A BAYESIAN APPROACH FOR ESTIMATING A PROPORTION USING A TWO-STAGE SAMPLING SCHEME INVOLVING IMPERFECT AND PERFECT CLASSIFICATION

BY

SCOTT P. NOWICKI

A Thesis
Submitted to the Faculty of Graduate Studies
in Partial Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Statistics
University of Manitoba
Winnipeg, Manitoba

ISBN 0-315-76873-8

A BAYESIAN APPROACH FOR ESTIMATING A PROPORTION USING A TWO-STAGE SAMPLING SCHEME INVOLVING IMPERFECT AND PERFECT CLASSIFICATION

BY

SCOTT P. NOWICKI

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 two-sided bound on this proportion. Two measuring devices are available to classify the items -- one device (imperfect) is relatively inexpensive but tends to misclassify items, while the second device (perfect) is an expensive procedure which does not misclassify items. A double sampling plan is used to provide information on the proportion of interest and on the effect of misclassification. Items in an initial sample are classified by the imperfect classifier into two groups -- those thought to be "nonconforming", and those thought to be "conforming". A sub-sample is taken from each group, and these items are reclassified by the perfect classifier.

Bayesian methods are used to obtain a posterior distribution for the proportion of nonconforming items in the population, which may then be used to obtain a point estimate and credibility bounds. Prior distributions for the proportion of nonconforming items and the misclassification rates are modelled by independent beta priors.

Neden (1986) also used a double sampling scheme and Bayesian methods to obtain a posterior distribution for the proportion of nonconforming items. However, she dealt with a one-sided "confirmatory" subsampling plan, (that is, a subsample was taken either from the "nonconforming" group or the "conforming" group, but not from both groups) and she produced a conservative one-sided credibility bound. Tenenbein (1970) also utilized a double sampling scheme, but his estimates and error bounds were obtained using classical asymptotic methods. His methods are not appropriate in situations commonly encountered in which the proportion of interest is quite small or in which the subsample sizes are small.

Economic considerations may necessitate that one choose between the one-sided "confirmatory" sampling plan of Neden, and the two-sided sampling plan considered here. That is, for a fixed sample size, this choice may be between taking all of the samples from one group or taking some from each group. It is shown that there are situations in which one sampling plan may be more appropriate than the other, depending on the form of the inferences to be drawn and on the prior information concerning the misclassification rates.

ACKNOWLEDGEMENTS

I am most grateful to my advisor, Dr. John F. Brewster for his advice and guidance not only in the development of this thesis, but throughout my entire graduate program. Particular thanks go to committee members Dr. Brian D. Macpherson and Dr. Darryl Kraft for their helpful comments and criticisms.

The interest of my friends and fellow classmates in the progress of my work was much appreciated.

Finally, I thank two special friends David Christopher and Peter Graceffo, and my family for their interest and encouragement throughout the duration of this thesis without which I may never have reached this point.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION AND SUMMARY

Taking a sample of items from a population, it is desired to estimate the proportion of items which have a certain characteristic and to place a one-sided or two-sided bound on this proportion. Two measuring devices are available to classify the items -- one device (fallible) is relatively inexpensive but tends to misclassify items, while the second device (infallible) is an expensive procedure which does not misclassify items.

Factors such as time and cost make it unreasonable to classify all the sample items by the infallible procedure. In the hope of reducing time and cost while providing an accurate estimate of the proportion of interest, it seems reasonable to consider a design which will incorporate both types of measuring devices.

A double sampling scheme may be used to provide information on the proportion of interest and on the effect of misclassification. At the first stage a random sample of items is taken from a lot and classified by the fallible procedure into one of two categories -- those thought to be conforming and those thought to be nonconforming. At this stage some of the items may be misclassified. For example, category 1 may contain both conforming and nonconforming items when in reality it was to contain only conforming items. Similarly, category 2 may also contain items from both groups when in fact it should only contain nonconforming items. Due to the fact that some items at this stage may be misclassified, a second stage sub-sample is taken from both categories. The items taken at this second stage are then classified by the infallible procedure. The idea of the two-stage sampling plan is to use the sub-samples to provide information on the accuracy of the fallible procedure.

In this thesis, Bayesian theory is used to find a solution. The Bayesian approach allows one to combine direct sample evidence with any prior information that may be available to form a posterior distribution for the proportion of interest.

In our case, three parameters are involved: p_o, which is the proportion of nonconforming items in the lot, p_1, which is the probability of a conforming item being correctly identified at the initial sampling stage, and p_2, which is the probability of a nonconforming item being correctly identified at the initial sampling stage.

Bayesian methods allow one to combine information obtained from the double sampling scheme with the joint prior distribution for p_o, p_1 and p_2 to form a joint posterior distribution of p_o, p_1 and p_2.

Bayesian methods also allow for the handling of the nuisance parameters p_1 and p_2. Inference on the parameter of interest, p_o, is based on its marginal posterior distribution. This marginal posterior distribution is obtained by integrating the joint posterior distribution of p_o, p_1 and p_2 over the nuisance parameters p_1 and p_2. The marginal posterior distribution of p_1 or p_2 can be obtained in a similar manner.

Bayesian methods require that prior distributions be specified for all of the parameters involved. When the prior information consists of previous sample evidence, the choice of a prior distribution may be immediate. If this is not the case, the decision maker can quantify his judgement with regards to the parameters of interest and form a subjective prior distribution.

Since the testing of items by an infallible method to determine it's correct classification may be expensive, it is desirable to incorporate any previous data that may be available. Information on the proportion of nonconforming items in previous lot shipments, information on the proportion of conforming items correctly classified by the inspection process, or information on the proportion of nonconforming items correctly classified by the inspection process may be available. If the process appears fairly stable, then it seems that this knowledge from previous samples should be utilized in the solution.

For this particular solution, knowledge about p_o, p_1 and p_2 will be modelled by independent beta priors. The use of independent beta distributions for the priors greatly simplifies the calculation of the posterior distribution. The use of beta distributions also provides us with a wide variety of shapes which can be used to approximate many reasonable prior distributions.

The motivation for this thesis was based on a problem that arose in the grain industry. A boxcar may contain two varieties of wheat: variety 1, which conforms to grading standards, and variety 2, which does not. If the proportion of nonconforming kernels is thought to exceed some specified value the carlot is down graded and the shipper is paid less.

The two varieties of wheat may be quite similar in appearance making correct visual identification difficult. Because of this, some kernels of wheat may be misclassified. Misclassification can occur in either direction: a variety 1 kernel may be classified as variety 2, or a variety 2 kernel may be classified as variety 1. Laboratory techniques that provide exact identification of the kernels are available, but this procedure is very costly and therefore only a small number of kernels can be classified by this procedure.

For example, a grain inspector may visually inspect a sample of 300 kernels and decide that 255 kernels are "conforming" and 45 kernels are "nonconforming". A sample of 10 kernels is taken from the 255 kernels thought to be "conforming" and a second sample of 10 kernels is also taken from the 45 kernels thought to be "nonconforming". These sub-samples are then classified by the laboratory technique in order to determine the correct classification. The laboratory analysis may have found 8 out of the 10 kernels taken from the "nonconforming" pile to be nonconforming and 10 out of the 10 kernels taken from the "conforming" pile to be conforming. The sampling scheme for this example can be seen in Figure 1.1. The problem is to combine all of this information in order to form inferences about p_o.

In this case, suitable beta prior distributions may consist of: $\beta(1,10)$ for p_o, $\beta(1,1)$ for p_1 and $\beta(20,1)$ for p_2. The prior for p_o represents the feeling that the proportion of nonconforming kernels in the carlot is small. The prior for p_1 represents a state of "ignorance" about the probability of a conforming kernel being correctly identified, and the prior for p_2 represents the feeling that the probability of correctly identifying a nonconforming kernel as nonconforming is very good. These prior distributions can be seen in Figure 1.2.

After combining the joint prior distribution for p_o, p_1 and p_2 with the sample evidence, and integrating with respect to p_1 and p_2, we obtain the posterior distribution for p_o in Figure 1.3. The mean of the posterior distribution provides a point estimate of p_o, and the posterior distribution can also used to provide a one-sided or two-sided bound for p_o. In this case, a point estimate for p_o would be 14.31% and a 95% credibility interval would be $9.76\% \leq p_o \leq 19.24\%$. The mathematical form of the posterior distribution can be found in Section 4.4.

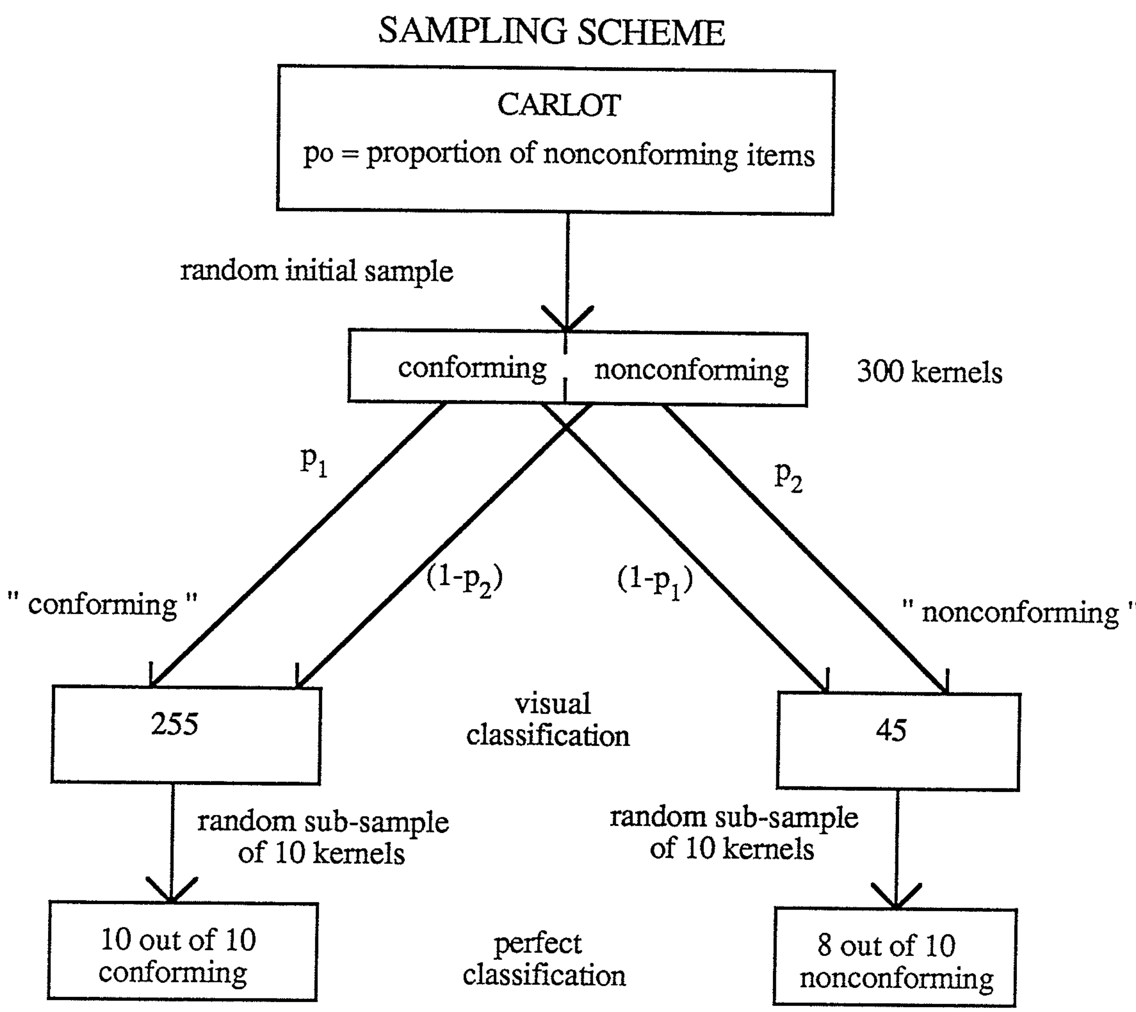

Figure 1.1: Sampling Scheme

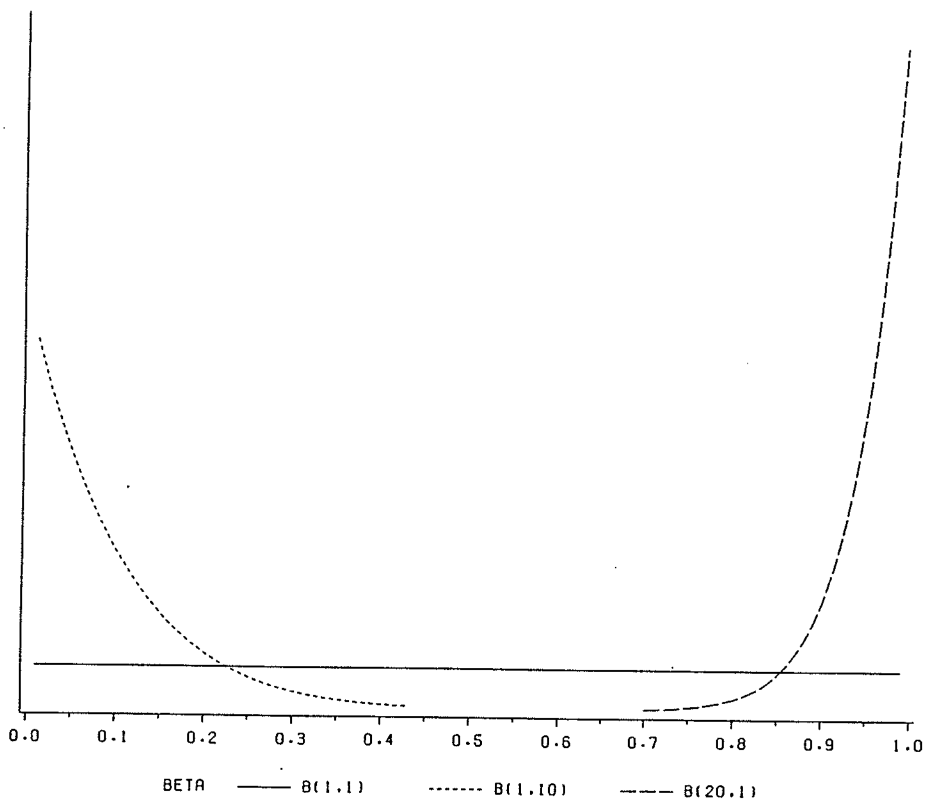

Figure 1.2: Beta Prior Density Functions for p_o, p_1 and p_2

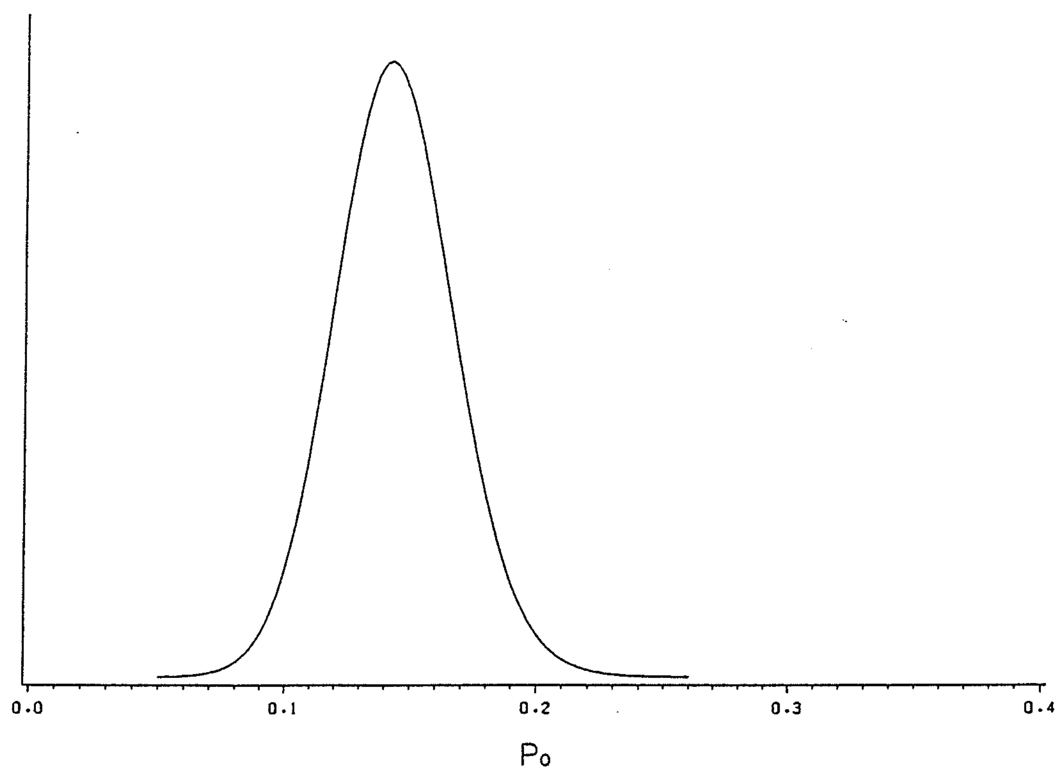

Figure 1.3: Posterior Density Function for p_o

Since varying the prior distributions will cause the posterior distribution to vary, it is of interest to see how sensitive the posterior distribution is to variations in the prior distributions. A look at the sensitivity of the posterior distribution can be found in Section 4.6.

In the grain industry a shipper is paid according to the proportion of kernels not meeting grading standards. It is therefore of interest to see if the proportion of nonconforming items p_o, exceeds some critical value, say p^*, and hence to test the following hypotheses:

$$H_0: p_o \leq p^*$$
$$H_1: p_o > p^* .$$

Credibility bounds permit the testing (in a Bayesian sense) of such hypotheses. If a lower bound for p_o is found to be greater than p^*, then the null hypothesis would be rejected in favor of the alternative.

Similarly, if the hypotheses to be tested is:

$$H_0: p_o \geq p^*$$
$$H_1: p_o < p^* ,$$

an upper bound for p_o can be found and used to test this hypotheses. If the upper bound is found to be less than p^*, then the null hypothesis would be rejected in favor of the alternative.

The work in this thesis is closely related to that of Neden (1986). Neden also used a double sampling scheme and Bayesian methods to obtain a posterior distribution for p_0. However, Neden dealt with a one-sided "confirmatory" sub-sample, that is, at the second stage of sampling, a sub-sample was taken from either the "nonconforming" group or the "conforming" group, -- the decision whether to draw a sample from the "nonconforming" or the "conforming" group depending on the size of the "nonconforming" group at the initial inspection stage. For example, if at the inspection stage a large number of nonconforming items are found, the inspector may conclude that the lot is unacceptable. One may then want to confirm that the lot does in fact contain an unacceptable number of nonconforming items. If this is the case, a sub-sample from the "nonconforming" group can be taken in order to obtain a lower bound for the proportion of nonconforming items.

Neden used a one-sided "confirmatory" sub-sampling plan in order to confirm if $p_o \le p^*$ or $p_o > p^*$. If the proportion of nonconforming items, p_o, in the lot was thought to be unacceptable at the initial sampling stage, a lower bound was desired by Neden in order to confirm that p_o was large. Neden assumed that no misclassification of nonconforming items could take place and therefore no sub-sample was taken from the "conforming" category. This assumption provides a conservative result, as any misclassification of the conforming items would cause the estimated value of p_o and the lower bound to become larger. Therefore, if H_o can be rejected under the assumption that no misclassification of nonconforming items can take place, it will also be rejected if misclassification of the nonconforming items can take place.

Similarly, if the "nonconforming" group is small at the initial sampling stage, the inspector may make the decision that the lot is acceptable. One may want to confirm that the lot is good by taking a sub-sample from the "conforming" group in order to obtain a upper bound for p_o. By assuming that no misclassification of conforming items can take place, no sub-sample is taken from the "nonconforming" category. This would provide a conservative result as any misclassification of conforming items would cause the estimated value of p_o and the upper bound to become smaller.

If some point estimate or two-sided interval estimate for the proportion of nonconforming items was desired, sub-samples would have to be taken from both of the visually inspected category 1 and category 2 groups (two-sided sub-sampling plan) in order to see the effect of misclassification in both directions. This thesis will compare the above two-sided sub-sampling plan with the one-sided "confirmatory" sampling plan used by Neden.

If a point estimate is not desired but one is interested in a lower bound there are several ways in which samples may be allocated. All samples could be taken from the "nonconforming" category (one-sided confirmatory sub-sample), or samples could be taken from both of the "conforming" and "nonconforming" categories (two-sided sub-sample).

In reality it is hoped that the probability of correctly classifying a conforming or nonconforming item be near 1. If prior distributions are used to reflect this information, there are situations where little difference in the lower regions for the one-sided and two-sided sampling plans can be seen. For example, such a situation arises when sub-samples of size 10 are taken from both categories. Similarly, there are situations where a difference in the lower region between the two sampling schemes can be seen.

If one is interested only in obtaining a lower bound, and resources permit that only 20 sub-samples may be taken, it seems that the one-sided "confirmatory" sampling plan with 20 samples taken from the "nonconforming" category may be more appropriate than the two-sided sampling plan with samples of size 10 taken from each category since it provides a higher lower bound.

Comparisons between the one-sided and two-sided sampling schemes on the lower tail region of the posterior distribution can be found in Section 4.7.

Tenenbein (1970), Diamond and Lilienfeld (1962) and Deming (1977) also considered designs which incorporated both fallible and infallible measuring devices. To estimate the proportion of items belonging to one of two possible categories, Tenenbein (1970) presented a double sampling scheme similar to that in this thesis. However, his estimates were obtained using asymptotic methods. Often the proportion of interest may be quite small, and together with small sub-samples, the use of asymptotic techniques based on normal theory do not seem appropriate. The use of Bayesian methods allows one to deal with the small sub-sample sizes. It will also allow us to deal with nuisance parameters and to incorporate any previous information that we may have available concerning the parameters of interest.

CHAPTER 2

THE PROBLEM

2.1 The General Problem

It is desired to estimate a proportion, p, of items which have a certain characteristic and to place a one-sided or two-sided bound on this proportion. Two measuring devices are available to classify the items -- one device is relatively inexpensive but tends to misclassify items, while the second device is an expensive procedure which does not misclassify items.

Due to factors such as time and cost it is unreasonable to classify all the sample items by the expensive procedure. If however the items are only classified by the cheaper procedure which tend to misclassify items, then we run the risk of inaccurate estimates and inferences. For example in Diamond and Lilienfeld (1962) a study involving cancer of the cervix and lack of circumcision is presented, and it is suggested that misclassification may produce an observed association between cancer of the cervix and lack of circumcision, when in fact no such association exists.

It therefore seems reasonable to consider a design in which both types of measuring devices can be used in the hope of reducing time and cost, while providing an accurate estimate of the proportion of interest. One such application is in the medical field where a relatively cheap but fallible procedure is an interview of a patient and an expensive but infallible procedure may involve a physical examination and laboratory test. Diamond and Lilienfeld (1962) discuss such a situation.

2.2 Effects of Misclassification

There are many practical problems when mistakes in classification are going to be made. In some situations classifications involve almost no risk of error, for example, classifications involving categories such as "lived" and "died". However many classifications involve considerable risk of error. For example, in deciding whether someone has a particular illness the misclassification rates could vary considerably depending on the facilities available, experience of the doctor and other factors.

Even when exact methods of classification are available they may not be feasible due to time and cost. This may then necessitate the use of cheaper and faster methods that are subject to errors. Such procedures can lead to misclassification in several directions. For example, in the electronics industry certain electrical components may be visually inspected rather that laboratory tested. The principle advantage of visual inspection would be its relatively high speed and low cost, but visual inspection in this case may cause a true defect to be classified as a non-defect, and may cause a non-defect to be classified as a defect.

It has been shown by Bross (1954) that, under misclassification, if the sample proportion $\frac{x}{n}$ is used as an estimate of p, the resulting estimate of p is biased, with the bias, being a function of the misclassification rate. (Here, n is the sample size and x is the number of items classified as being in a particular category)

2.3 The Double Sampling Scheme

In many situations there are both true and fallible measuring devices. The true classifier is an expensive procedure which does not misclassify items, while the fallible classifier is a relatively inexpensive procedure that tends to misclassify items. Tenenbein (1970) proposed a two-stage procedure as a compromise between the two extremes. A sample of size N was taken from the population of interest. A sub-sample of size n was taken from the original sample of size N and classified by both the fallible and true classifiers. The remaining (N-n) items were classified by the fallible classifier. An estimate was obtained by combining the information from both of the samples using asymptotic methods. The idea of the two-stage plan is to use sub-samples to provide information on the accuracy of the fallible procedure. In this thesis a double sampling scheme will be used along with Bayesian methods in order to obtain an estimate of the number of nonconforming items in a lot.

2.4 Sampling Scheme Notation

The following notation will be used throughout this paper, and is also used in the description of the sampling scheme (see Figure 2.1).

p_o = proportion of nonconforming items

p_1 = probability of a conforming item being correctly identified

p_2 = probability of a nonconforming item being correctly identified

n = total number of items in the initial sample

n_1 = number of items in the initial sample classified as being conforming

n_2 = number of items in the initial sample classified as being nonconforming

y = number of nonconforming items in the initial sample

y_1 = number of nonconforming items in the initial sample classified as conforming

y_2 = number of nonconforming items in the initial sample classified as nonconforming

m_1 = total number of items in the sub-sample taken from n_1

m_2 = total number of items in the sub-sample taken from n_2

x_1 = number of items found to be nonconforming in m_1 by a perfect technique

x_2 = number of items confirmed to be nonconforming in m_2 by a perfect technique

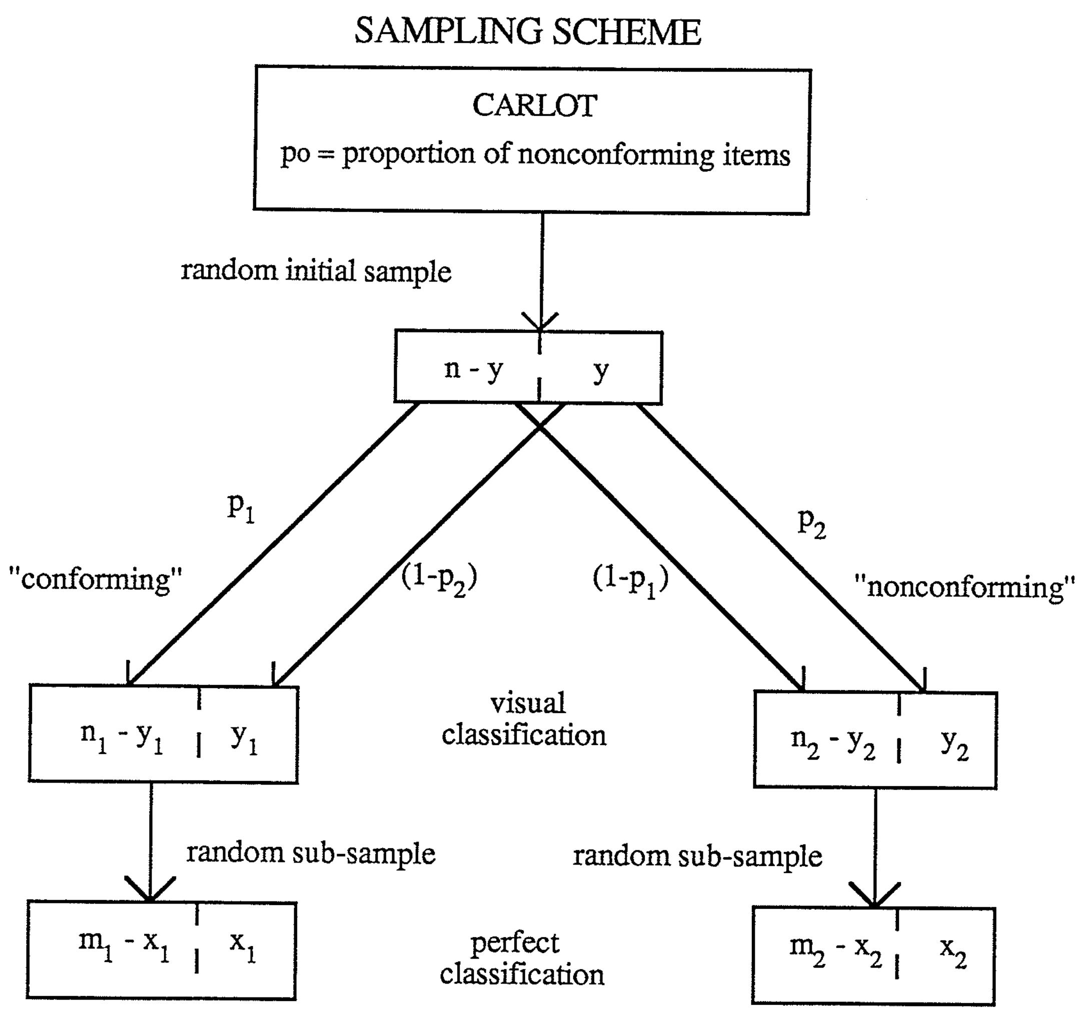

Figure 2.1: Double Sampling Scheme

A population contains a proportion p_0 of nonconforming items. An initial sample of size n is taken from this population, and this sample contains y nonconforming items. The sample is inspected and each item is classified into one of two categories, those items thought to conform to standard, n_1, or those items thought not to conform to standard, n_2.

It is assumed that the inspection process is not perfect. Due to the effect of misclassification the actual number of conforming items in the category thought to be conforming is $n_1 - y_1$, while the actual number of nonconforming items thought to be nonconforming is y_2, and therefore, the number of conforming items thought to be nonconforming is $n_2 - y_2$.

The probability of a nonconforming item being correctly classified as nonconforming is p_2, and the probability of a conforming item being correctly classified as conforming is p_1. The probabilities of p_1 and p_2 are not necessarily equal to one another. It is assumed that the probabilites of correct classification remain constant from item to item, independent of the classification of the previous items.

Random sub-samples of size m_1 and m_2 are taken from the two categories of size n_1 and n_2 thought to be conforming and nonconforming items respectively. It is assumed that a perfect technique of classification exists. The sub-samples of sizes m_1 and m_2 are examined using this technique and the actual number of nonconforming items x_1 and x_2 obtained.

The values n, n_1, n_2, m_1, m_2, x_1 and x_2 are observed directly whereas y, y_1 and y_2 are not observable. For example, a sample of n= 300 might be inspected and the number of conforming items observed to be $n_1 = 240$ and the number of nonconforming items observed to be $n_2 = 60$ items. Random sub-samples of size $m_1 = 10$ and $m_2 = 10$ may be taken from the 240 conforming items and the 60 nonconforming items respectively. Using a perfect technique to classify the items it may be found that 1 of the 10 sub-samples from the conforming items was actually nonconforming, $x_1 = 1$, while the sub-sample of size 10 taken from the nonconforming group may have found 8 nonconforming items, $x_2 = 8$, and $m_2 - x_2 = 2$ conforming items.

All of this information can now be combined with any prior information in order to form inferences about p_o.

CHAPTER 3

BAYESIAN ELEMENTS

3.1 Introduction

In Bayesian inference the parameter of interest is looked upon as a random variable having a prior distribution, that reflect one's opinion or knowledge about the parameter prior to the collection of the data. The main problem associated with Bayesian inference is the subjective nature of the prior distribution. The Bayesian approach consists of a mechanism which incorporates direct sample evidence with any prior information that may be available to form a posterior distribution. Once the posterior distribution of a parameter has been obtained, it can be used to produce estimates, or it can be used to make probability statements about the parameter of interest.

3.2 Prior Distribution

Suppose that someone is interested in making inferences about a parameter p, and it is assumed that p can take on any value from zero to one. It is also assumed that this information concerning p can be represented by a probability distribution, called the prior distribution of p.

The prior distribution allows one to base inferences and decisions on all available information. Often one may have some information about a parameter prior to taking a sample. This information may be based on previous sample results but this is not always the case. Rather, it may reflect the beliefs of the experimenter concerning the true value of the parameter before any data is obtained, and in this case the prior distribution will be based on subjective probability. In this section we will use h(p) to represent the prior density function for the parameter p.

Consider a Bernoulli process with parameter p, generating $y = (y_1, y_2, \ldots, y_n)$ independent random variables and where the likelihood of the sample outcome is

$$l(p \mid y) = p^k (1-p)^{n-k} \quad \text{where } 0<p<1, \quad y_i = 0 \text{ or } 1, \quad i=1, 2, \ldots, n \quad \text{and } k=\sum_{i=1}^{n} y_i.$$

To choose a prior h(p), a class of distributions is needed which lie in the interval (0,1). One class of distributions with this property is the beta family.

The probability density function for the beta distribution is given by

$$h(p) = \frac{\Gamma(r+s)}{\Gamma(r)\,\Gamma(s)}\, p^{r-1} (1-p)^{s-1},$$

for $0<p<1$, where $r>0$ and $s>0$.

Using a beta prior with parameters r and s, denoted by $\beta(r,s)$, one combines this with the sample information to obtain the posterior density function for p ,

$$g(p \mid y) = \frac{\Gamma(n+r+s)}{\Gamma(k+r)\,\Gamma(n-k+s)}\, p^{k+r-1} (1-p)^{n-k+s-1},$$

for $0<p<1$.

This posterior distribution also happens to belong to the beta family. A family of priors such that the posterior distribution also belongs to that family is known as a conjugate family. The concept of conjugate priors will be discussed in a later section.

The beta family provides a wide variety of shapes which can be used to approximate many reasonable prior distributions. Suppose someone is interested in the proportion p, of people with a rare disease in a particular city. To get information, a sample of individuals from the city is taken. Some previous information about p in similar cities of the country or previous beliefs about p may be expressed in the form of a prior. It may be felt with certainty that p is near zero which can be obtained by taking r=0.5 and s=9.5 in the $\beta(r,s)$ distribution. If it is believed that p concentrates near a small number such as 0.05 r and s can be chosen so that the mean is 0.05 and the variance is small, such a value is $\beta(6,114)$. If p is of interest but no information or belief concerning p is available, p may be taken to be uniformly distributed over (0,1) which corresponds to the beta distribution with r=s=1. Graphs for the above mentioned beta distributions can be seen in Figure 3.1.

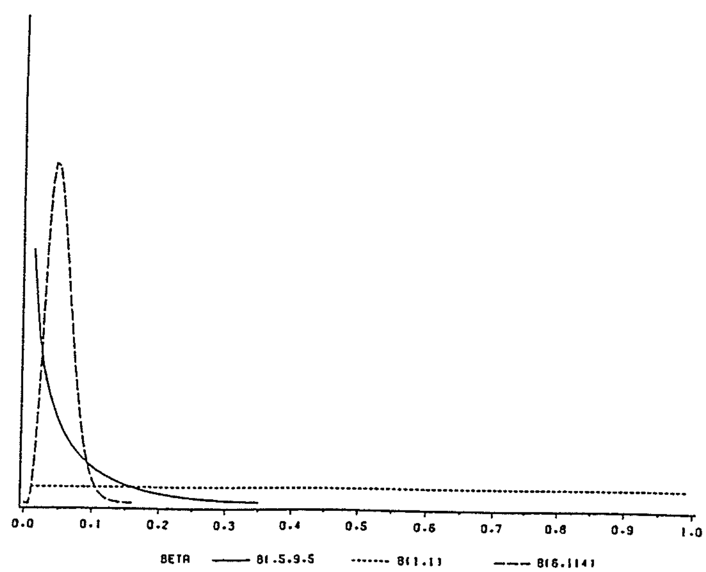

Figure 3.1: Examples of Beta Prior Density Functions

3.3 Posterior Distribution

The prior distribution is combined with the likelihood function through Bayes theorem to obtain the posterior distribution of the parameter(s) of interest. Given the observed data y, the conditional distribution of the parameter of interest p is given by

$$g(p \mid y) = \frac{l(p \mid y)\, h(p)}{\int l(p \mid y)\, h(p)\, dp} .$$

Here, $g(p \mid y)$ and $h(p)$ represent the posterior distribution and the prior distribution respectively, while $l(p \mid y)$ represents the likelihood function.

Given the data y, $l(p \mid y)$ is regarded as a function of p for fixed y. In terms of Bayes theorem the likelihood is the function through which the data modifies prior knowledge of p. The likelihood represents information about p coming from the data. The prior and posterior distributions are proper density functions, that is, both are nonnegative and integrate to one over the range of p.

3.4 Conjugate Priors

Due to the subjective nature of prior distributions there are many distributions which can be used to represent someone's prior beliefs. There are, however, some families of distributions which may be more desirable than others. The derivation of the posterior density function may be quite difficult to accomplish in practice, without the use of numerical methods. This is particularly true if both the prior, $h(p)$, and the likelihood, $l(p \mid y)$, do not have simple mathematical forms, making the integration in the denominator difficult.

In many problems there are families of prior distributions for which the determination of the posterior distribution is made computationally easier. Such a family is referred to as a conjugate family.

Three properties are desirable for conjugate families of distributions -- (1) analytic tractability, (2) richness, and (3) interpretability.

A prior distribution is analytically tractable if:

a) the posterior distribution is easy to determine given the prior distribution and the likelihood function

b) the prior is a member of a family then the posterior distribution is also a member of that family, and

c) expectations are easy to calculate.

The second property, richness, refers to the capability of expressing one's prior information and beliefs. Such a conjugate family of distributions should include distributions capable of different locations, dispersions and shapes. Finally, the third property, interpretability, refers to the ability to parameterize the conjugate family in a way that makes it easy to verify that a chosen prior agrees with the person's prior information.

Suppose x is a binomial random variable with likelihood $l(p \mid x)$, and the prior distribution is a beta distribution with parameters r and s, then

$$l(p \mid x) = \binom{n}{x} p^{x} (1-p)^{n-x} \quad , \quad x=0, 1, 2, \ldots, n,$$

and

$$h(p) = \frac{\Gamma(r+s)}{\Gamma(r)\,\Gamma(s)}\, p^{r-1}\,(1-p)^{s-1}, \text{ for } 0<p<1.$$

After applying Bayes theorem the posterior density function is given by

$$g(p \mid x) = \frac{\Gamma(n+r+s)}{\Gamma(r+x)\,\Gamma(n-x+s)}\, p^{x+r-1}\,(1-p)^{n-x+s-1}, \quad \text{for } 0<p<1,$$

which corresponds to a beta distribution with parameters x+r and n-x+s.

When sampling from a binomial process the beta family satisfies the properties of a conjugate family. Using the beta prior it is relatively easy to determine the posterior distribution given the sample results, and the posterior distribution belongs to the same family as the prior. Since the posterior distribution is a beta, the mean, variance and other moments may easily be determined. By varying the parameters of the beta distribution many shapes are possible, and therefore the property of richness is satisfied. The final property involving the ability to interpret the prior also applies in this example. The prior information can be interpreted as having roughly equivalent information as that contained in the sample from the binomial process.

The form of a conjugate family depends on the form of the likelihood function. In some cases no conjugate family may be found for the model, while in other cases it may not be possible to find a member of the conjugate family which satisfies all three properties.

3.5 Upper and Lower Bounds

Situations may arise where an upper or lower bound on a parameter is desired. For example, a market researcher may be concerned with the proportion, p, of a population that will buy a new product. The researcher would be interested in how bad things may be, that is, a lower bound for p.

The posterior distribution, g(p | y), may be used to construct lower and/or upper bounds for a parameter of interest. A 100(1-α) % upper credibility bound for a parameter, p, may be found by finding the value T_u such that

$$\int_0^{T_u} g(p \mid y)\, dp = 1-\alpha.$$

Similarly a 100(1-α) % lower credibility bound for a proportion, p, may be found by finding a value T_l such that

$$\int_0^{T_l} g(p \mid y)\, dp = \alpha.$$

A two-sided 100(1-α) % credibility interval involves the simultaneous specification of both a lower probability bound, T_l, and an upper probability bound T_u, such that the probability of being between T_l and T_u is 100(1-α) %. For example, a two-sided 95% credibility interval could be obtained by using the combination of a one-sided lower 97.5% credibility bound and a one-sided upper 97.5% credibility bound.

3.6 Sensitivity

Variation in the prior distribution will cause variation in the posterior distribution, thereby possibly causing totally different inferences to be drawn. Therefore, it is of interest to look at the effect on the posterior distribution by varying the prior distribution. If variations in the prior distribution have little or no effect on the posterior distribution, then we say that the posterior distribution is insensitive to changes in the prior distribution. However, if slight changes in the prior distribution have a large effect on the posterior distribution then it is said that the posterior distribution is sensitive to changes in the prior distribution. In situations where the posterior distribution is sensitive to the prior distribution it is vital that the experimenter adopt a prior that accurately represents his judgement.

3.7 Nuisance Parameters

In many problems, the underlying probability density function involves parameters in which there is no interest, these parameters are called nuisance parameters. In the Bayesian approach, inferences about the parameter of interest are based on the marginal posterior distribution of this parameter. To obtain the marginal posterior distribution of the parameter of interest we integrate the joint posterior distribution of the parameters over the nuisance parameters.

CHAPTER 4

THE SOLUTION

4.1 Introduction

Tenenbein (1970) used asymptotic methods in order to estimate the proportion of items of interest in a population where the items are subject to misclassification. Often the proportion of interest, p_o, may be quite small, and together with small sub-sample sizes, the use of asymptotic techniques based on normal theory are not always appropriate. The use of Bayesian methods allows one to deal with the small sub-sample sizes. It also allows one to deal with nuisance parameters and to incorporate any previous information that one may have available concerning the parameters of interest.

4.2 Sampling Distribution

From the sampling plan (see Figure 2.1) it can be seen that the random variables N_2, X_2 and X_1 are the only ones that may be observed. The probability distribution of these random variables is conditional on the values of p_o, p_1 and p_2.

Lemma 4.1:

$P(N_2=n_2, X_2=x_2, X_1=x_1 \mid p_o, p_1, p_2) =$

$$\sum_{y_1}\sum_{y_2}\left[\frac{n!}{(n-n_2-y_1)!\, y_1!\, (n_2-y_2)!\, y_2!}[(1-p_o)p_1]^{n-n_2-y_1}[p_o(1-p_2)]^{y_1} *\right.$$

$$\left.[(1-p_o)(1-p_1)]^{n_2-y_2}[p_op_2]^{y_2} * \frac{C(y_2,x_2)\, C(n_2-y_2,m_2-x_2)}{C(n_2,m_2)} * \frac{C(y_1,x_1)\, C(n-n_2-y_1,m_1-x_1)}{C(n-n_2,m_1)}\right],$$

(4.2.1)

for $x_1 = 0, 1, 2, \ldots, m_1$; $x_2 = 0, 1, 2, \ldots, m_2$;

$n_1 = m_1, m_1+1, \ldots, n$; $n_2 = m_2, m_2+1, \ldots, n$;

$y_1 = x_1, x_1+1, \ldots, n_1 - m_1 + x_1$; $y_2 = x_2, x_2+1, \ldots, n_2 - m_2 + x_2$;

$n_1 + n_2 = n$; $y_1 + y_2 = y$; and $0 \le p_o, p_1, p_2 \le 1$.

Proof:

It is assumed that there is an ongoing stream of items which can be considered an infinite population, that the probability of correctly classifying an item is independent of the classification of another item and that the probabilities of correctly classifying items, p_1 and p_2, are constant for all items.

Each item after classification from the initial sample can fall into one of four cells. Therefore the joint distribution of N_1-Y_1, Y_1, N_2-Y_2, and Y_2 is multinomial with observed values as shown below.

true classification

inspection classification	category 1	category 2	
category 1	$n_1 - y_1$	y_1	n_1
category 2	$n_2 - y_2$	y_2	n_2
	$n - y$	y	n

The associated cell probabilities are shown below.

true classification

inspection classification	category 1	category 2	
category 1	$(1-p_o)p_1$	$p_o(1-p_2)$	$p_1 + (1-p_1-p_2)p_o$
category 2	$(1-p_o)(1-p_1)$	$p_o p_2$	$(1-p_1) - (1-p_1-p_2)p_o$
	$1-p_o$	p_o	1

Therefore

$$P(N_2=n_2, N_1=n_1, Y_1=y_1, Y_2=y_2 \mid p_o, p_1, p_2) =$$

$$\frac{n!}{(n_1-y_1)!\, y_1!\, (n_2-y_2)!\, y_2!} [\,(1-p_o)p_1]^{n_1-y_1}[p_o(1-p_2)]^{y_1}[(1-p_o)(1-p_1)]^{n_2-y_2}[p_o p_2]^{y_2}, \tag{4.2.2}$$

where $n_2 = 0, 1, 2, \ldots, n$; $y_2 = 0, 1, 2, \ldots, n_2$; $n_1 + n_2 = n$.

The conditional distributions of the number of nonconforming items confirmed to be in the sub-samples by the true classification technique are hypergeometric. Thus,

$$P(X_2 = x_2 \mid N_2 = n_2, Y_2 = y_2) = \frac{C(y_2,x_2)\ C(n_2-y_2,m_2-x_2)}{C(n_2,m_2)},$$

(4.2.3)

and

$$P(X_1 = x_1 \mid N_1 = n_1, Y_1 = y_1) = \frac{C(y_1,x_1)\ C(n-n_2-y_1,m_1-x_1)}{C(n-n_2,m_1)}.$$

(4.2.4)

It follows that

$$\begin{aligned} &P(N_2=n_2, X_1=x_1, X_2=x_2, Y_1=y_1, Y_2=y_2 \mid p_o, p_1, p_2) = \\ &\qquad P(N_2=n_2, N_1=n_1, Y_1=y_1, Y_2=y_2 \mid p_o, p_1, p_2) * P(X_2=x_2 \mid N_2=n_2, Y_2=y_2) \\ &\qquad * \ P(X_1=x_1 \mid N_1=n-n_2, Y_1=y_1), \end{aligned}$$

where the expressions on the right hand side are given in Equations (4.2.2), (4.2.3) and (4.2.4). Since Y_2 and Y_1 are not observable the above probabilities are summed over y_1 and y_2 in order to obtain $P(N_2=n_2, X_2=x_2, X_1=x_1 \mid p_o, p_1, p_2)$.

4.3 Prior Distribution

Before the posterior distribution can be found, prior distributions need to be specified for p_o, p_1 and p_2. In this thesis the solution developed assumes that the parameters for p_o, p_1 and p_2 are independent. Beta priors were selected due to their ability to provide a wide variety of shapes which can be used to approximate many prior distributions, and because of their mathematical manageability. The assumption allows the joint distribution of p_o, p_1 and p_2 to be written as the product of three independent beta distributions.

Therefore

$$h\,(p_o, p_1, p_2) = h\,(p_o) * h\,(p_1) * h\,(p_2) \qquad \text{for} \quad 0 \leq p_o, p_1, p_2 \leq 1 \tag{4.3.1}$$

where

$$h\,(p_o) = \frac{\Gamma(a_o + b_o)}{\Gamma(a_o)\,\Gamma(b_o)}\, p_o^{a_o-1}\,(1 - p_o)^{b_o-1}$$

$$h\,(p_1) = \frac{\Gamma(a_1 + b_1)}{\Gamma(a_1)\,\Gamma(b_1)}\, p_1^{a_1-1}\,(1 - p_1)^{b_1-1}$$

and

$$h\,(p_2) = \frac{\Gamma(a_2 + b_2)}{\Gamma(a_2)\,\Gamma(b_2)}\, p_2^{a_2-1}\,(1 - p_2)^{b_2-1}$$

for constants $0 \leq a_o, b_o, a_1, b_1, a_2, b_2$.

Although the beta family is fairly rich, a wider variety of distributional shapes can be obtained by considering mixtures of beta distributions for the prior distributions of the parameters p_o, p_1 and p_2. These distributions may in fact be more realistic than single beta distributions. For example, items from a population may have come from several different locations, and therefore p_o may in fact vary because of location. The information from the various locations can be combined into a single prior distribution. The use of mixtures of beta distributions allows one the capability to have a number of modes in the prior distributions for p_o, p_1 and p_2. An example of a bimodal distribution is seen in Figure 4.1. Such mixtures are not considered in the thesis, but the work could easily be extended to handle such mixtures.

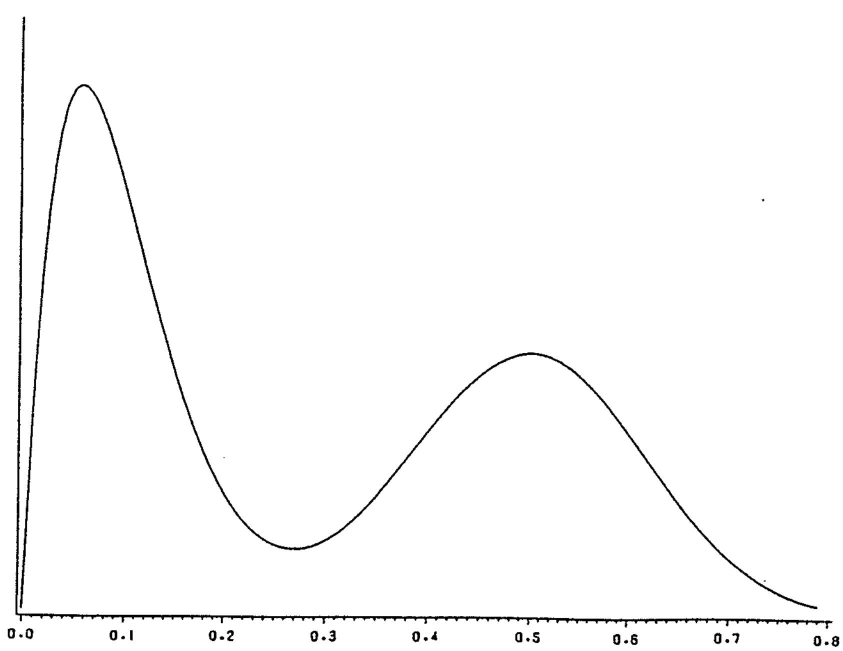

Figure 4.1: An Example of a Bimodal Prior Density Function -- A Mixture of Betas

4.4 Posterior Distribution

Lemma 4.2:

Using independent beta priors for p_o, p_1 and p_2, and having observed the values of N_2, X_2 and X_1, the posterior density function of p_o, $g(p_o \mid n_2, x_2, x_1)$ is

$$\frac{\displaystyle\sum_{y_1}\sum_{y_2}\left[\frac{(n-n_2-y_1+a_1-1)!\,(n_2-y_2+b_1-1)!\,(y_2+a_2-1)!\,(y_1+b_2-1)!\;\; p_o^{y_1+y_2+a_o-1}(1-p_o)^{n-y_1-y_2+b_o-1}}{(n-y_1-y_2+a_1+b_1-1)!\,(y_1+y_2+a_2+b_2-1)!\,(y_2-x_2)!\,(y_1-x_1)!\,(n_2-y_2-m_2+x_2)!\,(n-n_2-y_1-m_1+x_1)!}\right]}{\displaystyle\sum_{y_1}\sum_{y_2}\left[\frac{(y_1+y_2+a_o-1)!\,(n-y_1-y_2+b_o-1)!\,(n-n_2-y_1+a_1-1)!\,(n_2-y_2+b_1-1)!\,(y_2+a_2-1)!\,(y_1+b_2-1)!}{(n+a_o+b_o-1)!\,(n-y_1-y_2+a_1+b_1-1)!\,(y_1+y_2+a_2+b_2-1)!\,(n_2-y_2-m_2+x_2)!\,(n-n_2-y_1-m_1+x_1)!\,(y_2-x_2)!\,(y_1-x_1)!}\right]}, \tag{4.4.1}$$

for $0 \le p_o \le 1$;

where

$x_1 = 0, 1, 2, \dots, m_1$; $\quad x_2 = 0, 1, 2, \dots, m_2$;

$n_1 = m_1, m_1+1, \dots, n$; $\quad n_2 = m_2, m_2+1, \dots, n$;

$y_1 = x_1, x_1+1, \dots, n_1 - m_1 + x_1$; $\quad y_2 = x_2, x_2+1, \dots, n_2 - m_2 + x_2$.

Proof:

The joint posterior of p_o, p_1 and p_2 is

$$g(p_o, p_1, p_2 \mid n_2, x_2, x_1) = \frac{P(N_2=n_2, X_2=x_2, X_1=x_1 \mid p_o, p_1, p_2) * h(p_o,p_1,p_2)}{P(N_2=n_2, X_2=x_2, X_1=x_1)},$$

where $P(N_2=n_2, X_2=x_2, X_1=x_1 \mid p_o, p_1, p_2)$ is given by (4.2.1), $h(p_o,p_1,p_2)$ is given by (4.3.1), and where the normalizing constant is given by

$$P(N_2=n_2,X_2=x_2,X_1=x_1) = \int_0^1\int_0^1\int_0^1 P(N_2=n_2,X_2=x_2,X_1=x_1 \mid p_o,p_1,p_2) * h(p_o,p_1,p_2) \ dp_2 \ dp_1 \ dp_o$$

Therefore, it follows that the marginal posterior distribution for p_o is

$$g(p_o \mid n_2, x_2, x_1) = \int_0^1\int_0^1 g(p_o, p_1, p_2 \mid n_2, x_2, x_1) \ dp_2 \ dp_1$$

and is given by (4.4.1).

The cumulative posterior distribution function is given by

$$G(p_o \mid n_2, x_2, x_1) = \int_0^{p_o} g(p \mid n_2, x_2, x_1) \ dp =$$

$$\frac{\displaystyle\sum_{y_1}\sum_{y_2}\left[\frac{(n-n_2-y_1+a_1-1)!\,(n_2-y_2+b_1-1)!\,(y_2+a_2-1)!\,(y_1+b_2-1)!\displaystyle\int_0^{p_o} p^{y_1+y_2+a_o-1}\,(1-p)^{n-y_1-y_2+b_o-1}\,dp}{(n-y_1-y_2+a_1+b_1-1)!\,(y_1+y_2+a_2+b_2-1)!\,(y_2-x_2)!\,(y_1-x_1)!\,(n_2-y_2-m_2+x_2)!\,(n-n_2-y_1-m_1+x_1)!}\right]}{\displaystyle\sum_{y_1}\sum_{y_2}\left[\frac{(y_1+y_2+a_o-1)!\,(n-y_1-y_2+b_o-1)!\,(n-n_2-y_1+a_1-1)!\,(n_2-y_2+b_1-1)!\,(y_2+a_2-1)!\,(y_1+b_2-1)!}{(n+a_o+b_o-1)!\,(n-y_1-y_2+a_1+b_1-1)!\,(y_1+y_2+a_2+b_2-1)!\,(n_2-y_2-m_2+x_2)!\,(n-n_2-y_1-m_1+x_1)!\,(y_2-x_2)!\,(y_1-x_1)!}\right]}.$$

Examples of a posterior distribution and a cumulative posterior distribution can be seen in Figures 4.2 and 4.3.

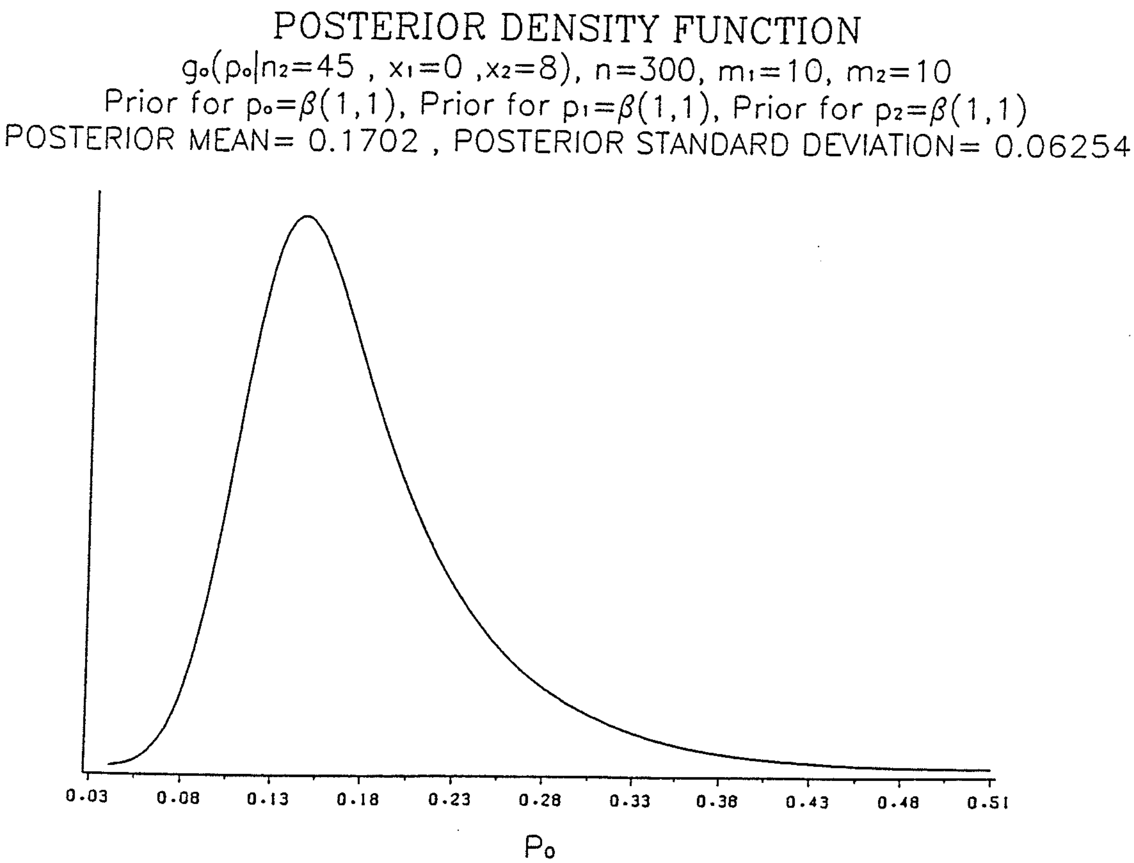

Figure 4.2: Posterior Density Function

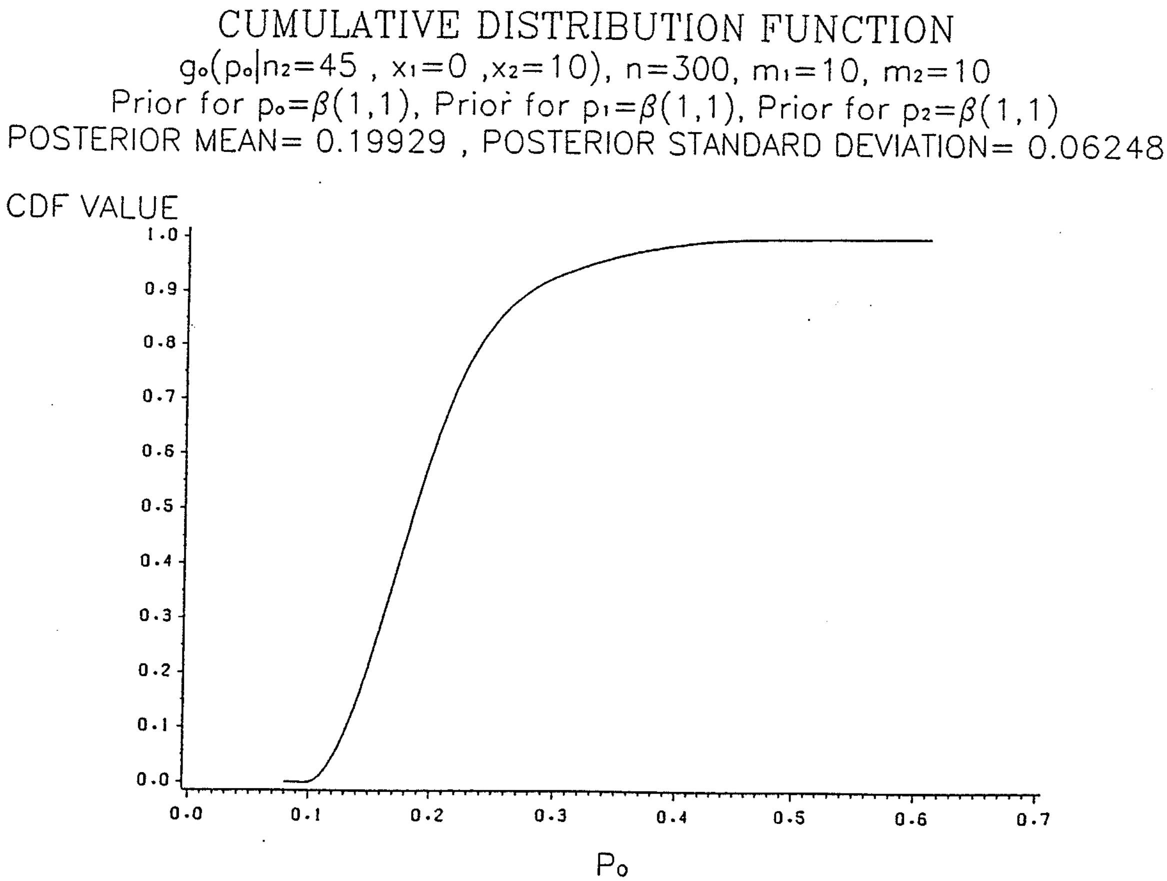

Figure 4.2: Cumulative Posterior Distribution Function for p_o

4.5 Credibility Bounds

Upper and lower credibility bounds for p_o may be found by inverting the cumulative posterior distribution. That is, an $100(1-\alpha)\%$ upper credibility bound is found by finding the value T_u such that

$$G(p_o \mid n_2, x_2, x_1) = \int_0^{T_u} g(p \mid n_2, x_2, x_1)\, dp = 1-\alpha$$

and an $100(1-\alpha)\%$ lower credibility bound is found by finding the value T_l such that

$$G(p_o \mid n_2, x_2, x_1) = \int_0^{T_l} g(p \mid n_2, x_2, x_1)\, dp = \alpha\ .$$

For example, 97.5 % upper and lower bounds for the posterior distribution of p_o, $g(p_o \mid n_2, x_2, x_1)$ with $n=300$, $n_2=45$, $m_1=m_2=10$, $x_2=10$, $x_1=x$ and with $\beta(1,1)$ priors for p_o, p_1 and p_2 can be found in Table 4.1.

4.6 Sensitivity Analysis

It is of interest to look at the effect of varying the prior distribution on the posterior for p_o. For the parameters p_1 and p_2, three priors are considered, $\beta(1,1)$ -- representing ignorance, $\beta(20,1)$ -- representing quite good inspector accuracy, and $\beta(100,1)$ -- representing nearly perfect inspector accuracy.

x_1	x_2	97.5% lower bound	97.5% upper bound
0	10	0.1161	0.3612
1	10	0.1425	0.4774
2	10	0.1763	0.5768
3	10	0.2190	0.6647
4	10	0.2711	0.7430
5	10	0.3327	0.8130
6	10	0.4040	0.8734
7	10	0.4860	0.9243
8	10	0.5808	0.9635
9	10	0.6940	0.9888
10	10	0.8378	0.9992

Table 4.1: 97.5% upper and lower bounds for $g(p_o \mid n_2, x_2, x_1)$ with $n=300$, $n_2=45$, $m_1=m_2=10$, $x_2=10$, $x_1=x$ and with $\beta(1,1)$ priors for p_o, p_1 and p_2 .

For the parameter p_o, three priors are also considered, $\beta(1,1)$: -- representing ignorance, $\beta(1.5,5)$ -- representing a unimodal prior density function with a mean at 0.23, and $\beta(1,10)$ -- representing the feeling that the proportion of nonconforming items is near zero.

Varying the prior distribution for p_2 does have an effect on both tail regions of the posterior distribution for p_o (Figures 4.4, 4.5 and 4.6). This effect seems to be slightly more pronounced in the upper tail region.

As the prior for p_1 is varied, we see only a slight effect on the posterior distribution for p_o in the lower tail region and very little effect in the upper tail region (Figures 4.6, 4.7 and 4.8).

Varying the prior for p_o also results in some change to the posterior distribution for p_o. This effect can be seen in both tails of the posterior distribution, but seems be larger in the upper tail region (Figures 4.9, 4.10 and 4.11).

As the number confirmed to be nonconforming decreases from x_2= 10 to x_2= 8 in the sub-sample there is some increase in the effect of varying the prior p_1 on the posterior distribution. This effect is not as noticeable when the priors for p_o and p_2 are varied.

In the sensitivity analysis that follows only selected examples can be presented as there are a number of quantities that can be varied. For selected samples we will consider the effect of varying the priors for p_o, p_1 and p_2 on the posterior for p_o, and at the same time consider some variation in the proportion of items found to be misclassified.

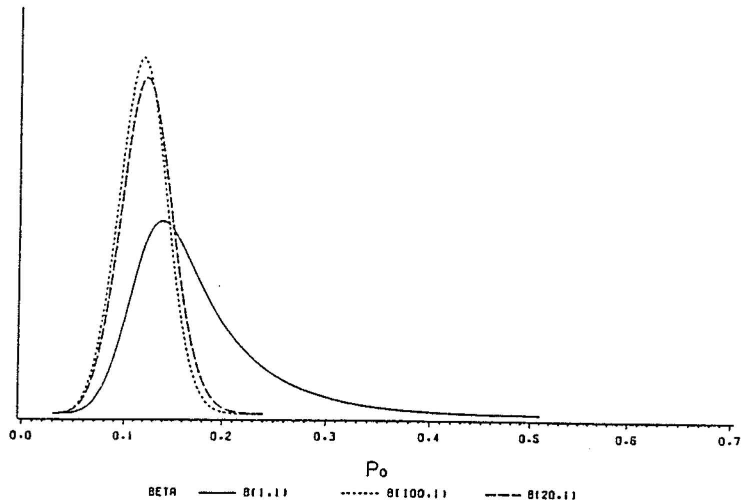

Figure 4.4: Effects of varying the prior for p_2 on the posterior for p_o with some confirmed misclassification, and where *a priori* p_o and p_1 are uncertain.
$g(p_o \mid n_2{=}45, x_1{=}0, x_2{=}8)$, $n{=}300$, $m_1{=}10$ and $m_2{=}10$
Prior for $p_o{=}\beta(1,1)$ and Prior for $p_1{=}\beta(1,1)$.

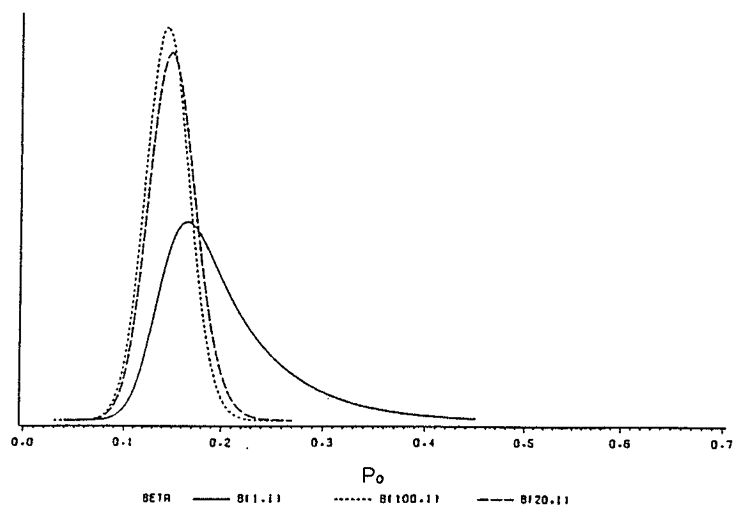

Figure 4.5: Effects of varying the prior for p_2 on the posterior for p_o with no confirmed misclassification, and where *a priori* p_o and p_1 are uncertain.
$g(p_o \mid n_2{=}45, x_1{=}0, x_2{=}10)$, $n{=}300$, $m_1{=}10$ and $m_2{=}10$
Prior for $p_o{=}\beta(1,1)$ and Prior for $p_1{=}\beta(1,1)$.

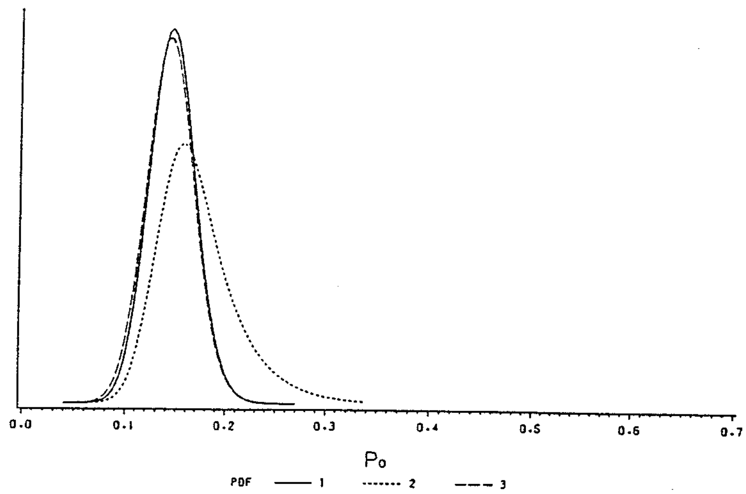

Figure 4.6: Effects of varying the priors for p_2 and p_1 on the posterior for p_o with no confirmed misclassification, and where *a priori* p_o is near zero.

$g(p_o \mid n_2=45, x_1=0, x_2=10)$, $n=300$, $m_1=10$ and $m_2=10$

pdf 1: Priors; $p_o= \beta(1,10)$, $p_1= \beta(20,1)$ and $p_2=\beta(20,1)$

pdf 2: Priors; $p_o= \beta(1,10)$, $p_1= \beta(20,1)$ and $p_2=\beta(1,1)$

pdf 3: Priors; $p_o= \beta(1,10)$, $p_1= \beta(1,1)$ and $p_2=\beta(20,1)$

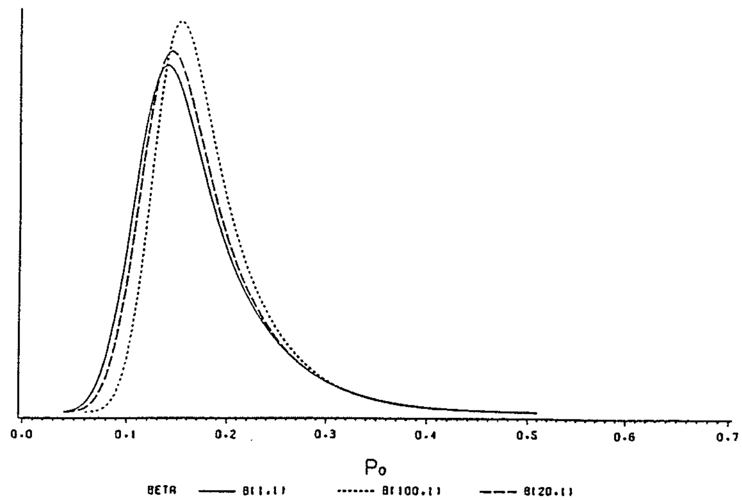

Figure 4.7: Effects of varying the prior for p_1 on the posterior for p_o with some confirmed misclassification, and where *a priori* p_o and p_2 are uncertain.
$g(p_o \mid n_2=45, x_1=0, x_2=8)$, $n=300$, $m_1=10$ and $m_2=10$
Prior for $p_o=\beta(1,1)$ and Prior for $p_2=\beta(1,1)$.

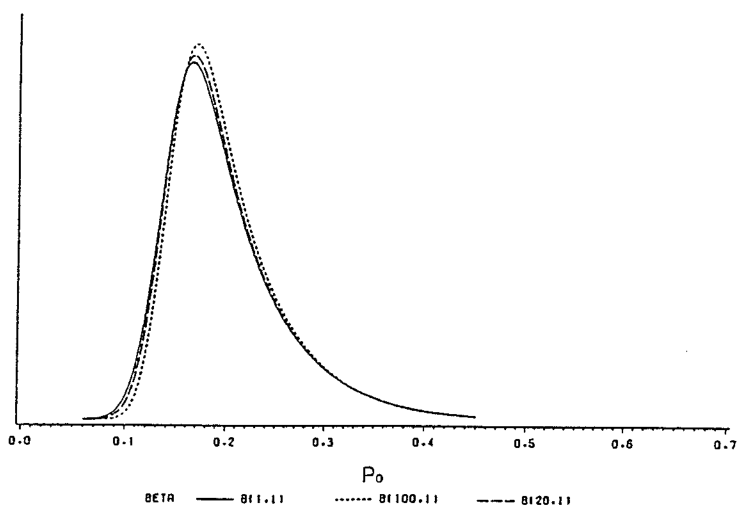

Figure 4.8: Effects of varying the prior for p_1 on the posterior for p_o with no confirmed misclassification, and where *a priori* p_o and p_2 are uncertain.
$g(p_o \mid n_2=45, x_1=0, x_2=10)$, $n=300$, $m_1=10$ and $m_2=10$
Prior for $p_o=\beta(1,1)$ and Prior for $p_2=\beta(1,1)$.

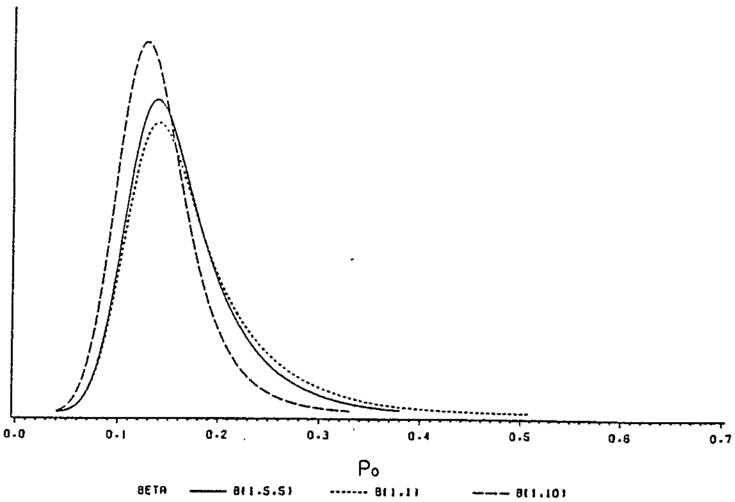

Figure 4.9: Effects of varying the prior for p_o on the posterior for p_o with some confirmed misclassification, and where *a priori* p_1 and p_2 are uncertain.
$g(p_o \mid n_2=45, x_1=0, x_2=8)$, $n=300$, $m_1=10$ and $m_2=10$
Prior for $p_1=\beta(1,1)$ and Prior for $p_2=\beta(1,1)$.

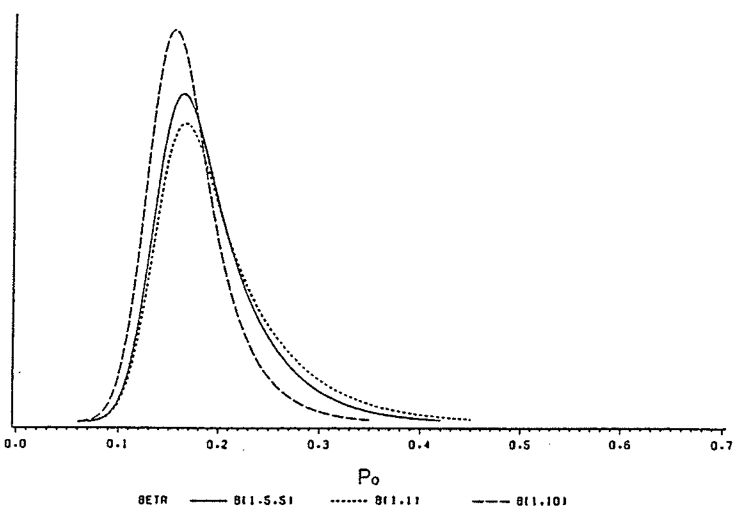

Figure 4.10: Effects of varying the prior for p_o on the posterior for p_o with no confirmed misclassification, and where *a priori* p_1 and p_2 are uncertain.
$g(p_o \mid n_2=45, x_1=0, x_2=10)$, $n=300$, $m_1=10$ and $m_2=10$
Prior for $p_1=\beta(1,1)$ and Prior for $p_2=\beta(1,1)$.

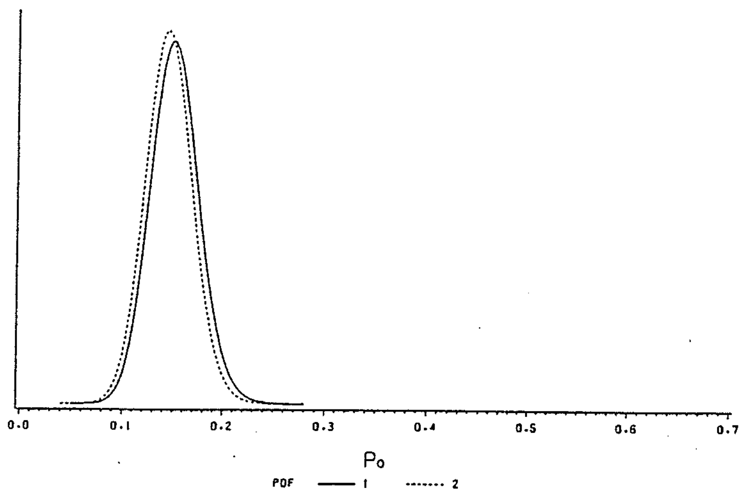

Figure 4.11: Effects of varying the prior for p_o on the posterior for p_o with no confirmed misclassification, and where *a priori* p_1 and p_2 are near one.

$g(p_o \mid n_2=45, x_1=0, x_2=10)$, $n=300$, $m_1=10$ and $m_2=10$
pdf 1: Priors; $p_o= \beta(1,1)$, $p_1= \beta(20,1)$ and $p_2=\beta(20,1)$
pdf 2: Priors; $p_o= \beta(1,10)$, $p_1= \beta(20,1)$ and $p_2=\beta(20,1)$

4.7 Comparison of One-sided and Two-sided Sampling Plans

Neden (1986) uses a one-sided "confirmatory" sub-sampling plan, that is, at the second stage of sampling a sub-sample is taken from either the "nonconforming" group or the "conforming" group. The decision to draw a sample from the "nonconforming" or the "conforming" group depends on the size of the "nonconforming" group at the initial inspection stage. If the "nonconforming" group is large at the initial sampling stage, the inspector may consider the lot to be unacceptable. In order to confirm that the lot does contain an unacceptable number of nonconforming items, a sub-sample from the "nonconforming" group can be taken in order to obtain a lower bound for p_o. Neden assumed that $p_2=1$, and because of this assumption no sub-sample was taken from the group classified as conforming. This assumption provides a conservative result as any misclassification of the nonconforming items would cause the lower bound to become larger.

Similarly, if the "nonconforming" group is small at the initial sampling stage, the inspector may make the decision that the lot is acceptable. One may want to confirm that the lot is good by taking a sub-sample from the "conforming" group in order to obtain a upper bound for p_o. By assuming that no misclassification of conforming items can take place, no sub-sample is taken from the "nonconforming" group. This would again provide a conservative result as any misclassification of conforming items would cause the estimated value of p_o and the upper bound to become smaller.

Due to time and cost, it may only be possible to sample a fixed number of items. If a point estimate for the proportion of nonconforming items was desired, samples would need to be taken from both groups. However, if a point estimate is not desired but one is interested in a lower bound, there are several ways in which samples may be allocated. All samples could be taken from the "nonconforming" group (one-sided confirmatory sub-sample), or samples could be taken from both of the "conforming" and "nonconforming" groups (two-sided sub-sample).

It is of interest to compare the two-sided sampling scheme to the one-sided confirmatory sampling scheme used by Neden (1986). From Figures 4.12 and 4.13 we see that when sub-samples of size 10 are taken from the "conforming" and "nonconforming" groups with ignorance priors representing the parameters, there is a large difference in the posterior distribution compared to the one-sided sampling plan where a sub-sample of size 10 is taken only from the "nonconforming" group.

When the prior for p_2 is taken to represent almost perfect inspection in the two-sided plan and when $x_1 = 0$, there is little difference in the lower region for both sampling plans. If the number of items found to be nonconforming, but initially classified as conforming, increases from $x_1 = 0$ to $x_1 = 1$, a difference in the lower region for both sampling plans can be seen (Figures 4.14 and 4.15).

It is also of interest to compare the two-sided sampling plan with sub-samples each of size 10, with the one-sided "confirmatory" sampling plan with a sub-sample of size 10 when an increase in the "nonconforming" group occurs. We see that when $n_2 = 45$ there is little difference in the lower region of the posteriors, but when n_2 is increased to 100 the difference increases slightly (Figures 4.16, 4.17, 4.18 and 4.19).

In all the cases observed above, when the sub-sample size for the one-sided sampling plan is increased to 20, there is a noticeable change in the lower region (Figures 4.15, 4.16, 4.17, 4.18 and 4.19).

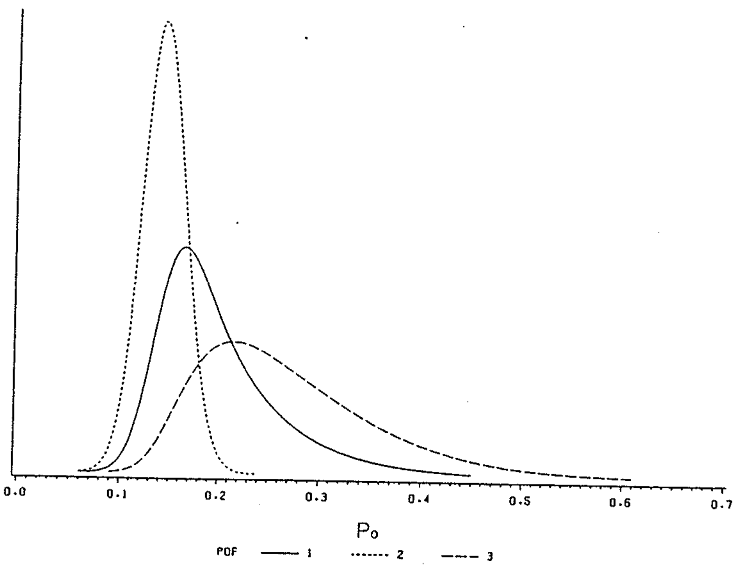

Figure 4.12: Effect of two-sided sampling versus one-sided confirmatory sampling on the posterior for p_o with some confirmed misclassification, and where *a priori* p_o, p_1 and p_2 are uncertain.

pdf 1: $g(p_o \mid n_2=45, x_1=0, x_2=10)$, $n=300$, $m_1=10$, $m_2=10$, $p_o = p_1 = p_2 = \beta(1,1)$

pdf 2: $g(p_o \mid n_2=45, x=10)$, $n=300$, $m=10$, $p_o = p_1 = \beta(1,1)$ and $p_2 = 1$

pdf 3: $g(p_o \mid n_2=45, x_1=1, x_2=10)$, $n=300$, $m_1=10$, $m_2=10$, $p_o = p_1 = p_2 = \beta(1,1)$

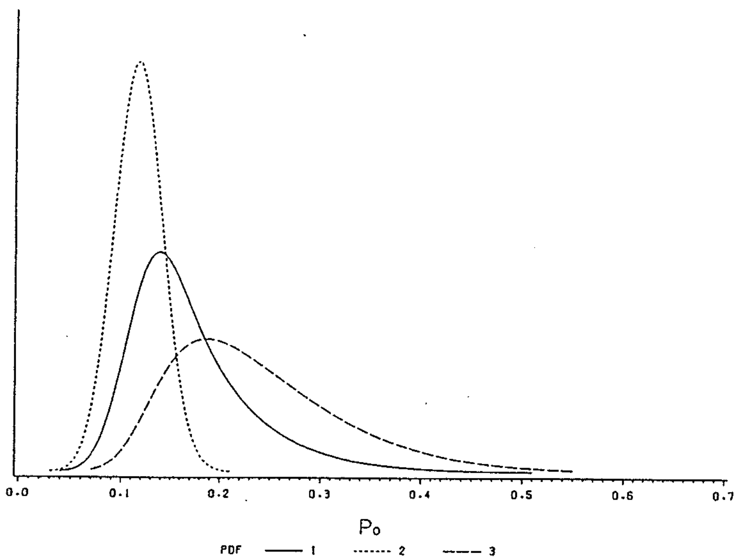

Figure 4.13: Effect of two-sided sampling versus one-sided confirmatory sampling on the posterior for p_o with some confirmed misclassification, and where *a priori* p_o, p_1 and p_2 are uncertain.

pdf 1: $g(p_o \mid n_2=45, x_1=0, x_2=8)$, $n=300$, $m_1=10$, $m_2=10$, $p_o= p_1= p_2= \beta(1,1)$

pdf 2: $g(p_o \mid n_2=45, x=8)$, $n=300$, $m=10$, $p_o= p_1= \beta(1,1)$ and $p_2= 1$

pdf 3: $g(p_o \mid n_2=45, x_1=1, x_2=8)$, $n=300$, $m_1=10$, $m_2=10$, $p_o= p_1= p_2= \beta(1,1)$

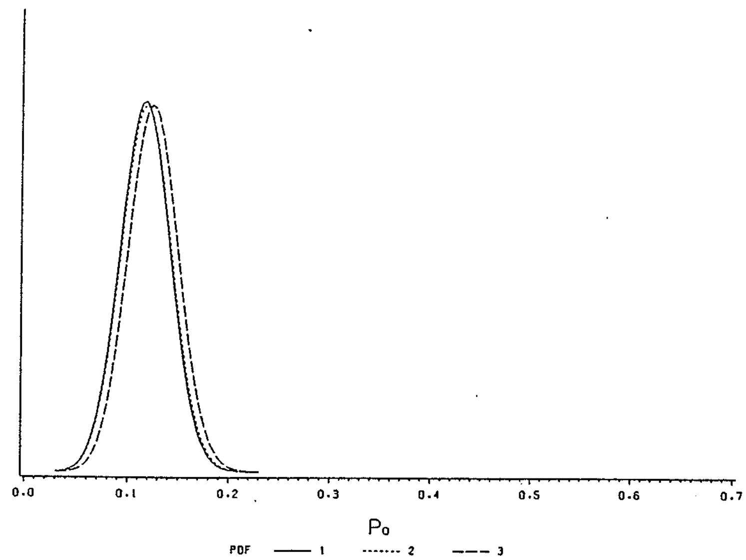

Figure 4.14: Effect of two-sided sampling versus one-sided confirmatory sampling on the posterior for p_o with some confirmed misclassification, and where *a priori* p_o and p_1 are uncertain, and p_2 is near one.

pdf 1: $g(p_o \mid n_2=45, x=8)$, $n=300$, $m=10$, $p_o= p_1= \beta(1,1)$ and $p_2= 1$

pdf 2: $g(p_o \mid n_2=45, x_1=0, x_2=8)$, $n=300$, $m_1=10$, $m_2=10$, $p_o= p_1= \beta(1,1)$, and $p_2= \beta(100,1)$

pdf 3: $g(p_o \mid n_2=45, x_1=1, x_2=8)$, $n=300$, $m_1=10$, $m_2=10$, $p_o= p_1= \beta(1,1)$, and $p_2= \beta(100,1)$

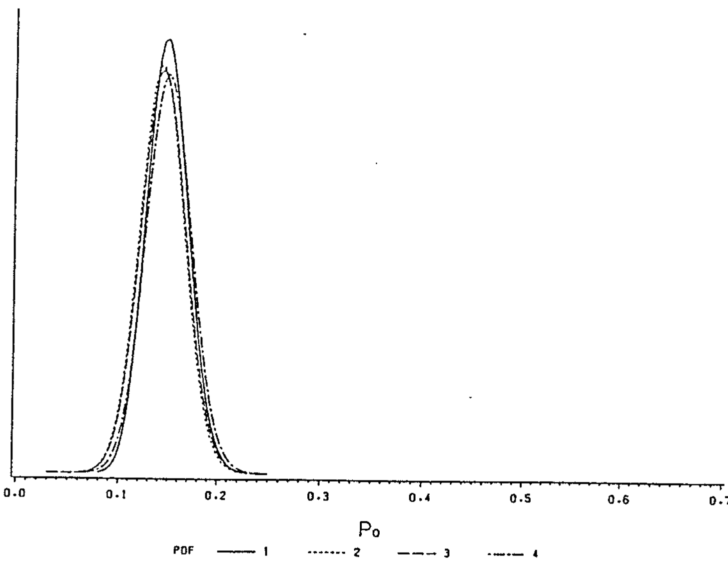

Figure 4.15: Effect of two-sided sampling versus one-sided confirmatory sampling on the posterior for p_o with some confirmed misclassification, and where *a priori* p_o and p_1 are uncertain, and p_2 is near one.

pdf 1: $g(p_o \mid n_2=45, x=20)$, $n=300$, $m=20$, $p_o= p_1= \beta(1,1)$ and $p_2= 1$

pdf 2: $g(p_o \mid n_2=45, x=10)$, $n=300$, $m=10$, $p_o= p_1= \beta(1,1)$ and $p_2= 1$

pdf 3: $g(p_o \mid n_2=45, x_1=0, x_2=10)$, $n=300$, $m_1=10$, $m_2=10$, $p_o= p_1= \beta(1,1)$, and $p_2= \beta(100,1)$

pdf 4: $g(p_o \mid n_2=45, x_1=1, x_2=10)$, $n=300$, $m_1=10$, $m_2=10$, $p_o= p_1= \beta(1,1)$, and $p_2= \beta(100,1)$

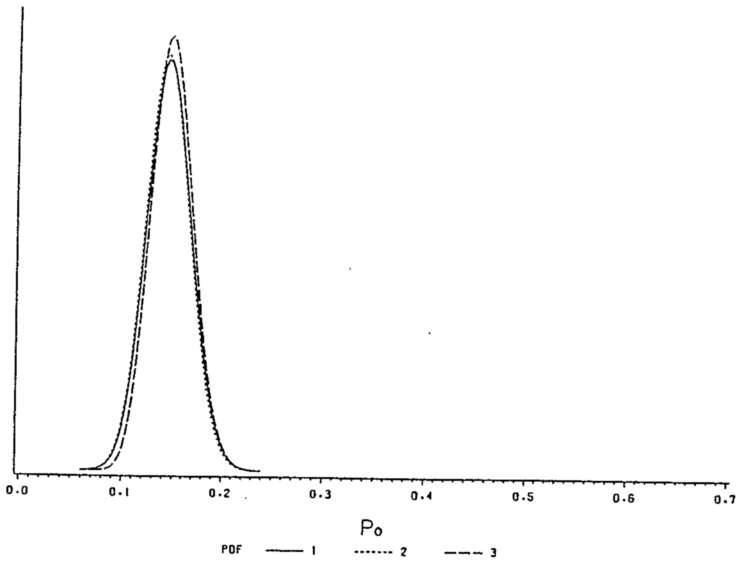

Figure 4.16: Effect of two-sided sampling versus one-sided confirmatory sampling on the posterior for p_o with no confirmed misclassification, and where *a priori* p_o is uncertain, p_1 near one, and p_2 near one.

pdf 1: $g(p_o \mid n_2{=}45, x_1{=}0, x_2{=}10)$, $n{=}300$, $m_1{=}10$, $m_2{=}10$, $p_o{=}\beta(1,1)$, $p_1{=}\beta(20,1)$ and $p_2{=}\beta(100,1)$

pdf 2: $g(p_o \mid n_2{=}45, x{=}10)$, $n{=}300$, $m{=}10$, $p_o{=}\beta(1,1)$, $p_1{=}\beta(20,1)$ and $p_2{=}1$

pdf 3: $g(p_o \mid n_2{=}45, x{=}20)$, $n{=}300$, $m{=}20$, $p_o{=}\beta(1,1)$, $p_1{=}\beta(20,1)$ and $p_2{=}1$

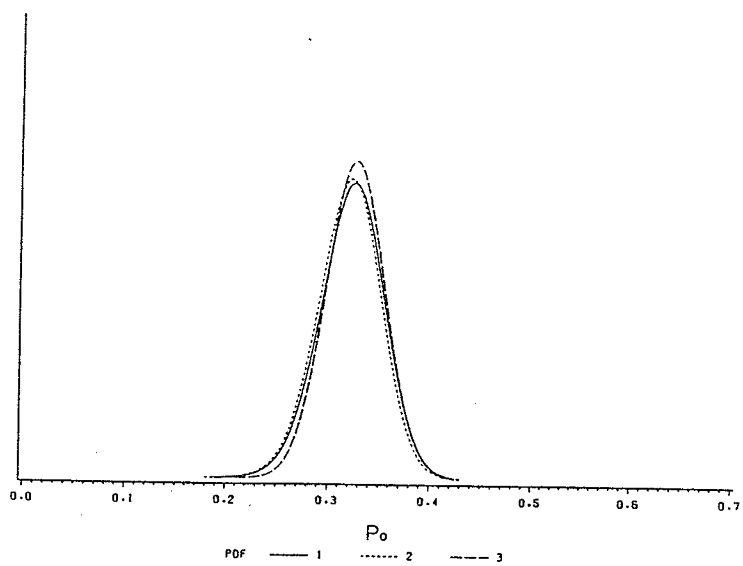

Figure 4.17: Effect of two-sided sampling versus one-sided confirmatory sampling on the posterior for p_o with no confirmed misclassification, and where *a priori* p_o is uncertain, p_1 near one, and p_2 near one.

pdf 1: $g(p_o \mid n_2=100, x_1=0, x_2=10)$, $n=300$, $m_1=10$, $m_2=10$, $p_o= \beta(1,1)$, $p_1= \beta(20,1)$ and $p_2= \beta(100,1)$

pdf 2: $g(p_o \mid n_2=100, x=10)$, $n=300$, $m=10$, $p_o= \beta(1,1)$, $p_1= \beta(20,1)$ and $p_2= 1$

pdf 3: $g(p_o \mid n_2=100, x=20)$, $n=300$, $m=20$, $p_o= \beta(1,1)$, $p_1= \beta(20,1)$ and $p_2= 1$

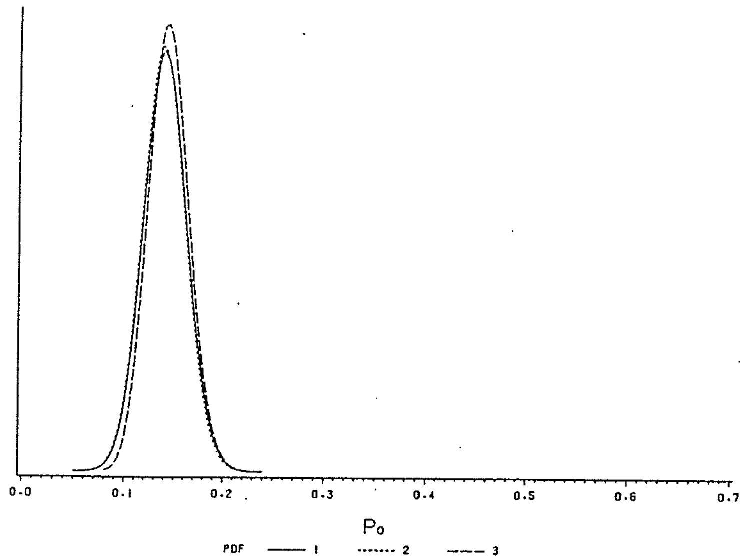

Figure 4.18: Effect of two-sided sampling versus one-sided confirmatory sampling on the posterior for p_o with no confirmed misclassification, and where *a priori* p_o is near zero, p_1 near one, and p_2 near one.

pdf 1: $g(p_o \mid n_2=45, x_1=0, x_2=10)$, $n=300$, $m_1=10$, $m_2=10$, $p_o= \beta(1,10)$, $p_1= \beta(20,1)$ and $p_2= \beta(100,1)$

pdf 2: $g(p_o \mid n_2=45, x=10)$, $n=300$, $m=10$, $p_o= \beta(1,10)$, $p_1= \beta(20,1)$ and $p_2= 1$

pdf 3: $g(p_o \mid n_2=45, x=20)$, $n=300$, $m=20$, $p_o= \beta(1,10)$, $p_1= \beta(20,1)$ and $p_2= 1$

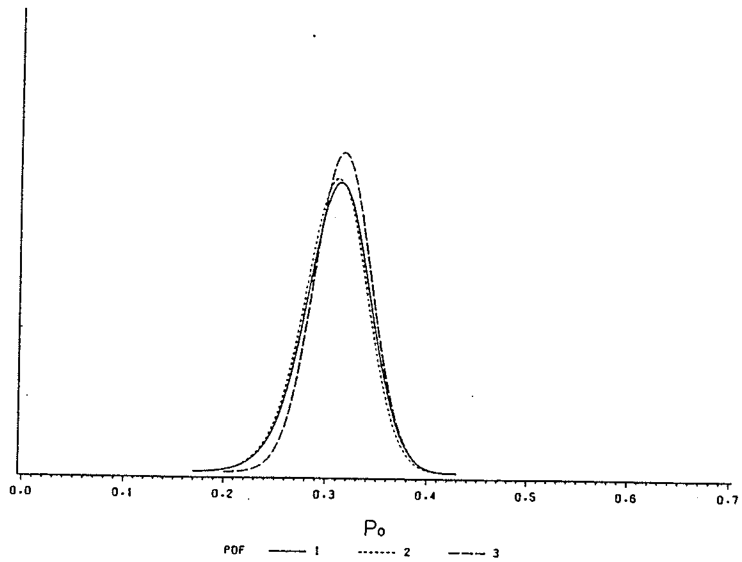

Figure 4.19: Effect of two-sided sampling versus one-sided confirmatory sampling on the posterior for p_o with no confirmed misclassification, and where *a priori* p_o is near zero, p_1 near one, and p_2 near one.

pdf 1: $g(p_o \mid n_2=100, x_1=0, x_2=10)$, $n=300$, $m_1=10$, $m_2=10$, $p_o=\beta(1,10)$, $p_1=\beta(20,1)$ and $p_2=\beta(100,1)$

pdf 2: $g(p_o \mid n_2=100, x=10)$, $n=300$, $m=10$, $p_o=\beta(1,10)$, $p_1=\beta(20,1)$ and $p_2=1$

pdf 3: $g(p_o \mid n_2=100, x=20)$, $n=300$, $m=20$, $p_o=\beta(1,10)$, $p_1=\beta(20,1)$ and $p_2=1$

4.8 Concluding Remarks

The posterior distribution developed in this thesis illustrates how a double sampling scheme and prior information may be combined in order to obtain a distribution upon which inferences about the proportion of interest may be based.

The one-sided "confirmatory" sampling plan of Neden and the two-sided sampling plan developed in this thesis were compared under various situations. It was shown that there are situations in which one sampling plan may be more appropriate than the other, depending on the form of the inference to be drawn and on the prior information concerning the misclassification rates.

Neden dealt with a one-sided "confirmatory" sub-sample, that is, at the second stage of sampling, a sub-sample was taken from either the "nonconforming" group or the "conforming" group -- the decision whether to draw a sample from the "nonconforming" or the "conforming" group depending on the size of the "nonconforming" group at the initial inspection stage. If the proportion of nonconforming items, p_o, in the lot was thought to be unacceptable at the initial sampling stage, a lower bound was desired by Neden in order to confirm that p_o was large. Neden assumed that no misclassification of nonconforming items could take place and therefore no sub-sample was taken from the "conforming" category. This assumption provides a conservative result, as any misclassification of the conforming items causes the estimated value of p_o and the lower bound to become larger. For example, if at the inspection stage a large number of nonconforming items are found, the inspector may conclude that the lot is unacceptable. One may then wish to confirm that the lot does in fact contain an unacceptable number of nonconforming items. If this is the case, a sub-sample from the "nonconforming" group can be taken in order to obtain a lower bound for the proportion of nonconforming items.

Similarly, if the "nonconforming" group is small at the initial sampling stage, the inspector may make the decision that the lot is acceptable. One may then want to confirm that the lot is good by taking a sub-sample from the "conforming" group in order to obtain a upper bound for p_o. By assuming that no misclassification of conforming items can take place, no sub-sample need be taken from the "nonconforming" category. This again provides a conservative result as any misclassification of conforming items would cause the estimated value of p_o and the upper bound to become smaller.

If some point estimate or two-sided interval estimate for the proportion of nonconforming items is desired, sub-samples would have to be taken from both of the visually inspected groups in order to see the effect of misclassification in both directions.

If a point estimate is not required but one is interested only in a lower bound, then there are several ways in which samples may be allocated. All samples could be taken from the "nonconforming" group (one-sided confirmatory sub-sample), or samples could be taken from both the "conforming" and "nonconforming" groups (two-sided sub-sample). For example, the advantage and disadvantage of (i) taking sub-samples of size 10 from each group, (ii) taking a sub-sample of size 10 only from the "nonconforming" group, and (iii) taking a sub-sample of size 20 only from the "nonconforming" group were examined in this thesis.

If the distribution of p_2 is represented by an ignorance prior then (i) is preferred to (ii). When sub-samples of size 10 are taken from the "conforming" and "nonconforming" groups, and compared to the one-sided sampling plan where a sub-sample of size 10 is taken only from the "nonconforming" group, a large difference in the lower regions of the posterior distributions is seen. In the situation where the prior distribution concerning p_2 is represented by ignorance, it would seem more appropriate to use the two-sided sampling plan in order to assess the effect of misclassification in both directions.

If *a priori* p_2 is near one then (i) and (ii) are approximately the same. It is hoped that the probability of correctly classifying a nonconforming item would be near 1, and, if prior distributions are used to reflect this information, little difference in the lower regions for the one-sided and two-sided sampling plans can be seen in many situations. However, as the size of nonconforming items in the initial sample increases, the difference becomes more pronounced. Moreover, if *a priori* p_2 is near one, then (iii) is better than (ii).

If only an upper bound is required then analogous conclusions can be drawn, but with p_1 replacing p_2.

If a one-sided "confirmatory" sampling plan is used on an ongoing basis, samples should still be taken occasionally from each group in order to assess the misclassification rates in both directions.

BIBLIOGRAPHY

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.

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.

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, 52, 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.

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

MUSTAFI, C. K. (1977), "On A Technique of Studying Misclassified Data," Sankhya, 39, Series B, 57-64.

NEDEN, LINDA R. (1986), "Bayesian One-sided Credibility Bounds for a Proportion using a Two-stage Sampling Plan Involving Imperfect and Perfect Classification," unpublished M.Sc. thesis, University of Manitoba, Department of Statistics.

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

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 for Estimating From Binomial Data with Misclassifications: Sample Size Determination," Biometrics, 27, 935-944.

WINKLER, ROBERT L. (1968), "The Consensus of Subjective Probability Distributions," Management Science, Series B, 15, 61-75.