Bureau of Business Research
Graduate School of Business Administration
University of Michigan

October 1972

BAYESIAN METHODS FOR ANALYZING NONRESPONSES WHEN ESTIMATING BINOMIAL PROPORTIONS

Working Paper No. 66

by

William J. Wrobleski
Associate Professor of Statistics
Graduate School of Business Administration
University of Michigan

ABSTRACT

In this paper, Bayesian allocation procedures for handling the problem of nonresponses in surveys with fixed sample size are examined for the purpose of estimating the true proportion to be found in one of two underlying population categories. These Bayesian estimation methods use certain families of conditional, conjugate-type prior density functions. In addition, explicit use is made of an allocation parameter which is associated with assigning the nonresponses observed in the sample to the underlying population categories whose (unknown) category proportions are being estimated.

CONTENTS

Introduction

Frequently in fixed-sample size market research surveys and other sample survey studies, the sampling data actually observed yield nonresponses such as refusals to answer, not at home, unrevealed preferences, and so on. In most large probability samples taken on a nationwide basis, sample sizes range from 1,000 to 3,000 respondents. For fixed sample size surveys within this range, the rate of nonresponse often turns out to be a significant fraction of the original fixed sample size chosen. Serious bias may result when statistical analyses of the observed data are made treating the reduced, random sample size actually observed as the fixed sample size originally chosen and the nonresponses are ignored.

In this paper Bayesian allocation procedures for handling the problem of nonresponses in fixed sample size surveys are examined for the purpose of estimating one of two underlying population proportions. These Bayesian estimation methods use certain families of conditional, natural, conjugate-type prior density functions and introduce explicitly an allocation parameter which is associated with assigning the nonresponses observed in the sample to the underlying population categories whose (unknown) category proportions are being estimated.

An illustration

Suppose one is interested in estimating the proportion of households whose annual income is $10,000 or more using data obtained from a simple random sample taken from a given target population, say, the population of all U. S. households. From this population a random sample of n households would be chosen, and each household so selected would be queried about whether or not its annual income was $10,000 or more. Thus <u>prior</u> to sampling the population can be represented by a two-category Bernoulli model in which one parameter represents the true (unknown) proportion of households in the U. S. population having an annual income of at least \$10,000, say, p, while the paramater q=1-p denotes the true (unknown) proportion of households having an annual income of less than \$10,000. Schematically the following representation is appropriate:

POPULATION CATEGORIES	
Category K_1	Category K_2
Income \$10,000 or more	Income less than \$10,000
p	q=1-p

To simplify the discussion, suppose that in sampling only one new category will occur, say, an "unrevealed" income category. Thus the process of obtaining information about household income through sampling can be expected to yield a certain (random) number of nonresponses, and the sampling process can be schematically represented as follows:

Sampling Process	Population Categories		Sampling Category
	Category K_1	Category K_2	Category C
	Income \$10,000 or more	Income less than \$10,000	Unrevealed Income
Parameters	p	q=1-p	p_o
Data	m_1	m_2	m_o

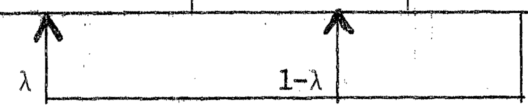
In other words, for a random sample of size n, m_1 households indicate they belong to the population category K_1 (that is, the households having an annual income of \$10,000 or more), and m_2 households give category K_2 as their income classification (annual income less than \$10,000), while m_o households do not reveal their annual income classification for one reason or another (for example, not at home, refusal to answer, do not know, and so on).

There is a simple probability model for describing the observed sample data. Denote by p_o the probability that a household randomly selected will not reveal its income classification. Besides this sampling parameter, p_o, and the population parameter, p, one additional parameter is required to specify the sampling model for the observed data.

Each household selected in the sample actually belongs to one of the two population categories, K_1 or K_2. However, m_o households selected in the sample for some reason do not reveal this information. Thus, it is natural to associate with each household which belongs to the sampling category called "unrevealed" income the conditional probability that such a household actually belongs to the population category K_1, as well as the conditional probability that the given household belongs to the second population category K_2. Denote the former conditional probability by λ. Then, if λ is assumed to be constant from household to household, λ is simply the conditional probability that a randomly selected household belongs to the population income category K_1, given that the household is classified as a nonresponse and did not reveal its income class when interviewed. Similarly, $1-\lambda$ denotes the conditional probability that such a household actually belongs to the second population income category K_2. Thus, for example, if λ were known to be 3/4, then in a sample of 1,000 households selected at random, if 180 did not reveal their income classification, it would be expected that 135 of these 180 households would have annual incomes of $10,000 or more and belong to the population income category K_1, while 45 of these households would be expected to have annual incomes below $10,000 and belong to the second population income category K_2.

With these definitions in mind for the probabilities p, q, p_o, and λ, the probability that a household selected at random reveals its income classification as K_1 is simply $p-\lambda p_o$, while the probability that a household selected at random reveals its income classification as K_2 is $q-(1-\lambda)p_o$. Schematically, the sampling model may be represented as follows:

Sampling Model	Population Categories		Sampling Category
	Category K_1	Category K_2	Category C
Category Probabilities	$p-\lambda p_o$	$q-(1-\lambda)p_o$	p_o
Data	m_1	m_2	m_o



This model enables one to explicitly recognize that income patterns different from those among the households revealing their income classification might be associated with households not responding, and that these distinctions can and should be maintained if one were to allocate such nonresponses back to the original population income categories for purposes of statistical analysis of these sampling data. The distinct feature of the sampling model is the explicit use of the conditional probability parameter λ as an allocation parameter for the nonresponses observed among the sample data.

Likelihood function

In summary, for a random sample of fixed sample size, n, were there to be no unclassified respondents (that is, when $p_o = 0$ and $m_o = 0$), the number of respondents in the sample, n_1, who are classified in one of the two population categories (for example, households having an annual income of $10,000 or more) and the number in the sample, n_2, who are classified in the other category (for example, households whose income is less than \$10,000) are distributed by the well-known binomial probability density

$$f(n_1, n_2|p) = \frac{n!}{n_1!\, n_2!} p^{n_1} q^{n_2}, \; 0 \leq n_1 \leq n, \; n = n_1 + n_2.$$

On the other hand, when $p_o > 0$, it is possible that some respondents in the sample who actually belong to one of the two population categories will remain unclassified (for example, households who will not reveal their income class when interviewed in the sample).

The observed sampling data, namely, m_1, m_2, and m_o, are distributed by the probability density

$$f(m_1, m_2, m_o|p, p_o, \lambda) = \frac{n!}{m_o!\, m_1!\, m_2!} p_o^{m_o} (p - \lambda p_o)^{m_1} (q - (1 - \lambda)p_o)^{m_2}$$

where

$$0 < p_o < 1; \; 0 < p < 1; \; 0 < \lambda < 1; \; q = 1 - p;$$

and

$$\lambda p_o \leq p \leq 1 - (1 - \lambda)p_o \text{ with } n = m_1 + m_2 + m_o.$$

The inequality $\lambda p_o \leq p \leq 1 - (1 - \lambda)p_o$ simply expresses analytically the requirement that the two sampling probabilities $p - \lambda p_o$ and $q - (1 - \lambda)p_o$ be non-negative.

Some point estimators for p

In general the marginal probability that an individual belongs to one of two mutually exclusive categories, say, category K_1, can be written using the law of total probability as

$$\begin{aligned} P(K_1) &= P(K_1 \& C) + P(K_1 \& \bar{C}) \\ &= P(K_1|C)P(C) + P(K_1|\bar{C})P(\bar{C}) \end{aligned}$$

where C denotes the event that an individual selected at random from the given population does not reveal which category he belongs to and $\overline{C}$ denotes the event that such a randomly selected individual does reveal his classification.

In the notation used for these probabilities

$$p = \lambda p_o + P(K_1|\overline{C})(1 - p_o)$$

where

p denotes $P(K_1)$; λ denotes $P(K_1|C)$; p_o denotes $P(C)$; and $1 - p_o$ denotes $P(\overline{C})$.

Similarly, for the other population category, say, K_2,

$$P(K_2) = P(K_2 \& C) + P(K_2 \& \overline{C})$$
$$= P(K_2|C)P(C) + P(K_2|\overline{C})P(\overline{C}),$$

or, in the notation used for these probabilities where $q = P(K_2)$ and $1 - \lambda = P(K_2|C$

$$q = (1 - \lambda)p_o + P(K_2|\overline{C})(1 - p_o)$$

From these expressions for $p = P(K_1)$ and $q = P(K_2)$ one immediately sees that whenever p_o is small

$$p \cong P(K_1|C) \text{ and } q \cong P(K_2|C)$$

and, consequently, even when the sample size n is not necessarily large, p and q can be estimated by the corresponding sample proportions observed among the classified responses and the nonresponses ignored.

On the other hand, suppose in a random sample of $n = 99$ households that $m_1 = 46$ revealed an annual income of \$10,000 or more, $m_2 = 21$ revealed an annual income of less than \$10,000, while $m_o = 32$ did not respond (that is, the households did not reveal their income classification).

For these data, nonresponses represent 32 percent of the sample, indicating that p_o is not small for the population of households sampled. Thus, it would no longer be adequate here to use the above approximations for p and q (which are valid when the proportion of nonresponses is small) and to estimate p and q by the corresponding sample proportions observed among the 67 households for which income data were obtained. How should p and q be estimated from these data?

In analyzing these data, it might be decided to disregard the $m_o = 32$ nonresponses on the assumption that these unclassified responses should be treated differently than the classified responses in order to avoid potentially serious bias which could occur should the income distribution of unclassified households be significantly different from the income distribution of those households revealing their income classification. Thus the proportion p having an annual household income of $10,000 or more would be estimated by the corresponding sample proportion $\hat{p}$ among those households in the sample who revealed their income classification, namely,

$$\hat{p} = \frac{m_1}{m_1 + m_2} = \frac{46}{46 + 21} = \frac{46}{67} = 0.69.$$

Alternately, it might be decided to divide the $m_o = 32$ nonresponses proportionately between the underlying population categories on the assumption that respondents not revealing their actual income class really have an income distribution similar to those stating which income class they belong to among the respondents sampled. Thus, p would be estimated as

$$\hat{p} = \frac{m_1 + 0.69m_o}{n} = \frac{46 + 0.69 \times 32}{99} = \frac{46 + 22}{96} = \frac{68}{96} = 0.69.$$

It is of interest to note that the same estimate for p is obtained whether one disregards the nonresponses or whether one allocates them to the basic population categories proportionate to those who indicate which income class they belong to, even though the underlying rationales for these two approaches are completely different.

Still another approach for analyzing these household income data would be to divide the nonresponses on a fifty-fifty basis between the two income classes being considered on the assumption that median annual household income is approximately \$10,000 and, hence, for any household selected at random there is a fifty-fifty chance that its income will be under \$10,000 or \$10,000 or more. Under this assumption the proportion p would be estimated as

$$\hat{p} = \frac{m_1 + 0.5m_o}{n} = \frac{46 + 0.5 \times 32}{99} = \frac{46 + 16}{99} = \frac{62}{99} = 0.63.$$

Estimates of p can also be obtained when certain exact a priori knowledge about the probabilities p, p_o, and/or λ is available. For example, if λ were known with certainty on an a priori basis, then the maximum likelihood estimator for the unknown proportion p would be $(m_1 + m_o)/n$. For these data, this estimator as a function of λ takes the linear form $0.46 + 0.32\lambda$, giving a minimum estimate of .46 if λ were 0 and a maximum estimate of .78 if λ were 1. This is obvious since $\lambda = 0$ suggests allocating none of the m_o unclassified responses back to the \$10,000 or more income class, whereas $\lambda = 1$ suggests allocating all of the m_o responses back to this income class.

In most surveys of this type certainly some knowledge of the probabilities p, p_o, and λ, especially in the form of <u>a priori judgments expressed as prior probability densities</u> would be expected. Thus a Bayesian analysis of such data will be presented in this paper for the case when prior probability densities are selected from the family of natural conjugate densities.

Standardized and nonstandardized beta densities

The family of natural conjugate densities to be used in this Bayesian analysis includes products of certain standardized and nonstandardized beta densities. For convenience these densities and some of their moments are presented here.

Standardized Beta Density

on Interval (0, 1) with Parameters s > r > 0:

$$g(x|r, s) = \frac{x^{r-1}\ (1 - x)^{s-r-1}}{B(r, s)},\ 0 < x < 1,\ s > r > 0$$

where

$$B(r, s) = \int_0^1 x^{r-1}(1 - x)^{s-r-1}\ dx = \frac{\Gamma(r)\ \Gamma(s - r)}{\Gamma(s)}.$$

In turn, the mean, mode, variance, and coefficient of variation for a random variable having a beta density are:

Mean of X	:	$E(x\|r, s) = \frac{r}{s}$
Mode of X	:	$M(x\|r, s) = \frac{r - 1}{s - 2}$
Variance of X	:	$\sigma^2(x\|r, s) = \frac{r(s - r)}{s^2(s + 1)}$
Coefficient of Variation of X	:	$CV(x\|r, s) = \frac{\sigma(x\|r, s)}{E(x\|r, s)} = \left(\frac{s - r}{r(s + 1)}\right)^{\frac{1}{2}}$

Nonstandardized Beta Density

on Interval (a, b), b > a > 0 with Parameters s > r ≥ 0:

$$g(x|a, b, r, s) = \frac{(x - a)^{r-1}\ (b - x)^{s-r-1}}{(b - a)^{s-1}\ B(r, s)},\ a < x < b,\ s > r > 0.$$

The mean and variance of a random variable having a nonstandardized beta density can be shown to be, respectively,

$$E(x|a, b, r, s) = \frac{br + a(s - r)}{s}$$

and

$\sigma^2(x|a, b, r, s) = (b - a)^2\ \sigma^2(x|r, s)$, where $\sigma^2(x|r, s)$ is given above.

Family of natural conjugate densities

As previously noted, the observed sampling data, namely, m_1, m_2, and m_o, are distributed by the probability density

$$f(m_1, m_2, m_o | p, p_o, \lambda) = \frac{n!}{m_o!\, m_1!\, m_2!} p_o^{m_o} (p - \lambda p_o)^{m_1} (q - (1 - \lambda)p_o)^{m_2}$$

where

$$o < p_o < 1;\ 0 < p < 1;\ 0 < \lambda < 1;\ q = 1 - p$$

and

$$p_o \leq p \leq 1 - (1 - \lambda)p_o \text{ with } n = m_1 + m_2 + m_o.$$

Upon inspecting the above likelihood function, it is immediately evident that its kernel and residual are simply the functions

$$L(p, p_o, \lambda | m_1, m_2, m_o) = K(p, p_o, \lambda | m_1, m_2, m_o)\ R(m_1, m_2, m_o)$$

where

$$K(p, p_o, \lambda | m_1, m_2, m_o) = p_o^{m_o} (p - \lambda p_o)^{m_1} (q - (1 - \lambda)p_o)^{m_2}$$

and

$$R(m_1, m_2, m_o) = \frac{n!}{m_o!\, m_1!\, m_2!}.$$

The kernel function can be expressed as products of standardized and nonstandardized beta densities as follows:

$$K(p, p_o, \lambda | m_1, m_2, m_o) = p_o^{m_o} (p - \lambda p_o)^{m_1} (q - (1 - \lambda)p_o)^{m_2}$$

$$= (p_o^{m_o} (1 - p_o)^{m_1 + m_2 + 1}) \left(\frac{(p - a)^{m_1} (b - p)^{m_2}}{(1 - p_o)^{m_1 + m_2 + 1}}\right)$$

where

$$a = \lambda p_o,\ b = 1 - (1 - \lambda)p_o, \text{ and } a < p < b.$$

Consider in turn the following beta density funtions, namely,

$$g(p|a, b, m_1 + 1, m_1 + m_2 + 2) = \frac{(p - a)^{m_1} (b - p)^{m_2}}{(1 - p_o)^{m_1 + m_2 + 1} B(m_1 + 1, m_1 + m_2 + 2)}$$

for

$$a < p < b \text{ where } a = \lambda p_o \text{ and } b = 1 - (1 - \lambda)p_o,$$

$$g(p_o|m_o + 1, n + 3) = \frac{p_o^{m_o} (1 - p_o)^{m_1 + m_2 + 1}}{B(m_o + 1, n + 3)}, \; 0 < p_o < 1,$$

and

$$g(\lambda|1, 2) = 1 \text{ for } 0 < \lambda < 1, \text{ i.e., the uniform density on } (0, 1).$$

The kernel function is seen to be proportional to the product of these three beta densities, namely,

$$K(p, p_o, \lambda|m_1, m_2, m_o) \propto g(p, p_o, \lambda|m_1, m_2, m_o)$$

where

$g(p, p_o, \lambda|m_1, m_2, m_o) =$

$g(p|a, b, m_1 + 1, m_1 + m_2 + 2) \; g(p_o|m_o + 1, n + 3) \; g(\lambda|1, 2).$

Family I: Nonstandard Beta Priors for p

$$f(p|p_o, \lambda, r, s) = \frac{(p - a)^{r-1} (b - p)^{s-r-1}}{(1 - p_o)^{s-1} B(r, s)}$$

where

$$a = \lambda p_o, \; b = 1 \; (1 - \lambda)p_o, \; a < p < b, \; s > r > 0 .$$

Family II: Standardized Beta Priors for p_o

$$f(p_o|t, u) = \frac{p_o^{t-1} (1 - p_o)^{u-t-1}}{B(t, u)}, \; 0 < p_o < 1, \; u > t > 0.$$

Family III: Standardized Beta Priors for λ

$$f(\lambda|v, w) = \frac{\lambda^{v-1} (1 - \lambda)^{w-v-1}}{B(v, w)}, \; 0 < \lambda < 1, \; w > v > 0.$$

Since the kernel function is proportional to a density formed by multiplying together certain beta densities from these families of beta densities, the general form for a natural conjugate density is obtained as the density function

$$f(p, p_o, \lambda | r, s, t, u, v, w) = f(p|p_o, \lambda, r, s)\ f(p_o|t, u)\ f(\lambda|v, w)$$

where $f(p|p_o, \lambda, r, s)$, $f(p_o|t, u)$ and $f(\lambda|v, w)$ are chosen, respectively, from among the beta probability densities belonging to Families I, II, and III.

<u>Notion of conditional prior invariance</u>

Natural conjugate prior densities for p, p_o, and λ are products of densities in which the prior densities for p_o and λ are independent standardized beta densities; the conditional prior density for p, given p_o and λ, is a nonstandardized beta density defined on the interval $a \leq p \leq b$, where $a = \lambda p_o$ and $b = 1 - (1 - \lambda)p_o$.

This inequality for p corresponds to the obvious probability statements that

$$P(K_1 \ \&\ \overline{C}) \leq P(K_1) \leq P(K_1) + P(K_2 \ \&\ C)$$

$$P(K_1 \ \&\ \overline{C}) \leq P(K_1) \leq (P(K_1) + P(K_2)) + (P(K_2 \ \&\ C) - P(K_2)).$$

Since, however, $P(K_1) + P(K_2) = 1$, while $P(K_2) = P(K_2 \ \&\ C) + P(K_2 \ \&\ \overline{C})$, this inequality may be written

$$P(K_1 \ \&\ \overline{C}) \leq P(K_1) \leq 1 - P(K_2 \ \&\ \overline{C})$$

or, equivalently,

$$P(K_1|\overline{C})\ P(\overline{C}) \leq P(K_1) \leq 1 - P(K_2|\overline{C})\ P(\overline{C}),$$

which is the inequality $a \leq p \leq b$ where $a = \lambda p_o$ and $b = 1 - (1 - \lambda)p_o$ when expressed in terms of the alternate notation used for these probabilities.

Thus, since the range of p depends on p_o and λ, any prior probability density for p in general must be a conditional probability density. In addition, however, for a natural conjugate prior density, the only dependency on p_o and λ exhibited by any conditional natural conjugate prior density for p is the analytical dependency which corresponds to the general condition that the range for p is restricted by the above probability inequalities which involved the probabilities p_o and λ.

The use of a natural conjugate prior density for p, p_o, and λ, therefore, greatly simplifies the prior assessments of these probabilities in a given application. One reason for this simplification is that judgments about p_o and λ need only be made so that the corresponding personalistic probability densities which are used to represent these prior beliefs about p_o and λ are assessed so they are independent (in the probability sense) of one another.

A second reason for this simplification is that whenever a natural conjugate prior density is used, an assessment of one's prior judgments about p is made conditionally, for the given values of p_o and λ, according to the principle of "conditional prior invariance." In other words, no matter what values of p_o and λ are chosen for the purpose of assessing a conditional natural conjugate beta prior density for p, the values of r and s which determine such a nonstandardized beta density must be independent of the particular values of p_o and λ being used, and, therefore, the shape of such conditional beta natural conjugate priors for p must be invariant of p_o and λ, except insofar as the range of p depends on these two probabilities. Thus the assessment of a conditional natural conjugate prior density for p can be carried out under the assumption that $p_o = 0$.

Consequently, when using a natural conjugate prior for p, p_o, and λ, the prior assessment of the population proportion p is in effect separated and made independently from the assessments of the proportions p_o and λ, which are probabilities associated with the particular sampling process.

Joint posterior density

Using a natural conjugate density as a prior density for p, p_o, and λ, the posterior density of p, p_o, and λ--given the observed data m_1, m_2, and m_o--can be obtained. The posterior density, of course, is related to the likelihood and the selected prior density by Bayes' theorem, namely,

$$f(p, p_o, \lambda | m_1, m_2, m_o) = \frac{f(m_1, m_2, m_o | p, p_o, \lambda)\ f(p, p_o, \lambda)}{f(m_1, m_2, m_o)},$$

and, therefore,

$$f(p, p_o, \lambda | m_1, m_2, m_o) \ \alpha\ f(m_1, m_2, m_o | p, p_o, \lambda)\ f(p, p_o, \lambda).$$

Using a natural conjugate density for $f(p, p_o, \lambda)$,

$$f(m_1, m_2, m_o | p, p_o, \lambda)\ f(p, p_o, \lambda)\ \alpha$$

$$\left[\frac{(p - a)^{m_1+m_2-1}\ (b - p)^{m_2+s-r-1}}{(1 - p_o)^{m_1+m_2+s-1}}\right] \left[p_o^{m_o+t-1}\ (1 - p_o)^{m_1+m_2+u-t-1}\right]$$

$$\left[\lambda^{v-1}\ (1 - \lambda)^{w-v-1}\right]$$

where $a = \lambda p_o$, $b = 1 - (1 - \lambda)p_o$, $a < p < b$, $0 < p_o < 1$, $0 < \lambda < 1$.

Consider the following beta densities:

$$f(p | p_o, \lambda, r^*, s^*) = \frac{(p - a)^{r^*-1}\ (b - p)^{s^*-r^*-1}}{(1 - p_o)^{s^*-1}\ B(r^*, s^*)}$$

where

$a = \lambda p_o$; $b = 1 - (1 - \lambda)p_o$; $a < p < b$; $r^* = m_1 + r$; and $s^* = m_1 + m_2 + s$,

$$f(p_o | t^*, u^*) = \frac{p_o^{t^*-1}\ (1 - p_o)^{u^*-t^*-1}}{B(t^*, u^*)},\quad 0 < p_o < 1$$

where $t^* = m_o + t$ and $u^* = n + u$, and

$$f(\lambda|v^*, w^*) = \frac{\lambda^{v^*-1}(1-\lambda)^{w^*-v^*-1}}{B(v^*, w^*)}, \; 0 < \lambda < 1$$

where $v^* = v$ and $w^* = w$.

Thus the posterior density of p, p_o, and λ--given the data m_1, m_2, and m_o--is the density function

$$f(p, p_o, \lambda|m_1, m_2, m_o) = f(p, p_o, \lambda|r^*, s^*, t^*, u^*, v^*, w^*)$$

where

$$f(p, p_o, \lambda|r^*, s^*, t^*, u^*, v^*, w^*) = f(p|p_o, \lambda, r^*, s^*)\, f(p_o|t^*, u^*)\, f(\lambda|v^*, w^*)$$

with

$$r^* = m_1 + r, \qquad s^* = m_1 + m_2 + s$$

$$t^* = m_o + t, \qquad u^* = n + u$$

$$v^* = v, \qquad w^* = w.$$

Other posterior densities

The joint posterior density of the probabilities p_o and λ associated with the sampling nonresponses is of interest too. This joint posterior density is obtained from the joint posterior density of p, p_o, and λ upon integration of $f(p, p_o, \lambda|m_1, m_2, m_o)$ with respect to p. Thus,

$$f(p_o, \lambda|m_1, m_2, m_o) = \int_a^b f(p, p_o, \lambda|m_1, m_2, m_o)\, dp,$$

where $a = \lambda p_o$ and $b = 1 - (1 - \lambda)p_o$. Since

$$f(p, p_o, \lambda|m_1, m_2, m_o) = f(p|p_o, \lambda, r^*, s^*)\, f(p_o|t^*, r^*)\, f(\lambda|v^*, w^*),$$

where $f(p_o|t^*, r^*)$ and $f(\lambda|v^*, w^*)$ are standardized beta densities on the interval $(0, 1)$, while $f(p|p_o, \lambda, r^*, s^*)$ is a nonstandardized beta density on the interval (a,b), the joint posterior density of p_o and λ is obtained immediately as

$$f(p_o, \lambda|m_1, m_2, m_o) = f(p_o|t^*, r^*)\, f(\lambda|v^*, w^*) \int_a^b f(p|p_o, \lambda, r^*, s^*)dp.$$

The integral appearing on the right hand side of this expression is equal to 1, of course, since $f(p|p_o, \lambda, r^*, s^*)$ is a probability density on the interval (a,b). Consequently,

$$f(p_o, \lambda|m_1, m_2, m_o) = f(p_o|m_1, m_2, m_o)\ f(\lambda|m_1, m_2, m_o)$$

where

$$f(p_o|m_1, m_2, m_o) = \frac{p_o^{m_o + t - 1}\ (1-p_o)^{m_1 + m_2 + u-t-1}}{B(m_o + t, n + u)} = f(p_o|t^*, u^*)$$

and

$$f(\lambda|m_1, m_2, m_o) = \frac{\lambda^v (1 - \lambda)^{w-v-1}}{B(v, w)} = f(\lambda|v^*, w^*).$$

Since the joint posterior density of p_o and λ is the product of their marginal posterior densities, one important observation is that, given the observed data m_1, m_2, and m_o, p_o and λ continue to be stochastically independent as they were before sampling (whenever a natural conjugate prior density is used in order to represent prior judgments about p, p_o, and λ).

Further, since $f(\lambda|m_1, m_2, m_o) = f(\lambda)$, a second important observation is that the sample data m_1, m_2, and m_o do not contain any intrinsic information about the conditional probability allocation parameter λ when a natural conjugate prior is used to assess prior beliefs held about p, p_o, and λ jointly.

Finally, the posterior mean and variance of p_o are given as

$$E(p_o|m_1, m_2, m_o) = \frac{t^*}{u^*} = \frac{m_o + t}{n + u}$$

and

$$\sigma^2(p_o\ m_1, m_2, m_o) = \frac{t^*\ (u^* - t^*)}{(u^*)^2\ (u^* + 1)} = \frac{(m_o + t)\ (m_1 + m_2 + (t-u))}{(n + u)^2\ (n + u + 1)},$$

while the posterior mean and variance of λ are identical to the prior mean and variance of λ, namely,

$$E(\lambda|m_1, m_2, m_o) = \frac{v^*}{w^*} = \frac{v}{w} = E(\lambda)$$

and

$$\sigma^2(\lambda|m_1, m_2, m_o) = \frac{v^*(w^* - v^*)}{(w^*)^2\ (w^* + 1)} = \frac{v(w-v)}{w^2(w + 1)} = \sigma^2(\lambda).$$

Besides the joint and marginal posterior densities of the parameters p_o and λ associated with the sampling nonresponses, the joint posterior densities of p and p_o and p and λ, as well as the marginal posterior density of p, are of interest too. These posterior densities, however, are difficult to obtain in analytical form, and a Monte Carlo analysis of them may be necessary.

Marginal density of the observed data

The marginal density of the observed data m_1, m_2, and m_o can be obtained from the relationship

$$f(m_1, m_2, m_o) = \frac{f(p, p_o, \lambda, m_1, m_2, m_o)}{f(p, p_o, \lambda | m_1, m_2, m_o)},$$

and it is readily seen to be the density

$$f(m_1, m_2, m_o) = \left(\frac{n!}{m_o!\, m_1!\, m_2!}\right) \left(\frac{B(r^*, s^*)}{B(r, s)}\right) \left(\frac{B(t^*, u^*)}{B(t, u)}\right)$$

or, equivalently,

$$f(m_1, m_2, m_o) =$$

$$\frac{\Gamma(n+1)\Gamma(m_o+t)\Gamma(m_1+r)\Gamma(m_2+s-r)\Gamma(m_1+m_2+u-t)\Gamma(s)\Gamma(u)}{\Gamma(n+u)\Gamma(m_o+1)\Gamma(m_1+1)\Gamma(m_2+1)\Gamma(m_1+m_2+s)\Gamma(r)\Gamma(s-r)\Gamma(t)\Gamma(u-t)},$$

sometimes called a "beta-binomial" density or a "hyper-binomial" density.

Posterior mean of $f(p|m_1, m_2, m_o)$

Since the Bayesian point estimator of p against quadratic loss is the posterior mean of $f(p|m_1, m_2, m_o)$, namely,

$$E(p|m_1, m_2, m_o) = \int_0^1 pf(p|m_1, m_2, m_o)dp,$$

the posterior density of p--given the observed data m_1, m_2, and m_o--is required. This posterior density, however, is difficult to obtain in analytical form.

The posterior mean $E(p|m_1, m_2, m_o)$ may be derived without explicitly obtaining the posterior density of p since an alternate expression for the posterior mean is

$$E(p|m_1, m_2, m_o) = E(E(p|p_o, \lambda, m_1, m_2, m_o)),$$

where the outermost expectation appearing on the right-hand side of this identity is understood to be formed with respect to the joint posterior density $f(p_o, \lambda|m_1, m_2, m_o)$ of p_o and λ, given the observed sample data.

The conditional density of p--given p_o, λ, m_1, m_2, and m_o--is a nonstandardized beta density defined on the interval (a,b) where $a = \lambda p_o$ and $b = 1 - (1 - \lambda)p_o$, namely,

$$f(p|p_o, \lambda, m_1, m_2, m_o) = g(p|a, b, r^*, s^*)$$

where

$$g(p|a, b, r^*, s^*) = \frac{(p - a)^{r^*-1} (b - p)^{s^*-r^*-1}}{(1 - p_o)^{s^*-1} B(r^*, s^*)},$$

and $a \leq p \leq b$ with $r^* = m_1 + r$ and $s^* = m_1 + m_2 + s$. Thus the conditional mean of p--given p_o, λ, m_1, m_2, and m_o--is simply

$$E(p|p_o, \lambda, m_1, m_2, m_o) = E(p|a, b, r^*, s^*)$$

where

$$E(p|a, b, r^*, s^*) = \frac{br^* + a(s^* - r^*)}{s^*} = a + \left(\frac{r^*}{s^*}\right) \left(b - a\right).$$

Substitution of $a = \lambda p_o$, $b = 1 - (1 - \lambda)p_o$, $r^* = m_1 + r$, and $s^* = m_1 + m_2 + s$ into this expression gives the conditional mean of p--given p_o, λ, m_1, m_2, and m_o--explicitly as

$$E(p|p_o, \lambda, m_1, m_2, m_o) = \lambda p_o + \left(\frac{m_1 + r}{m_1 + m_2 + s}\right)(1 - p_o).$$

In turn, according to the identity previously given, the posterior mean of p--given only the observed data m_1, m_2, and m_o, namely, $E(p|m_1, m_2, m_o)$--can be obtained by using the right-hand side of this expression for $E(p|p_o, \lambda, m_1, m_2, m_o)$ and forming its conditional expectation with respect to the joint posterior density of p_o and λ, given the observed sample data, namely,

$$E(E(p|p_o, \lambda, m_1, m_2, m_o)) = E(\lambda p_o + (\frac{m_1 + r}{m_1 + m_2 + s})(1 - p_o) \mid m_1, m_2, m_o).$$

Applying this identity when the prior density of p, p_o, and λ is a natural conjugate density yields the expression

$$p^* = E(p|m_1, m_2, m_o)$$

where

$$p^* = \left(\frac{v^*}{w^*}\right)\left(\frac{t^*}{u^*}\right) + \left(\frac{r^*}{s^*}\right)\left(1 - \frac{t^*}{u^*}\right)$$

or, substituting for v^*, w^*, t^*, u^*, r^*, and s^*,

$$v^* = v, \; w^* = w$$

$$t^* = m_o + t, \; u^* = n + u$$

$$r^* = m_1 + r, \; s^* = m_1 + m_2 + s.$$

$$p^* = \left(\frac{v}{w}\right)\left(\frac{m_o + t}{n + u}\right) + \left(\frac{m_1 + r}{m_1 + m_2 + s}\right)\left(\frac{m_1 + m_2 + u-t}{n + u}\right).$$

Structure of Bayesian point estimator

The Bayesian estimator for p against quadratic loss using a natural conjugate prior density has a very simple and appealing structure. As previously noted, the marginal probability that an individual belongs to one of two mutually exclusive categories, say, category K_1, can be written, using the law of total probability, as

$$\begin{aligned} P(K_1) &= P(K_1 \;\&\; C) + P(K_1 \;\&\; \overline{C}) \\ &= P(K_1|C)P(C) + P(K_1|\overline{C})P(\overline{C}), \end{aligned}$$

where C denotes the event that an individual selected at random from the given population does not reveal which category he belongs to, while $\overline{C}$ denotes the event that such a randomly selected individual does reveal his classification.

In the notation used for these probabilities

$$p = \lambda p_o + P(K_1|\overline{C})(1 - p_o)$$

An examination of the Bayesian point estimator for p reveals that p^* can be expressed as

$$p^* = \lambda^* p_o^* + P^*(K_1|\overline{C})(1 - p_o^*)$$

where λ^*, p_o^*, and $P^*(K_1|\overline{C})$ are simply estimators for the probabilities λ, p_o, and $P(K_1|\overline{C})$ which appear in the expression connecting the probability p with these other population probabilities.

The estimators λ^*, p_o^*, and $P^*(K_1|\overline{C})$ of λ, p_o, and $P(K_1|\overline{C})$ can be shown to be weighted averages of the prior opinions held about these probabilities (as expressed through their expected values determined from the prior probability densities being used) and the sample estimators of those probabilities based on the observed data m_1, m_2, and m_o.

In other words, considering λ^* for example,

$$\lambda^* = E(\lambda|m_1, m_2, m_o) = \frac{v^*}{w^*} = \frac{v}{w} = E(\lambda) = \lambda^{(o)}.$$

Thus the posterior estimate of λ--given the observed data m_1, m_2, and m_o--is not affected or changed by the data and, therefore, it remains the same as the prior estimate of λ expressed as the mean $\lambda^{(o)} = E(\lambda)$ of the prior density assigned to λ. In other words, the sample data do not contain any intrinsic information about the conditional probability allocation parameter when a natural conjugate prior is used to assess prior beliefs held about p, p_o, and λ.

On the other hand, the posterior estimate p_o^* of p_o, namely,

$$p_o^* = E(p_o|m_1, m_2, m_o) = \frac{t^*}{u^*},$$

can be expressed as

$$p_o^* = w_1 \left(\frac{m_o}{n}\right) + w_2 E(p_o) = \frac{m_o + t}{n + u}$$

where

$$w_1 = \frac{n}{n + u},\ w_2 = \frac{u}{n + u},\ \text{and } E(p_o) = \frac{t}{n}.$$

Thus, p_o^* is simply a weighted average of the sample estimate of p_o, namely, $\hat{p}_o = \frac{m_o}{n}$, and the prior estimate of p_o as expressed by the mean $p_o^{(\sigma)} = E(p_o)$ of the prior density assigned to p_o.

Finally, the posterior estimate of $P(K_1|\overline{C})$, the conditional probability that a classified observation belongs to the population category K_1, is simply the posterior mean of p--given the observed data m_1, m_2, and m_o and given that $p_o = 0$ which, of course, simply expresses analytically the fact that $P(K_1|\overline{C})$ is a conditional probability formed among classified respondents among whom, therefore, the probability that an unclassified response will be found (namely, p_o) must be 0. Thus,

$$P^*(K_1|\overline{C}) = E(p|p_o = 0, \lambda, m_1, m_2, m_o) = \frac{r^*}{s^*}.$$

A prior estimate of $P(K_1|\overline{C})$, however, is simply the mean of the prior conditional density assigned to p, given $p_o = 0$, namely,

$$P^{(o)}(K_1|\overline{C}) = \frac{r}{s} = E(p|p_o = 0, \lambda),$$

while the sample estimate of $P(K_1|\overline{C})$ is merely the proportion among the classified observations belonging to the given population category, namely,

$$\hat{P}(K_1|\overline{C}) = \frac{m_1}{m_1 + m_2}.$$

It can be seen immediately that the posterior estimate $P^*(K_1|\overline{C})$ of $P(K_1|\overline{C})$--given the observed data m_1, m_2, m_o--is just a weighted average of these two estimates (the first reflecting prior judgments about this probability, while the second is based only on the sample data obtained), namely,

$$P^*(K_1|\overline{C}) = w_1 \left(\frac{m_1}{m_1 + m_2}\right) + w_2\, E(p|p_o = 0, \lambda) = \frac{m_1 + r}{m_1 + m_2 + s}$$

where

$$w_1 = \frac{m_1 + m_2}{m_1 + m_2 + s}, \; w_2 = \frac{s}{m_1 + m_2 + s}, \; E(p|p_o = 0, \lambda) = \frac{r}{s}.$$

Evaluating the risk of the Bayes estimator

For quadratic loss, the risk of using the Bayes estimator $p^* = E(p|m_1, m_2, m_o)$ when the data m_1, m_2 and m_o are observed is proportional the conditional variance $\sigma^2(p|m_1, m_2, m_o)$ of the marginal posterior density to of p. When using a particular joint prior density, ξ, for expressing prior judgements held about p, p_o, and λ, the Bayes risk associated with choosing $p^* = E(p|m_1, m_2, m_o)$ is given (up to a constant of proportionality) as

$$\rho^*(\xi) = E(\sigma^2(p|m_1, m_2, m_o)),$$

namely, the risk of the Bayes estimator, p^*, against quadratic loss averaged over the observed data m_1, m_2, and m_o using the joint marginal density, $f(m_1, m_2, m_o)$, of the data.

The required conditional variance $\sigma^2(p|m_1, m_2, m_o)$ can be obtained indirectly from the identity

$$\sigma^2(p|m_1, m_2, m_o) = \sigma^2(E(p|p_o, \lambda, m_1, m_2, m_o)) + E(\sigma^2(p|p_o, \lambda, m_1, m_2, m_o))$$

where again, as when $p^* = E(p|m_1, m_2, m_o)$ was determined, the outermost expected values required for the two expressions which appear on the right-hand side of this identity are understood to be made with respect to the joint posterior density of p_o and λ, given the observed sample data, namely, $f(p_o, \lambda|m_1, m_2, m_o)$.

Previously the conditional expectation $E(p|p_o, \lambda, m_1, m_2, m_o)$, whose conditional variance is required as the first of the two terms of this identity, was determined as

$$E(p|p_o, \lambda, m_1, m_2, m_o) = \lambda p_o + (\frac{r^*}{s^*})(1-p_o)$$

where $r^* = m_1 + r$ and $s^* = m_1 + m_2 + s$. To obatin its conditional variance, recall the general expression for the variance of the product of two stochastically independent random variables X and Y.

$$\sigma^2(XY) = \sigma^2(X)\sigma^2(Y) + E^2(X)\sigma^2(Y) + E^2(Y)\sigma^2(X)$$

which, of course, can be extended should other conditioning random variables Z be used

$$\sigma^2(XY|Z) = \sigma^2(X|Z)\sigma^2(Y|Z) + E^2(X|Z)\sigma^2(Y|Z) + E^2(Y|Z)\sigma^2(X|Z).$$

This expression for $\sigma^2(XY|Z)$ can be immediately applied to obtain the variance $\sigma^2(E(p|p_o, \lambda, m_1, m_2, m_o))$ conditioned only on the observed data m_1, m_2, and m_o of the conditional expectation $E(p|p_o, \lambda, m_1, m_2, m_o)$. Thus let X denote $\lambda - c$ and Y denote p_o, where--given m_1, m_2, and m_o--c denotes the constant, r^*/s^* or $(m_1+r)/(m_1 + m_2 + s)$. Then $E(p|p_o, \lambda, m_1, m_2, m_o)$ can be written as the product of X and Y shifted by the addition of a constant c, namely,

$$E(p|p_o, \lambda, m_1, m_2, m_3) = XY + c$$

where $X = \lambda - c$ and $Y = p_o$, for the givens m_1, m_2, and m_o are stochastically independent random variables. Consequently, taking Z as the conditioning random variables m_1, m_2, and m_o, it is seen that

$$\sigma^2(E(p|p_o, \lambda, m_1, m_2, m_o)|m_1, m_2, m_o) = \sigma^2(XY+c|Z) = \sigma^2(XY|Z)$$

where, of course, $c = \frac{r^*}{s^*} = \frac{m_1 + n}{m_1 + m_2 + s}$ does not alter the conditional variance of XY, given Z, since for given Z (that is, for given m_1, m_2, and m_o) the linear shift c can be treated as a constant. Therefore,

$$\sigma^2(XY|Z) = \sigma^2(X|Z)\sigma^2(X|Z) + E^2(X|Z)\ \sigma^2(Y|Z) + E^2(Y|Z)\sigma^2(X|Z)$$

or, alternatively,

$$\sigma^2(E(p|p_o, \lambda, m_1, m_2, m_o)|m_1, m_2, m_o) = \sigma^2(\lambda-c|m_1, m_2, m_o)\sigma^2(p_o|m_1, m_2, m_o) +$$
$$E^2(\lambda-c|m_1, m_2, m_o)\sigma^2(p_o|m_1, m_2, m_o) + E^2(p_o|m_1, m_2, m_o)\sigma^2(\lambda-c|m_1, m_2, m_o).$$

Again, conditioned on m_1, m_2, and m_o, c may be treated as a constant and, therefore,

$$\sigma^2(\lambda-c|m_1, m_2, m_o) = \sigma^2(\lambda|m_1, m_2, m_o)$$

and the conditional variance $\sigma^2(E(p|p_o, \sigma, m_1, m_2, m_o))$ can be expressed as

$$\sigma^2(E(p|p_o, \lambda, m_1, m_2, m_o)) = \sigma^2(\lambda|m_1, m_2, m_o)\sigma^2(p|m_1, m_2, m_o) +$$

$$E^2(\lambda-\frac{r^*}{s^*}|m_1, m_2, m_o)\sigma^2(p_o|m_1, m_2, m_o) + E^2(p_o|m_1, m_2, m_o)\sigma^2(\lambda|m_1, m_2, m_o).$$

The second term required to obtain the conditional variance $\sigma^2(p|m_1, m_2, m_o)$ from the given identity involves the expectation of the conditional variance $\sigma^2(p|p_o, \lambda, m_1, m_2, m_o)$ formed with respect to the joint marginal posterior density of p_o and λ--given the observed data m_1, m_2, and m_o. Since the conditional density of p--given p_o, λ, m_1, m_2, and m_o--is a nonstandardized beta density with parameters $r^* = m_1 + r$ and $s^* = m_1 + m_2 + s$ on the interval (a,b) where $a = \lambda p_o$ and $b = 1 - (1-\lambda)p_o$, while the variance of such a nonstandardized data random X variable is given generally as

$$\sigma^2(X|a,b, r^*,s^*) = (b-a)^2\sigma^2(X|r^*,s^*)$$

with $\sigma^2(X|r^*,s^*)$ denoting the variance of a standardized random variable with parameters r^* and s^*, namely,

$$\sigma^2(X|r^*,s^*) = \frac{r^*(s^*-r^*)}{(s^*)^2(s^*+1)},$$

the conditional variance $\sigma^2(p|p_o, \lambda, m_1, m_2, m_o)$ is given as

$$\sigma^2(p|p_o, \lambda, m_1, m_2, m_o) = (b-a)^2\sigma^2(p|r^*,s^*) = (\frac{r^*(s^*-r^*)}{(s^*)^2(s^*+1)})(1-p_o)^2,$$

since $b - a = (1-(1-\lambda)p_o) - \lambda p_o = 1 - p_o$.

Thus the conditional expectation of $\sigma^2(p|p_o, \lambda, m_1, m_2, m_o)$--given only the observed data m_1, m_2, and m_o--is

$$E(\sigma^2(p|p_o, \lambda, m_1, m_2, m_o)|m_1, m_2, m_o) = (\frac{r^*(s^*-r^*)}{(s^*)^2(s^*+1)})E((1-p_o)^2|m_1, m_2, m_o).$$

In turn $E((1-p_o)^2|m_1, m_2, m_o)$ may be expressed in terms of the conditional expectation and conditional variance of p_o, given m_1, m_2, m_o, namely,

$$E((1-p_o)^2|m_1, m_2, m_o) = \sigma^2(1-p_o|m_1, m_2, m_o) + E^2(1-p_o|m_1, m_2, m_o)$$

or, equivalently,

$$E((1-p_o)^2|m_1, m_2, m_o) = \sigma^2(p_o|m_1, m_2, m_o) + (1-E(p_o|m_1, m_2, m_o))^2.$$

Using this identity, the expression

$$E(\sigma^2(p|p_o, \lambda, m_1, m_2, m_o)) = \left(\frac{r^*(s^*-r^*)}{(s^*)^2(s^*+1)}\right)\left(\sigma^2(p_o|m_1, m_2, m_o) + (1-E(p_o|m_1,m_2,m_o))^2\right)$$

is obtained for this conditional expectation.

Finally, since the risk of using the Bayes estimator $p^* = E(p|m_1, m_2, m_o)$, when the observed sample data m_1, m_2, and m_o are obtained from the sample survey, is given (up to a constant of proportionality for quadratic loss) by $\sigma^2(p|m_1, m_2, m_o)$ from the identity

$$\sigma^2(p|m_1, m_2, m_o) = \sigma^2(E(p|p_o, \lambda, m_1, m_2, m_o)) + E(\sigma^2(p|p_o, \lambda, m_1, m_2, m_o)),$$

the risk of this Bayes decision takes the explicit form up to a constant of proportionality

$$\begin{aligned}\sigma^2(p|m_1, m_2, m_o) &= \sigma^2(\lambda|m_1, m_2, m_o)\sigma^2(p_o|m_1, m_2, m_o) + \\ &E^2(\lambda-\tfrac{r^*}{s^*}|m_1,m_2,m_o)\sigma^2(p_o|m_1, m_2, m_o) + E^2(p_o|m_1, m_2, m_o)\sigma^2(\lambda|m_1,m_2,m_o) + \\ &\left(\frac{r^*(s^*-r^*)}{(s^*)^2(s^*+1)}\right)\left(\sigma^2(p_o|m_1,m_2,m_o) + (1-E(p_o|m_1,m_2,m_o))^2\right).\end{aligned}$$

The various conditional means and conditional variances required for p_o and λ in this expression for $\sigma^2(p|m_1,m_2,m_o)$ are summarized below:

Conditional Mean or Variance of Posterior Densities	Expressed in Terms of Parameters of Posterior Densities	Expressed in Terms of Data and Parameters of Prior Densities
$\sigma^2(\lambda\|m_1,m_2,m_o)$	$\frac{v^*(w^*-v^*)}{(w^*)^2(w^*+1)}$	$\frac{v(w-v)}{w(w+1)}$
$\sigma^2(p_o\|m_1,m_2,m_o)$	$\frac{t^*(u^*-t^*)}{(u^*)^2(u^*+1)}$	$\frac{(m_o+t)(m_1+m_2+u-t)}{(n+u)^2(n+u+1)}$
$\sigma^2(p\|p_o = 0, \lambda, m_1,m_2,m_o)$	$\frac{r^*(s^*-r^*)}{(s^*)^2(s^*+1)}$	$\frac{(m_1+r)(m_2+s-r)}{(m_1+m_2+s)^2(m_1+m_2+s+1)}$
$E(\lambda\|m_1,m_2,m_o)$	$\frac{v^*}{u^*}$	$\frac{v}{w}$
$E(p_o\|m_1,m_2,m_o)$	$\frac{t^*}{u^*}$	$\frac{m_o+t}{n+u}$
$E(p\|p_o=o,\lambda,m_1,m_2,m_o)$	$\frac{r^*}{s^*}$	$\frac{m_1+r}{m_1+m_2+s}$

From these expressions the following may also be obtained, namely,

$$\sigma^2(p_o|m_1,m_2,m_o) + (1-E(p_o|m_1,m_2,m_o))^2 = \frac{(u^*-t^*)(u^*-t^*+1)}{u^*(u^*+1)} = \frac{(m_1+m_2+u-t+1)(m_1+m_2+u-t)}{(n+u)(n+u+1)}.$$

Using these various expressions $\sigma^2(p|m_1,m_2,m_o)$ may be written in terms of the parameters of the posterior densities of p, p_o, and λ as

$$\sigma^2(p|m_1,m_2,m_o) = \left(\frac{v^*(w^*-v^*)}{(w^*)^2(w^*+1)}\right)\left(\frac{t^*(u^*-t^*)}{(u^*)^2(u^*+1)}\right) +$$

$$\left(\frac{v^*}{w^*} - \frac{r^*}{s^*}\right)^2\left(\frac{t^*(u^*-t^*)}{(u^*)^2(u^*+1)}\right) + \left(\frac{t^*}{u^*}\right)^2\left(\frac{v^*(w^*-v^*)}{(w^*)^2(w^*+1)}\right) + \left(\frac{r^*(s^*-r^*)}{(s^*)^2(s^*+1)}\right)\left(\frac{(u^*-t^*)(u^*-t^*+1)}{u^*(u^*+1)}\right).$$

Alternately, $\sigma^2(p|m_1,m_2,m_o)$ may be written in terms of the observed data m_1, m_2, and m_o and the parameters of the prior densities of p, p_o, and λ as

$$\sigma^2(p_o|m_1,m_2,m_o) = (\frac{v(w-v)}{w^2(w+1)})(\frac{(m_o+t)(m_1+m_2+u-t)}{(n+u)^2(n+u+1)}) +$$

$$(\frac{v}{w} - \frac{m_1+r}{m_1+m_2+s})^2(\frac{(m_o+t)(m_1+m_2+u-t)}{(n+u)^2(n+u+1)}) + (\frac{m_o+t}{n+u})^2(\frac{v(w-v)}{w^2(w+1)}) +$$

$$(\frac{(m_1+r)(m_2+s-r)}{(m_1+m_2+s)^2(m_1+m_2+s+1)})(\frac{(m_1+m_2+u-t+1)(m_1+m_2+u-t)}{(n+u)(n+u+1)}).$$

From this last expression for $\sigma^2(p_o|m_1,m_2,m_o)$, the Bayes risk under quadratic loss of using the estimator $p^* = E(p|m_1,m_2,m_o)$ can be evaluated up to a constant of proportionality by forming the expectations of the various terms involving the observed data m_1, m_2, m_o with respect to their joint marginal density $f(m_1,m_2,m_o)$. This calculation will not be undertaken here; instead, various asymptotic expressions for $\sigma^2(p|m_1,m_2,m_o)$, which are valid when the sample size n is large, will be explored.

Asymptotic expansions for $\sigma^2(p|m_1,m_2,m_o)$

First, suppose the sample size n is large relative to the parameters t and u which determine the prior natural conjugate beta density for p_o. Then $\frac{t}{n} \cong o$, $\frac{u}{n} \cong o$, and $\frac{u-t}{n} \cong o$, and $\sigma^2(p|m_1,m_2,m_o)$ may be approximated as

$$\sigma^2(p|m_1,m_2,m_o) \cong (\sigma^2(\lambda)+(E(\lambda) - (\frac{m_1+r}{(m_1+m_2+s)})^2)(\frac{\hat{p}_o(1-\hat{p}_o)}{n}) +$$

$$\sigma^2(\lambda)(\hat{p}_o)^2 + (\frac{(m_1+r)(m_2+s-r)}{(m_1+m_2+s)^2(m_1+m_2+s+1)})(1-\hat{p}_o)^2$$

where $\hat{p}_o = \frac{m_o}{n}$ is the proportion of nonresponses actually observed in the sample and $1 - \hat{p}_o = \frac{m_1 + m_2}{n}$ is the proportion of classified responses observed in the sample.

Alternately, suppose the sample size n is large with respect to the parameters r and s which determine the prior natural conjugate conditional beta density for p. In this situation $\frac{r}{n} \cong o$, $\frac{s}{n} \cong o$, and $\frac{s-r}{n} \cong o$; therefore an approximate expression for $\sigma^2(p|m_1,m_2,m_o)$ is

$$\sigma^2(p|m_1,m_2,m_o) \cong (\sigma^2(\lambda)+(E(\lambda) - \frac{\hat{p}_1}{\hat{p}_1+\hat{p}_2})^2)\frac{(m_o+t)(m_1+m_2+u-t)}{(n+u)^2(n+u+1)} +$$

$$\sigma^2(\lambda)(\frac{m_o+t}{n+u})^2 + (\frac{(m_1+m_2+u-t+1)(m_1+m_2+u-t)}{(n+u)(n+u+1)})(\frac{\hat{p}_1\hat{p}_2}{n(\hat{p}_1+\hat{p}_2)^3})$$

where $\hat{p}_1 = \frac{m_1}{n}$ denotes the sample proportion of classified responses belonging to the first population category, and $\hat{p}_2 = \frac{m_2}{n}$ denotes the sample proportion of classified responses belonging to the second population category.

As a third situation, suppose the sample size n is large with respect to each of the four parameters t, u, r, and s related to the prior natural conjugate beta densities assigned to p_o and p. In this case the two approximations previously given for $\sigma^2(p|m_1,m_2,m_o)$ become the approximation

$$\sigma^2(p|m_1,m_2,m_o) \cong (\sigma^2(\lambda)+(E(\lambda) - \frac{\hat{p}_1}{\hat{p}_1+\hat{p}_2})^2)(\frac{\hat{p}_o(1-\hat{p}_o)}{n}) + \sigma^2(\lambda)(\hat{p}_o)^2 + \frac{\hat{p}_1\hat{p}_2}{n(\hat{p}_1+\hat{p}_2)} .$$

Finally, when the sample size n dominates r, s, t, and u in the sense that $\frac{r}{n}$, $\frac{s}{n}$, $\frac{t}{n}$, and $\frac{u}{n}$ are small, then whenever the sample size is large and $\hat{p}_1 + \hat{p}_2$ is not exceeding small (or, equivalently, whenever the proportion of nonresponses observed in the sample, $\hat{p}_o$, is not almost 1 and the sample does not yield almost all nonresponses), the conditional variance $\sigma^2(p|m_1,m_2,m_o)$, which gives the risk of using $p^* = E(p|m_1,m_2,m_o)$ as an estimate for p (except, perhaps, for a multiplicative constant), is approximated simply by

$$\sigma^2(p|m_1,m_2,m_o) \cong \sigma^2(\lambda)(\hat{p}_o)^2.$$

In other words, this asymptotic expression for $\sigma^2(p|m_1,m_2,m_o)$ indicates that whenever the sample taken is large enough so that the sample data related to p and p_o are no longer overly influenced by whatever prior judgments may have been held about them, then, provided most of the observations in the actual sample are not nonresponses, the risk of using the Bayes estimator $p^* = E(p|m_1,m_2,m_o)$ to estimate p is determined by (1) the square of the observed proportion of nonresponses in the sample, and (2) the variance of the prior natural conjugate beta density for the allocation parameter λ (this prior density, of course, expressing the only information which is available about the allocation parameter).

In summary, when the sample size n is large and the observed sample information about p and p_o outweighs the prior judgements made about them, should the proportion of nonresponses observed in the sample be small, the risk in using the Bayes estimator $p^* = E(p|m_1,m_2,m_o)$ for purposes of estimating p is negligible. Alternately, if the observed proportion of nonresponses is large, but not so large that almost the entire sample consists of non-responses, then should the prior judgements held about the allocation parameter λ reflect the nearly certain belief that λ is a particular value so that essentially the entire unit mass of the prior density is placed at this point, again for practical purposes the risk of using this Bayes estimator can be ignored.