# Bayesian Analysis of Binary Data Subject to Misclassification

M. Evans I. Guttman Y. Haitovsky T. Swartz *

**Abstract**

This paper considers estimation of success probabilities of categorical binary data subject to misclassification errors from the Bayesian point of view. It has been shown by Bross (1954) that sample proportions are in general biased estimates. This bias is a function of the amount of misclassification and can be substantial. Tenenbein (1970) proposed to eliminate the bias by subjecting a portion of the sample to both true and fallible classifiers, resulting in a 2 x 2 table, from which the misclassification rates can be estimated. The rationale is that fallible classifiers are inexpensive relative to infallible ones. Hence if only a part of the sample is measured by the infallible classifier one can obtain a more efficient estimate, for a given sampling budget, than by measuring the whole sample using the infallible classifier.

In many contexts an infallible classifier is unavailable or prohibitively expensive. Bayesian methods then provide a useful approach for dealing with the consequent nonidentifiability problems which arise when we want to carry out inference.

In this paper we treat both the single measurement and the repeated measurements (where the former is a special case of the latter) from a Bayesian point of view. The posterior analyses are carried out using both Gauss-Jacobi quadrature and Gibbs sampling. Through examples it is shown that in most cases Gauss-Jacobi quadrature produces very good approximations, both in terms of accuracy and speed of computation. The Gibbs sampler requires more computation to reach the same level of accuracy as the Gauss-Jacobi.

## 1. Introduction

In various applications and particularly medical contexts binary data may be subject to misclassification errors. For example, a healthy patient may be incorrectly diagnosed as sick by a physician. Conversely a sick patient may be incorrectly diagnosed as healthy. The effects of ignoring misclassification were first noted by Bross(1954) who showed that classical estimators based on proportions may be profoundly biased. He also investigated the loss of power resulting from the standard analysis of a 2x2 contingency table. Others have extended his work to consider the effects of misclassification on generalized contingency tables and the derivation of optimal designs. Fleiss(1973) and Kleinbaum, Kupper and Morgenstern(1982) are both reference sources surveying the problem of binary misclassification.

As for the effect of misclassified categorical data on the estimation of success probabilities, Bross(1954) showed that the knowledge of both false positive and false negative rates is needed for correcting the bias

*M. Evans and I. Guttman are Professors, Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S1A1. Y. Haitovsky is Professor, Departments of Economics and Statistics, Hebrew University, Jerusalem 91905, Israel. T. Swartz is Assistant Professor, Department of Mathematics and Statistics, Simon Fraser University, Burnaby, British Columbia, Canada V5A1S6. The authors thank 2 referees and the Associate Editor for helpful comments.

resulting from estimation based on observed proportions. Tenenbein(1970, 1971, 1972) proposed to obtain this information by double sampling; i.e. by subjecting a portion of the sample to an infallible classifier. Thus, this portion of the sample is cross-classified by both fallible and correct classifiers, enabling the estimation of false positive and false negative rates. The rationale of Tenenbein's double sampling procedure is that inexpensive classification procedures will often result in misclassifications, whereas true classifiers are more expensive and often substantially so.

But what if an infallible device is not available or its use is prohibitively expensive, as is often the case with medical problems? One approach is a Bayesian approach where we assume prior knowledge on the misclassification parameters, and then incorporate this prior knowledge with sample information via Bayes rule. It is this approach which we consider in this paper. Gaba and Winkler(1992) also discuss Bayesian inference for this model when a single measurement is available. Our paper extends their analysis to the multiple measurement case and the two sample problem. In addition, we consider a more general class of priors, we derive expressions for posterior distribution functions and we obtain an expression for the posterior probability of encountering $k_1$ false positives and $k_2$ false negatives. Most significantly, however, we focus on developing computational approaches for the practical implementation of Bayesian techniques. Matchar, Simel, Geweke and Feussner(1990) discuss aspects of Bayesian inference when full classification information is available.

In section 2, we formulate the basic problem from the Bayesian point of view and then indicate immediate generalizations to our model. In section 3, we discuss the implementation of the posterior analysis via two approaches: quadrature using Gauss-Jacobi rules and Monte Carlo integration using the Gibbs sampler. The two-sample problem is considered in section 4 and the posterior probability of encountering $k_1$ false positives and $k_2$ false negatives is given in section 5. Section 6 presents numerical examples and some conclusions are offered in section 7.

## 2. The Bayesian Formulation

In keeping with the medical analogy and the original framework developed by Bross(1954) we consider a random patient having probability $p$ of being sick, error probability $\theta$ of being misdiagnosed as healthy and error probability $\phi$ of being misdiagnosed as sick. In symbols we write

$$\begin{aligned} p &= P(\eta = 1) \\ \theta &= P(X = 0 \mid \eta = 1) \\ \phi &= P(X = 1 \mid \eta = 0) \end{aligned}$$

where $\eta = 1(0)$ is a latent indicator variable indicating sick(healthy) and X = 1(0) is the observed diagnosis indicator variable indicating sick(healthy). For a sample of n patients we obtain the likelihood

$$[p(1-\theta) + (1-p)\phi]^t [p\theta + (1-p)(1-\phi)]^{n-t} \tag{1}$$

where $t = \sum_{i=1}^{n} x_i$.

The probability of a single individual being diagnosed sick for fixed error probabilities $(\theta, \phi)$ is just $f(p|\theta, \phi) = p(1-\theta) + (1-p)\phi$. It is natural to require that this be an increasing function of $p$ and this entails that $\phi < 1 - \theta$ or equivalently that $\theta + \phi < 1$. Hence we impose this restriction hereafter. The appeal of this restriction can also be seen from $Corr[X, \eta] = p(1-p)(1-\theta-\phi)/SD[X]SD[\eta]$ and hence the correlation is positive if and only if the restriction applies; see Haitovsky and Zelen(1990)

and Deming(1977). With this condition $P(\eta = 1 \mid X = 1) > P(\eta = 1 \mid X = 0)$ or equivalently $P(\eta = 0 \mid X = 0) > P(\eta = o \mid X = 1)$. Deming(1977) uses the latter.

We note that $f(p|\theta,\phi) = f(1-p|1-\phi, 1-\theta)$ and therefore $1 - f(p|\theta,\phi) = 1 - f(1-p|1-\phi, 1-\theta)$. Hence the model, without the restriction, is nonidentifiable for every value of $n$. The restriction eliminates this redundancy in the parameterization but we can still have $f(p|\theta_1,\phi_1) = f(p|\theta_2,\phi_2)$ for some $p$ when $(\theta_1,\phi_1) \neq (\theta_2,\phi_2)$. Hence the nonidentifiability is reduced but not eliminated by the restriction. ¿From a statistical viewpoint the nonidentifiability arises because we have no sample information that allows us to distinguish amongst the values of $(p,\phi,\theta)$ which lead to a fixed value of $f(p|\theta,\phi)$. Thus, to eradicate the nonidentifiability problem, more information is necessary. In double sampling this additional information is attained by subjecting a subsample to both fallible and infallible classification devices. The Bayesian approach has the virtue that nonidentifiability is not a problem for inference; one simply integrates to obtain the desired marginal posterior distributions.

We consider a more general problem than that just discussed. We allow for the situation where $m_i$ physicians examine the $i^{th}$ patient assuming each physician has the same probability of misdiagnosis and that $x_i$ of the physicians diagnose the $i^{th}$ patient as sick. This could also correspond to the situation where multiple readings are observed from some measuring device. The likelihood for data $x_1,\ldots,x_n$ is then

$$L(p,\theta,\phi \mid \underline{x}) \propto \prod_{i=1}^{n} [p(1-\theta)^{x_i}\theta^{m_i-x_i} + (1-p)\phi^{x_i}(1-\phi)^{m_i-x_i}]. \tag{2}$$

A natural choice for the prior is a Beta prior for $p$ and an independent Dirichlet prior for $(\theta,\phi)$ leading to the joint posterior

$$\pi(p,\theta,\phi \mid \underline{x}) \propto L(p,\theta,\phi \mid \underline{x}) p^{v_1-1}(1-p)^{v_2-1}\theta^{v_3-1}\phi^{v_4-1}(1-\theta-\phi)^{v_5-1} \ .$$

In the case of prior ignorance one might choose $v_i = 1$, $i = 1,\ldots,5$ yielding independent uniform priors on $p$ and $(\theta,\phi)$. In section 6 a more detailed discussion is given of the effect of the choice of the prior.

Assuming a Dirichlet prior for $(\theta,\phi)$ imposes the additional restriction that $\theta$ and $\phi$ are negatively correlated. In certain contexts this may be unnatural and hence a Dirichlet would not be appropriate. In the domain of application we are addressing, however, a negative correlation is quite reasonable. When a physician overestimates the amount of sickness (health) in a population, the physician will tend to make more (fewer) misdiagnoses of healthy people as sick, and at the same time will tend to make fewer (more) misdiagnoses of sick people as healthy. Such circumstances would arise when a physician's errors are largely caused by a substantial error in their prior beliefs concerning the prevalence of the disease in the population. A larger class of distributions on the simplex which contains the Dirichlet family and which allows for positive correlations is given by the class of generalized Liouville distributions, see Devroye(1986). Bayesian inference for this family together with the associated computational problems is a subject for further research.

As in the case where diagnosis errors do not arise, our primary concern is drawing inferences on the parameter $p$ (the probability that a random patient is sick). The marginal posterior density for $p$ takes the form of a positive polynomial of degree $n + v_1 + v_2$ when $v_1$ and $v_2$ are integers; a structure which we exploit in the next section. Further, in the Bayesian context, the problem of comparing 2 (or more) independent populations is not unduly complicated. For example we simply multiply the marginal posterior densities of $p_1$ and $p_2$ to obtain their joint posterior density from which quantities such as $\gamma = p_1 - p_2$ can be studied. The two-sample problem is taken up in section 4.

## 3. Computations

In carrying out the posterior analysis, based on the general model of section 2, we need to evaluate 3-dimensional integrals of the form

$$I(m) = \int_0^1 \int_0^{1-\phi} \int_0^1 m(p,\theta,\phi) L(p,\theta,\phi \mid \underline{x}) \pi_p(p) \pi_{\phi,\theta}(\phi,\theta) dp d\theta d\phi \tag{3}$$

where the functions $\pi_p$ and $\pi_{\phi,\theta}$ are the densities of the Beta$(v_1, v_2)$ and Dirichlet$(v_3, v_4, v_5)$ distributions respectively. For example $I^{-1}(1)$ is the norming constant for the joint posterior density and $I(p^i)/I(1)$ is the $i^{th}$ posterior moment for the probability that a random patient is sick.

A simple transformation is helpful in evaluating (3); namely $p \rightarrow p, \phi \rightarrow \phi$ and $\theta \rightarrow \alpha = \theta/(1-\phi)$. With this transformation (3) is proportional to

$$\int_0^1 \int_0^1 \int_0^1 m(p,(1-\phi)\alpha,\phi) L(p,(1-\phi)\alpha,\phi \mid \underline{x}) p^{v_1-1}(1-p)^{v_2-1} \alpha^{v_3-1}(1-\alpha)^{v_5-1} \times \phi^{v_4-1}(1-\phi)^{v_3+v_5-1} dp d\alpha d\phi. \tag{4}$$

Note that with this transformation the ranges of integration become 0 to 1 for all variables and the prior density becomes a product of Beta densities. Then a natural multiple quadrature rule to be used in approximating $I(m)$ is a particular type of Gauss rule known as a product Jacobi rule. It is given by

$$I^*(m) = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} w_{1,i_1} w_{2,i_2} w_{3,i_3} m(t_{1,i_1}, t_{2,i_2}(1-t_{3,i_3}), t_{3,i_3}) L(t_{1,i_1}, t_{2,i_2}(1-t_{3,i_3}), t_{3,i_3} \mid \underline{x})$$

where $n_1 n_2 n_3$ is the order of the rule and the $w_{j,i_j}$ and the $t_{j,i_j}$ $i_j = 1, \ldots, n_j$, $j = 1,2,3$ are referred to as the weights and points respectively. These weights and points are completely determined by $v_1, v_2, v_3, v_4, v_5$ and can be obtained using the Fortran programs provided in Stroud and Secrest(1966) or from such libraries as the NAG and IMSL subroutines. For a mathematical definition of Jacobi weights and points the reader is referred to Davis and Rabinowitz(1984).

The appeal of the product Jacobi rule is that the approximation of $I(m)$ by $I^*(m)$ is exact if $m(p,\theta,\phi)$ is a polynomial of degree $u_1$, $u_2$ and $u_3$ in $p$, $\theta$ and $\phi$ respectively where $u_1 + n \leq 2n_1 - 1$, $u_2 + \sum_{i=1}^n m_i \leq 2n_2 - 1$ and $u_2 + u_3 + \sum_{i=1}^n m_i \leq 2n_3 - 1$. Hence the norming constant and more generally the posterior expectations of polynomials in $p$, $\theta$ and $\phi$ can all be exactly evaluated. In practice the approximation may be quite good even if $n_1$, $n_2$ and $n_3$ are not sufficiently large to satisfy the above 3 inequalities. The approximation may also be quite good in the case where $m(p,\theta,\phi)$ is not a polynomial but can be well-approximated by a polynomial.

A difficulty with the product Jacobi rule arises when evaluating posterior probability contents such as $P(p \leq p_0 \mid \underline{x})$ for some $p_0$. The reason for this is that $m(p,\theta,\phi)$ then takes the form of an indicator function which is not well-approximated by a polynomial. Clearly the product Jacobi rule will yield the same value for $P(p \leq p_0 \mid \underline{x})$ and $P(p \leq p_0^* \mid \underline{x})$ whenever $t_{1,i_1} \leq p_0 \leq p_0^* < t_{1,i_1+1}$ for some $i_1$. A solution to this difficulty is to express the marginal posterior density of $p$ as the sum

$$\pi_{p|\underline{x}}(p \mid \underline{x}) = \sum_{i=0}^{n} c_i h_i^{(v_1,v_2)}(p) \pi_p(p)$$

where $h_i^{(v_1,v_2)}(p)$ is the degree $i$ orthonormal polynomial with respect to the $Beta(v_1, v_2)$ density and $c_i = \int_0^1 h_i^{(v_1,v_2)} \pi_{p|\underline{x}}(p \mid \underline{x})dp$. This is the orthogonal expansion of $\pi_{p|\underline{x}}$ with respect to the orthonormal polynomials of $\pi_p$. It is exact because $\pi_{p|\underline{x}}$ takes the form of a polynomial times $\pi_p$. Note that the orthonormal polynomials $h_i^{(v_1,v_2)}(p)$ can be generated through Jacobi recurrence formulae (see Davis and Rabinowitz(1984), p. 38) and that the $c_i$ can be exactly calculated using Gauss-Jacobi quadrature. An application of Rodrigue's formula, which expresses $h_i^{(v_1,v_2)}(p)\pi_p(p)$ as a derivative which can be integrated in closed form (see Davis and Rabinowitz (1984), p. 38), leads to the following exact expression for the posterior distribution function of $p$,

$$\begin{aligned} P(p \le p_0 \mid \underline{x}) &= \sum_{i=0}^{n} c_i \int_0^{p_0} h_i^{(v_1,v_2)}(p)\pi_p(p)dp \\ &= c_0\, Beta(v_1, v_2, p_0) + p_0^{v_1}(1-p_0)^{v_2} \sum_{i=1}^{n} \frac{c_i d_i^{(v_1,v_2)}}{i d_{i-1}^{(v_1+1,v_2+1)}} h_{i-1}^{(v_1+1,v_2+1)}(p_0) \end{aligned}$$

where $Beta(v_1, v_2, p_0)$ is the distribution function of the $Beta(v_1, v_2)$ distribution evaluated at $p_0$ and

$$d_i^{(v_1,v_2)} = \left[\frac{\Gamma(v_1)\Gamma(v_2)\Gamma(i+v_1+v_2)(2i+v_1+v_2-1)i!}{\Gamma(v_1+i)\Gamma(v_2+i)\Gamma(v_1+v_2)(i+v_1+v_2-1)}\right]^{1/2}.$$

Note that this expression holds for all $p_0$ and no further integrations are required once we have computed all the $c_i$.

A second computational approach which can be easily applied here is the Gibbs sampling algorithm introduced by Geman and Geman(1984). The Gibbs sampler is a Monte Carlo algorithm whereby repeated sampling from updated conditional distributions asymptotically yields marginal samples. Following the notation of Gelfand and Smith(1990) and Gelfand et. al.(1990) we denote densities by square brackets, so that joint, conditional and marginal densities appear for example as $[X, Y]$, $[X \mid Y]$ and $[Y]$. In the Bayesian context of our problem the Gibbs sampling algorithm takes the following form:

(1) Initialize $p$, $\theta$ and $\phi$ with $p^0$, $\theta^0$ and $\phi^0$
(2) For $i = 1, \ldots, M$
- generate $p^i$ from $[p \mid \theta^{i-1}, \phi^{i-1}, \underline{x}]$
- generate $\theta^i$ from $[\theta \mid p^i, \phi^{i-1}, \underline{x}]$
- generate $\phi^i$ from $[\phi \mid p^i, \theta^i, \underline{x}]$
(3) Record $(p^M, \theta^M, \phi^M)$
(4) Repeat steps (1)-(3) $N$ times

Note that although $p^0$, $\theta^0$ and $\phi^0$ are arbitrary we typically set them equal to the values of $p^M$, $\theta^M$ and $\phi^M$ from the previous iteration of step (2).

The Gibbs algorithm yields a 3-dimensional sample of size $N$ which we treat as a sample from the posterior distribution. The choice of cycle length $M$ to minimize dependence between the sample vectors and the choice of $N$ to ensure posterior convergence depend on the particular application.

The generation of variates from $[p \mid, \theta, \phi, \underline{x}]$, $[\theta \mid p, \phi, \underline{x}]$ and $[\phi \mid p, \theta, \underline{x}]$ is not trivial as in each case we need to efficiently sample from a density taking the form of a positive polynomial times a beta density. For example the rejection method is difficult to implement as it is not clear how to find an efficient bounding density from which to sample.

Instead we consider the data augmentation approach whereby the introduction of additional variables results in straightforward sampling. This is done at the expense of increasing the number of conditional distributions from which to sample. In our problem we introduce $n$ additional variables $\eta_i$ for $i = 1, \ldots, n$ where $\eta_i = 1(0)$ indicates that the $i^{th}$ patient is sick(healthy). We also make the transformation $(\theta, \phi) \rightarrow (\alpha, \phi)$ where $\alpha = \theta/(1-\phi)$, sample from the conditional of $\alpha$ given the remaining variables and then put $\theta = (1-\phi)\alpha$. Using the notation $\underline{\eta}_{(i)} = (\eta_1, \ldots, \eta_{i-1}, \eta_{i+1}, \ldots, \eta_n)$ it is not difficult to show that

$$
\begin{aligned}
[\eta_i \mid p, \alpha, \phi, \underline{\eta}_{(i)}, \underline{x}] &\propto [p(1-\theta)^{x_i}\theta^{m_i - x_i}]^{\eta_i}[(1-p)\phi^{x_i}(1-\phi)^{m_i - x_i}]^{1-\eta_i} \\
[p \mid \alpha, \phi, \underline{\eta}, \underline{x}] &\propto p^{\sum \eta_i + v_1 - 1}(1-p)^{(1-\sum \eta_i) + v_2 - 1} \\
[\alpha \mid p, \phi, \underline{\eta}, \underline{x}] &\propto \sum_{j=0}^{\sum x_i \eta_i} \left( \begin{array}{c} \sum x_i \eta_i \\ j \end{array} \right) \phi^j [\alpha^{\sum m_i \eta_i - \sum x_i \eta_i + j + v_3 - 1}(1-\alpha)^{\sum x_i \eta_i - j + v_5 - 1}] \\
[\phi \mid p, \alpha, \underline{\eta}, \underline{x}] &\propto \sum_{j=0}^{\sum x_i \eta_i} \left( \begin{array}{c} \sum x_i \eta_i \\ j \end{array} \right) \alpha^j (1-\alpha)^{-j} [\phi^{\sum x_i - \sum x_i \eta_i + j + v_4 - 1}(1-\phi)^{\sum m_i - \sum x_i + v_3 + v_5 - 1}].
\end{aligned}
$$

The conditional sampling is now straightforward as the 4 density forms are Bernoulli, Beta, a mixture of Betas and a mixture of Betas respectively. We note, however, that sampling from the conditional distributions for $\theta$ and $\phi$ becomes progressively more difficult as $n$ rises.

## 4. The Two-Sample Problem

A problem which is of particular interest to experimenters is the comparison of rates of disease in two independent populations. Typically these two populations are the cases and the controls. Letting $\pi_1(p_1 \mid \underline{x}_1)$ and $\pi_2(p_2 \mid \underline{x}_2)$ be the marginal posterior densities of $p_1$ and $p_2$ respectively we are interested in drawing inference on $\gamma = p_1 - p_2$ where the density of $\gamma$ is given by

$$f(\gamma) = \int_U \pi_1(\gamma + u \mid \underline{x}_1)\pi_2(u \mid \underline{x}_2)du$$

and $U = \{u \mid 0 < u < 1,\ 0 < \gamma + u < 1\}$. This general formulation allows for different priors to be placed on the parameters of the two separate populations. Splitting the domain of integration into $-1 < \gamma \leq 0$ and $0 < \gamma < 1$ and changing variables in each domain according to $w = (u+\gamma)/(1+\gamma)$ and $w = u/(1-\gamma)$ respectively it is not difficult to show that

$$f(\gamma) = \begin{cases} r_1(\gamma) = \int_0^1 (1+\gamma)\pi_1(w + w\gamma \mid \underline{x}_1)\pi_2(w(1+\gamma) - \gamma \mid \underline{x}_2)dw & -1 < \gamma \leq 0 \\ r_2(\gamma) = \int_0^1 (1-\gamma)\pi_1(\gamma + w(1-\gamma) \mid \underline{x}_1)\pi_2(w - w\gamma \mid \underline{x}_2)dw & 0 < \gamma < 1 \end{cases}$$

Note that $f(\gamma)$ is particulary amenable to Gauss quadrature if the Beta prior parameters for $p_1$ and $p_2$ are integers. Specifically, $r_1(\gamma)$ and $r_2(\gamma)$ are each polynomials of degree $n^* = n_1 + v_1^{(1)} + v_2^{(1)} + n_2 + v_1^{(2)} + v_2^{(2)} - 3$ where $n_i$ and $(v_1^{(i)}, v_2^{(i)})$ are the sample size and Beta prior parameters in population $i$, $i = 1, 2$. The calculation of posterior moments of $\gamma = p_1 - p_2$ is then straightforward using the expression for $f(\gamma)$ and Gauss-Legendre quadrature.

The calculation of posterior probability contents of $\gamma = p_1 - p_2$ can also be carried out exactly. Letting $h_i$ be the $i^{th}$ degree orthonormal polynomial with respect to the Uniform(0,1) density, we write $r_1(\gamma) = \sum_{i=0}^{n^*} c_i h_i(\gamma + 1)$ and $r_2(\gamma) = \sum_{i=0}^{n^*} c_i^{'} h_i(\gamma)$ where $c_i = \int_0^1 h_i(\gamma) r_1(\gamma - 1) d\gamma$ and $c_i^{'} = \int_0^1 h_i(\gamma) r_2(\gamma) d\gamma$. Then

$$P(\gamma \leq \gamma_0 \mid \underline{x}_1, \underline{x}_2) = \begin{cases} \sum_{i=0}^{n^*} c_i \int_0^{1+\gamma_0} h_i(\gamma) d\gamma & -1 < \gamma_0 \leq 0 \\ c_0 + \sum_{i=0}^{n^*} c_i^{'} \int_0^{\gamma_0} h_i(\gamma) d\gamma & 0 < \gamma_0 < 1 \end{cases}$$

An application of Rodrigue's formula then yields the expression

$$P(\gamma \leq \gamma_0 \mid \underline{x}_1, \underline{x}_2) = \begin{cases} c_0(1+\gamma_0) + (1+\gamma_0)(-\gamma_0) \sum_{i=1}^{n^*} c_i \left[\frac{6}{i(i+1)}\right]^{1/2} h_{i-1}^{(2,2)}(1+\gamma_0) & -1 < \gamma_0 \leq 0 \\ c_0 + c_0^{'}(\gamma_0) + \gamma_0(1-\gamma_0) \sum_{i=1}^{n^*} c_i^{'} \left[\frac{6}{i(i+1)}\right]^{1/2} h_{i-1}^{(2,2)}(\gamma_0) & 0 < \gamma_0 < 1 \end{cases}$$

In the case of Gibbs sampling, nothing new is required in the two-sample problem. We simply sample as before in the two independent cases and combine our variates according to $\gamma = p_1 - p_2$.

## 5.The Posterior Probability of $k_1$ False Positives and $k_2$ False Negatives

A false positive occurs when a diagnosis of disease is made when an individual is healthy and a false negative occurs when a diagnosis of health is made when an individual is diseased. Of course the actual number of these is latent in the data. Suppose we pretend, however, that these are observable and denote the number of false positives by $k_1$ and the number of false negatives by $k_2$. We consider first the problem where $m_1 = \ldots = m_n = 1$ to avoid excessive combinatorial problems.

Expanding (1) we obtain

$$\begin{aligned} & \sum_{k_1=0}^{t} \binom{t}{k_1} [(1-p)\phi]^{k_1} [p(1-\theta)]^{t-k_1} \sum_{k_2=0}^{n-t} \binom{n-t}{k_2} [p\theta]^{k_2} [(1-p)(1-\phi)]^{n-t-k_2} \\ = & \sum_{k_1=0}^{t} \sum_{k_2=0}^{n-t} \binom{t}{k_1} \binom{n-t}{k_2} p^{t-k_1+k_2} (1-p)^{n-t-k_2+k_1} \theta^{k_2} (1-\theta)^{t-k_1} \phi^{k_1} (1-\phi)^{n-t-k_2} \end{aligned} \quad (5)$$

Note that $(1-p)\phi$ is the probability that a false positive is obtained and $p\theta$ is the probability that a false negative is obtained for a single randomly selected individual. If we multiply (5) by the prior of $p$, $\theta$ and $\phi$ and recall that the inverse of the norming constant is $I(1)$, then the joint posterior probability of having obtained $k_1$ false positives and $k_2$ false negatives is

$$I(1)^{-1} \binom{t}{k_1} \binom{n-t}{k_2} \int_0^1 \int_0^1 \int_0^{1-\phi} p^{t-k_1+k_2} (1-p)^{n-t+k_1-k_2} \quad \theta^{k_2} (1-\theta)^{t-k_1} \phi^{k_1} (1-\phi)^{n-t-k_2} \times \pi_p(p) \pi_{\theta,\phi}(\theta, \phi) d\theta d\phi dp.$$

Again making the transformation $\theta \rightarrow \alpha/(1-\phi)$ we see that this posterior probability can be calculated exactly using Gauss-Jacobi quadrature.

In the general case the answer is somewhat more complicated. Let $\mathcal{A}_k$ denote the set of all subsets of $\{1, \ldots, n\}$ of cardinality $k$, where $k$ is the latent number of sick patients and put $t(A) = \sum_{i \epsilon A} x_i$ and

$m(A) = \sum_{i \epsilon A} m_i$ for $A \epsilon \mathcal{A}_k$. Note that $t(A)$ equals the number of correct diagnoses of sickness, $m(A) - t(A)$ equals the number of false negatives, $t(A^c)$ equals the number of false positives and $m(A^c) - t(A^c)$ equals the number of correct negatives. Then (2) can be written as

$$\begin{aligned} & \sum_{k=0}^{n} p^k (1-p)^{n-k} \sum_{A \epsilon \mathcal{A}_k} \prod_{i \epsilon A} (1-\theta)^{x_i} \theta^{m_i - x_i} \prod_{i \epsilon A^c} \phi^{x_i} (1-\phi)^{m_i - x_i} \\ = & \sum_{k=0}^{n} p^k (1-p)^{n-k} \sum_{A \epsilon \mathcal{A}_k} (1-\theta)^{t(A)} \theta^{m(A) - t(A)} \phi^{t(A^c)} (1-\phi)^{m(A^c) - t(A^c)} \ . \end{aligned} \tag{6}$$

Hence the posterior probability of having obtained $k_1$ false positives and $k_2$ false negatives is

$$I(1)^{-1} \sum_{k=0}^{n} \sum_{A \epsilon \mathcal{A}_k(k_1, k_2)} \int_0^1 \int_0^1 \int_0^{1-\phi} p^k (1-p)^{n-k} \theta^{k_2} (1-\theta)^{t(A)} \quad \phi^{k_1} (1-\phi)^{m(A^c) - t(A^c)} \times$$

$$\pi_p(p) \pi_{\theta, \phi}(\theta, \phi) d\theta d\phi dp \tag{7}$$

where $\mathcal{A}_k(k_1, k_2) \subseteq \mathcal{A}_k$ contains those subsets $A$ for which $m(A) - t(A) = k_2$ and $t(A^c) = k_1$. Hence to calculate (7) we must first determine the $\mathcal{A}_k(k_1, k_2)$ for $k = 0, \ldots, n$ and this may be difficult for large $n$, $k_1$ and $k_2$. For low values of $k_1$ and $k_2$, however, (7) can be easily and exactly evaluated using the previously discussed transformation and quadrature methods.

## 6. Examples

**Example 1**. We generated $x_1, \ldots, x_{20}$ with $p = .12, \theta = .05, \phi = .2$ and $m_i = 10, i = 1, \ldots, 20$. The generated data, with their frequencies recorded in parentheses, are given by 10(1), 9(1), 5(1), 4(2), 3(2), 2(5), 1(7) and 0(1). Note that if these data were generated without repeated measurements the asymptotic relative bias would have been $[P(X = 1) - p]/p = [(1-\theta)p + (1-p)\phi - p]/p = 1.417$. Using these data and the restriction $\theta + \phi < 1$, the maximum likelihood (ML) and noninformative Bayes estimates of the parameters were obtained. For the Bayes estimates a Uniform(0,1) prior was placed on $p$ and an independent Dirichlet(1,1,1) prior was placed on $(\theta, \phi)$. The following ML estimates were obtained from these data:

$$p_{ML} = .1000 \qquad \theta_{ML} = .0500 \qquad \phi_{ML} = .2000.$$

The posterior means, for the noninformative prior, are recorded in column 1 of Table 1 as well as their posterior standard deviations. In column 1 of Table 2 we have recorded the prior means and standard deviations of these quantities. As expected the posterior means are convex combinations of the prior means and the ML estimates. The marginal posterior densities for $p, \theta$ and $\phi$ and the joint density for $(\theta, \phi)$ are given in Figures 1, 2, 3 and 4 and were computed using Gauss-Jacobi quadrature.

In this example the quadrature approach for computing posterior quantities was quite successful. For example, in calculating the posterior mean of $p$, we require $(n_1, n_2, n_3) = (11, 101, 101)$ for exact computations. We find that with $(n_1, n_2, n_3) = (10, 22, 22)$ we obtain at least 5 digit accuracy and require only .25 minutes of computing time on a Sun Sparcstation.

The Gibbs sampling algorithm was also implemented and several different choices of $M$ and $N$ were tried. When uniform priors were employed and we asked for approximately 3 decimals of accuracy the

following results were obtained. With $M = 1$, and $N = 50,000$ Gibbs sampling required approximately 55 minutes of computing time and produced the estimate .13824 of the posterior mean of $p$ with estimated standard error .00033. With $M = 5$ and $N = 10,000$ again 55 minutes of computing time were required and gave the estimate .13835 with estimated standard error .00072. The estimates of the standard error indicate that very little autocorrelation is present. Hence a considerable amount of computation is necessary to obtain a reasonable level of accuracy.

While Gibbs sampling is much less efficient than Gauss-Jacobi quadrature in this example, it has the virtue of being much simpler to implement. In many contexts this is a significant argument in its favour. If the computation time is a relevant consideration, however; e.g. we have many such analyses to do or we are putting together a package of subroutines which will be accessed by many users, then clearly it is worth the extra effort to implement the quadrature approach. Another argument that might be advanced in favour of the Gibbs sampling algorithm is the dependence of the Gauss-Jacobi approach on the class of priors we have used. The Gibbs algorithm has the appearance of being more flexible as the basic algorithm is applicable no matter what prior is selected. This somewhat overstates the case, however, as efficient algorithms to generate from the conditionals, which we have with the priors used in this paper, are not necessarily available and, more seriously, the convergence of the algorithm can be a difficulty. For example, early experimentation with this model when particular, independent Beta priors were placed on each of $p$, $\theta$ and $\phi$, without the restriction $\theta + \phi < 1$, lead to a multimodal posterior and the consequent lack of convergence within meaningful computation times for Gibbs sampling. As is well-known, Gibbs sampling can have difficulty when the posterior is multimodal, see for example Evans, Guttman and Olkin(1993). By contrast the Gauss-Jacobi approach is again applicable and has no difficulty with such a posterior.

Of course we are not arguing in favour of a quadrature approach generally to the integration problems faced in implementing Bayesian inference. In many cases, e.g. characterized by high-dimensionality, quadrature is out of the question and a Monte Carlo integration algorithm is an absolute necessity. Implementing any integration algorithm so that it produces reliable results over a wide class of situations is difficult. When the structure of the problem can be exploited to avoid such difficulties then it seems natural to exploit this. For this reason we advocate the Gauss-Jacobi approach for the class of integration problems we have been discussing.

Next we experimented with different choices of prior means and variances of $(p, \theta, \phi)$ by appropriate choices of the parameters $v_1, \ldots, v_5$. The posterior means and standard deviations, as calculated using Gauss-Jacobi quadrature, are listed in columns 2-9 of Table 1. The prior means and standard deviations of these quantities are given in the corresponding columns of Table 2. Note that we can recover the parameters of the Beta for $p$ and the Dirichlet for $(\theta, \phi)$ from these quantities.

The contents of Table 1 show, as expected, that higher prior means produced higher posterior means. For instance a value of prior mean for $p$ equal to 1/2 produced posterior means in the range of .1364-.1393, while a prior mean of 1/9 produced posterior means in the range of .1035-.1097. Likewise, for a given prior mean of 1/2 or 1/9 for $p$, the smaller the prior mean placed on $\theta$, the smaller the posterior mean for $\theta$. The same is true for $\phi$. Note that the posterior means are convex combinations of the ML estimates and the corresponding prior means with the minor exception of $E(\theta)$ in column (4). Also the smaller the prior variance, the closer is the posterior mean to the corresponding prior mean.

**Example 2**. We consider the two-sample problem where 5 patients are diagnosed sick in a sample of

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| E($p$) | .1393 | .1049 | .1372 | .1366 | .1364 | .1035 | .1095 | .1091 | .1097 |
| SD($p$) | .0734 | .0565 | .0721 | .0717 | .0716 | .0556 | .0296 | .0296 | .0251 |
| E($\theta$) | .0990 | .0963 | .0666 | .0507 | .0500 | .0500 | .0962 | .0500 | .0500 |
| SD($\theta$) | .0703 | .0667 | .0462 | .0349 | .0172 | .0172 | .0664 | .0172 | .0172 |
| E($\phi$) | .2023 | .2027 | .1948 | .2000 | .2000 | .2000 | .2027 | .2000 | .2000 |
| SD($\phi$) | .0300 | .0297 | .0287 | .0283 | .0224 | .0224 | .0297 | .0224 | .0224 |

Table 1: Posterior means and standard deviations for example 1.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| E(p) | .5000 | .1111 | .5000 | .5000 | .5000 | .1111 | .1111 | .1111 | .1111 |
| SD(p) | .2887 | .0994 | .2887 | .2887 | .2887 | .0994 | .0329 | .0329 | .0269 |
| E($\theta$) | .3333 | .3333 | .1000 | .0500 | .0500 | .0500 | .3333 | .0500 | .0500 |
| SD($\theta$) | .2357 | .2357 | .0905 | .0476 | .0184 | .0184 | .2357 | .0184 | .0184 |
| E($\phi$) | .3333 | .3333 | .1000 | .2000 | .2000 | .2000 | .3333 | .2000 | .2000 |
| SD($\phi$) | .2357 | .2357 | .0905 | .0873 | .0337 | .0337 | .2357 | .0337 | .0337 |

Table 2: Prior means and standard deviations for example 1.

8 from population 1 and 3 patients are diagnosed sick in a sample of 11 from population 2. For both populations we specify a Beta(2,2) prior for $p$ and a Dirichlet(1,1,9) prior for $(\theta, \phi)$.

We calculated several posterior quantities of interest for $\gamma = p_1 - p_2$. Here the Gauss-Jacobi quadrature approach gives instantaneous and accurate results. For example, the posterior mean and standard deviation of $\gamma$ are given by .24688 and .21813 respectively and $P(\gamma \leq .2 \mid \underline{x}_1, \underline{x}_2)$ is given by .40252. These values were obtained following the method of the discussion of section 4 and by noting that the prior parameters $v_1$ and $v_2$ are integers for both populations. Experimentation with different priors lead to conclusions similar to those in Example 1.

## 7. Conclusions

It is seen that the analysis of binary data subject to misclassification is well handled by Bayesian methods with the prior structure we have specified. Further the calculations associated with this are efficiently carried out using Gauss-Jacobi quadrature. Exact expressions are available for both posterior moments and posterior probability contents. When very high order rules are required for exact expressions, it is often the case that lower order rules provide adequate approximations. A feature of quadrature is that the accuracy of a particular rule can be assessed by repeating the calculations with a higher order rule. The Gibbs sampling algorithm is easily implemented for this class of problems. Our analysis here indicates that Gibbs sampling works well in these problems but is considerably less efficient than the Gauss-Jacobi approach. Fortran software used in the analysis of binary data subject to misclassification is available from the authors upon request.

## References


Bross, I. (1954). Misclassification in 2x2 tables. Biometrics 10, 478-486.

Davis, P.J. and Rabinowitz, P. (1984). Methods of Numerical Integration, Second Edition. Academic Press.

Deming, W.E.(1977). An essay on screening, or two phase sampling applied to surveys of a community. Int. Stat. Rev., 45, 29-37.

Devroye, L. (1986). Non-Uniform Random Variate Generation. Springer-Verlag, New York.

Evans, M., Guttman, I. and Olkin, I. (1993). Numerical aspects in estimating the parameters of a mixture of normal distributions. J. of Comp. and Graph. Stat., Vol. 1, 4, 351-365.

Fleiss, J.L. (1973). Statistical Methods for Rates and Proportions. John Wiley and Sons, New York.

Gaba, A. and Winkler, R.L. (1992). Implications of errors in survey data: a Bayesian model. Manag. Sci., 38, 7, 913-925.

Gelfand, A.E. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. JASA, 85, 398-409.

Gelfand, A.E., Hills, S.E., Racine-Poon, A. and Smith, A.F.M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. JASA, 85, 972-985.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence. 6, 721-741.

Haitovsky, Y. and Zelen, M. (1990). Conditional resampling for estimating success probabilities from binomial data with misclassification. Research report.

Kleinbaum, D.G., Kupper, L.I. and Morgenstern, H. (1982). Epidemiological Research. LifeTime Learning Publications, London.

Matchar, D.B., Simel, D.L., Geweke, J.F. and Feussnor, J.R. (1990). A Bayesian method for evaluating medical test operating characteristics when some patients fail to be diagnosed by the reference standard. Medical Decision Making, Hanley and Belfus, Inc.

Stroud, A.H. and Secrest, D.H. (1966). Gaussian Quadrature Formulas. Prentice-Hall, Englewood Cliffs, New Jersey.

Tenenbein, A. (1970). A double sampling sheme for estimating from binomial data with misclassifications. JASA, 65, 1350-1361.

Tenenbein, A. (1971). A double sampling scheme for estimating from binomial data with misclassifications : sample size determination. Biometrics, 27, 935-944.

Tenenbein, A. (1972). A double sampling scheme for estimating from misclassified multinomial data with applications to sampling inspection. Technometrics, 14, 187-202.