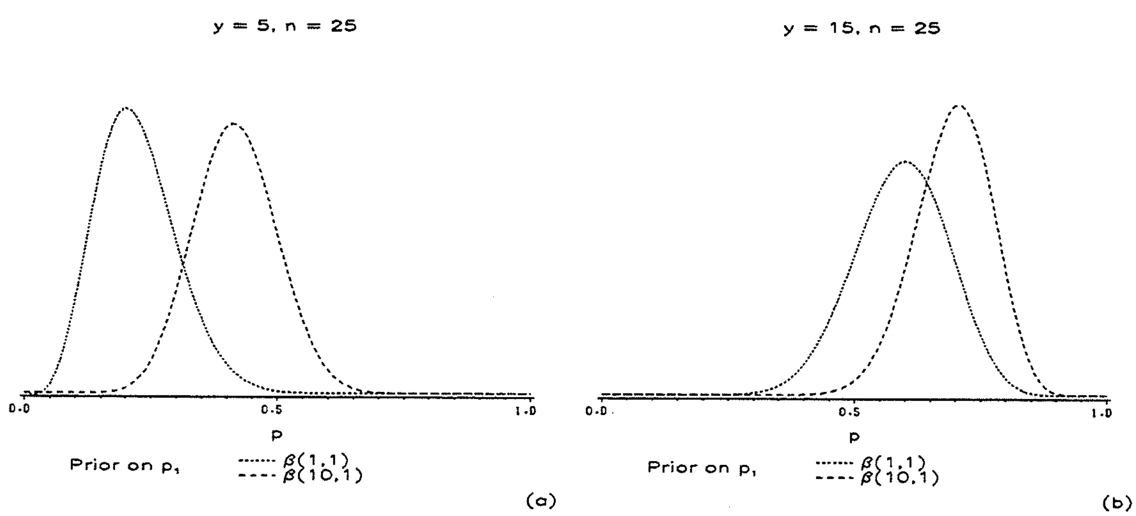


Figure 3.3: POSTERIOR PROBABILITY DENSITY FUNCTIONS



3.2 PRIOR DISTRIBUTION

Information known about p, without knowledge of the data, is expressed in a prior probability distribution. For convenience, the distribution will be assumed to be continuous and the prior density function will be denoted by h(p). The use of a prior distribution allows assumptions about the distribution of the parameter of interest to be formally considered.

One possible prior distribution for p in the binomial example being considered is a beta distribution, $\beta(a,b)$. The probability density function for the beta distribution, $\beta(a,b)$ is

$$h(p) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1} (1-p)^{b-1}, \quad 0 \le p \le 1,$$

for $0 < a,b$. Changes to the two parameters to this distribution will allow approximation of a wide variety of distributions.

An initial choice of parameters for the beta prior might be $a = 1$, $b = 1$, which reduces the beta distribution to the uniform distribution. Use of the uniform distribution can be thought of as modelling "ignorance", "any value of p being equally likely" prior to sampling.

Box and Tiao (1973) suggest that Jeffrey's prior, $\beta(0.5,0.5)$, is the appropriate non-informative prior for a binomial mean, but that for moderate sample sizes "the posterior density is not very sensitive to the precise choice of a prior" (page 36), and the use of $\beta(1,1)$ is approximately equivalent to using $\beta(0.5,0.5)$.

In many problems the use of Bayesian procedures with specific non-informative priors will produce answers that agree with classical answers (having a frequency interpretation). In some cases the appropriate prior may be improper, in the sense that the integral of the prior over the parameter space is infinite. The choice of the prior used to get classical answers partially depends on the goal. For example, when the lower bound is of interest, a 95% lower credibility bound for a binomial proportion found using an improper prior of $\beta(0,1)$ is the same as the classical 95% lower confidence bound. Alternatively when the upper bound is of interest the use of the improper prior, $\beta(1,0)$, leads to a 95% upper credibility bound that corresponds to a classical 95% upper confidence bound.

If more precise information is available about which values of p are more likely to occur, perhaps from earlier similar experiments, then alternate values for a and b can be chosen. For example a = 10, b = 1 might be chosen to model prior knowledge that $P(p > 0.7402) = 0.95$. Priors $\beta(1,1)$ and $\beta(10,1)$ are illustrated in Figure 3.2.

3.3 POSTERIOR DISTRIBUTION

Using Bayes Theorem the sampling information and prior information are combined into the conditional distribution of p, given the sample results, known as the posterior distribution. The posterior density function is given by

$$g(p|y) = \frac{f(y|p)\cdot h(p)}{f_1(y)} ,$$

where the marginal distribution $f_1(y) = {}_0\int^1 f(y|p)\cdot h(p)\, dp$ 'normalizes' the product so that $g(p|y)$ is a proper probability density function. That is, ${}_0\int^1 g(p|y)\, dp = 1$.

The family of beta distributions is the conjugate family for binomial sampling (Raiffa and Schlaifer). That is, when a beta prior is combined with binomial sample information through Bayes Theorem the resultant posterior distribution is also a member of the family of beta distributions. For the binomial example considered here , with a beta prior, the posterior distribution is the beta distribution, $\beta((a+y),(n-y+b))$.

In the numerical example, with $n = 25$ and $y = 5$, if the prior is $\beta(1,1)$ then the posterior distribution is $\beta(6,21)$. In this same example if $y = 15$ then the posterior distribution is $\beta(16,11)$. Similarly for a prior of $\beta(10,1)$ the posterior for $y = 5$ is $\beta(15,21)$ and for $y = 15$ is $\beta(25,11)$.

The plots in Figures 3.3.(a) and 3.3.(b) show that with the ignorance prior, $\beta(1,1)$, there is no shift from the likelihood, whereas the prior $\beta(10,1)$ has a stronger influence.

Figures 3.4, 3.5 and 3.6 illustrate that when the sample size increases from $n = 25$ to $n = 200$, with the proportion of successes remaining constant, (that is, $y = 15$ for $n = 25$ and $y = 120$ for $n = 200$), changing the parameters for the prior has less effect on the posterior distribution. The data then "dominates" the prior for large sample sizes.

Figure 3.4: LIKELIHOOD FUNCTIONS

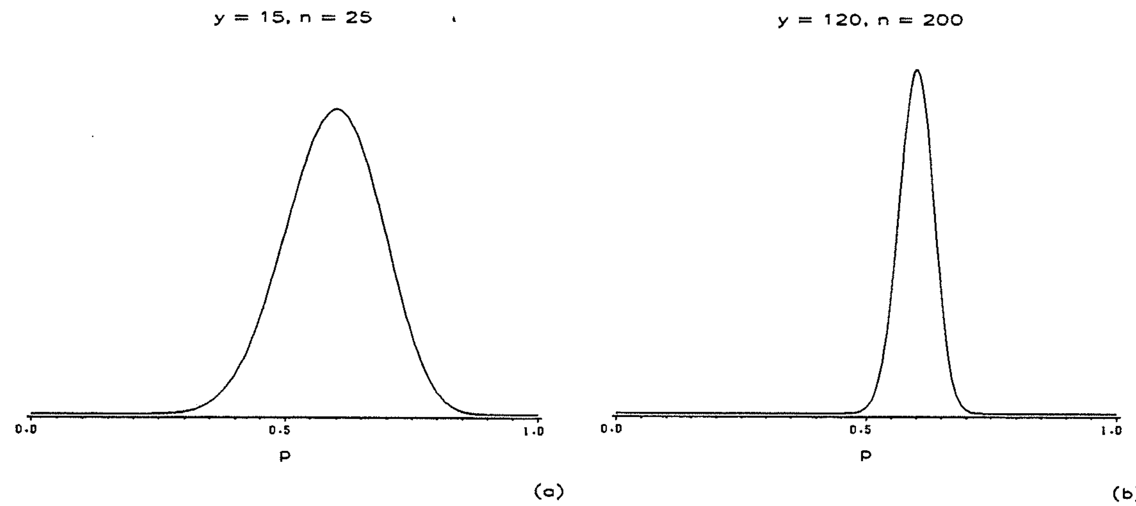


Figure 3.5: PRIOR PROBABILITY DENSITY FUNCTIONS

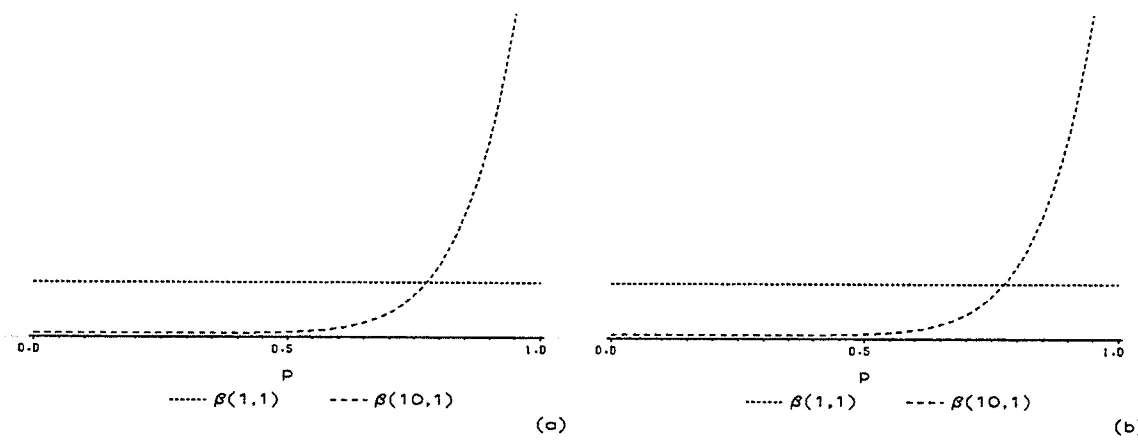
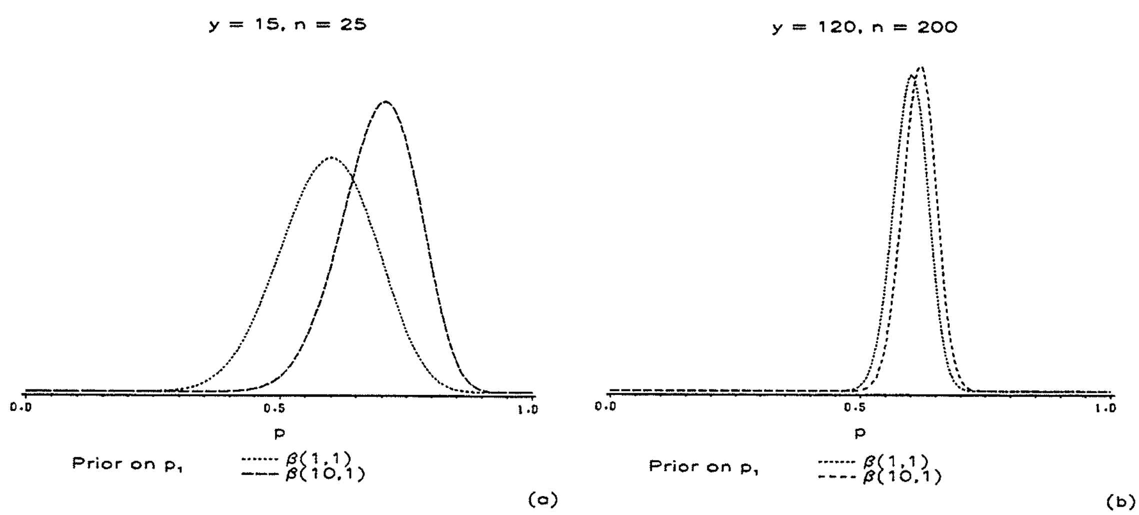


Figure 3.6: POSTERIOR PROBABILITY DENSITY FUNCTIONS



3.4 BAYES CREDIBILITY REGION

Inferences about the parameter of interest are made based on the posterior distribution. One way to summarize the information presented by the posterior distribution is to report a credibility interval which contains a stated amount of the probability. A $(1-\alpha)\cdot 100\%$ lower credibility bound for the population proportion is that value of p which divides the area under the posterior density curve into two regions, that below p contains $\alpha\cdot 100\%$ of the area (or α of the probability) and that above p contains $(1-\alpha)\cdot 100\%$ of the area.

For the numerical example being considered, lower credibility bound for p, for α = 0.10, 0.05 and 0.01 are given in Table 3.1.

TABLE 3.1

			$(1-\alpha)\cdot 100\%$ lower bounds		
PRIOR	y	n	90%	95%	99%
$\beta(1,1)$	5	25	0.1260	0.1056	0.0733
$\beta(10,1)$	5	25	0.3129	0.2858	0.2380
$\beta(1,1)$	15	25	0.4707	0.4357	0.3716
$\beta(10,1)$	15	25	0.5944	0.5637	0.5057
$\beta(1,1)$	120	200	0.5546	0.5418	0.5178
$\beta(10,1)$	120	200	0.5730	0.5605	0.5370

3.5 NUISANCE PARAMETERS

If the distribution of the observed value depends not only upon the parameter of interest, but also on other parameters, then these are known as nuisance parameters. In the Bayesian approach inferences about the parameter of interest are based on the marginal posterior distribution that is obtained by integrating, over the parameter space of the nuisance parameters, the joint posterior distribution of the parameter of interest and the nuisance parameters. That is, the nuisance parameters are "integrated out".

3.6 SENSITIVITY

In drawing conclusions based on a posterior distribution, the sensitivity of the posterior to changes to the parameters of the priors should be considered. If the posterior distribution shows little change under different prior distributions this indicates that the information coming from the sample dominates the results. In this case the choice of parameters for the prior is not very critical. However if the posterior is sensitive to changes to the parameters of the prior then it is important to recognize this fact. In some cases this may indicate the advisability of obtaining more data.

3.7 RELATIONSHIP TO CLASSICAL RESULTS

As mentioned in Section 3.2, with ignorance priors the Bayesian results may agree with classical results. If there is a discrepancy, then it will tend to be smaller for large sample sizes. In binomial sampling the $(1-\alpha)\cdot 100\%$ classical lower confidence bound is that value of p for which

$$\sum_{i=y}^{n} C(n,i)\, p^{i} (1-p)^{n-i} = \alpha, \quad 0 \le p \le 1.$$

By examining Table 3.2 it can be seen that the classical lower confidence bounds are equal to the lower credibility bounds with an improper prior, $\beta(0,1)$.

TABLE 3.2

		CLASSICAL LOWER BOUNDS		
y	n	90%	95%	99%
5	25	0.1006	0.0822	0.0542
10	25	0.4523	0.4168	0.3524
120	200	0.5525	0.5397	0.5156

			BAYES LOWER BOUNDS		
PRIOR	y	n	90%	95%	99%
$\beta(0,1)$	5	25	0.1006	0.0822	0.0542
$\beta(0,1)$	10	25	0.4523	0.4168	0.3524
$\beta(0,1)$	120	200	0.5525	0.5397	0.5156
$\beta(1,1)$	5	25	0.1256	0.1056	0.0734
$\beta(1,1)$	15	25	0.4707	0.4357	0.3716
$\beta(1,1)$	120	200	0.5547	0.5419	0.5178
$\beta(10,1)$	5	25	0.3129	0.2859	0.2383
$\beta(10,1)$	15	25	0.5944	0.5637	0.5059
$\beta(10,1)$	120	200	0.5730	0.5605	0.5371

Chapter IV

THE SOLUTION

4.1 WHY A BAYESIAN SOLUTION?

The grain grading problem was first considered using a classical approach. However difficulties arose in developing an estimator that would combine the information coming from the initial sample with that coming from the subsample, and in particular in the placing of confidence bounds on the estimator. The small subsample size and the interest in testing for a small proportion suggested that asymptotic methods such as those proposed by Tenenbien (1970) would not be suitable. Use of Bayesian methods allows for handling of the nuisance parameters p_1 and p_2 (the probability of correctly classifying Variety 1 and Variety 2 kernels). It also allows explicit inclusion of information external to the data, such as opinions about likely values of the parameters or estimates of parameter distributions from earlier samples.

4.2 CHOICE OF PRIOR DISTRIBUTIONS

It has been indicated in the literature (Biegel 1974, Sinclair 1978) that p_0, the parameter of interest, and the nuisance parameters, p_1 and p_2, may be dependent upon one another. Despite this, a solution to the specific problem presented in this thesis will be developed that considers these parameters to be independent. This assumption of independence allows the joint prior distribution of these parameters to be written as the product of independent prior distributions. As was mentioned in Section 2.4, an assumption will be made that no misclassification of nonconforming kernels occurs. Thus the prior distribution of p_2 is modelled by the delta function $p_2 = 1$. The prior distributions of p_0 and p_1 will be modelled by independent beta distributions. The choice of beta distributions as priors for these parameters is further discussed in Sections 5.2.2 and 5.2.3. As noted in Section 3.2 the beta is a flexible distribution with an appropriate range which can approximate many other distributions. Because of its mathematical manageability, use of beta priors in this problem permits clear illustration of the use of priors in general and their effect on inferences. If more suitable prior distributions can be found, or empirical distributions can be developed from a series of samples, then they can be incorporated into the analysis in a manner similar to that used for the independent beta priors. If the prior is more complex, then additional numerical difficulties will undoubtedly arise. On the other hand, the beta distributions could be replaced by finite mixtures of betas without increasing the complexity a great deal. This would permit approximation to a wider class of prior distributions.

4.3 DISTRIBUTION OF THE SAMPLE

From the sampling plan in Figure 2.1, in Chapter II, it is seen that only values of the random variables N_2 and X may be observed. Since it is assumed, for reasons given in Section 2.4, that $p_2 = 1$, the probability of observing values of these random variables is only conditional on the values of p_0 and p_1.

Lemma 4.1.

$P(N_2=n_2, X=x \mid p_0, p_1) =$

$$\sum_{y_2} \left[\frac{n!}{(n-n_2)!(n_2-y_2)!y_2!} \cdot [p_1(1-p_0)]^{n-n_2} [(1-p_1)(1-p_0)]^{n_2-y_2} p_0^{y_2} \cdot \frac{C(y_2,x) \cdot C(n_2-y_2,m-x)}{C(n_2,m)} \right] \tag{4.3.1}$$

for $x = 0,1,\ldots,m$; $n_2 = m,m+1,\ldots,n$; and where the summation is over $y_2 = x,x+1,\ldots,n_2-m+x$. [Here, $0 \le p_0,p_1 \le 1$]

Proof of Lemma 4.1. The necessary assumptions are that the carlot approximates an infinite population, that the probability of correctly classifying any one kernel is independent of how any other kernel is classified, and that the probabilities of correctly classifying kernels, p_1 and p_2, are constant for all kernels. It can then be shown that after classification of the initial sample the joint distribution of $N_1 - Y_1$, Y_1, $N_2 - Y_2$ and Y_2 is multinomial, with observed values as shown in Table 4.1 below.

TABLE 4.1

INSPECTION CLASSIFICATION	TRUE CLASSIFICATION VARIETY 1	TRUE CLASSIFICATION VARIETY 2	
VARIETY 1	n_1-y_1	y_1	n_1
VARIETY 2	n_2-y_2	y_2	n_2
	$n-y$	y	n

The associated probabilities are given in Table 4.2

TABLE 4.2

INSPECTION CLASSIFICATION	TRUE CLASSIFICATION VARIETY 1	TRUE CLASSIFICATION VARIETY 2	
VARIETY 1	$p_1(1-p_0)$	$(1-p_2)p_0$	$p_1+(1-p_1-p_2)p_0$
VARIETY 2	$(1-p_1)(1-p_0)$	p_2p_0	$(1-p_1)-(1-p_1-p_2)p_0$
	$1-p_0$	p_0	1

As only a lower credibility bound is desired in this particular case, and as a subsample is taken only from the group classified as Variety 2, an assumption is made that $p_2 = 1$. Under this assumption $y_1 = 0$, and the joint distribution of N_1, $N_2 - Y_2$ and Y_2 is multinomial with cell probabilities as given below.

cell	n_1	n_2-y_2	y_2	n
cell probability	$p_1(1-p_0)$	$(1-p_1)(1-p_0)$	p_0	1

Therefore,

$P(N_1=n_1, N_2=n_2, Y_2=y_2 | p_0, p_1) =$

$$\frac{n!}{n_1!(n_2-y_2)!y_2!} \cdot [p_1(1-p_0)]^{n_1} [(1-p_1)(1-p_0)]^{n_2-y_2} p_0^{y_2}, \qquad (4.3.2)$$

where $n_2 = 0,1,\ldots,n$; $y_2 = 0,1,\ldots,n_2$ and $n_1 + n_2 = n$.

The conditional distribution of the number of Variety 2 kernels confirmed to be in the subsample by the laboratory technique is hypergeometric. In other words,

$$P(X=x | N_2=n_2, Y_2=y_2) = \frac{C(y_2,x)C(n_2-y_2,m-x)}{C(n_2,m)}, \qquad (4.3.3)$$

where $x = MAX(0,(m-n_2+y_2)),\ldots,MIN(m,y_2)$, and it follows that

$P(N_2=n_2, Y_2=y_2, X=x | p_0, p_1) =$

$$P(N_1=n_1, N_2=n_2, Y_2=y_2 | p_0, p_1) \cdot P(X=x | N_2=n_2, Y_2=y_2),$$

where the expressions on the right hand side are given in Equations 4.3.2 and 4.3.3. As Y_2 is not observable, in order to obtain $P(N_2=n_2, X=x | p_0, p_1)$ the above probabilities are summed over values of y_2 to obtain Equation 4.3.1.

Note. If Equation 4.3.1 is summed over the stated values of N_2 and X then the result is equal to

$$P(N_2 \geq m) = \sum_{n_2} \left[C(n,n_2) \, [1-(1-p_0)p_1]^{n_2} [(1-p_0)p_1]^{n-n_2} \right] < 1,$$

where the summation is over $n_2 = m, m+1, \ldots, n$. The over-all sum being less than one is a consequence of no subsample being taken if $n_2 < m$. (Notice that N_2 follows a binomial distribution.)

4.4 PRIOR DISTRIBUTION

For a Bayesian solution a prior needs to be specified for p_0 and p_1. For simplicity the joint distribution of p_0 and p_1 will be modelled as two independent beta distributions, so that

$$h(p_0,p_1) = h_0(p_0) \cdot h_1(p_1), \quad 0 \leq p_0,p_1 \leq 1. \qquad (4.4.1)$$

where

$$h_0(p_0) = \frac{\Gamma(a_0+b_0)}{\Gamma(a_0)\Gamma(b_0)} \cdot p_0^{a_0-1} (1-p_0)^{b_0-1}, \quad 0 \leq a_0,b_0$$

and

$$h_1(p_1) = \frac{\Gamma(a_1+b_1)}{\Gamma(a_1)\Gamma(b_1)} \cdot p_1^{a_1-1} (1-p_1)^{b_1-1}, \quad 0 \leq a_1,b_1 .$$

4.5 POSTERIOR DISTRIBUTION

Lemma 4.2. Having observed $N_2=n_2$ and $X=x$, and using independent beta priors for p_0 and p_1, the posterior density function of p_0, $g_0(p_0|n_2,x)$, is given by

$$\frac{\displaystyle\sum_{y_2}\left[\frac{(n_2-y_2+b_1-1)!\,p_0^{\,y_2+a_0-1}(1-p_0)^{n-y_2+b_0-1}}{(y_2-x)!\,(n_2-y_2-m+x)!\,(n-y_2+a_1+b_1-1)!}\right]}{\displaystyle\frac{1}{(n+a_0+b_0-1)!}\sum_{y_2}\left[\frac{(n_2-y_2+b_1-1)!\,(n-y_2+b_0-1)!\,(y_2+a_0-1)!}{(y_2-x)!\,(n_2-y_2-m+x)!\,(n-y_2+a_1+b_1-1)!}\right]}, \quad (4.5.1)$$

$0 \le p_0 \le 1$. Here, $x = 0,1,\ldots,m$; $n_2 = m,m+1,\ldots,n$ and the summations are over $y_2 = x,x+1,\ldots,n-m+x$.

Proof of Lemma 4.2. The joint posterior of p_0 and p_1 is

$$g(p_0,p_1|n_2,x) = \frac{P(N_2=n_2,X=x|p_0,p_1)\cdot h(p_0,p_1)}{P(N_2=n_2,X=x)}$$

where the expression in the numerator is the product of Equations 4.3.1 and 4.3.2, and where the normalizing constant is given by

$$P(N_2=n_2,X=x) = \int_0^1\int_0^1 P(N_2=n_2,X=x|p_0,p_1)\cdot h(p_0,p_1)\,dp_0\,dp_1.$$

It follows that the marginal posterior for p_0 is

$$g_0(p_0,n_2,x) = \int_0^1 g(p_0,p_1|n_2,x)\,dp_1,$$

as given by Equation 4.5.1.

||

The associated posterior cumulative distribution function is given by

$$G_0(p_0|n_2,x) = \int_0^{p_0} g_0(p|n_2,x)\ dp\ ,$$

$$= \sum_{y_2} \int_0^{p_0} A_{y_2}\ p^{y_2+a_0-1}\ (1-p)^{n-y_2+b_0-1}\ dp\ ,$$

where the summation is over $y_2 = x, x+1, \ldots, n-m+x$.

Here

$$A_{y_2} = \frac{(n_2-y_2+b_1-1)!\ (n+a_0+b_0-1)!}{(y_2-x)!(n_2-y_2-m+x)!(n-y_2+a_1+b_1-1)!\ B}\ ,$$

where

$$B = \sum_{y_2} \left[\frac{(n_2-y_2+b_1-1)!(n-y_2+b_0-1)!(y_2+a_0-1)!}{(y_2-x)!(n_2-y_2-m+x)!(n-y_2+a_1+b_1-1)!}\right]\ .$$

As $G_0(p_0|n_2,x)$ is a linear combination of incomplete beta functions, available computer algorithms for solving incomplete betas can be utilized.

Figure 4.1 is an example of a particular posterior density function and its associated posterior cumulative distribution function.

Remark. Although it is not of primary interest in this thesis, the marginal posterior for p_1 can be found in a similar manner. That is

$$g_1(p_1|N_2=n_2,X=x) = \int_0^1 g(p_0,p_1|n_2,x)\ dp_0 =$$

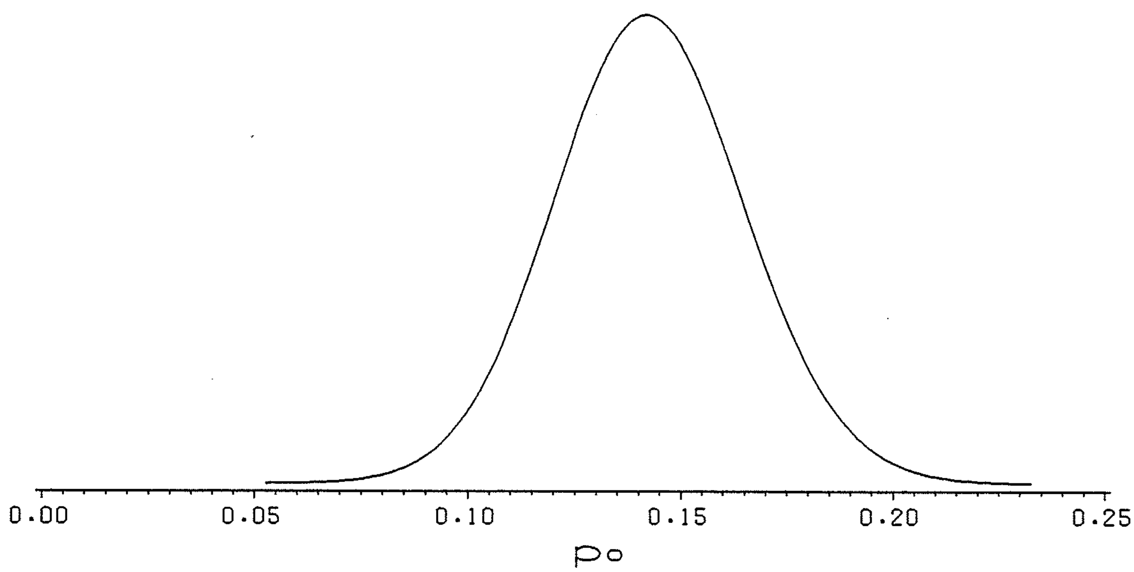
Figure 4.1: POSTERIOR DISTRIBUTION FOR p_0

Posterior Density Function

$g_0(p_0|n_2 = 45, x = 10)$, $n = 300$, $m = 10$

Prior for $p_0 = \beta(1,1)$, Prior for $p_1 = \beta(1,1)$

Posterior Mean = 0.14278, Posterior Standard Deviation = 0.02254

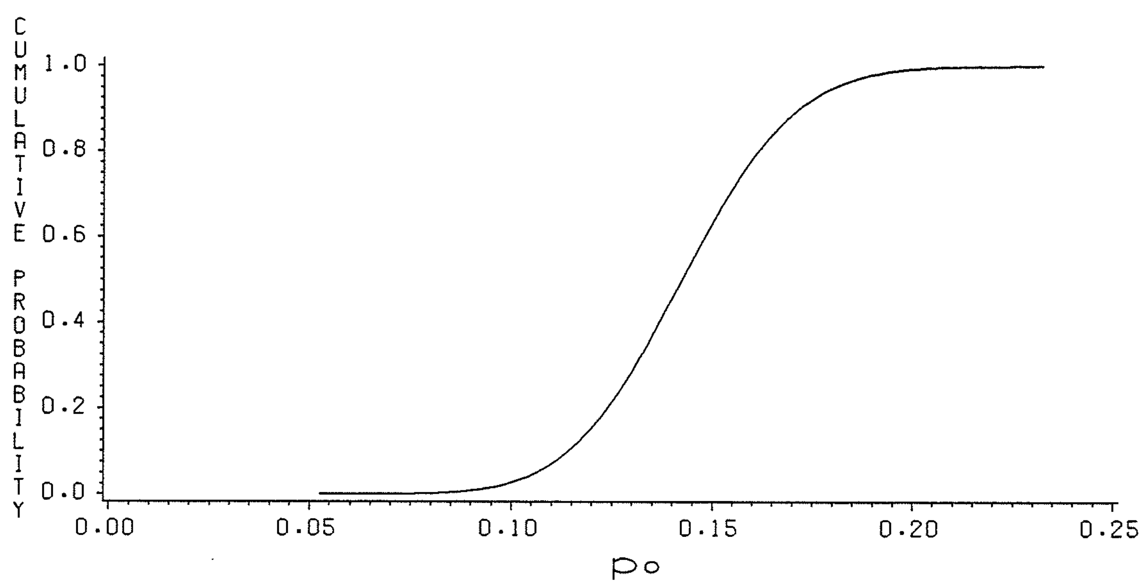


Posterior Cumulative Distribution Function

$G_0(p_0|n_2 = 45, x = 10)$, $n = 300$, $m = 10$

Prior for $p_0 = \beta(1,1)$, Prior for $p_1 = \beta(1,1)$

Lower Bounds: 90% = 0.114, 95% = 0.106, 99% = 0.091



$$\frac{\sum_{y_2}\left[\frac{(n-y_2+b_0-1)!(y_2+a_0-1)!}{(y_2-x)!(n_2-y_2-m+x)!}\, p_1^{\,n-n_2+a_1-1}(1-p_1)^{n_2-y_2+b_1-1}\right]}{(n-n_2+a_1-1)!\sum_{y_2}\left[\frac{(n_2-y_2+b_1-1)!(n-y_2+b_0-1)!(y_2+a_0-1)!}{(n-y_2+a_1+b_1-1)!(y_2-x)!(n_2-y_2-m+x)!}\right]},$$

$0 \le p_1 \le 1$. Here, $x = 0,1,\ldots,m$; $n_2 = m,m+1,\ldots,n$ and the summations are over $y_2 = x,x+1,\ldots,n-m+x$.

4.6 CREDIBILITY BOUNDS

A $(1-\alpha)\cdot 100\%$ lower credibility bound for p_0 is found by inverting the posterior cumulative distribution. That is, a value of $p_0 = LB$ is found such that

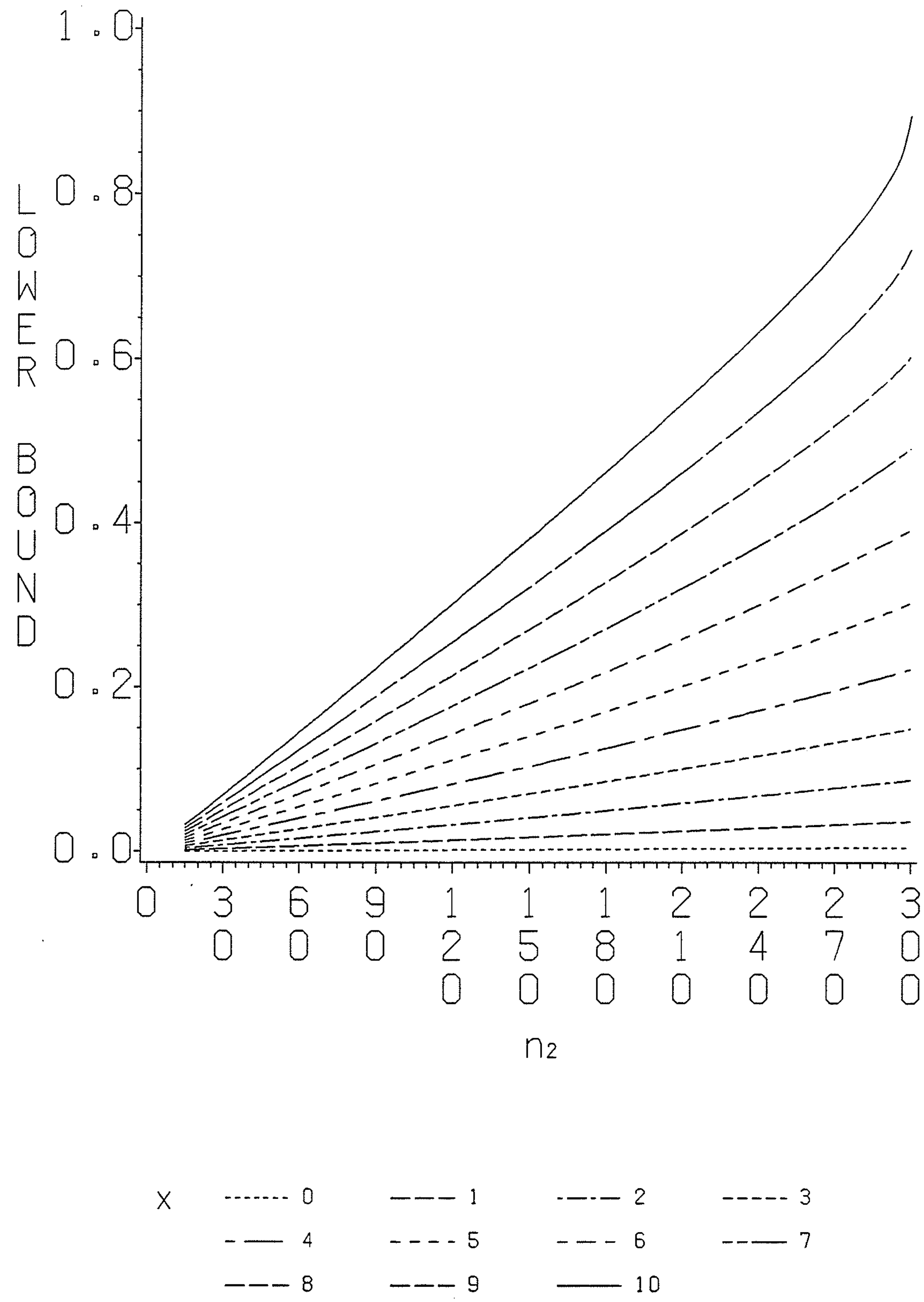
$$G_0(LB) = \alpha.$$

This value can be found using a search procedure such as that used in the computer program presented in Appendix A. The results of such a search are shown in Figure 4.2.

The Figure presents graphically the 95% lower credibility bounds for p_0 for $n_2 = 15,\ldots,300$ and $x = 0,\ldots,10$ when the priors on p_0 and p_1 are $\beta(1,1)$. For observed values of n_2 and x, a 95% lower bound for p_0 can be found from the diagram using graphical interpolation.

Alternatively equations could be fit to the lines in Figure 4.2 and these equations could then be used to estimate 95% lower credibility bounds for observed sets of n_2 and x. Over the smaller values of n_2 (those that are of most interest in this thesis) it can be seen that there is a nearly linear relationship between the values of the lower bound

Figure 4.2: 95% LOWER BOUNDS



and the size of n_2. Table 4.3 presents the coefficients for the linear relationship

$$LB = A + B \cdot n_2$$

for $x = 0,\ldots,10$ and $n_2 = 15,\ldots 165$, for the plots in Figure 4.2.

TABLE 4.3

x	A	B
0	-0.0000114	0.00001624
1	-0.0001260	0.00011642
2	-0.0003781	0.00027555
3	-0.0007704	0.00047218
4	-0.0012772	0.00069679
5	-0.0019412	0.00094557
6	-0.0027517	0.00121627
7	-0.0037809	0.00151112
8	-0.0050879	0.00183150
9	-0.0068985	0.00218621
10	-0.0099915	0.00259971

For posterior distributions that are not highly skewed the lower credibility bound can be estimated using a normal approximation. The posterior mean can be found from

$$E(p_0|n_2,x) = \int_0^1 p_0 \cdot g_0(p_0|n_2,x) \cdot dp_0$$

and the standard deviation, $Std(p_0)$,from the square root of

$$Var(p_0|n_2,x) = E(p_0{}^2|n_2,x) - (E(p_0|n_2,x))^2$$

where

$$E(p_0{}^2|n_2,x) = \int_0^1 p_0{}^2 \cdot g_0(p_0|n_2,x) \cdot dp_0$$

These integrals can be easily evaluated as linear combinations over y_2 of complete beta integrals, see Appendix A (Main Program).

With these values, a $(1 - \alpha)$ lower credibility bound can be approximated as

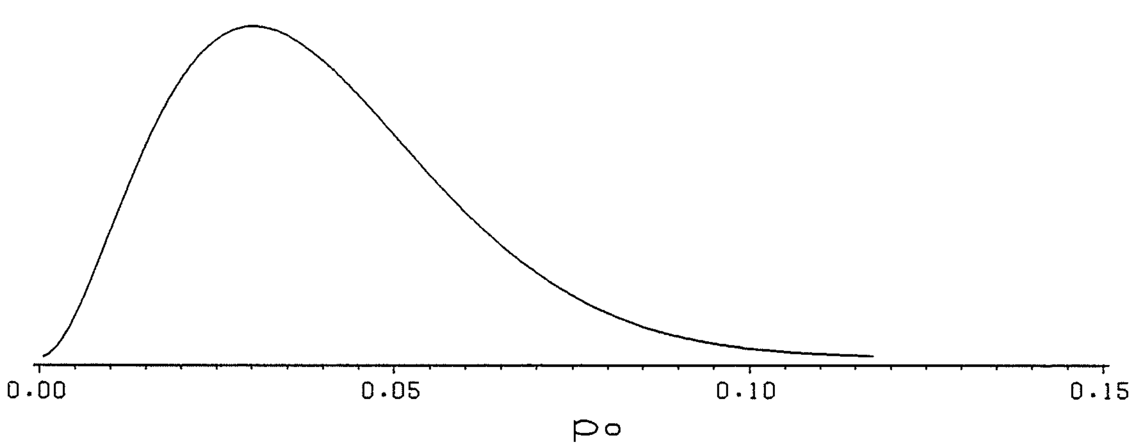
$$LB = E(p_0 | n_2, x) - c \cdot Std(p_0 | n_2, x)$$

where c is that value such that $P(Z<c) = \alpha$ for a standard normal distribution.

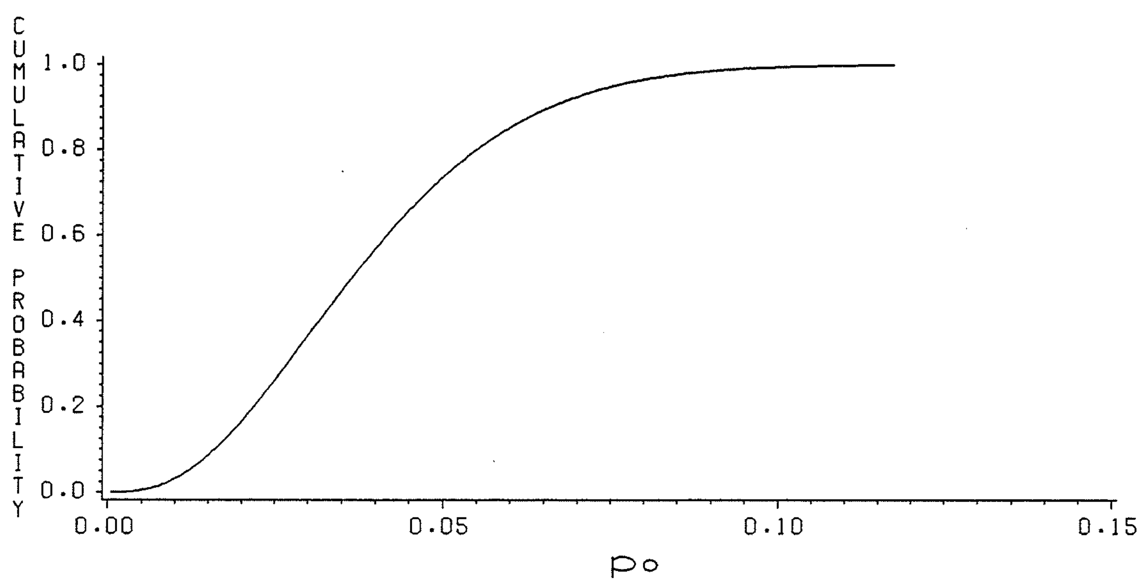
It should be noted that although the central region of the posterior distribution (the region around the mode) may be roughly normally distributed, the left hand tail of the distribution is typically "fatter" than the right hand tail. Hence an estimate of the lower bound calculated using a normal approximation may not be very accurate. For example, for the skewed posterior density function shown in Figure 4.3 a 95% lower credibility bound found using a normal approximation is 0.03918 - (1.645)(0.01965) = 0.007 whereas the true 95% lower credibility bound is 0.012.

Figure 4.3: HIGHLY SKEWED POSTERIOR DISTRIBUTION

Posterior Density Function
$g_0(p_0|n_2 = 45, x = 2)$, $n = 300$, $m = 10$
Prior for $p_0 = \beta(1,1)$, Prior for $p_1 = \beta(1,1)$
Posterior Mean = 0.03918, Posterior Standard Deviation = 0.019654



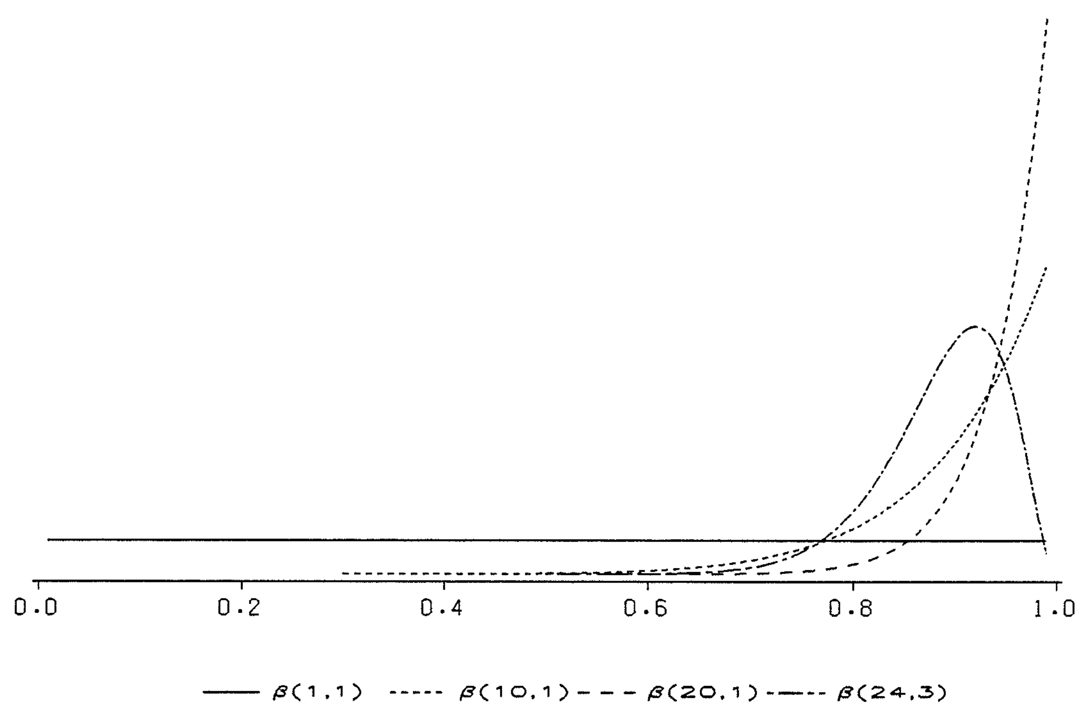
Posterior Cumulative Distribution Function
$G_0(p_0|n_2 = 45, x = 2)$, $n = 300$, $m = 10$
Prior for $p_0 = \beta(1,1)$, Prior for $p_1 = \beta(1,1)$
Lower Bounds: 90% = 0.016, 95% = 0.012, 99% = 0.006



4.7 SENSITIVITY CHECKS

Information about the inspection process will be acquired over time and therefore priors on p_1, the probability of correctly classifying a conforming kernel, can be used which will reflect this knowledge. Since it is likely that the inspection process will be fairly accurate in most instances, priors reflecting this assumption will be examined in this section for their effects on the posterior distribution. Four priors will be examined: $\beta(1,1)$ (representing ignorance), $\beta(10,1)$ and $\beta(20,1)$ (representing increasingly greater inspection accuracy) and $\beta(2.43,4)$ (representing inspection that is generally good, but often not perfect). Figure 4.4 shows plots of these beta distributions.

Figure 4.4: BETA DISTRIBUTIONS USED AS PRIORS FOR p_1



Posterior distributions for p_0 with these priors on p_1 are shown in Figures 4.5 and 4.6. It can be seen that the effect of the various priors is amplified by increasing values of n_2. There is also some increase in the effect of the prior as the number of kernels confirmed to be nonconforming in the subsample decreases from $x = 10$ to $x = 8$.

The effect of changes to the prior for p_0 is also examined. Four priors for p_0 are examined for their effect on the posterior distribution. The priors looked at are; $\beta(1,1)$ (representing ignorance about the probable proportion of nonconforming kernels in the carlot), $\beta(1,5)$ and $\beta(1,10)$ (representing decreasingly smaller proportions of the nonconforming variety expected to be in the carlot) and $\beta(1.5,5)$ (representing the belief that there are at least some nonconforming kernels in the carlot, but they are not expected to be a large proportion of them present). These priors are displayed in Figure 4.7.

The effect of these changes to the prior for p_0 is examined in conjunction with two priors for p_1, $\beta(1,1)$ (representing ignorance) and $\beta(10,1)$ (representing better than average inspector accuracy). The posterior distributions under these various priors are displayed in Figures 4.8 to 4.11. It can be seen that when there are only a small number of kernels classified as nonconforming that the choice of priors for p_0 and p_1 has negligible effect.

Figure 4.5: EFFECTS OF VARYING THE PRIOR FOR p_1 ON THE POSTERIOR FOR p_0

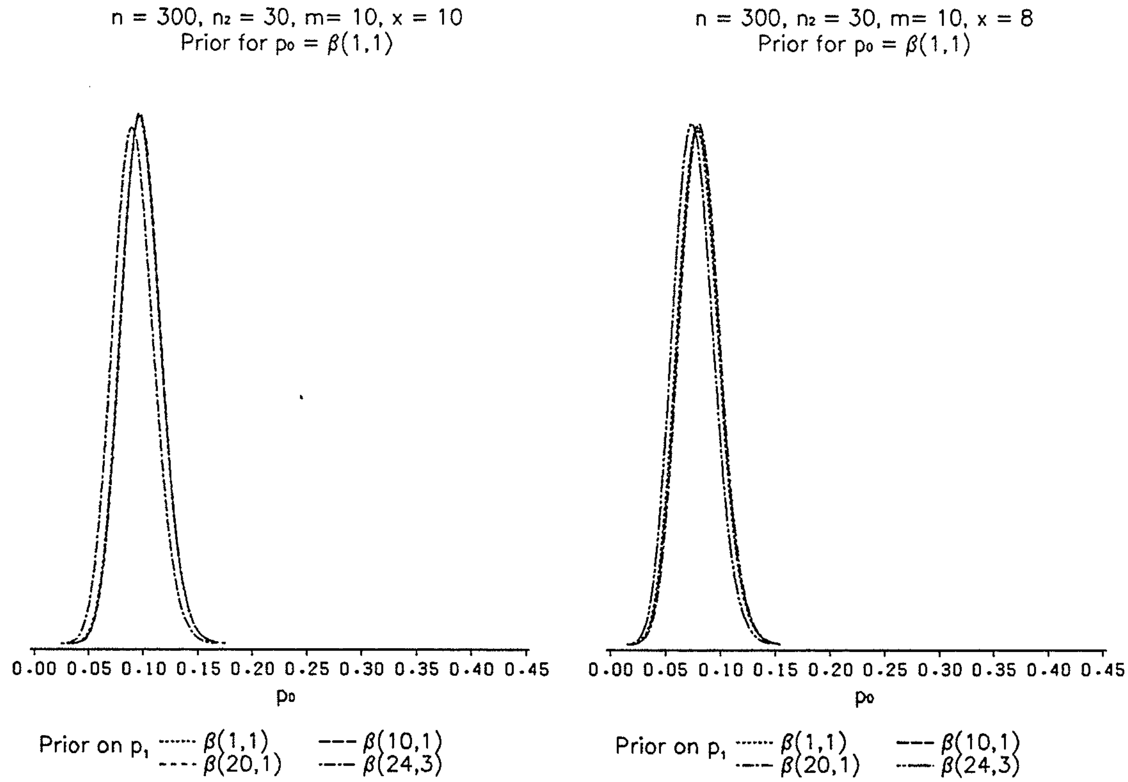


Figure 4.6: EFFECTS OF VARYING THE PRIOR FOR p_1 ON THE POSTERIOR FOR p_0

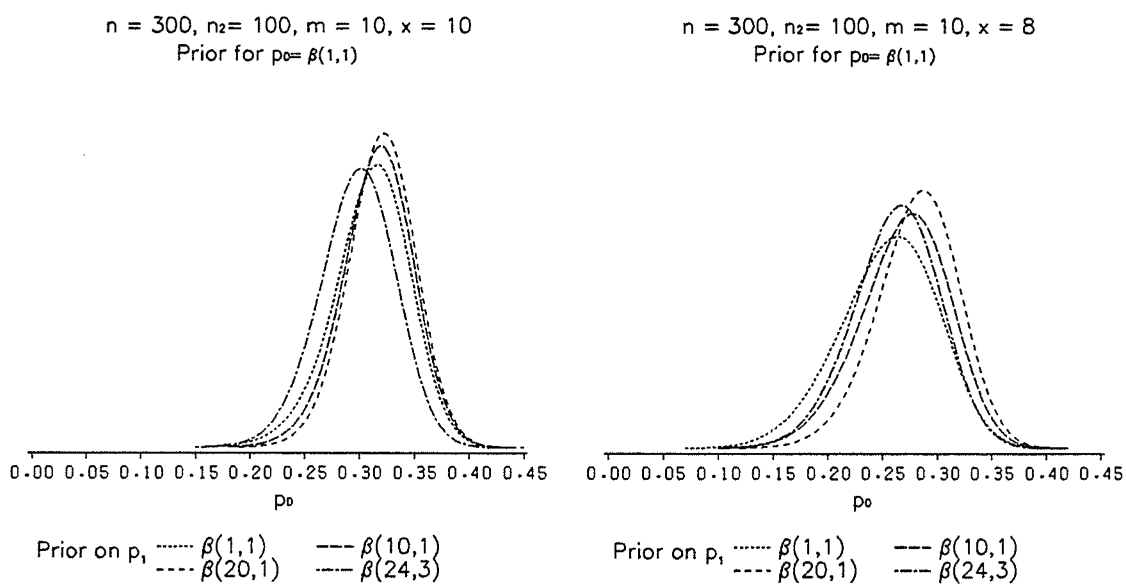


Figure 4.7: BETA DISTRIBUTIONS USED AS PRIORS FOR p_0

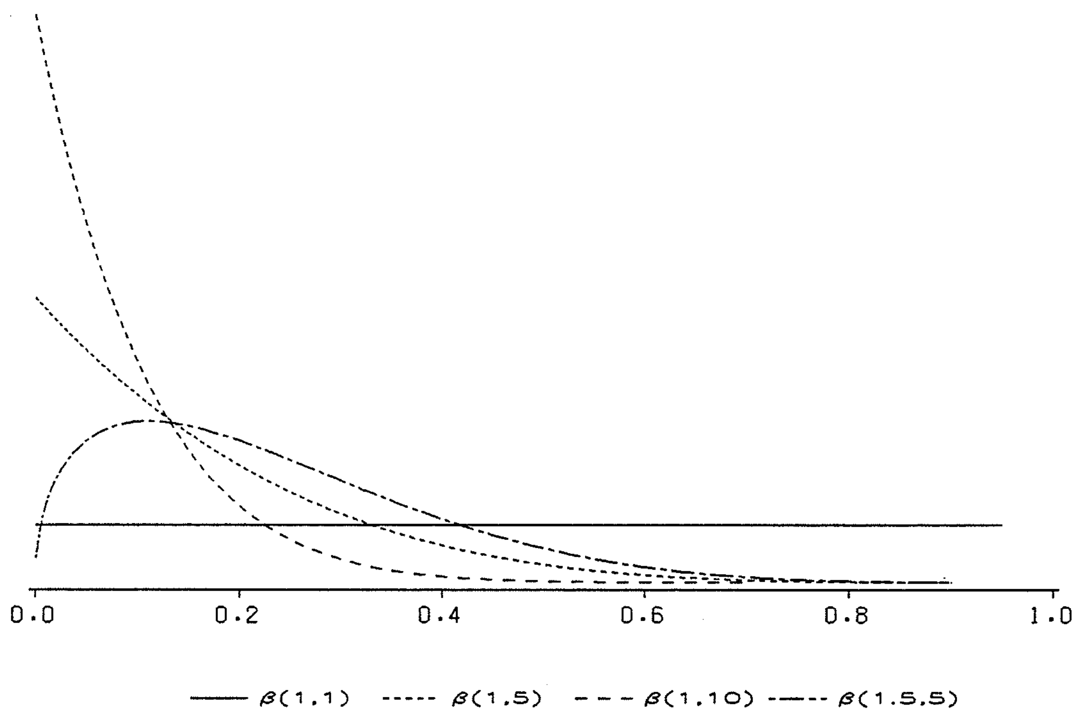


Figure 4.8: EFFECTS OF VARYING THE PRIOR FOR p_0 ON THE POSTERIOR FOR p_0

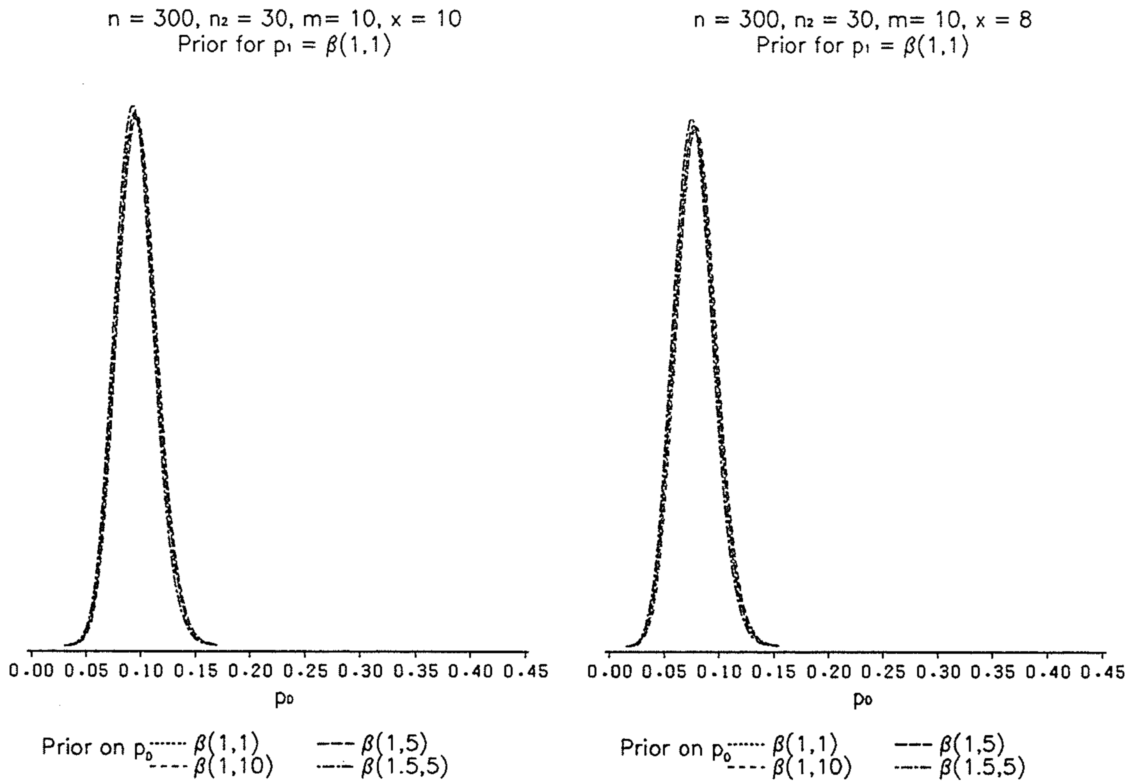


Figure 4.9: EFFECTS OF VARYING THE PRIOR FOR p_0 ON THE POSTERIOR FOR p_0

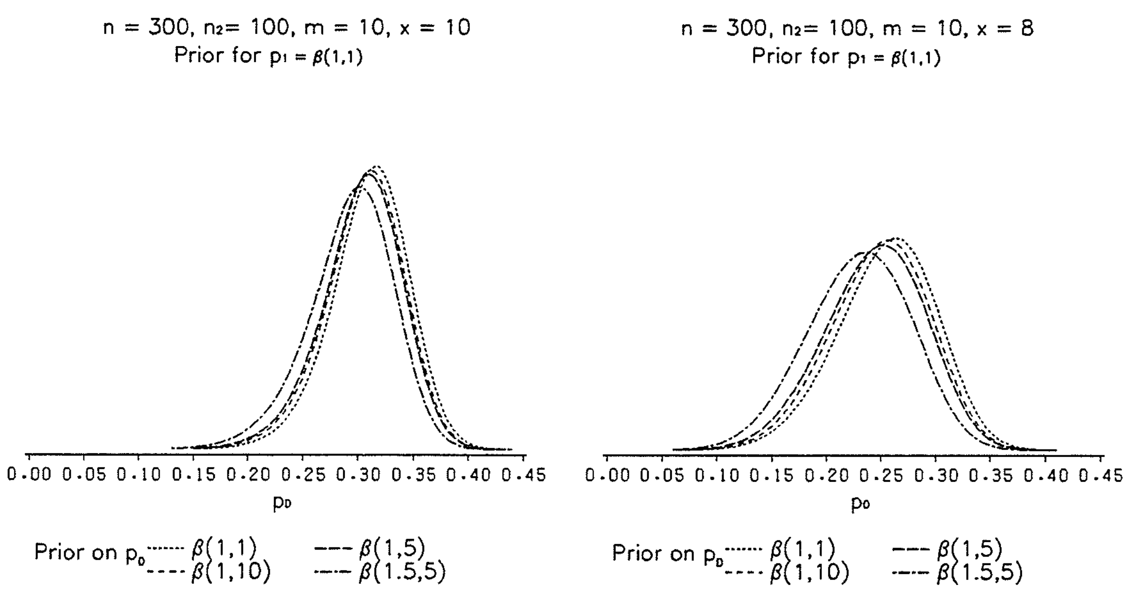


Figure 4.10: EFFECTS OF VARYING THE PRIOR FOR p_0 ON THE POSTERIOR FOR p_0

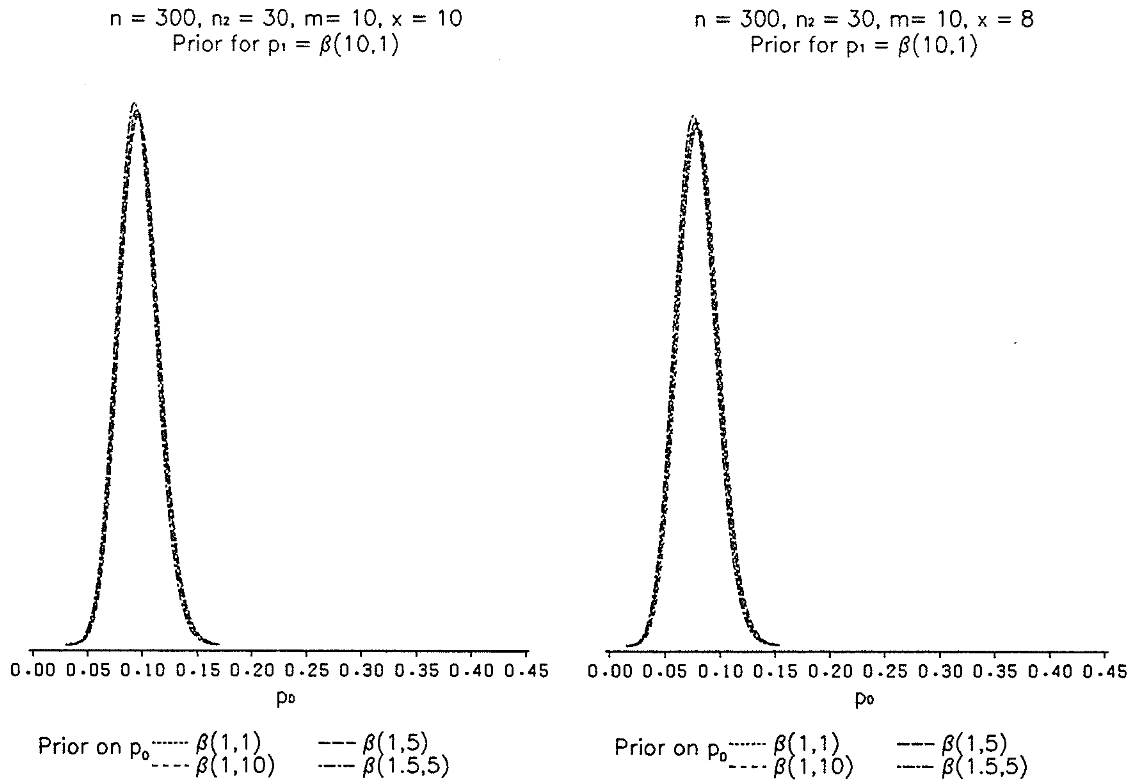
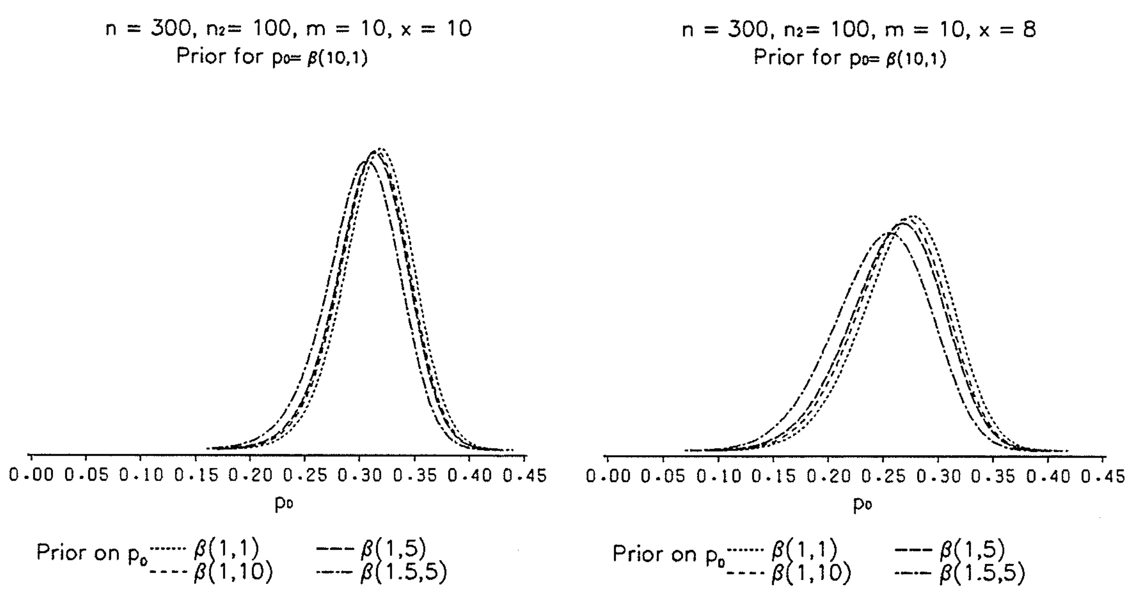


Figure 4.11: EFFECTS OF VARYING THE PRIOR FOR p_0 ON THE POSTERIOR FOR p_0



4.8 EFFECTS OF CHANGES TO SAMPLE OR SUBSAMPLE SIZE

Figures 4.12 and 4.13 illustrate the effect of changes to sample size or subsample size. Priors for p_0 and p_1 were maintained as independent ignorance priors, [$\beta(1,1)$]. It can be seen that doubling the sample size from n = 300 to n = 600 kernels when the subsample size is held constant at m = 10 or m = 20 has minimal effect on the lower tail region of the posterior distribution (the region affecting lower credibility bounds). However doubling the subsample size from m = 10 to m = 20 (when the sample size is held constant at n = 300 or n = 600) does affect the lower tail region. This effect is magnified by increases in n_2. The effects observed above are fairly similar for cases when x/m = 1.0 and when x/m = 0.8.

It should be noted that, based on the above observations, the greatest improvement to the power of a test based on a lower credibility bound would come from increasing the size of the subsample.

Figure 4.12: EFFECT OF CHANGES TO SAMPLE OR SUBSAMPLE SIZES ON THE POSTERIOR FOR p_0

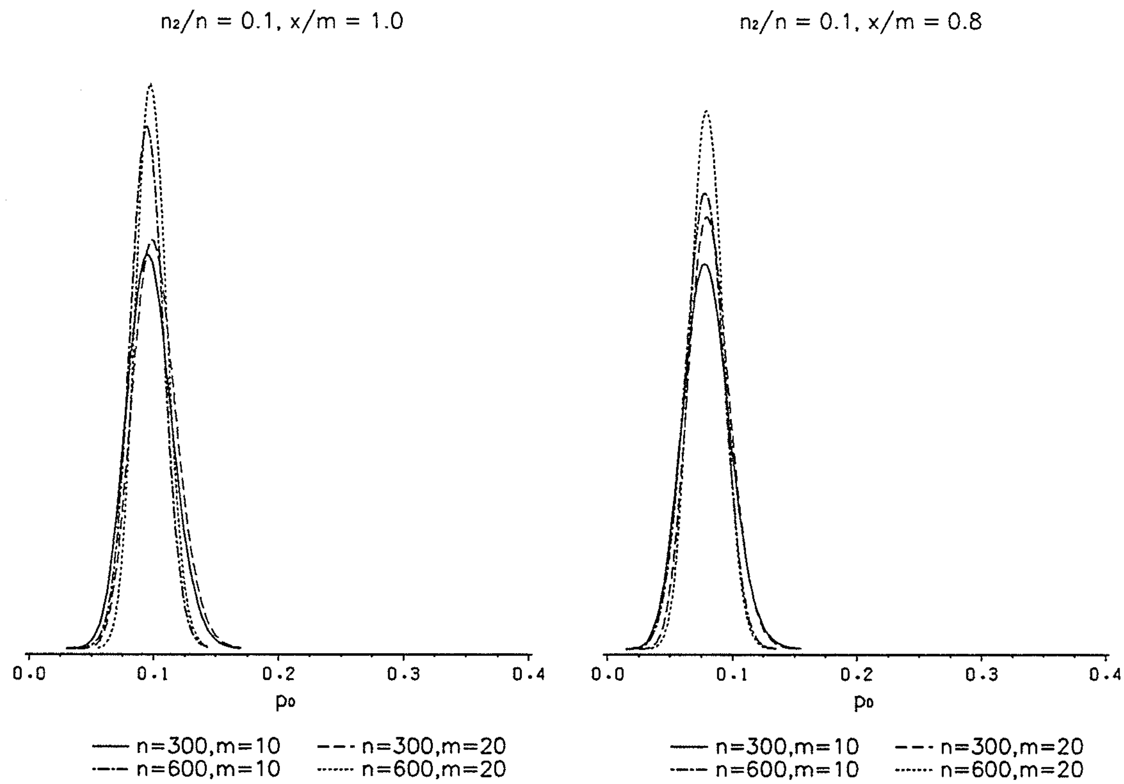
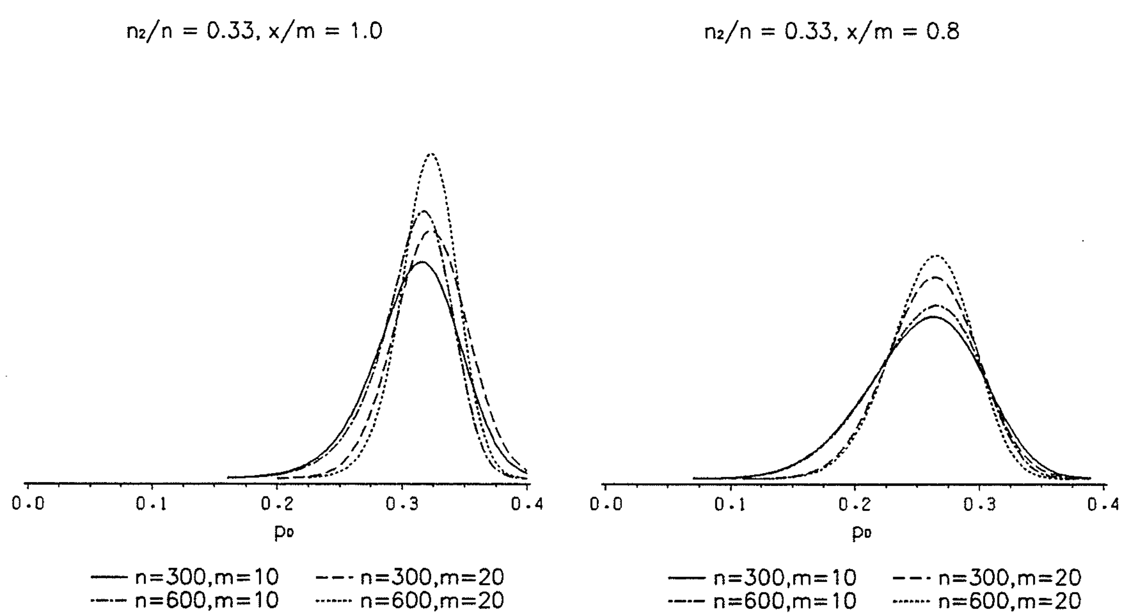


Figure 4.13: EFFECT OF CHANGES TO SAMPLE OR SUBSAMPLE SIZES ON THE POSTERIOR FOR p_0



Chapter V

LITERATURE SURVEY

Literature examined as background to this thesis was principally in two broad categories: the role of misclassification in reaching statistical conclusions, and Bayesian methods in sampling.

5.1 MISCLASSIFICATION

5.1.1 MISCLASSIFICATION IN QUALITY CONTROL

The effect of misclassification or 'inspector error' is of importance in the field of quality control. A survey paper by Dorris and Foote (1978) gives a good outline of proposed models for inspection error for attribute data, count data and variables data. It summarizes work done on the effects of inspector error in statistical quality control procedures, measuring inspector accuracy and modelling the causes of inspector error. The paper concludes by pointing out areas where further research should be considered. The authors note, that even if there is a reasonable model for misclassification "it is not possible to assess the extent of the difficulty or design a compensating plan without knowledge of the error probabilities" and "in practice, of course, these values (for error probabilities) are rarely if ever known and probably change over time" (page 190).

This need for knowing the error probabilities is exemplified by many of the articles examined. There are a large number of articles (Beainy and Case, 1981; Carr, 1982; Case, Bennett and Schmidt, 1973,1975; Collins, Case and Bennett, 1973; Hoag, Foote and Mount-Campbell, 1975; Minton, 1972; Wooding, 1979) in which particular quality control procedures are examined and adjustments for known error probabilities are proposed. In the reverse direction Healy (1981) develops a model for classification subject to error so that the classification mechanism can be designed to have a known, acceptable error probability.

5.1.2 TWO-STAGE SAMPLING PLANS

In order to incorporate estimates of the error probabilities into the estimate of a population proportion Tenenbien (1970) proposed a two-stage sampling plan. In this plan an initial sample is drawn. From this a smaller subsample is removed. The units in the subsample are examined by both the fallible and infallible classifiers, whereas the remaining units in the initial sample are only classified by the fallible classifier. This differs from the problem being considered in this thesis wherein information about only one misclassification probability is acquired. Tenenbien obtains maximum likelihood estimates of the population proportion and both error probabilities. The asymptotic variance of the estimate of population proportion is also derived. In the thesis problem the subsample size is very small, and the population proportion of interest may be very small; therefore the asymptotic results from Tenenbien's paper cannot be applied. In Tenenbien (1971) procedures for determining optimal sample or subsample sizes are presented and cost comparisons are made to sampling using only the infallible classifier. Tenenbien (1972) extends the results from binomial data to multinomial data.

A similar two-stage sampling plan, applied to a health care survey, is given in Deming (1977). Information per unit cost is examined for two methods of allocating the subsample size (proportional and Neyman) and using the infallible classifier only. Deming concludes that the two-stage plan is only cost effective if there is a large difference per item between the cost of classifying using the fallible method and the cost using the infallible method. He suggests the break-even point is roughly 1:6. In the problem under consideration in this thesis the cost difference is considerably larger. He also notes (page 36)

> sampling to measure the prevalence of a rare characteristic is a subject all by itself, beyond the scope of this paper,... Statistical procedures to determine with a prescribed probability the prevalence of a certain rare disease does not exceed some small proportion such as $p \leq 1/50$ call forth still further theory, also not covered here.

A further example of use of a two-stage procedure is given in Fleiss (1981), Chapter 12.

Hochberg (1977) uses Tenenbien's work to develop a generalized approach for multidimensional cross-classified data, where there is no structure to the error probabilities. He proposes a combined maximum likelihood, least squares methodology, and a least squares methodology, but finds that unless some restrictions can be placed on the structure of the error probabilities that there are "too many degrees of freedom for the misclassification error space" (page 920).

5.1.3 OTHER APPROACHES

Other approaches to incorporating misclassification into the model include using log-linear models (Chen, 1979), classifying items by two characteristics (Diamond and Lilienfeld, 1962; Chiacchurini and Arnold, 1977), doing repeated classifications of the items (Sutcliffe, 1965a,1965b; Koch, 1969) and using a game theoretic approach (Rahali and Foote, 1982).

5.1.4 MODELS FOR MISCLASSIFICATION RATES

Much of the work on developing models for misclassification rates has been done in the fields of psychology and signal detection theory. Sinclair (1978) provides a survey of some of the more mathematically oriented of these models. Tiemstra (1981) looks at measures of inspector effectiveness. In the wheat classification problem under consideration in this thesis not only are many of the variables mentioned by Sinclair (such as working conditions or number of 'defects' already found) affecting inspector performance, but also the items being examined, the wheat kernels, vary in how identifiable they are.

5.1.5 SAMPLING DISTRIBUTION OF THE NUMBER CLASSIFIED NONCONFORMING

Collins and Case (1976) show that when sampling with replacement from an "infinite" population, where misclassification is present, the distribution of the number classified nonconforming is binomial. In a group of related papers, Johnson, Kotz and Rodrigues (1985), Johnson, Kotz and Sorkin (1980) and Kotz and Johnson (1982) the distribution of the number classified nonconforming when sampling without replacement from a finite population is shown to be compound binomial, confounded with the hypergeometric distribution of the true number of nonconforming items in the sample.

5.2 BAYESIAN METHODS

Box and Tiao (1973) and Raiffa and Schlaifer (1961) were used as general references on Bayesian methods.

5.2.1 BAYESIAN SAMPLING PROCEDURES

Godfrey and Neter (1984) develop a sampling procedure to determine an upper credibility bound for a proportion in a problem in accounting. They do not consider misclassification. In Coombs and Stephens (1980) an upper credibility bound for a proportion is also determined, in this case it is assumed that the maximum value of this upper bound is known to be something less than one. Hald (1968) and Guild and Raka (1980) provide examples of Bayesian sampling plans in quality control.

5.2.2 PRIORS FOR THE POPULATION PROPORTION

Case and Keats (1982) examine distributions which "describe the number of defectives from lot to lot" (page 10). They recommend the use of a mixed binomial distribution as a prior to describe situations where items in a lot come from a variety of sources. They also look at the Polya distribution (a discrete form of the beta), the uniform distribution (which they feel is unrealistic but suitable as an "ignorance prior") and the binomial distribution (which they show "renders sampling useless and inappropriate" (page 10)).

Because of its ability to approximate a wide variety of distributions and because of the simplification of the mathematics which it provides, the beta distribution is frequently chosen as a prior to model the distribution of the population proportion from lot to lot. Hald (1968) and Lauer (1978) provide examples of single sample attribute sampling plans using a beta prior on the population proportion. Trader (1983) shows that, in sampling from a bernoulli process where misclassification is present, the conjugate family of distributions for the population proportion is an infinite mixture of beta distributions. Use of conjugate distributions as priors generally leads to mathematical tractability, however in this case fitting an infinite mixture is not feasible. Trader therefore looks at a "stretched" truncated beta distribution as a prior. She finds this difficult to handle computationally, and concludes that a beta distribution, (the conjugate distribution when inspection is perfect) is the most reasonable prior to use. The Bayesian analysis here has only one unknown parameter as the probabilities of misclassification are assumed to be known.

5.2.3 PRIORS FOR THE NUISANCE PARAMETERS

Menzefriche (1984) uses a product of betas to model the joint distribution of the misclassification probabilities. He finds that a normal approximation to this joint distribution is adequate, where there is a moderate to large population size, a moderate sample size, and either a moderate population proportion or very small misclassification probabilities.

5.2.4 ROBUSTNESS

Box and Tiao (1962) provide arguments for the robustness of inferences based on the use of prior distributions and Bayesian methods. The papers by Godfrey and Neter (1984) and Pfanzagl (1963) are good examples of thorough examination of the robustness of inferences to the choice of prior.

5.2.5 ESTIMATING PARAMETERS FOR PRIORS

Skellam (1948) and Weiler (1965) provide methods for estimating the parameters for a beta prior from past records of estimates of the population proportion. Their examples do not include the effect of misclassification.

5.2.6 BAYESIAN ACCEPTANCE SAMPLING PLANS UNDER MISCLASSIFICATION

Two papers were found in which the effects of misclassification on Bayesian acceptance sampling plans are considered. Moskowitz and Fink (1977) develop a recursive algorithm to find the optimal single sample acceptance plan, with a discrete prior on the population proportion, when misclassification is present. Kittler and Pau (1980) also develop a Bayesian quality control scheme. Their scheme requires the use of a test set to determine the misclassification probabilities, but they are unable to determine good guidelines for deciding how large this test set should be in order to have acceptablely accurate estimates of the misclassification probabilities. There appears to have been little work done in the area of obtaining information about misclassification from the data in Bayesian acceptance sampling plans.

Chapter VI

CONCLUDING REMARKS

6.1 POSSIBLE IMPROVEMENTS TO THE SAMPLING PLAN

A $(1-\alpha)\cdot 100\%$ lower credibility bound for p_0 based on the posterior distribution developed in Chapter IV has been proposed in this thesis as a method of testing the hypotheses:

$$H_0: p_0 \le p' \text{ vs. } H_1: p_0 > p'.$$

There are some changes to the sampling plan that could be looked at that might lead to making a correct decision more frequently and provide a more reliable estimate of p_0.

For practical reasons the size of the initial sample and the subsample in the thesis problem were fixed. As discussed in Section 4.8 increasing the size of the initial sample and/or the subsample might be considered.

Alternatively the size of the subsample might be allowed to vary with the size of the groups classified as nonconforming, and an optimal subsample size might be determined (taking costs into account) as in Deming (1977).

A multistage sampling procedure could also be used at the subsample stage, with second and subsequent samples being taken depending upon the number of kernels confirmed to be nonconforming by the laboratory test, and the size of the group classified as nonconforming.

The group classified as conforming could also be subsampled and this information could be used to develop an upper credibility bound. Having an upper credibility bound would allow the implementation of a quality assurance plan for meeting grading standards when the carlots of wheat are sold. Sampling from this group should also be done in order to acquire information about the probability of misclassifying a nonconforming kernel.

6.2 AREAS FOR FURTHER RESEARCH

Even though the group classified as conforming is not sampled, the posterior distribution developed in this thesis could be modified so that the assumption that $p_2 = 1$ could be changed to $p_2 = c$, where c is some fixed value.

For simplicity, independent beta priors (for the population proportion and the probability of misclassifying a conforming kernel) were used in the development of the posterior distribution in Section 4.5. As mentioned in Chapter V it might be reasonable to consider independent mixtures of beta distributions as priors. This would not greatly increase the computational difficulties and would increase the number of other distributions which could be approximated. Other families of distributions might be considered as priors but their choice may be limited by their tractability.

From the material in Chapter IV the sample covariance between p_0 and p_1 could be examined. This might provide justification for the use of independent priors or permit the development of a more suitable joint prior. A bivariate beta distribution might be considered.

From the posterior distribution developed in Chapter IV and the capability of finding lower credibility bounds provided by the computer program in Appendix A, acceptance sampling procedures could be developed for testing the hypothesis $p_0 \leq p'$. Based on the work in this thesis the rule is to accept the carlot as high grade wheat if the lower credibility bound is $\leq p'$. An alternative way of displaying this answer might be to develop a table of all values of n_2 and x. For each n_2 there would be some test value, q(n2), $0 \leq q(n2) \leq m$. If it was observed that $x \leq q(n2)$, the carlot would be graded as low grade wheat. It is possible that there will be large values of n_2 for which all values of x would lead to the carlot being graded as low grade wheat and therefore the subsample would not need to be taken.

Once such an acceptance rule was established its operating characteristic curve could be examined, and comparisons made to existing or standard acceptance rules.

Methods still need to be developed to extract information about inspection performance from a series of samples in order to estimate parameters to the prior for p_1, the probability of correctly classifying conforming kernels. It would be helpful if this was a dynamic method as there is likely to be a "training" effect on the the inspectors performance over one crop year, as well as possibly a seasonal effect due to deterioration of kernels during storage or an annual effect due to changes to growing conditions in different crop years.

6.3 CONCLUSION

The posterior distribution developed for the sampling problem presented in this thesis illustrates how information available from the sampling plan and any prior information can be combined to provide a distribution upon which conclusions about the proportion on nonconforming kernels in the carlot can be based.

Strengths and weaknesses of the sampling plan have been examined and improvements suggested. Incorporation of prior information has been illustrated and alternative specification of prior information considered. It is hoped that the information in this thesis provides a base for further research and development of applied procedures.

Appendix A

The following program will, for one user supplied set of sample values, produce plots of the posterior density function and the posterior cumulative distribution function developed in Section 4.5. It also prints out the posterior mean and posterior standard deviation. It searches for and prints out the lower bound for three user selected credibility coefficients.

The user is to supply the following values.

- From the sample, values of n, m, n_2 and x. $n \geq n_2$, $n \geq m$, $n_2 \geq x$.
- Parameters for the priors for p_0 and p_1, a_0, b_0, a_1, and b_1. $a_0, b_0, a_1, b_1 > 0$.
- The maximum number of times that $G_0(p_0)$ can be calculated in any one search for a lower bound, MAXFN. MAXFN > 0.
- How close the value of $G_0(p)$ is to the desired $(1 - \alpha)$ value before the search stops, ACC. ACC < 1.
- Number of significant digits of accuracy desired in the estimate of the lower bound, NSIG. NSIG $\geq$ 1.
- For three desired $(1 - \alpha)$ lower credibility bounds provide the corresponding α levels, ALPHA1, ALPHA2 and ALPHA3. 0 < ALPHA1,ALPHA2,ALPHA3 < 1

The values are entered in the program after the CARDS statement, in the order given, separated by blanks.

The program is written in SAS Version 5.08 (Statistical Analysis System). This programming package was chosen because of the availability of built-in functions for calculating beta probabilities and inverse beta values, its use of 16 bit precision and the integration of a programming language and graphical procedures.

No formal test of the accuracy of results from the built-in functions was done, but in all cases where results could be cross checked by other methods the results were acceptable. Because of the limitations on the range of values that various built-in functions will accept, this program checks for out of range values, prints an error message and then tries to continue. Therefore there may be some input values for which valid results are not available.

This program has been used over a wide variety of input values, but has not been tested for extreme values. It is the responsibility of the user to provide valid input data, as the program does not check input values.

The program consists of four modules. The first calculates one value of the posterior density function, given a value of p. The next calculates one value of the posterior cumulative distribution function, given a value of p. The search module uses Brent's algorithm (Brent, 1977) to search for the zero of a function. In this case it is searching for that value of p such that $G_0(p) - \alpha = 0$. It calls the module which calculates the posterior cumulative distribution.

The main program performs several functions. It calculates the posterior mean, posterior standard deviation and that portion in the denominator of both the posterior density function and posterior cumulative distribution function that is the summation B of Section 4.5. From the posterior mean and standard deviation a range of values is determined for which the posterior density function is sufficiently larger than zero to show in a plot. Values for plotting the posterior probability function and the posterior cumulative distribution function are calculated by calling the appropriate modules. The search module is called to determine the lower bounds corresponding to the input α values. Plotting procedures are called to display the results. The GOPTIONS may be specific to the University of Manitoba system.

The program is presented next, and sample output for input values of:

```
N = 300, M = 10, N2 = 45, X = 8,
A0 = 1, B0 = 1, A1 = 1, B1 = 1,
MAXFN = 8, ACC = 0.001, NSIG = 3,
ALPHA1 = 0.10, ALPHA2 = 0.05, ALPHA3 = 0.01
```

follows in Figure A.1.

```
OPTIONS DQUOTE NOCAPS;
***********************************************************;
* MACRO GP CALCULATES ONE VALUE OF THE PDF;
***********************************************************;
%MACRO
  GP( P,    /* VALUE OF P AT WHICH TO CALCULATE THE PDF    */
    PDF)    /* CALCULATED VALUE OF PDF TO BE RETURNED      */;
%LOCAL
  NUM       /* NUMERATOR ACCUMULATOR                       */
  Y2        /* SUMMATION INDEX                             */
  TEMP      /* LOG OF ONE VALUE IN SUMMATION               */;
*---------------------------------------------------------*;
*ZERO ACCUMULATOR;
  NUM = 0;
* CALCULATE NUMERATOR;
  DO Y2 = X TO MAXY2;
    TEMP = ((N - Y2 + B0 - 1)*LOG(1 - &P)
            + (A0 + Y2 - 1)*LOG(&P)
            + LGAMMA(N2 + B1 - Y2) - LGAMMA(Y2 - X + 1)
            - LGAMMA(N2 - Y2 - M + X + 1)
            - LGAMMA(N - Y2 + A1 + B1) + CONST);
    NUM = NUM + EXP(TEMP);
  END;
* CALCULATE VALUE OF PDF;
  &PDF = NUM/DENOM;
*---------------------------------------------------------*;
%MEND;
```

```
**********************************************************;
* MACRO FP CALCULATES ONE VALUE OF THE CDF;
**********************************************************;
%MACRO
  FP( P,    /* VALUE OF P AT WHICH TO CALCULATE THE CDF   */
  CDF)      /* CALCULATED VALUE OF CDF TO BE RETURNED     */;
%LOCAL
  NUM       /* NUMERATOR ACCUMULATOR                      */
  Y2        /* SUMMATION INDEX                            */
  PB        /* INCOMPLETE BETA RESULT                     */
  PREVINCR /* LOG OF PREVIOUS VALUE IN SUMMATION          */
  INCR      /* LOG OF CURRENT VALUE IN SUMMATION          */
  T2        /* TOLERANCE                                  */
  T3        /* MINIMUM POSSIBLE VALUE FOR BETA FUNCTION   */;
*--------------------------------------------------------*;
* SUFFIX FOR LABELS IN CURRENT CALL TO THIS MACRO;
  %LET ID = &SYSINDEX;
* CALCULATE TOLERANCES;
  T2 = T/100000000; T3 = 10 ** (-60);
* FOR SMALL VALUES OF P RETURN CDF=0,
  FOR LARGE VALUES OF P RETURN CDF=1;
  IF (&P < T2) THEN &CDF = 0;
  ELSE IF (&P > 1 - T2) THEN &CDF = 1;
* ELSE CALCULATE VALUE OF FUNCTION AT P;
  ELSE DO;
* ZERO ACCUMULATOR;
    NUM = 0; INCR = -180;
* CALCULATE NUMERATOR;
```

```
    DO Y2 = X TO MAXY2;
* IF INCOMPLETE BETA VALUES ARE CLOSE TO ZERO THIS
  TERM IN THE SUMMATION CONTRIBUTES LITTLE TO THE
  SUM. PASS ON TO THE NEXT TERM;
      PB = PROBBETA(&P, Y2+A0, N-Y2+B0);
      IF (PB < T3) THEN GOTO NEXT&ID;
* ELSE CALCULATE INCREMENT TO SUM;
      ELSE DO;
        PREVINCR = INCR;
        INCR = LGAMMA(N2 - Y2 + B1) + LGAMMA(Y2 + A0)
               + LGAMMA(N - Y2 + B0) + LOG(PB)
               - LGAMMA(Y2 - X + 1)
               - LGAMMA(N2 - Y2 - M + X + 1)
               - LGAMMA(N - Y2 + A1 + B1);
* CATCH  VALUES OF INCR OUT OF RANGE FOR EXP FUNCTION;
        IF (INCR > 174 OR INCR < -180) THEN GOTO NEXT&ID;

* ADD INCREMENT TO SUM;

        NUM = NUM + EXP(INCR);

    END;
    NEXT&ID: END;
  LAST&ID: &CDF = NUM/DENOM;
  END;
*---------------------------------------------------------*;
%MEND;
```

```
************************************************************;
* MACRO SEARCH USES BRENT'S ALGORITHM TO FIND THE VALUE
  OF P FOR WHICH THE CDF IS EQUAL TO ALPHA.;
************************************************************;
%MACRO
  SEARCH(
   LB,      /* LOWER BOUND VALUE, RETURNED                   */
   ALPHAI) /*   OF THE (1 - ALPHA) LOWER BOUND               */;
%LOCAL
  COUNT     /* COUNTER TO LIMIT THE NUMBER OF TIMES THE     */
            /* FUNCTION CAN BE CALLED                       */
  A B C     /* VALUES OF P THE SEARCH IS CHECKING           */
  FA FB FC /* VALUE OF FUNCTION AT A,B OR C                 */
  D E       /* TEMPORARY HOLDERS OF A B OR C IN EXCHANGE    */
  MID       /* MID WAY BETWEEN B AND C                      */
  P Q R S   /* CALCULATED VALUES USE IN INTERPOLATION       */
  TEMP      /* RATIO OF FUNCTION VALUES                     */
  TOL       /* CURRENT DESIRED TOLERANCE                    */
  Z         /* STANDARD NORMAL PERCENTILE VALUE             */
            /* GREATER THAN (1 - ALPHA)                     */;
*----------------------------------------------------------*;
* SUFFIX FOR LABELS IN CURRENT CALL TO THIS MACRO;
  %LET IDX = &SYSINDEX;
* CALCULATE AN APPROXIMATE LOWER BOUND USING NORMAL
  DESIRED VALUE APPROXIMATION AS ONE END POINT FOR
  THE SEARCH. IT MUST LIE BETWEEN THE INITIAL ENDPOINTS;
      ALPHA = &ALPHAI;
      COUNT = 0;
```

```
      Z = - PROBIT(ALPHA);
      B = MAX((EP0 - Z*SD0),ACC);
      %FP(B,FB);
      FB = FB - ALPHA;
* IF FUNCTION DOES NOT EXIST AT THIS POINT HALT PROCEDURE;
      IF (FB=.) THEN DO;
        PUT 'VALUE NOT FOUND';
        B = 0;
        GO TO LOOP&IDX;
      END;
*CHOOSE APPROPRIATE END OF RANGE AS OTHER END POINT;
      IF (FB<0) THEN DO;
        A=1; FA = 1 - ALPHA;
      END;
      ELSE DO;
        A=0; FA = - ALPHA;
      END;
* USE ALGORITHM TO FIND THE ZERO OF A FUNCTION;
      TOL = T*MAX(ABS(B),.1);
      C=A; FC = FB;
      IF (FA*FB > 0) THEN
        PUT 'FA AND FB HAVE THE SAME SIGN ';
      ELSE DO;
* DETERMINE WHICH TWO POINTS OUT OF A,B AND C ARE CLOSEST
  TO THE DESIRED POINT;
        DO WHILE ((ABS(FB) > ACC) AND (ABS(C-B) > TOL)
                   AND COUNT < MAXFN);
          IF (FB*FC>0) THEN DO;
```

```
  C=A; FC=FA;
  D=B-C; E=D;
END;
IF (ABS(FC) < ABS(FB)) THEN DO;
  A=B; B=C; C=A;
  FA=FB; FB=FC; FC=FA;
END;
MID =(C-B)/2;
TOL = T*MAX(ABS(B),0.1);
IF ((ABS(E) >= TOL) AND (ABS(FA) > ABS(FB)))
  THEN DO;
  S=FB/FA;
*INVERSE QUADRATIC INTERPOLATION;
  IF (A NE C) THEN DO;
     TEMP = FA/FC;
     R = FB/FC;
     P = S*((C-B)*TEMP*(TEMP-R) - (B-A)*(R-1));
     Q = (TEMP -1)*(R-1)*(S-1);
  END;
*LINEAR INTERPOLATION;
  ELSE DO;
     P=(C-B)*S;
     Q=(1-S);
  END;
  IF (P>0) THEN Q = -Q;
  ELSE P = -P;
  IF (2*P >= 3*MID*Q OR P>= ABS(E*Q*.5)) THEN
     GOTO JUMP&IDX;
```

```
        E=D; D=P/Q;
      END;
     *BISECTION;
      ELSE DO;
JUMP&IDX:   E=MID;
        D=E;
      END;
      A=B; FA=FB;
      IF (ABS(D) <= TOL/2) THEN
         TEMP = ABS(TOL/2)*SIGN(MID);
      ELSE TEMP =D;
      B = B + TEMP;
      %FP(B,FB);
      FB = FB - ALPHA;
      IF (FB=.) THEN DO;
        PUT 'VALUE NOT FOUND IN SEARCH';
        GO TO LOOP&IDX;
      END;
      COUNT = COUNT +1;
    END;
    IF (COUNT >= MAXFN) THEN
      PUT 'STOPPED DUE TO MAXIMUM ITERATIONS ';
    IF ABS(C-B) < TOL THEN
      PUT 'C AND B CLOSE ';
  END;
 LOOP&IDX: &LB = B;
*---------------------------------------------------------*;
%MEND;
```

```
***********************************************************;
* MAIN PROGRAM;
***********************************************************;
DATA PDFDATA(KEEP=P0 GP0)
               /* DATA FOR PLOTTING DENSITY FUNCTION    */
     CDFDATA(KEEP=P0 FP0)
               /* DATA FOR CUMULATIVE DISTRIBUTION FCT.  */;
  INPUT
  N      /* NUMBER IN INITIAL SAMPLE                    */
  M      /* NUMBER IN SUBSAMPLE                         */
  N2     /* NUMBER CLASSIFIED VARIETY 2 IN INITIAL SAMPLE*/
  X      /* NUMBER CONFIRMED TO BE VARIETY 2 IN SUBSAMPLE*/
  A0     /* PARAMETER OF PRIOR FOR P0                   */
  B0     /* PARAMETER OF PRIOR FOR P0                   */
  A1     /* PARAMETER OF PRIOR FOR P1                   */
  B1     /* PARAMETER OF PRIOR FOR P1                   */
  MAXFN  /* MAXIMUM NUMBER OF TIMES FUNCTION IS TO BE   */
         /* CALLED IN SEARCH FOR LOWER BOUND            */
  ACC    /* ACCURACY TO WHICH THE AREA UNDER THE CDF    */
         /* EQUALS THE DESIRED ALPHA BEFORE SEARCH STOPS */
  NSIG   /* NUMBER OF SIGNIFICANT FIGURES OF AGREEMENT  */
         /* BETWEEN DESIRED VALUE FOR THE LOWER BOUND   */
         /* AND THE ONE FOUND                           */
  ALPHA1 /* ALPHA VALUE FOR SPECIFIC LOWER BOUND DESIRED */
  ALPHA2 /* ALPHA VALUE FOR SPECIFIC LOWER BOUND DESIRED */
  ALPHA3 /* ALPHA VALUE FOR SPECIFIC LOWER BOUND DESIRED */;
```

```
* GLOBAL VARIABLES
  DENOM /* SUMMATION IN DENOMIMATOR OF PDF AND CDF          */
  MAXY2 /* UPPER LIMIT OF SUM                               */
  EP0   /* EXPECTED VALUE OF P0                             */
  EP0SQ /* EXPECTATION SQUARED OF TERMS IN SUMMATION        */
  SD0   /* STANDARD DEVIATION OF P0                         */
  R     /* CONSTANT IN NUMERATOR                            */
  V     /* CONSTANT IN DENOMINATOR                          */
  TEMP  /* ONE TERM IN SUMMATION OF DENOMINATOR             */
  LOW   /* LOWER LIMIT FOR P0                               */
  LOW   /* LOWER LIMIT FOR P0                               */
  HIGH  /* UPPER LIMIT                                      */
  DIFF  /* DIFFERENCE BETWEEN HIGH AND LOW                  */
  STEP  /* INCREMENT                                        */
  CONST /* LOG OF CONSTANT IN DENOMINATOR, PLACED IN        */
        /* NUMERATOR TO PREVENT VALUES BEING OUT OF         */
        /* RANGE FOR EXP FUNCTION.                          */
  T     /* TOLERANCE VALUE                                  */
  P0    /* VALUE OF P0 THAT FUNCTION IS CALCULATED AT       */
  GP0   /* DENSITY AT P0                                    */
  FP0   /* CUMULATIVE DENSITY AT P0                         */
  LB1   /* LOWER BOUND OF 1 - ALPHA1 CREDIBILITY INTERVAL*/
  LB2   /* LOWER BOUND OF 1 - ALPHA2 CREDIBILITY INTERVAL*/
  LB3   /* LOWER BOUND OF 1 - ALPHA3 CREDIBILITY INTERVAL*/;
*-----------------------------------------------------------*;
* CALCULATE THE SUMMATION IN THE DENOMINATOR OF BOTH THE
  PROBABILITY  DENSITY FUNCTION AND THE CUMULATIVE DIST-
  RIBUTION FUNCTION. CALCULATE THE EXPECTED VALUE AND THE
```

```
  STANDARD DEVIATION.;
* ZERO ACCUMULATORS;
  DENOM = 0;
  EP0 = 0;
  EP0SQ = 0;
* CALCULATE UPPER LIMIT OF SUMMATION;
  MAXY2 = N2 - M + X;
* CALCULATE SUMMATION;
  DO Y2 = X TO MAXY2;
    TEMP =  EXP(LGAMMA(N2 + B1 - Y2) + LGAMMA(N + B0 - Y2)
                + LGAMMA(A0 + Y2) - LGAMMA(Y2 - X + 1)
                - LGAMMA(N2 - Y2 - M + X + 1)
                - LGAMMA(N - Y2 + A1 + B1));
    DENOM = DENOM + TEMP;
    R = Y2 + A0;
    EP0 = EP0 + TEMP*R;
    EP0SQ = EP0SQ + TEMP*R*(R+1);
  END;
  V = N + A0 + B0;
  EP0 = ROUND((EP0 /(V*DENOM)),.00001);
  SD0 = ROUND((SQRT(EP0SQ/(V*(V+1)*DENOM) - EP0*EP0),
                                               .00001);
* CALCULATE A +/- 4 STANDARD DEVIATION BOUND AS A
  REASONABLE RANGE TO CALCULATE VALUES OF THE
  PDF AND CDF OVER;
  LOW = ROUND(MAX(0,(EP0 - 4*SD0)),.0001);
  HIGH = ROUND(MIN(1,(EP0 + 4*SD0)),.0001);
* IF THIS RANGE IS NARROW, THIS IS A DEGENERATE FUNCTION;
```

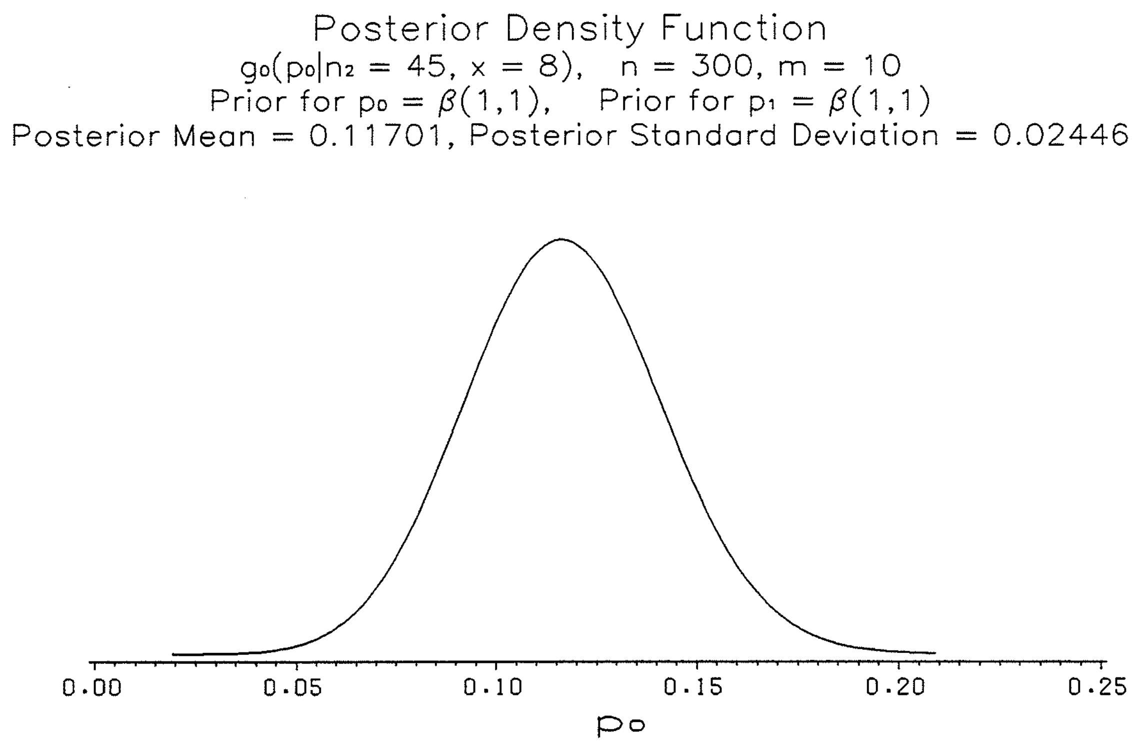
```
  DIFF = HIGH - LOW;
  IF DIFF < 0.005 THEN DO;
    PUT "DEGENERATE FUNCTION";
    GOTO LAST;
  END;
* DETERMINE AN INCREMENT FINE ENOUGH TO GIVE GOOD PLOTS;
  STEP = ROUND(MIN(0.01,(DIFF/10)),0.0001);
* SELECT RANGE OF VALUES WITHIN (0,1);
  LOW = MAX(STEP/2,LOW);
  HIGH = MIN((1 - STEP/2),HIGH);
* CALCULATE CONSTANT USED IN PDF;
    CONST = LGAMMA(V);
* CALCULATE ACCURACY FACTOR;
  T = 10**(-NSIG);
* GENERATE VALUES OF THE PDF AND CDF;
  DO P0 = LOW TO HIGH BY STEP;
    %GP(P0,GP0);
    %FP(P0,FP0);
    OUTPUT PDFDATA;
    OUTPUT CDFDATA;
  END;
* FIND LOWER BOUNDS CORRESPONDING TO THE INPUT ALPHAS ;
  %SEARCH(LB1,ALPHA1);
  %SEARCH(LB2,ALPHA2);
  %SEARCH(LB3,ALPHA3);
* ROUND RESULTS TO THE ACCURACY WITH WHICH THEY WERE FOUND;
  LB1 = ROUND(LB1,T);
  LB2 = ROUND(LB2,T);
```

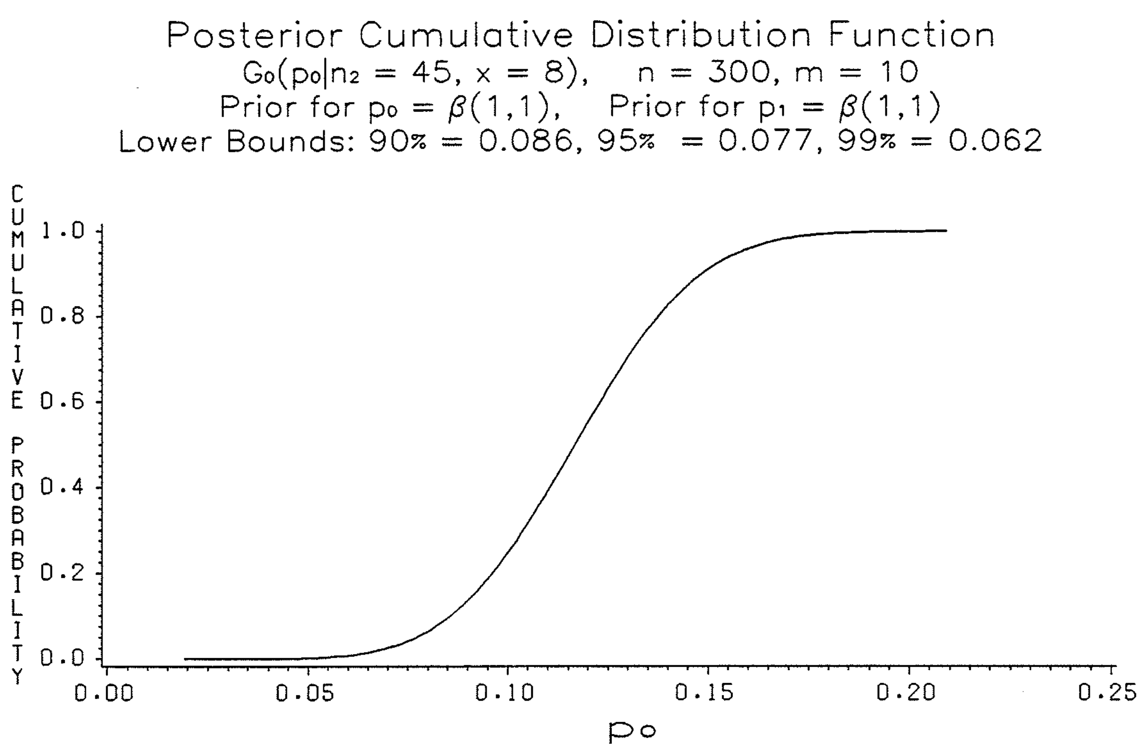
```
  LB3 = ROUND(LB3,T);
* STORE VALUES FOR PRINTING TITLES;
  CALL SYMPUT('NN',LEFT(N));
  CALL SYMPUT('MM',LEFT(M));
  CALL SYMPUT('NN2',LEFT(N2));
  CALL SYMPUT('XX',LEFT(X));
  CALL SYMPUT('AA0',LEFT(A0));
  CALL SYMPUT('BB0',LEFT(B0));
  CALL SYMPUT('AA1',LEFT(A1));
  CALL SYMPUT('BB1',LEFT(B1));
  CALL SYMPUT('EEP0',LEFT(EP0));
  CALL SYMPUT('SSD0',LEFT(SD0));
  CALL SYMPUT('AAPH1',LEFT((1 - ALPHA1)*100));
  CALL SYMPUT('AAPH2',LEFT((1 - ALPHA2)*100));
  CALL SYMPUT('AAPH3',LEFT((1 - ALPHA3)*100));
  CALL SYMPUT('LLB1',LEFT(LB1));
  CALL SYMPUT('LLB2',LEFT(LB2));
  CALL SYMPUT('LLB3',LEFT(LB3));
  CALL SYMPUT('LLOW',
          LEFT(MAX(0,(ROUND(LOW,0.01) - 0.05))));
  CALL SYMPUT('HHIGH',
          LEFT(MIN(1,(ROUND(HIGH,0.01) + 0.05))));
  IF HIGH - LOW > 0.2 THEN CALL SYMPUT('BBY',0.1);
  ELSE CALL SYMPUT('BBY',0.05);
  LABEL FP0 = "CUMULATIVE PROBABILITY";
LAST: STOP;
CARDS;
300 10 45 8 1 1 1 1 8 .001 3 .10 .05 .01
```

```
***********************************************************;
GOPTIONS NOTEXT82 ROTATE DEVICE=XEROX COLORS=(BL);
PROC GPLOT DATA=PDFDATA GOUT=GRAPHLIB;
  TITLE1 H=2.4 "POSTERIOR DENSITY FUNCTION";
  TITLE2 F=SIMPLEX H=2.0 "g" H=1.0 "0" H=2.0 "(p" H=
       1.0 "0" H=2.0 "|n" H=1.0 "2" H=
       2.0 " = &NN2., x = &XX.), n = &NN., m = &MM.";
  TITLE3 F=SIMPLEX H=2.0 "Prior for p" H=1.0 "0" H=
       2.0 " = " F=GREEK "b(&AA0.,&BB0)" F=
       SIMPLEX "  Prior for p" H=1.0 "1" H=
       2.0 " = " F=GREEK "b(&AA1.,&BB1.)";
  TITLE4 F=SIMPLEX H=1.0 OCC=2 "E(p" H=2.0 "0" H=
       2.0 " = &EEP0., STD(p" H=1.0 "0" H=
       2.0 " = &SSD0.";
  AXIS1 LABEL=(F=SIMPLEX H=2 "P" H=1 "0") VALUE=(H=1.5)
      ORDER=(&LLOW TO &HHIGH BY &BBY);
  AXIS2 LABEL=NONE MAJOR=NONE VALUE=NONE STYLE=0;
  PLOT GP0*P0 /HAXIS=AXIS1 VAXIS=AXIS2;
  SYMBOL1 I=SPLINE L=1;
```

```
***********************************************************;
PROC GPLOT DATA=CDFDATA GOUT=GRAPHLIB;
  TITLE1 H=2.4 "POSTERIOR CUMULATIVE DISTRIBUTION FUNCTION";
  TITLE2 F=SIMPLEX H=2.0 "G" H=1.0 "0" H=2.0 "(p" H=
         1.0 "0" H=2.0 "|n" H=1.0 "2" H=
         2.0 " = &NN2., x = &XX.), n = &NN., m = &MM.";
  TITLE3 F=SIMPLEX H=2.0 "Prior for p" H=1.0 "0" H=
         2.0 " = " F=GREEK "b(&AA0.,&BB0)" F=
         SIMPLEX "  Prior for p" H=1.0 "1" H=2.0 " = " F=
         GREEK "b(&AA1.,&BB1.)";
  TITLE4 F=SIMPLEX H=2.0 "Lower Bounds: &AAPH1.% = " H=
         2.0 "&LLB1., &AAPH2.%  = &LLB2., &AAPH3.% = " H=
         2.0 "&LLB3.";
  AXIS1 LABEL=(F=SIMPLEX H=2.0 "P" H=1.0 "0") VALUE=(H=1.5);
        ORDER=(&LLOW TO &HHIGH BY &BBY) MINOR=(N=9);
  AXIS2 ORDER=(0 TO 1 BY 0.1) MINOR=(N=3) VALUE=(H=1.5);
  PLOT FP0*P0 /HAXIS=AXIS1 VAXIS=AXIS2;
  SYMBOL1 I=SPLINE L=1;
***********************************************************;
PROC GREPLAY IGOUT=GRAPHLIB;
  REPLAY 1 2;
```

Figure A.1: SAMPLE OUTPUT PLOTS





BIBLIOGRAPHY

BEAINY, ILHAM and CASE, KENNETH E. (1981). "A Wide Variety of AOQ and ATI Performance Measures with and without Inspection Error." **Journal of Quality Technology** 13, 1-9.

BIEGEL, JOHN E. (1974). "Inspector Errors and Sampling Plans." **American Institute of Industrial Engineers, Transactions** 6, 284-287.

BOX, G.E.P. and TIAO, G.C. (1962). "A Further Look at Robustness via Bayes's Theorem." **Biometricka** 49, 419-432.

BOX, GEORGE E.P. and TIAO, GEORGE C. (1973). **Bayesian Inference in Statistical Analysis**, Addison-Wesley Publishing Co : Reading, Massachusetts.

BROSS, IRWIN (1954). "Misclassification in 2 x 2 Tables." **Biometrics** 10,478-486.

BRENT, R.P. (1971). "An Algorithm with Guaranteed Convergence for Finding a Zero of a Function." **The Computer Journal** 14, 422:425.

CARR, WENDELL E. (1982). "Sampling Plan Adjustment for Inspection Error and Skip-Lot Plan." **Journal of Quality Technology** 14, 105-114.

CASE, KENNETH E., BENNETT, G. KEMBLE and SCHMIDT, J. W. (1973). "The Dodge CSP-1 Continuous Sampling Plan Under Inspection Error." **American Institute of Industrial Engineers, Transactions** 5, 193-202.

CASE, KENNETH E., BENNETT, G. KEMBLE and SCHMIDT, J. W. (1975). "The Effect of Inspection Error on Average Outgoing Quality." **Journal of Quality Technology** 7, 28-33.

CASE, KENNETH E. and KEATS, J. BERT (1982). "On the Selection of a Prior Distribution in Bayesian Acceptance Sampling." **Journal of Quality Technology** 14, 10-18.

CHEN, T. TIMOTHY (1979). "Log-Linear Models for Categorical Data with Misclassification and Double Sampling." **Journal of the American Statistical Association** 74, 481-488.

CHIACCHIERINI, RICHARD P. and ARNOLD, JESSE C. (1977). "A Two-Sample Test for Independence in 2 x 2 Contingency Tables with Both Margins Subject to Misclassification." Journal of the American Statistical Association 72, 170-174.

COLLINS, RUFUS D., Jr. and CASE, KENNETH E. (1976). "The Distribution of Observed Defectives in Attribute Acceptance Sampling Plans Under Inspection Error." American Institute of Industrial Engineers, Transactions 8, 375-378.

COLLINS, RUFUS D., Jr., CASE, KENNETH E. and BENNETT, G. KEMBLE (1973). "The Effects of Inspection Error on Single Sampling Inspection Plans." International Journal of Production Research 11, 289-298.

COMBS, CHARLES A. and STEPHENS, LARRY J. (1980). "Upper Bayesian Confidence Limits on the Proportion Defective." Journal of Quality Technology 12, 196-200.

DEMING, W.EDWARDS (1977). "An essay on Screening, or Two-phase Sampling, Applied to Surveys of a Community". International Statistical Review 45, 29-37.

DIAMOND, EARL L. and LILIENFELD, ABRAHAM M. (1962). "Effects of Errors in Classification and Diagnosis in Various Types of Epidemiological Studies. American Journal of Public Health 2, 1137-1144.

DORRIS, ALAN L. and FOOTE, BOBBIE L. (1978). "Inspection Errors and Statistical Quality Control: A Survey. American Institute of Industrial Engineers, Transactions 10, 184-192.

FLEISS, JOSEPH L. (1981). Statistical Methods for Rates and Proportions, John Wiley and Sons : New York

GODFREY, JAMES and NETER, JOHN (1984). "Bayesian Bounds for Monetary Unit Sampling in Accounting and Auditing." Journal of Accounting Research 22, 497-525.

GUILD, RICHARD D. and RAKA, IDA I. D. G. (1980). "Effective Sampling Plans Based on a Prior Distribution." Journal of Quality Technology 12, 88-93.

HALD, A. (1968). "Bayes Single Sampling Attribute Plans for Continuous Prior Distributions." Technometrics 10, 667-683.

HEALY, J. D. (1981). "The Effects of Misclassification Error on the Estimation of Several Population Proportions." The Bell System Technical Journal 60, 697-705.

HOAG, LAVERNE L., FOOTE, BOBBIE L. and MOUNT-CAMPBELL, CLARK (1975). "The Effect of Inspector Accuracy on the Type I and Type II Errors of Common Sampling Techniques." Journal of Quality Technology 7, 157-164.

HOCHBERG, YOSEF (1977). "On the Use of Double Sampling Schemes in Analyzing Categorical Data with Misclassification Errors." **Journal of the American Statistical Association** 72, 914-921.

JAECH, JOHN L. (1980). "Determination of Acceptance Numbers and Sample Sizes for Attribute Sampling Plans." **Journal of Quality Technology** 12, 187-190.

JOHNSON, NORMAN L., KOTZ, SAMUEL and RODRIGUES, ROBERT N. (1985). "Statistical Effects of Imperfect Inspection Sampling: I. Some Basic Distributions." **Journal of Quality Technology** 17, 1-31.

JOHNSON, N. L., KOTZ, SAMUEL and SORKIN, H. L. (1980). "“Faulty Inspection” Distributions." **Communications in Statistics A** 9, 917-922.

KITTLER, J. and PAU, L. F. (1980). "Automatic Inspection by Lots in the Presence of Classification Errors." **Pattern Recognition** 12, 237-241.

KOCH, GARY G. (1969). "The Effect of Non-Sampling Errors on Measures of Association in 2 X 2 Tables." **Journal of the American Statistical Association** 64, 851-864.

KOTZ, SAMUEL and JOHNSON, NORMAN L. (1982). "Errors in Inspection and Grading: Distributional Aspects of Screening and Hierarchal Screening." **Communications in Statistics A** 11, 1997-2016.

LAUER, G. NICHOLAS (1978). "Acceptance Probabilities for Sampling Plans Where the Proportion Defective Has a Beta Distribution." **Journal of Quality Technology** 10, 52-55.

MENZEFRICKE, ULRICH (1984). "The Effects of Variability in Inspector Error." **Journal of Quality Technology** 16, 131-135.

MINTON, GEORGE (1972). "Verification Error in Single Sampling Inspection Plans for Processing Survey Data." **Journal of the American Statistical Association** 67, 46-54.

MOSKOWITZ, HERBERT and FINK, RAYMOND (1977). " A Bayesian Algorithm Incorporating Inspector Errors for Quality Control and Auditing." **TIMS Studies in the Management Sciences** 7, 79-104.

PFANZAGL, J. (1963). "Sampling Procedures Based on Prior Distributions and Costs." **Technometrics** 5, 47-61.

RAIFFA, HOWARD and SCHLAIFER, ROBERT (1961). **Applied Statistical Decision Theory.** Graduate School of Business Administration, Harvard University : Boston.

RAHALI, BOUBEKEUR and FOOTE, B. L. (1982). "An Approach to Compensate for Uncertainty in Knowledge of Inspector Error." **Journal of Quality Technology** 14, 190-185.

SINCLAIR, M.A. (1978). " A Collection of Modelling Approaches for Visual Inspection in Industry." **International Journal of Production Research** 16, 275-292.

SKELLAM, J.G. (1948). "A Probability Distribution Derived From the Binomial Distribution by Regarding the Probability of Success as Variable Between the Sets of Trials." **Journal of the Royal Statistical Society B** 10, 257-261.

SUTCLIFFE, J. P. (1965a). "A Probability Model for Errors of Classification. I. General Considerations" **Psychometrika** 30, 73-96.

SUTCLIFFE, J. P. (1965b). "A Probability Model for Errors of Classification, II. Particular Cases" **Psychometrika** 30, 129-155.

TENENBEIN, AARON (1970). "A Double Sampling Scheme for Estimating from Binomial Data with Misclassifications." **Journal of the American Statistical Association** 65, 1350-1361.

TENENBEIN, AARON (1971). "A Double Sampling Scheme for Estimating From Binomial Data with Misclassifications: Sample Size Determination." **Biometrics** 27,935-944.

TENENBEIN, AARON (1972). "A Double Sampling Scheme for Estimating from Misclassified Multinomial Data with Applications to Sampling Inspection." **Technometrics** 14, 187-202.

TIEMSTRA, PETER J. (1981). "Measuring the Performance of Inspection and Sorting Systems." **Journal of Quality Technology** 13, 149-204.

TRADER, RAMONA L. (1983). "A Bayesian Analysis of Imperfect Inspection Models." **Communications in Statistics, Theory and Methods** 12, 397-408.

WEILER, H. (1965). "The Use of Incomplete Beta Functions for Prior Distributions in Binomial Sampling." **Technometrics** 7, 335-347.

WOODING, W. M. (1979). "A Source of Bias in Attributes Testing, and a Remedy." **Journal of Quality Technology** 11, 169-176.