# INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

# Bayesian Approaches To Modeling The Conditional Dependence Between Multiple Diagnostic Tests

Nandini Dendukuri
Department of Epidemiology and Biostatistics
McGill University, Montréal
September, 1998

A Thesis submitted to the
Faculty of Graduate Studies and Research
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy

# CONTENTS

# List of Figures

# List of Tables

# STATEMENT OF ORIGINALITY

Although there is an enormous literature on statistical methods for diagnostic test data, there are no frequentist solutions that directly address the problem of estimating parameters in the presence of three or less correlated tests. To our knowledge there is also no literature discussing a Bayesian solution to this problem, even for identifiable cases. This thesis has addressed this gap in the literature. Therefore, the models developed in Chapters 4 and 5 appear for the first time, as does the analysis of the Strongyloides data set in Chapter 6.

Of special note is the intended audience for which this thesis was written. In keeping with the spirit of a multi-disciplinary Department of Epidemiology and Biostatistics, we have included sufficient background material on Bayesian analysis so that the thesis could be read by a practicing epidemiologist. Conversely, we have also provided introductory definitions of terms relating to the diagnostic testing situation, so that the interested statistician with no epidemiology background should have no trouble.

# ABSTRACT

The differential diagnosis of a disease is often based on the information obtained from multiple diagnostic tests or multiple applications of the same test. The usual assumption in such situations is that the test results are statistically independent within each subject conditional on knowing the true disease status. This assumption greatly simplifies the statistical analysis of such data. In practice, however this assumption may be violated, as for example when there is a certain subject-related characteristic that may increase or decrease the probability of detection in two or more tests. The classical or frequentist solutions that account for the correlation between tests require a minimum of four different tests to obtain an identifiable solution. However, it is not always possible to have results from four different tests, particularly when tests are expensive, time consuming or invasive. Our objective in this thesis is to draw simultaneous inferences about the prevalence and test parameters while adjusting for the possibility of conditional dependence between tests, particularly in the situation when we have three or fewer tests, leading to a non-identifiable problem. We do so by way of a Bayesian approach, which utilizes available information about the prevalence and test parameters summarized in the form of prior distributions. The first of the two methods we propose models the dependence as a direct effect between each pair of tests. The second method uses a random effects model and simulates the dependence between tests via their sensitivity and specificity which are modeled as functions of a latent, subject-specific 'disease intensity'. Both models are based on dichotomous tests and the parameters are estimated using a Gibbs Sampler. It was found that ignoring the conditional dependence between tests could lead to misleading estimates of the sensitivities and specificities of the tests and of disease prevalence. Therefore, the methods presented here may improve inferences from surveys which are designed to provide estimates of the prevalence of disease in a particular population, when correlations among the diagnostic tests used may be present.

# RÉSUMÉ

Le diagnostic d'une maladie est souvent basé sur l'information obtenue de plusieurs tests ou plusieurs applications d'un même test. Dans un tel cas on assume généralement que les résultats de tests sont statistiquements indépendants pour chaque sujet à condition de connaître le véritable état de la maladie. Cette supposition simplifie grandement l'analyse statistique de telles données. En pratique, toutefois, cette supposition peut être violée lorsqu'il y a, par exemple, une certaine caractéristique reliée au sujet qui peut augmenter ou diminuer la probabilité de détection dans le cas de deux tests ou plus. Les solutions classiques ou fréquentistes qui tiennent compte de la corrélation entre ces tests requièrent un minimum de quatre tests différents pour obtenir une solution identifiable. Cependant, il n'est pas toujours possible d'obtenir les résultats de quatre tests différents, surtout lorsque ces tests sont onéreux, longs, ou stressants. Notre objectif dans cette thèse est de tirer des conclusions simultanées au sujet de la prévalence de la maladie étudiée et des paramètres des tests en tenant compte de la dépendance conditionnelle possible entre les tests, particulièrement dans la situation où nous avons un problème non-identifié. Nous procédons selon l'approche bayesienne, laquelle utilise l'information disponible au sujet de la prévalence et des paramètres de tests résumése sous la forme de distributions *a priori*. La première des deux méthodes que nous proposons modélise la dépendance comme un effet direct entre chaque paire de tests. La seconde utilise un modèle à effet "random effects" et simule la dépendance entre les tests de par leur sensibilité et leur spécifité lesquelles sont modélisées comme des fonctions de la latence ou la sverité de la maladie de chaque sujet. Les deux modèles sont basés sur des tests dichotomiques et les paramètres sont estimés en utilisant un échantilloneur de Gibbs. Ne pas prendre, en consideration la dépendance conditionnelle entre les tests peut mener à des estimations erronées de la sensibilité et spécifité des tests ainsi que de la prévalence de la maladie. Par conséquent, les méthodes présentées ici peuvent donner une meilleure inférence á partir de données de sondage conçus pour estimer de la prévalence d'une maladie dans une population donnée, lorsque des corrélations peuvent être présentes.

# ACKNOWLEDGMENTS

First and foremost, I would like to express my deep sense of gratitude to my supervisor, Dr. Lawrence Joseph, for his invaluable advice and unstinted encouragement which enabled me to complete the Ph.D. program.

I am sincerely grateful to McGill University for awarding me the honour and privilege of the Max Stern Recruitment Fellowship, which provided me financial support for three years (1995-98).

I would like to record my thanks to the Faculty at the Department of Epidemiology and Biostatistics, McGill University. In particular, I would like to mention Dr. James Hanley, Dr. Jean-Francois Boivin and Dr. Theresa Gyorkos, who have counselled me on several occasions. I extend my thanks to the students of the Department for creating a professional and friendly atmosphere that made my stay there truly memorable.

I express my sincere appreciation to Dr. Harold Roussel for his timely assistance in all the computer-related aspects of the work involved in this thesis.

Last but not the least, I am truly thankful to my family, particularly my parents, who provided me sustained motivation to undertake and complete this course of study.

# 1

# INTRODUCTION: THE DIAGNOSTIC TESTING PROBLEM

Historically, medical diagnoses have been made on the basis of subjective knowledge gathered from the medical history and observed symptoms of the patient. In recent decades this has been augmented by the more methodical process of obtaining objective information from diagnostic tests. Thus, in the framework of medical decision making, diagnostic tests have come to occupy an important role and consequently their appropriate analysis is a very active area of biostatistical research today.

In this thesis we take up the specific problem of modeling the conditional dependence between multiple diagnostic tests, especially in the absence of a gold standard or reference test. Two or more diagnostic tests may be conditionally dependent when their results are related due to a factor other than the disease status. Such a dependence could occur, for instance, between tests which are based on the same underlying principle or between results from the same test at two different points in time. Several authors have demonstrated that it is important to account for this dependence while analyzing the results from diagnostic tests, in order to obtain unbiased estimates of the prevalence of disease and test parameters (Fryback, 1978; Vacek, 1985; Brenner, 1996; Torrance-Rynard and Walter, 1997). While there has been much recent work in this area, the problem of how to analyze diagnostic test data when the tests are correlated and when there is no perfect (gold standard) test remains to be solved,

especially when the number of tests available is less than 5, as is usually the case in practice. This problem is especially difficult since, as we will see, it is non-identifiable, *i.e.* there is not sufficient information available to obtain a unique estimate of all the parameters involved. Nevertheless, the frequency at which it occurs (often unrecognized) motivated a serious look at the problem in this thesis.

This introductory chapter covers the basic concepts behind diagnostic tests, the parameters used to evaluate their performance, the utility of multiple testing and the conditional dependence between tests. We also provide the objectives of the thesis and an outline of the forthcoming chapters.

## *1.1 Diagnostic Tests: Definition and performance parameters*

Diagnostic tests are routinely used in public health, community medicine and clinical medical practice to help gain more information about a patient's or a group of patients' condition, and to separate subjects into classes with different probabilities of disease. A test is typically determined by:

1. A **separator** variable, which is a measurable property of the subject, associated with the disease of interest, and
2. A **positivity criterion**, which is a particular value of the separator variable that divides subjects into different disease categories.

In order to diagnose the presence of hypertension, for example, a possible separator variable is the average diastolic blood pressure over three successive readings. A diastolic blood pressure of 90mm of Hg could be used as the positivity criterion so that patients with an average diastolic blood pressure greater than this value would

be diagnosed as hypertensive. It is common to dichotomize a continuous separator variable, using a single positivity criterion, such that there are only two possible test results - positive or negative. Throughout this thesis, we will be dealing only with dichotomous tests. Apart from simplifying the exposition of the problem, the use of dichotomous tests is motivated by the fact that in reality most medical decisions are dichotomous: to operate or not to operate, to prescribe a drug or not to prescribe a drug, etc (Weinstein et al., 1980). Our methods, however, could straightforwardly be extended to the case where the test results are multi-categorical, or even continuous.

Diagnostic tests can be of varied formats - questionnaires, biochemical tests, genetic tests, radiographic tests, and so on. Whatever their format, tests are seldom perfect. That is, they do not always correctly diagnose the subject's true disease status. This is usually because it is not possible to find a separator variable which clearly demarcates subjects into diseased and non-diseased categories. For instance, we could hypothesize that the separator variable follows a different distribution for the diseased and non-diseased subjects as illustrated in Fig 1.1. Subjects who fall

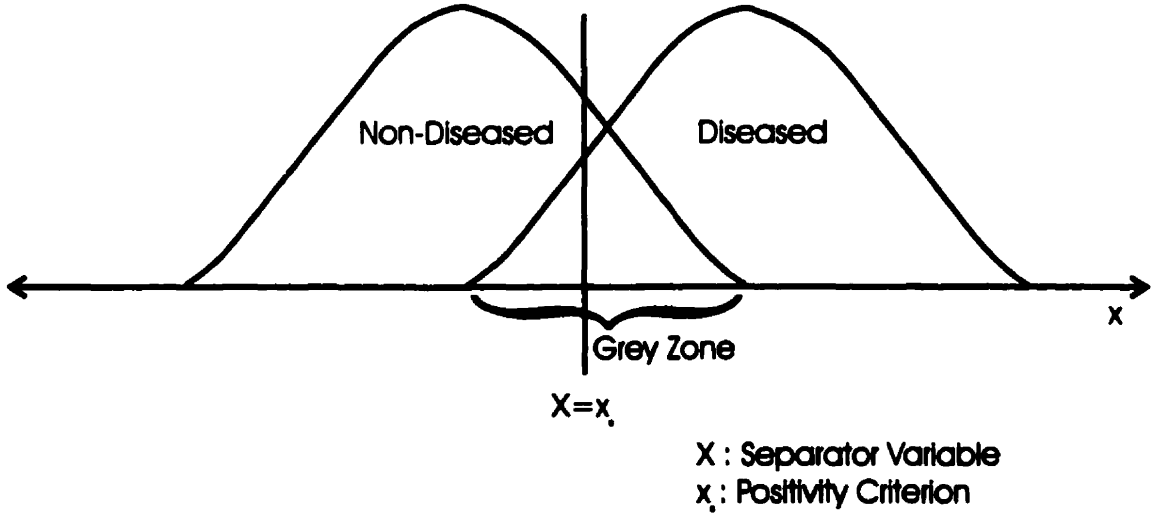


Figure 1.1: Grey zone where subjects are likely to be misclassified.

in the 'grey zone', where the two distributions overlap, may be misclassified because both diseased and non-diseased subjects can have test results that lie in this range.

Misclassification can also occur due to random or systematic measurement error. Random errors can occur even when we have a technically perfect test. In the case of a stool examination, for instance, a subject is diagnosed as positive if the disease-causing parasite is detected in the stool. An error might occur because the subject's diet makes it difficult to detect the parasite, because the technician analyzing the stool makes an error, or because the instrument used to detect the parasite is deficient. Sometimes, however, the test itself is invalid. For example, when the level of serum glucose, a separator variable for myocardial infarction, is measured using a miscalibrated instrument that adds on 5 units to every measurement, subjects who fall below the positivity criterion of 100 mg/100ml may be systematically misclassified as positive.

Despite their imperfection, many tests are routinely used in clinical diagnosis, screening and epidemiological studies. The correct interpretation of test results is dependent on knowledge of the population under study and the parameters which characterize test performance. The parameters of primary interest in the diagnostic setup are the prevalence of the disease in the population, and test properties such as sensitivity and specificity of the test, which are defined as follows:

The **prevalence** is the proportion of truly diseased subjects in the population of interest. Throughout this thesis it will be denoted by $\pi = P(D = 1)$, where $D = 1$ denotes positive disease status. The probability of being non-diseased, $D = 0$, is then given by $P(D = 0) = 1 - P(D = 1) = 1 - \pi$.

The **sensitivity** is the probability that a subject who is truly diseased will be correctly diagnosed by the test as being positive. In other words, it is the conditional probability of testing positive given that the disease is present. In probabilistic terms

this can be written as $S = P(T = 1|D = 1)$, where $T = 1$ indicates that the result of the test is positive.

The **specificity** of the disease is the probability that a subject who is not diseased will test negative or the conditional probability of testing negative given that the disease is absent. This is denoted by $C = P(T = 0|D = 0)$, where $T = 0$ indicates that the result of the test is negative.

Ideally we would like both the sensitivity and specificity of a test to be as high as possible. Theoretically a perfect test would have $S = C = 1$. In practice however, even the most accurate test available does not have $S = C = 1$. The best test is usually called the **gold standard** test, even though it may not be 100% accurate. Detecting diseased subjects is clearly of importance, but it is also of interest to reduce the number of false positives especially when a positive test is followed by a costly or risky intervention. However, if either the sensitivity or specificity is increased, usually by altering the positivity criterion, the other is almost always automatically decreased. The following example illustrates the estimation of the prevalence and test parameters in the situation when the test results can be compared to those from a 'perfect' gold standard test.

**Example 1.1:** Shapiro et al., 1974 obtained the data presented in Table 1.1 from a study on screening for breast cancer. The extreme right column contains the results of a screening test which consisted of a physical examination and mammography. The last row contains the 'gold standard' results which were obtained when these women were followed for a year subsequent to the screening and some of them were diagnosed with breast cancer. We can see, from the cross-classification in Table 1.1, that there are subjects who were incorrectly classified as negative by the screening test and a large number of subjects who were falsely classified as positive by the screening test, when in fact they never went on to develop the disease. The population prevalence and test parameters for the screening test can be estimated, by assuming the gold

| | | Breast Cancer | | Total |
|---|---|---|---|---|
| | | Cancer confirmed | Cancer not confirmed | |
| Screening test | Positive | 132 | 983 | 1115 |
| | Negative | 45 | 63650 | 63695 |
| | Total | 177 | 64633 | 64810 |

Table 1.1: Cross-classification of results from screening test and the 'true' diagnosis at the end of one year of follow-up.

standard to represent the truth, as follows:

$$\text{Prevalence} = \frac{\text{No. of subjects who were truly diseased}}{\text{Total no. of subjects}} = \frac{177}{64810} = 0.27\,\%,$$

$$\text{Sensitivity} = \frac{\text{No. of truly diseased subjects who tested positive}}{\text{No. of subjects who were truly diseased}} = \frac{132}{177} = 74.6\,\%,$$

$$\text{Specificity} = \frac{\text{No. of truly non-diseased subjects who tested negative}}{\text{No. of subjects who were truly non-diseased}} = \frac{63650}{64633} = 98.5\,\%.$$

It is important to note at this point, that it would not have been possible to obtain these estimates in the absence of a gold standard test. What could we infer, for instance, when the gold standard was known not to be the truth or when we have two or more non-gold standard tests whose results are in conflict? Further, what if these tests are correlated and do not provide independent information? In this thesis

we will deal simultaneously with all these problems.

The prevalence, sensitivity and specificity are parameters of greater interest to public health practitioners and policy makers. Two other related parameters, which are more meaningful to the clinician in interpreting test results for a single patient, are the positive predictive value and the negative predictive value, which are defined as follows:

The **positive predictive value** is the probability that a subject who has tested positive actually has the disease, and is denoted by $PV+ = P(D = 1|T = 1)$. In other words, the $PV+$, is the conditional probability of being truly diseased given that the subject tested positive. In some instances we use the notation $PVP$ for the predictive value positive.

The **negative predictive value** is the probability that a subject who has tested negative is in fact disease-free and is denoted by $PV- = P(D = 0|T = 0)$. Hence the $PV-$ is the conditional probability of being truly non-diseased given that the subject tested negative. In some instances we use the notation $PVN$ for the predictive value negative.

The positive and negative predictive values of a test can be shown to be functions of the prevalence, sensitivity and specificity as follows:

$$PV+ = P(D = 1|T = 1) = \frac{P(D = 1)P(T = 1|D = 1)}{P(D = 1)P(T = 1|D = 1) + P(D = 0)P(T = 1|D = 0)}$$

$$= \frac{P(D = 1)P(T = 1|D = 1)}{P(D = 1)P(T = 1|D = 1) + (1 - P(D = 1))(1 - P(T = 0|D = 0))}$$

$$= \frac{\pi S}{\pi S + (1 - \pi)(1 - C)}.$$

Similarly,

$$PV- = P(D=0|T=0) = \frac{P(D=0)P(T=0|D=0)}{P(D=1)P(T=0|D=1)+P(D=0)P(T=0|D=0)}$$

$$= \frac{(1-P(D=1))P(T=0|D=0)}{P(D=1)(1-P(T=1|D=1))+(1-P(D=1))P(T=0|D=0)}$$

$$= \frac{(1-\pi)C}{\pi(1-S)+(1-\pi)C}.$$

Hence, when interpreting the result of a test for a single individual, it is important to know the prevalence of the disease and the sensitivity and specificity of the test, in the population to which the individual belongs.

**Example 1.1 continued:** Using the breast cancer data in Table 1.1 we have:

$$PV+ = \frac{132}{1115} = 11.8\%, \quad \text{and}$$

$$PV- = \frac{63650}{63695} = 99.9\%.$$

It is interesting to note that this test has been designed such that it has a fairly high sensitivity of 74.6%. This is probably because breast cancer is a fatal disease and it is important that as few cases as possible go undetected. This, however, comes at the cost of a poor positive predictive value. A subject in this population who has a positive result on the screening test has only an 11.8% chance of actually developing breast cancer.

## 1.2 *Multiple diagnostic tests*

The differential diagnosis of disease is rarely based on the results of a single test. In order to improve the accuracy of the diagnosis, physicians often order more than one

diagnostic test or more than one application of the same diagnostic test. The gold standard procedure may in fact be a set of two or more tests which together provide more accurate information.

**Example 1.2:** In order to estimate the prevalence of Strongyloides infection among a group of refugees, Joseph et al., 1995 used two commonly available tests with complementary characteristics - a stool examination and a serological test. The stool examination has a very poor sensitivity of 24%, and a high specificity of 95% The serology test, on the other hand, has a relatively higher sensitivity of 81% but a lower specificity of 72%. Though neither test is a gold standard, their combined results help to improve the accuracy of diagnosis. We will return to this example later in the thesis.

When multiple tests are used, the performance of two or more tests may be related due to a variable other than the disease status. This could happen, for instance, when two tests are based on the same biological phenomenon, when two questionnaires contain overlapping items, or when the two tests are in fact replications of the same test at two different times. Such a similarity between a pair of tests may be measured using their **covariance** within each disease class. We denote the covariance between two tests, $T_1$ and $T_2$, among the diseased and non-diseased subjects as $covp_{12}$ and $covn_{12}$, respectively. In probabilistic terms this would be expressed as:

$$covp_{12} = E(T_1T_2|D=1) - E(T_1|D=1)E(T_2|D=1), \text{ and}$$
$$covn_{12} = E(T_1T_2|D=0) - E(T_1|D=0)E(T_2|D=0),$$

where $E(X)$ represents the expectation of the random variable $X$. In the event when two tests are conditionally *independent*, *i.e.* independent within a disease class, the covariance between them in that disease class is 0. The concept of conditional dependence between tests is dealt with in greater detail in the next sub-section. Throughout the thesis we use the terms conditional dependence and correlation interchangeably.

### 1.2.1 *Identifiability*

The parameter $\theta$ which indexes the probability distribution $F_\theta$ is said to be identifiable if $F_{\theta_1} \neq F_{\theta_2}$ when $\theta_1 \neq \theta_2$. More simply, a problem is said to be identifiable when it has a unique solution, *i.e.* when the number of degrees of freedom of the observed data is equal to or greater than the number of unknown parameters to be estimated.

Figure 1.2 is a diagrammatic representation of all the parameters involved in the diagnostic testing problem in the general situation when we have p tests - $\pi$ denotes

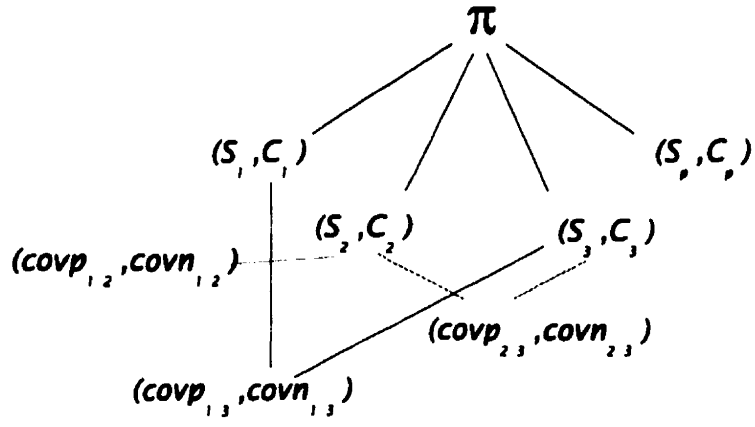


Figure 1.2: Parameters involved in the multiple diagnostic tests problem.

the population prevalence, $S_j$ and $C_j$ denote the sensitivity and specificity of the $j^{th}$ test, $j = 1, \ldots, p$, and $covp_{kl}$ and $covn_{kl}$, $k, l = 1, \ldots, p$, $k \neq l$, denote the covariance between the tests $k$ and $l$ among the positively and negatively diseased subjects, respectively. The methods developed in this thesis are in the context of a community situation where two or more tests are applied to each of $N$ ($> 1$) subjects.

The total number of parameters involved in the diagnostic testing problem illustrated in Figure 1.2 is $(2 \times p) + (2 \times {}^pC_2) + 1$. The number of degrees of freedom available, which is determined by the number of possible cross-classifications of test

results, is $2^p - 1$. Assuming that none of the $p$ tests is a gold standard, then, in order for the problem to be identifiable, it is required that,

$$(2 \times p) + (2 \times {}^pC_2) + 1 \leq 2^p - 1,$$
$$\Rightarrow p \geq 5,$$

*i.e.* results from a minimum of 5 tests should be available. However, results from five tests may not always be available, especially when the tests have to be applied to several subjects and when they are expensive, invasive or time-consuming. In this thesis we develop methods which can be used even when the problem is non-identifiable.

### *1.2.2 Conditional independence*

From elementary probability theory we know that when two events are independent, the probability that they occur jointly is given by the product of their individual probabilities. Mathematically, if $A$ and $B$ are independent events, $P(A \cap B) = P(A)P(B)$. This definition can be extended to the situation when the events of interest are conditional on another event. For instance, we may be interested in the results of two diagnostic tests conditional on the disease status, *i.e.* $P(T_1 = t_1|D = d)$ and $P(T_2 = t_2|D = d)$. If the two tests are conditionally independent, given $D = d$ it means that

$$P(T_1 = t_1, T_2 = t_2|D = d) = P(T_1 = t_1|D = d)P(T_2 = t_2|D = d),\ \forall t_1, t_2.$$

This means that among the subjects for whom $D = d$ the results of $T_1$ have no bearing on the results of $T_2$. This need not necessarily hold if the disease status changes, *i.e.* the two tests may be independent among the diseased but not among the non-diseased subjects, or vice-versa.

When analyzing the results from multiple tests their joint distribution is often needed. For example, suppose that a subject must test positive on all $p$ tests used

in a study in order to qualify for the intervention. To estimate the total cost of this intervention in a certain population, we would need to know the probability of obtaining a positive result on all tests. From the law of total probability, this is given by

$$\begin{aligned} & P(T_1 = 1, T_2 = 1, \ldots, T_p = 1) \\ & = P(D = 1)P(T_1 = 1, T_2 = 1, \ldots, T_p = 1|D = 1) \\ & \quad + P(D = 0)P(T_1 = 1, T_2 = 1, \ldots, T_p = 1|D = 0). \end{aligned} \tag{1.1}$$

To calculate the probability in equation (1.1), we need the joint distribution of the $p$ tests conditional on the disease status. This information is seldom available and in order to circumvent the need for it, it is common to assume the tests to be conditionally independent. In this case, equation (1.1) can be written in terms of the sensitivity and specificity of each test, as follows

$$\begin{aligned} & P(T_1 = 1, T_2 = 1, \ldots, T_p = 1) \\ & = P(D = 1)P(T_1 = 1|D = 1)P(T_2 = 1|D = 1) \ldots P(T_p = 1|D = 1) \\ & \quad + P(D = 0)P(T_1 = 1|D = 0)P(T_2 = 1|D = 0) \ldots P(T_p = 1|D = 0) \\ & = \pi S_1 S_2 \ldots S_p + (1 - \pi)(1 - C_1)(1 - C_2) \ldots (1 - C_p). \end{aligned} \tag{1.2}$$

However, the conditional independence assumption is not always realistic, since test results may be correlated within each disease class. When the conditional independence assumption is invalid, it could result in biased estimates of the parameters of interest as the following example illustrates.

**Example 1.3:** This example is modified from one used in the book *Clinical Decision Making* by Weinstein et al., 1980 (page 155). At a certain hospital, reno-vascular disease among hypertensive patients is diagnosed using two tests - an intravenous pyelogram ($IVP$) and a renogram ($RG$). The sensitivity and specificity of the two tests are listed in Table 1.2.

| Test | $Sensitivity(S)$ | $Specificity(C)$ |
|---|---|---|
| $IVP$ | 0.78 | 0.89 |
| $RG$ | 0.85 | 0.90 |

Table 1.2: Sensitivity and specificity of tests used to diagnose reno-vascular disease.

Patients testing positive on both tests are subjected to a costly and invasive surgical intervention. Hence it is of importance to assess the probability of a positive result. In the absence of any information about their joint distribution, the probability of obtaining a positive result on both tests may be estimated using equation (1.2), assuming they are conditionally independent. Using the information that the prevalence of renovascular disease among hypertensive patients is 10%, we have

$$\begin{aligned}
&P(IVP = 1, RG = 1) \\
&= \pi P(IVP = 1, RG = 1|D = 1) + (1 - \pi)P(RG = 1, IVP = 1|D = 0) \\
&= \pi P(IVP = 1|D = 1)P(RG = 1|D = 1) \\
&\qquad + (1 - \pi)P(IVP = 1|D = 0)P(RG = 1|D = 0) \\
&= \pi S_{IVP}S_{RG} + (1 - \pi)(1 - C_{IVP})(1 - C_{RG}) \\
&= (0.1)(0.85)(0.78) + (1 - 0.1)(1 - 0.9)(1 - 0.89) \\
&= 0.0762.
\end{aligned}$$

In the case of $IVP$ and $RG$, however, the information on their joint distribution happens to be available and is presented in Table 1.3. So the probability of obtaining a positive result on both tests would in fact be

$$\begin{aligned}
P(IVP = 1, RG = 1) &= P(D = 1)P(IVP = 1, RG = 1|D = 1) \\
&\qquad + P(D = 0)P(RG = 1, IVP = 1|D = 0) \\
&= (0.1)(0.69) + (0.9)(0.08) \\
&= 0.141.
\end{aligned}$$

| | $D = 1$ | | $D = 0$ | |
|---|---|---|---|---|
| Test result | $IVP = 1$ | $IVP = 0$ | $IVP = 1$ | $IVP = 0$ |
| $RG = 1$ | 0.69 | 0.16 | 0.08 | 0.02 |
| $RG = 0$ | 0.09 | 0.22 | 0.03 | 0.87 |

Table 1.3: Cross-classification of results from the intravenous pyelogram test and the renogram test.

This means that the cost of the study intervention will be almost twice as great as the conditional independence assumption would suggest.

We will return to this issue later in the thesis in Chapters 4 and 5, where we show how assuming tests are conditionally independent when it is not the case, could lead to biased estimates of the prevalence and test parameters.

## 1.3 *Objectives of the thesis*

Commonly used methods for the analysis of results from multiple diagnostic tests assume that these tests are conditionally independent as this simplifies the process. As mentioned in the previous section, this may not always be the case. A more realistic model would take into account the dependence between tests within each disease class. Such a model, however, is non-identifiable when we have four or fewer tests, as seen in Section 1.2.1. Hence, classical frequentist solutions to this problem require that a minimum of four tests be used in order to solve the problem directly (See Chapter 3). But results from four tests may not always be possible in practice. With this in mind, the main objective of this thesis is stated as follows:

To develop methodology for Bayesian inference about the prevalence and all test parameters in the situation where multiple tests or replications of tests are used, while adjusting for the conditional dependence between them, particularly when the

problem is non-identifiable. More specifically, we will:

1. Develop a fixed effects model for conditionally dependent diagnostic tests.
2. Develop a random effects model for conditionally dependent diagnostic tests.
3. Demonstrate how these models may work in practice by both simulations and application to real data. This step is especially important given the non-identifiable nature of the problem.

While the main focus of this thesis involves the development of statistical methodology, we have tried to present the material here in a manner such that it can be read by biostatisticians and epidemiologists alike. In particular, this chapter has reviewed the basic notions of diagnostic tests, while the next chapter will provide the necessary statistical background.

## 1.4 *Organization of the thesis*

The outline of the thesis is as follows:

Chapter 2 provides background material on the statistical concepts which form the foundation of the work developed here. We introduce the concept of Bayes' rule and some Bayesian computational techniques, namely the Sampling Importance-Resampling ($SIR$) algorithm and the Gibbs sampler, which will be used later in the thesis.

Chapter 3 presents a brief review of the literature on methods used to analyze results from diagnostic tests. These include the Bayes conditional independence method and more recent methods using latent class analysis. The problem of non-identifiablility and the advantage of the Bayesian approach in providing a solution

for such problems is discussed. We then describe some frequentist methods that have been developed to model the conditional dependence between diagnostic tests when five or more tests are available. Finally we present a summary of some Bayesian methods that have been used in the analysis of diagnostic test data.

In Chapter 4 we describe how conditional dependence between tests affects test results and how it may be measured using the covariance between pairs of tests. We then formulate a fixed effects model using the covariance to model the dependence between tests, and describe a Bayesian approach for its solution. In Chapter 5 we use random effects to model the conditional dependence between tests and once again propose a Bayesian approach to draw inferences for the parameters of this model. Here, the test sensitivities and specificities are taken to be functions of a subject-specific 'intensity' which is a latent or unobserved variable. The dependence between test results is induced by this additional variable without explicit reference to the covariance.

In Chapter 6, the methods of Chapters 4 and 5 are applied to the results from two diagnostic tests conducted in a group of Cambodian refugees to determine the prevalence of Strongyloides infection. Finally, we end with a summary chapter on our conclusions and suggestions for future research.

# 2

# PRELIMINARIES: BAYESIAN ANALYSIS AND COMPUTATIONAL TECHNIQUES

In this chapter we provide a brief discussion of Bayesian analysis which is fundamental to the methods developed in this thesis. This is followed by a section describing some computational tools for Bayesian inference with examples illustrating their usage.

## *2.1 Bayesian analysis*

Over the last three decades, Bayesian statistical analysis has come to represent an important alternative to the classical or frequentist school of thought. A primary motivation for the use of Bayesian techniques is that they facilitate a common-sense interpretation of statistical conclusions. Further, these methods are flexible and can be used to model very complex problems, where recent computational advances have made Bayesian inference feasible (see section 2.2).

Frequentist statistics considers unknown parameters as fixed, and examines the behavior of data-based statistics as the data are imagined to change over the sample space. For example, a frequentist 95% *confidence* interval around a parameter is interpreted as 'a random interval (the randomness coming from its endpoints varying across different data sets other than that observed) which would capture the true parameter 95% of the time in repeated applications across different experiments'.

Thus, inferences are indirect at best, since the interval at hand is based on a single realization of the experiment and it is not known whether it falls among the 95% of intervals which correctly capture the true parameter.

The Bayesian approach, on the other hand, treats unknown parameters as random and involves drawing inferences conditional on the observed data and quantifying the uncertainty in the inference using probability. A Bayesian 95% *probability* interval is literally an interval where the parameter of interest has a 95% probability of being located. This direct interpretation, as we will see below, comes at the cost of requiring the specification of a prior distribution over all unknown parameters, which means that its probability statements are interpreted subjectively. A detailed comparison of the two schools of thought can be found in Berger, 1985. See Gelman et al., 1995 for an introduction to data analysis from a Bayesian point of view.

### 2.1.1 *Bayes' Theorem*

Bayes' rule, which is pivotal to the use of Bayesian methods, can be summarized as follows: Let us consider the general situation where we are interested in drawing inferences about $\theta$, which is a parameter characterizing the distribution of the observable variable $Y$. In the Bayesian framework, $\theta$ is treated as a random variable having a distribution $f_\theta(\theta)$. This distribution, which is termed the **prior distribution**, represents the information available on $\theta$ prior to observing $Y = y$. Statistical conclusions about $\theta$ are then expressed in terms of the probability of $\theta$ conditional on the observed values of $y$, $f_{\theta|Y}(\theta|y)$, which is called the **posterior distribution**.

The basic idea behind Bayesian thinking is to pool together the information from the prior distribution $f_\theta(\theta)$ and the **likelihood** function $f_{Y|\theta}(Y = y|\theta)$ of the observed value of $Y = y$, to obtain the posterior distribution $f_{\theta|Y}(\theta|Y = y)$ using the following

relation which is called Bayes' theorem of conditional probability:

$$f_{\theta|Y}(\theta|y) = \frac{f(\theta,y)}{f(y)} = \frac{f_\theta(\theta)f_{Y|\theta}(y|\theta)}{\sum_\theta f_\theta(\theta)f_{Y|\theta}(y|\theta)}. \tag{2.1}$$

In the case when $\theta$ is continuous, $f(y) = \int_{-\infty}^{\infty} f(\theta)f(y|\theta)d\theta$. Thus the primary task in developing a Bayesian solution to any specific problem is to select a suitable model $f(\theta, y)$ and to find $f(\theta)$ which accurately summarizes the available information about $\theta$. An equivalent form of equation (2.1) omits the factor $f(y)$, which does not depend on $\theta$ and can hence be considered a normalizing constant. We then have

$$f(\theta|y) \propto f(\theta)f(y|\theta) \tag{2.2}$$

For the sake of brevity we will often use this *unnormalized* form of the posterior density in this thesis.

### 2.1.2 *Prior distributions*

The use of prior distributions has been controversial, and is the leading issue in the debate between the frequentist and Bayesian schools of statistical thought. On the one hand, in theory it is very attractive to summarize all past information into a prior distribution, and update it with the information in the current data set to arrive at a posterior density which summarizes all that is known about the set of parameters under investigation. The great problem, of course, is that in practice the choice of prior distributions is virtually never unique and different choices of prior distributions will lead to different posterior densities and possibly different conclusions. Spiegelhalter et al., 1994, suggest providing posterior densities across the range of reasonable prior densities in the context of reporting results from clinical trials. This common sense approach can be applied to areas other than clinical trials as well, offering a partial solution to the problem.

In non-identifiable problems, however, one has the choice of either carrying out a Bayesian analysis, or changing the problem to an identifiable one, and solving

the simpler problem via a frequentist approach. As we will see, the main problem addressed in this thesis is non-identifiable. In particular, we will address the problem of estimating the test properties and population prevalence from two tests in the presence of conditional dependence between the tests. Therefore, only the Bayesian approach offers a direct solution to this problem. Nevertheless, it is particularly important to select prior densities carefully in non-identifiable problems since their effects can be 'everlasting'. From the central limit theorem, Bayesian and frequentist approaches offer numerically similar inferences as the sample size increases across a wide class of problems. This is because the information in the prior density becomes overwhelmed by that in the data as the sample size increases. However, this does **not** happen in the case of non-identifiable problems, where final inferences greatly depend on the choice of prior density, even with large sample sizes.

*Elicitation of prior distributions*

As discussed above, one of the most important steps in any Bayesian analysis is the process of determining the distribution that accurately summarizes the information available prior to the experiment. This information is typically gathered from previous studies or from subjective, expert opinion. Since prior elicitation tends to be elictee- and application-specific, it is difficult to automate this process. Several methods that have been suggested to streamline the process are reviewed in Chaloner, 1996 and Wolfson, 1995.

One simple method is to divide the range of the parameter of interest, $\theta$, into intervals and assign relative probabilities to each of these intervals in a manner that reflects the experimenter's beliefs. While this method might seem natural when $\theta$ is discrete, it quickly becomes complicated as the range of $\theta$ increases or when it is continuous. Alternatively, we could assume that the prior distribution comes from a family of well known parametric distributions $f(\theta|\eta)$, where $\eta$ is fixed so that the

resulting distribution matches our prior beliefs as closely as possible. For example, if $\eta$ were two dimensional, as in the case of the *Normal* distribution, then specification of two moments (such as the mean and the variance) or two quantiles (such as the $5^{th}$ and $95^{th}$ percentiles) would suffice to determine $\eta$. Sometimes a prior belonging to a particular distributional family may simplify the computation of the posterior distribution. In particular, when the likelihood of the observed data belongs to the exponential family of distributions, it is possible to select a *conjugate* prior which would result in a posterior distribution which belongs to the same family as the prior.

**Example 2.1** Suppose that $X_1, \ldots, X_n$ are independent and identically distributed as $Binomial(N, p)$. The likelihood of the data can be written as

$$\begin{aligned} L(X_1, \ldots, X_n|p) &\propto \prod_{i=1}^{n} p^{X_i}(1-p)^{N-X_i}, \\ &= p^{\sum_{i=1}^n X_i}(1-p)^{nN-\sum_{i=1}^n X_i} \end{aligned}$$

A reasonable choice of a prior density for $p$ is the $Beta(\alpha, \beta)$ prior density such that

$$f(p) \propto p^{\alpha-1}(1-p)^{\beta-1}, \qquad 0 \leq p \leq 1,\ \alpha > 0, \beta > 0.$$

This prior distribution is defined over the entire range of possible values of $p$ and has the further advantage of taking on several shapes over this range. Using Bayes' theorem we obtain the posterior density of $p$ as follows

$$\begin{aligned} f(p|X_1, \ldots, X_n) &\propto L(X_1, \ldots, X_n|p) f(p), \\ &\propto p^{\sum_{i=1}^n X_i}(1-p)^{nN-\sum_{i=1}^n X_i} p^{\alpha-1}(1-p)^{\beta-1}, \\ &= p^{\sum_{i=1}^n X_i+\alpha-1}(1-p)^{nN-\sum_{i=1}^n X_i+\beta-1}. \end{aligned} \tag{2.3}$$

The above equation is proportional to a *Beta* density with parameters $\alpha' = \sum_{i=1}^n X_i + \alpha$ and $\beta' = N^2 - \sum_{i=1}^n X_i + \beta$. Since this is the only function proportional to equation (2.3) which integrates to 1, the posterior density of $p$ is indeed $Beta(\alpha', \beta')$ and the *Beta* is the conjugate family for the *Binomial* likelihood. See Berger, 1985 and Carlin and Louis, 1996 for more examples of conjugate families.

Given that all the parameters in the diagnostic testing problem are continuous, for the methods developed in this thesis we have selected priors of standard distributional forms, utilizing conjugate distributions wherever feasible. While this method facilitates the conversion of prior information into parameters, a drawback is that in attempting to force the available information into the form of a standard distribution, one may end up with a prior distribution that may not exactly match the available information. Also, there may be more than one distribution, belonging to different families, which conform to the prior beliefs of the experimenter but result in very different posterior densities. In such a case it may not always be clear which prior should be used, so that reporting results across a range of reasonable prior densities is again indicated.

When there is no prior information available on $\theta$ or when we want to draw inferences based on the data alone, we could use a *non-informative* or *diffuse* prior, *i.e.* one where the data dominates any information in the prior. Several methods which have been proposed for the construction of such priors are discussed in (Carlin and Louis, 1996). One of the most common methods, which we shall employ later in the thesis, is to use a uniform distribution over the range of $\theta$.

## 2.2 *Bayesian computational techniques*

Until recently, Bayesian analysis was not frequently used in practice, because it often involved the integration of complex functions for which there is no analytical solution. Sophisticated numerical analysis techniques used to solve such problems require lengthy calculations. Over the last decade, however, Bayesian analysis has gained increasing importance in applied statistical analysis because of the availability of numerical Monte Carlo methods for sampling from the distribution of interest without actually having to first derive the exact posterior density. The availability of fast com-

puters which can execute these methods within a reasonable time frame has resulted in an explosion of applied Bayesian research. In this section we discuss two such methods which are used later in the thesis - the Sampling Importance Resampling (SIR) algorithm and the Gibbs sampler.

### 2.2.1 *The Sampling-Importance-Resampling (SIR) algorithm*

The SIR algorithm, which was discussed by Rubin, 1988, is particularly useful when it is difficult to obtain the analytical form of the distribution of interest, $g(x)$, or even to simulate a sample from it, but where there exists a distribution, $h(x)$, which is absolutely continuous with respect to $g(x)$ and is easier to sample from. The SIR consists of the following steps:

1. Draw a sample of size $n$, $(x_1, x_2, \ldots, x_n)$ from the 'proposal' distribution $h(x)$.
2. Assign weights $w_1, w_2, \ldots, w_n$ to each corresponding $x_i$ such that $w(x_i) = \frac{g(x_i)}{h(x_i)}$, $i = 1, \ldots, n$.
3. Finally, draw a new random sample of size $m$, $x_1^*, x_2^*, \ldots, x_m^*$ with replacement from the discrete distribution over $x_1, x_2, \ldots, x_n$ with probabilities proportional to $w_1, w_2, \ldots, w_n$.

The resulting sample $x_1^*, x_2^*, \ldots, x_m^*$ is approximately an independent sample from $g(x)$. For a quick proof of this, see Smith and Gelfand, 1992. Clearly, increasing $n$ and $m$ increases the accuracy of the estimates. Theoretically, any distribution, including a uniform density over the range of $x$, can be used as the 'proposal' distribution $h(x)$. However, the more closely $h(x)$ resembles $g(x)$, the smaller the value of $n$ required to obtain a good approximation of $g(x)$, that is, the closer $x_1^*, x_2^*, \ldots, x_m^*$ will resemble a random sample from $g(x)$.

**Example 2.2:** In a hypothetical community study it was of interest to estimate the prevalence $\pi$, of the disease $X$. For the purpose of illustration we assume the exact sensitivity and specificity of the test are known to be $S = 0.95$ and $C = 0.85$. Of the 1000 subjects tested for disease $X$, 270 tested positive. The probability of obtaining a positive test is given by:

$$\begin{aligned} P(T=1) &= P(D=1)P(T=1|D=1) + P(D=0)P(T=1|D=0) \\ &= \pi S + (1-\pi)(1-C) \end{aligned}$$

The likelihood function can now be written as:

$$\begin{aligned} L &\propto P(T=1)^{270}(1-P(T=1))^{1000-270} \\ &= (0.95\pi + (1-\pi)(1-0.85))^{270}(1-(0.95\pi + (1-\pi)(1-0.85))^{730} \end{aligned}$$

A review of earlier studies, say, suggested that the prior distribution of $\pi$ in the population was $Beta(10, 90)$. Therefore the posterior distribution of $\pi$ is given by

$$\begin{aligned} g(\pi|data) \propto\ & (0.95\pi + (1-\pi)(1-0.85))^{270}(1-(0.95\pi + (1-\pi)(1-0.85))^{730} \\ & \times \pi^{10-1}(1-\pi)^{90-1} \end{aligned} \tag{2.4}$$

This function can of course be integrated, numerically or otherwise, to obtain the posterior density or a close approximation. Nevertheless, we will use the SIR algorithm for illustrative purposes. The function in equation (2.4) is not of the form of any common distribution, but a SIR algorithm can be used to obtain a sample from it. We select the $Uniform(0, 1)$ as the proposal distribution, which means $h(\pi) = 1$. This is not necessarily the best choice, but is adequate for the purpose of this example and guarantees absolute continuity with respect to $g(\pi|data)$. We then proceed as follows:

1. Draw $m$ values $\pi_1, \ldots, \pi_m$ from $Uniform(0, 1)$

2. Assign a weight $w(i)$ to each $\pi_i$, such that

$$w(i) = g(\pi_i) = \pi_i^{10-1}(1-\pi_i)^{90-1}(\pi_i 0.95 + (1-\pi_i)(1-0.85))^{270}$$
$$\times(1-(0.95\pi_i + (1-\pi_i)(1-0.85))^{730} \quad i = 1\ldots, m$$

3. Sample $n$ values $\pi_1^*, \ldots, \pi_n^*$ from $\pi_1, \ldots, \pi_m$ with weights proportional to $w_1, \ldots, w_m$.

It was found that the posterior median and 95% posterior probability interval of $\pi$ were 0.1404 and (0.1118,0.1662), respectively. The proposal distribution $h(\pi)$ and prior and posterior distributions for $\pi$ are illustrated in Figure 2.1, where $m$ and $n$ were taken to be 1000 and 2000, respectively. The plot for the posterior density was obtained by smoothing the histogram of the posterior sample using the *ksmooth* function in S-plus.

### 2.2.2 *The Gibbs sampler*

The Gibbs sampler, see for example, Geman and Geman, 1984, comes from the class of Markov Chain Monte Carlo (MCMC) procedures. These methods are based on Monte Carlo integration using Markov chains and have been used to simplify a wide class of high-dimensional integration problems, especially in Bayesian analysis. The Gibbs sampler is the most commonly used of the MCMC techniques, and is the fundamental tool for the inferential methods developed in this thesis.

Consider the following situation: We are interested in the marginal posterior distributions of $n \geq 2$ variables $X_1, X_2, \ldots, X_n$. However, it is their so-called 'full conditional distributions' of the form $f(X_i|X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$, $i = 1, \ldots, n$ that are known or are easier to sample from. This form is usually easily available, up to a normalizing constant, for each $X_i$, $i = 1, \ldots, n$ by multiplying its prior density

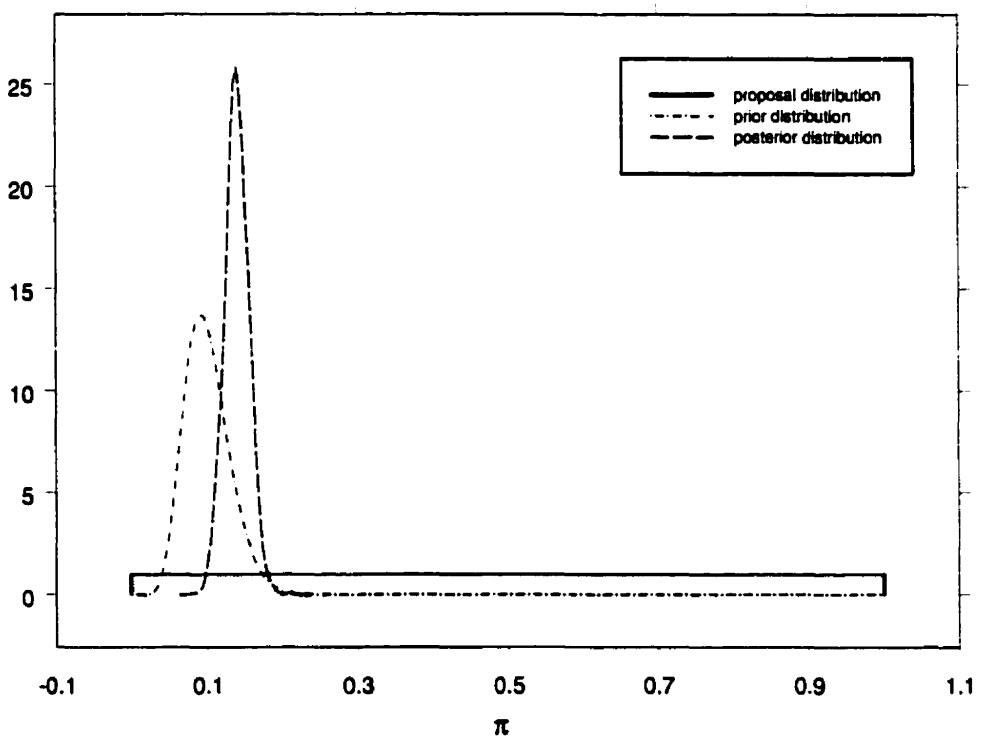


Figure 2.1: Proposal and prior distributions overlaid by posterior distribution obtained using a SIR.

with the likelihood function as in equation (2.2) and considering all variables except $X_i$ as constants.

The Gibbs sampler can be used to break down a multivariate problem into a series of smaller dimensional problems. It works by simulating a sample from each of the marginal posterior distributions using the following steps:

1. Assign starting values to all variables such that $X_1 = x_1^{(1)}$, $X_2 = x_2^{(1)}$, ..., $X_n = x_n^{(1)}$.

2. Draw a single value $x_1^{(2)}$ for $X_1$ from the full conditional distribution $f(X_1|X_2 =$

$x_2^{(1)}, \ldots, X_n = x_n^{(1)})$, *i.e.*, the distribution of $X_1$ conditional on the remaining variables $X_2, \ldots, X_n$.

3. Repeat Step 2 for all $n$ variables. For example, draw $x_i^{(2)}$ for $X_i$ from the distribution $f(X_i|X_1 = x_1^{(2)}, \ldots, X_{i-1} = x_{(i-1)}^{(2)}, X_{i+1} = x_{(i+1)}^{(1)}, \ldots, X_n = x_n^{(1)})$

A single iteration is completed when a value has been drawn for each of the $n$ variables. This procedure is repeated, say, $N$ times where $N$ is typically several thousand. The sample $(x_{i1}, \ldots, x_{iN})$, $i = 1, \ldots, n$ thus obtained is a possibly correlated random sample from the marginal distribution of $X_i$, and can be used to obtain posterior inferences about the unknown parameters. For a description of the Markov chain concepts related to the Gibbs sampler see Roberts, 1996.

**Example 2.3:** To illustrate the Gibbs sampler we use the example from the classic paper on data augmentation by Tanner and Wong, 1987. Based on a genetic linkage model, 197 animals are considered to be distributed into 4 categories following a multinomial distribution such that $y = (y_1, y_2, y_3, y_4) = (125, 18, 20, 34)$, with corresponding probabilities $(\frac{1}{2} + \frac{\theta}{4}, \frac{(1-\theta)}{4}, \frac{1-\theta}{4}, \frac{\theta}{4})$, where $0 \leq \theta \leq 1$. The variable $y$ is transformed by splitting the first cell into two cells with probabilities $\frac{1}{2}$ and $\frac{\theta}{4}$. The transformed variable $x = (x_1, x_2, x_3, x_4, x_5)$, is such that $x_1 + x_2 = y_1 = 125$, $x_3 = y_2 = 18$, $x_4 = y_3 = 20$ and $x_5 = y_4 = 34$. We can think of $x_2$ as having a Binomial distribution such that, $x_2 \sim Bin(\frac{\theta}{\theta+2}, 125)$. The likelihood function of the observed data is

$$f(y|\theta) \propto (2+\theta)^{y_1}(1-\theta)^{y_2+y_3}\theta^{y_4},$$

which can be rewritten in terms of the $x$ as

$$f(x|\theta) \propto \theta^{x_2+x_5}(1-\theta)^{x_3+x_4}$$

If the prior distribution of $\theta$ is taken to be $Uniform(0, 1)$, the posterior distribution

of $\theta$ would be equal to the normalized likelihood, and could be expressed as

$$f(\theta|x) \propto \theta^{x_2+x_5}(1-\theta)^{x_3+x_4}$$
$$\Rightarrow \theta|x \sim Beta(x_2+34, 38)$$

Thus, we know the distribution of $x_2$ conditional on $\theta$ and the distribution of $\theta$ conditional on $x_2$, but not the marginal distribution of either variable. We can use the Gibbs sampler to obtain random samples from the marginal distributions of $x_2$ and $\theta$ as follows:

1. Start with an arbitrary initial value $\theta = \theta^{(1)}$
2. Draw $x_2 = x_2^{(1)}$ from $Bin(125, \frac{\theta^{(1)}}{\theta^{(1)}+2})$
3. Draw $\theta^{(2)}$ from $Beta(x_2^{(1)}+34, 38)$

Repeat steps two and three until a sufficiently large sample from the full conditional distribution of $\theta$ is obtained. The posterior median and 95% posterior probability interval of $\theta$ were found to be 0.6273 and (0.5241,0.7221), respectively. The prior and posterior distributions for the parameter $\theta$ are seen in Figure 2.2. Here $x_2$ is considered as 'latent' or unobserved data. We will use a similar technique for deriving marginal posterior densities of latent parameters in Chapters 4 and 5.

### 2.2.3 *Diagnostics for the Gibbs sampler*

As with many other statistical methods, once a technique has been selected and the parameters of interest have been estimated, it is important to ensure that the technique has operated as expected. For an MCMC chain this means assessing when the procedure has 'converged' and when it can be safely terminated. In other words, here we want to be sure that the samples obtained come from the true stationary

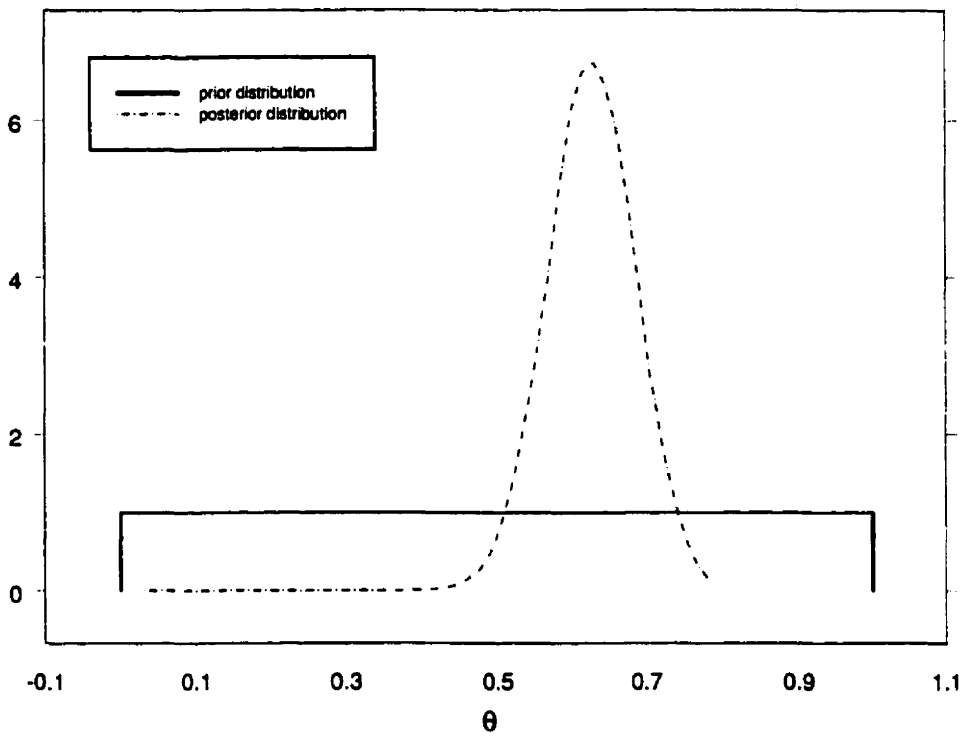


Figure 2.2: Prior distribution overlaid by posterior distribution obtained using a Gibbs sampler

distribution, *i.e.* the true joint posterior distribution, and that a sufficiently large sample is collected for an accurate approximation of the joint and marginal posterior densities. This is difficult because what is produced by the algorithm at convergence is not a single number or even a distribution, but a *sample* from a distribution. Furthermore, the algorithm is complex, and, of course, we usually do not have the true posterior densities against which to compare our approximations. Theoretically, an accurately modeled problem will eventually converge after an infinite number of iterations (Gelfand and Smith, 1990). However, given the complexity of most MCMC problems, each iteration is expensive in terms of computer time and it is desired to minimize the required number of iterations. Although several methods

have been proposed, 'MCMC convergence diagnostics' remains an area which is still being actively researched. A comparative review of some commonly used methods is presented in the paper by Cowles and Carlin, 1996.

In this section we summarize two such methods which we employed to assess the performance of the methods developed in this thesis. Both these methods are available as part of a software package called CODA (developed by the MRC Biostatistics unit at the University of Cambridge). In addition to these two methods, the auto-correlation within each chain and the cross-correlation between different parameters can be utilized to evaluate convergence.

*Method I: Raftery and Lewis*

Raftery and Lewis, 1992, have proposed a method which will detect slow convergence to the stationary distribution as well as provide a way of bounding the precision of the variance of the posterior quantiles. This method determines (a) the number of *burn-in* iterations, $M$, to be discarded in the initial part of the chain before it converges, (b) the number of further iterations, $N$, to be run in order to obtain the desired precision and (c) the *thinning interval*, $k$, which is the number of iterations to be discarded between two successive, retained, independent iterations to obtain a sequence of independent random iterates. While considerations (a) and (c) are not absolutely mandatory for the operation of a Gibbs sampler algorithm, they help use the computer memory more efficiently. The first step involves running a pilot chain of length $N_{min}$, which is the minimum number of iterations that would be required for the desired accuracy if the samples were independent. The value of $N_{min}$ is determined as a function of the quantile $q$, that is of interest to be estimated with precision $r$, with $s$ being the probability of attaining the specified accuracy, such that

$$N_{min} = \{\Phi^{-1}(\frac{s+1}{2})\}^2 \frac{q(1-q)}{r^2}$$

where $\Phi(.)$ is the cumulative distribution function for the standard normal distribution. The results from the pilot run are entered into CODA to obtain estimates of $M$, $N$ and $k$. A large value of $M$ indicates slow convergence and a large value of $N$ which is greater than $N_{min}$ or equivalently a value of $k$ greater than 1 suggests strong autocorrelations within the chain. In addition we can also calculate the dependency factor

$$I = \frac{M+N}{N_{min}}$$

which measures the increase in the number of iterations due to the dependence in the sequence. Values of $I$ much greater than 1 suggest a high level of dependence in the model and values greater than 5 suggest that the implementation (often the parameterization) of the problem may need to be changed. The diagnostics of Raftery and Lewis, 1992, while helpful, do not guarantee that convergence has occurred by $M$ iterations, or that $N$ iterations will necessarily have the desired accuracy. This is because the method assumes that a function of the quantiles of interest follow a Markov chain, and uses the ergodic theorem to derive a 'sample size' for the accurate estimation of each marginal posterior density, and uses the BIC criterion (see Kass and Raftery, 1995) to determine the required burn-in. Both of these are approximations, so that this method should be used with caution.

*Method II: Gelman and Rubin*

The method developed by Gelman and Rubin, 1992 addresses the problem of undiagnosed slow convergence (Gelman and Meng, 1996). This could happen, for instance, when successive observations in a chain are highly correlated or when the model is overparametrized. This in turn may prevent proper 'mixing' of the chain, giving a false impression of convergence. Such a problem may not be visible by viewing a single trace plot which is the plot of the successive observations vs iteration numbers. The method consists of two parts as follows:

1. Observe the trace plots of multiple, say $m = 5$, parallel sequences of length $n$ starting at pre-determined points which are well dispersed over the range of the target distribution. By ensuring that the starting points are well dispersed it will be possible to detect if the MCMC eventually identifies the correct mode(s) of the stationary distribution each time. All $n$ trace plots are then overlaid to see if the individual sequences can be distinguished after eliminating the burn-in iterations.

2. The second step involves the calculation of a quantitative measure which checks if the empirical distribution of simulations obtained separately from each chain is approximately the same as the empirical distribution obtained when the sequences are mixed together. This is done by comparing the within-sequence and between-sequence variance for each parameter. For each parameter of interest, say $\psi$, the parallel chains are labeled $\psi_{ij}, i = 1, \ldots, m, j = 1, \ldots, n$ and the two quantities of interest are calculated as

$$B = \frac{n}{m-1}\sum_{i=1}^{m}(\bar{\psi}_{i.} - \bar{\psi}_{..})^2, \text{ where } \bar{\psi}_{i.} = \frac{1}{n}\sum_{j=1}^{n}\psi_{ij}, \quad \bar{\psi}_{..} = \frac{1}{m}\sum_{i=1}^{m}\bar{\psi}_{i.} .$$

$$W = \frac{1}{m}\sum_{i=1}^{m}s_i^2, \text{ where } s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(\psi_{ij} - \bar{\psi}_{i.})^2 .$$

The two variance components are used to construct the ratio of an overestimate and an underestimate of the posterior variance as follows

$$\hat{R} = \frac{\frac{n-1}{n}W + \frac{1}{n}B}{W} .$$

As $n \to \infty$, or in other words when convergence is reached in all sequences, $\hat{R}$ will be equal to or very close to 1.

When the overlaid sequences are distinguishable or $\hat{R}$ is very different from 1 (say ¿ 1.2) the model may require reformulation. There has been an enormous amount of research into the properties of MCMC algorithms since the popularizing paper of

Gelfand and Smith, 1990. See the recent book by Gilks et al., 1996, and the references therein for a path into this literature.

## 2.3 Summary

In this chapter we have seen the basic concepts behind the use of Bayesian analysis as they will be applied in Chapters 4 and 5 of this thesis. An important step in any Bayesian analysis is collecting accurate prior information and then determining the distribution which most closely matches our prior beliefs. This step is of particular importance when we have a non-identifiable problem since in these problems, even with a very large sample size, the influence of the informative prior distributions can remain great. Also important to any Bayesian analysis is the process of determining the posterior distributions of the parameters of interest. This hitherto complicated process has now been simplified by computational techniques such as the SIR and the Gibbs sampler.

In the next chapter we review some of the methods which have been developed for the analysis of diagnostic tests, particularly those which provided the background for the methods developed in this thesis.

# 3

# A Review of the Literature

Over the last four decades a great deal of research has been done to develop statistical methods for the analysis of diagnostic test data. The result is a choice of several models using a variety of statistical techniques including Bayesian methods, generalized linear models and latent class analysis. This chapter is a brief review of this literature. Some of the techniques presented here provide the motivation for the methodologies developed in the remaining chapters of the thesis while others are mainly of historical interest.

## *3.1 The Bayesian independence model*

The Bayesian independence model, introduced by Ledley and Lusted, 1959, was among the earliest formal methods for the analysis of diagnostic tests. Since then it has enjoyed great popularity in computer-aided medical diagnosis algorithms and other areas of statistical pattern recognition. Though this method has no direct relation to those developed in this thesis, we discuss it here since it involves Bayes' theorem and purports to analyze simultaneously results from multiple, independent tests.

The outline of the method is as follows: Let $d_1, \ldots, d_q$ denote $q$ mutually exclusive and exhaustive, well-defined disease conditions. The tests $T_1, \ldots, T_p$ are $p$ tests used

to diagnose a patient's disease condition, such that the $j^{th}$ test takes on $l_j$ distinct values, $1, 2, \ldots, l_j$. For example, if $T_1$ is 'chest pain', with $l_1 = 3$ it could take values 1, 2 and 3 indicating none, radiating and non-radiating pain. Each new patient is thus a realization of the random variable $(D, T)$ where $D$ is the true disease status and $T = (T_1, \ldots, T_p)$ is the vector of test results. The diagnostic problem now is to observe $T = t$ and infer from it the value of $D$. This means observing

$$t = (t_1, \ldots, t_p),$$
$$t \in \{1, \ldots, l_1\} \times \ldots \times \{1, \ldots, l_p\}.$$

and estimating the probability $P(D = d_k|t)$, $k = 1, \ldots, q$. The desired probability can be calculated using Bayes' theorem as follows

$$P(D = d_k|T = t) = \frac{P(D = d_k)P(T = t|D = d_k)}{\sum_{k'=1}^{q} P(D = d_{k'})P(T = t|D = d_{k'})}, \quad k = 1, \ldots, q. \qquad (3.1)$$

It is assumed that valid estimates of the prevalence of each disease condition $P(D = d_k)$, $k = 1, \ldots, q$, and the test performance parameters $P(T_j = t_j|D = d_k)$, $j = 1, \ldots, p$, $k = 1, \ldots, q$ are available from studies conducted earlier. A fundamental feature of the algorithm is that it hypothesizes that results from the $p$ tests are conditionally independent. This implies

$$P(T = t|D = d_k) = P(T_1 = t_1|D = d_k)P(T_2 = t_2|D = d_k) \ldots P(T_p = t_p|D = d_k),$$
$$k = 1, \ldots, q,$$

which simplifies the calculation of the joint probability of the $p$ tests. In practice, it means that if the true disease state is known, then the test results are independent of each other, *i.e.*, knowing the result of Test $j$ provides no information about Test $j'$, $j \neq j'$, if the disease state is known. Equation (3.1) can now be rewritten as

$$P(D = d_k|y) = \frac{P(D = d_k)P(T_1 = t_1|D = d_k) \ldots P(T_p = t_p|D = d_k)}{\sum_{k'=1}^{q} P(D = d_{k'})P(T_1 = t_1|D = d_{k'}) \ldots P(T_p = t_p|D = d_{k'})},$$

$$k = 1, \ldots, q.$$

The early literature on this method demonstrated much confusion about the use of the conditional independence assumption which was understood by many researchers to be a necessary condition for use of this method (Feinstein, 1977). Since conditional independence between diagnostic tests is seldom exactly achieved in practice, the method was criticized for being unrealistic. This gave way to much of the literature assessing the effects of the assumption of conditional independence on the parameter estimates of interest. Several authors have shown that the assumption can be relaxed, in that in many practical situations it does not materially affect the estimated probabilities of disease (Lincoln and Parker, 1967, Russek et al., 1983, Hilden, 1984).

In the light of the methods developed later in this thesis, it is of interest to note that though this method makes use of Bayes' theorem, it is not Bayesian in its inferential approach. In fact, there is not even the use of an interval to give an idea of the variability in the estimated probabilities. Point estimates are used for the prior information on disease prevalence $P(D = d_k)$ and test parameters $P(T_j = t_j|D = d_k)$, which are assumed exactly correct. As will be explained in Section 3.2.5, it is difficult to know these values exactly and a fully Bayesian approach would assign a prior distribution over the range of possible values for the prevalence and test properties which accounts for the uncertainty about the values of these parameters.

Early methods, such as this one, sought to identify symptoms associated with the presence or absence of disease, in order to develop an algorithm for predicting the probability of disease following an individual test result. Such an algorithm is usually developed by comparing test results with the true disease status or gold standard test results. Later methods, such as those described in the following sections, attempted to model data more realistically for situations where no gold standard test was available. The latter methods focussed more on the estimation of population level quantities such as disease prevalence.

## 3.2 *Latent class analysis*

Often we are interested in measuring a variable which cannot be observed directly. Consider, for example, 'religious commitment'. Such an unobservable or *latent* variable can be estimated using one or more observable indicators, for example 'frequency of church visits', 'perceived importance of religious beliefs', and so on (McCutcheon, 1990). *Latent Structure Analysis*, which was first introduced by Lazarsfeld in 1950, has been widely used by social scientists to model such problems and particularly to study the relations between the observed indicator variables (Lazarsfeld and Henry, 1968).

In recent years these methods have found application in modeling the results from diagnostic tests to estimate the prevalence and test parameters. In the diagnostic test situation, the observed variables would be the results from the diagnostic tests and the latent variable would be the true disease status, which in the absence of a gold standard test is unobservable. *Latent Class (LC)* problems are a subset of Latent Structure problems where the latent variable is discrete, taking a finite number of distinct values. For ease of exposition and notation, and, keeping in mind the context of this thesis, the discussion here is focussed on situations where both the observed and latent variables are dichotomous. The results are easily extended to address problems with more than two latent classes.

The basic premise of latent class analysis is that the relations between two or more observed variables are explained entirely by the latent variable. This can be stated in statistical terms as: two or more observed variables are independent conditional on the latent variable. In the case of the diagnostic testing problem, this would amount to saying: the results of two or more tests are independent within groups of the diseased and non-diseased subjects, or

$$P\left(T_1 = t_1, T_2 = t_2, \ldots, T_p = t_p | D = d\right)$$

$$= P(T_1 = t_1|D = d)\, P(T_2 = t_2|D = d) \ldots P(T_p = t_p|D = d).$$

Note that we do not expect independence to hold between groups, *i.e.* we do not expect that

$$P(T_1 = t_1, T_2 = t_2, \ldots, T_p = t_p) = P(T_1 = t_1)\, P(T_2 = t_2) \ldots P(T_p = t_p),$$

as it would imply poor performance of the diagnostic tests.

Following the notation described in Chapter 1, the probability of observing the vector $(t_1, \ldots, t_p)$ of test results can now be expressed as:

$$\begin{aligned} & P(T_1 = t_1, T_2 = t_2, \ldots, T_p = t_p), \qquad t_j = 0 \text{ or } 1,\ j = 1, \ldots, p \\ &= \sum_{d=0}^{1} P(D = d) P(T_1 = t_1, T_2 = t_2, \ldots, T_p = t_p|D = d) \\ &= \sum_{d=0}^{1} P(D = d) P(T_1 = t_1|D = d)\, P(T_2 = t_2|D = d) \ldots P(T_p = t_p|D = d). \end{aligned} \tag{3.2}$$

Equation (3.2) follows from the fact that there are two mutually exclusive, latent disease classes, $D = 0$ and $D = 1$, and, conditional on the disease status the results of the $p$ tests are independent. Hereafter we refer to the model with two latent classes denoting diseased and non-diseased as the $2LC$ or 'two latent class' model. In the next section we discuss how to set up the likelihood function for this model using these equations.

### 3.2.1 *Observed and complete likelihoods*

For ease of illustration let us consider the situation when we have $p = 2$ tests. Again, the results can easily be extended to the situation where results from more than two tests are available. The cross-classification of the observed results from two tests $T_1$ and $T_2$ can be summarized as in Table 3.1. Following the basic assumption of

| | $T_1 = 1$ | $T_1 = 0$ |
|---|---|---|
| $T_2 = 1$ | $n_{11}$ | $n_{01}$ |
| $T_2 = 0$ | $n_{10}$ | $n_{00}$ |

Table 3.1: Cross-classification of observed results from two dichotomous tests.

latent class analysis, the two tests are considered to be independent conditional on the disease status. Thus from equation (3.2) we see that the probability of observing $(T_1 = 1, T_2 = 1)$ is

$$\begin{aligned} P(T_1 = 1, T_2 = 1) &= P(D = 1)P(T_1 = 1|D = 1)P(T_2 = 1|D = 1) \\ &\quad + P(D = 0)P(T_1 = 1|D = 0)P(T_2 = 1|D = 0), \\ &= \pi S_1 S_2 + (1 - \pi)(1 - C_1)(1 - C_2). \end{aligned}$$

Similarly the probabilities of observing the other three cells are

$$\begin{aligned} P(T_1 = 1, T_2 = 0) &= \pi S_1(1 - S_2) + (1 - \pi)(1 - C_1)C_2, \\ P(T_1 = 0, T_2 = 1) &= \pi(1 - S_1)S_2 + (1 - \pi)C_1(1 - C_2), \\ P(T_1 = 0, T_2 = 0) &= \pi(1 - S_1)(1 - S_2) + (1 - \pi)C_1C_2. \end{aligned}$$

Using the above equations, the likelihood function of the observed data, $L_o$, can now be written as

$$\begin{aligned} L_o = \frac{(n_{11} + n_{10} + n_{01} + n_{00})!}{n_{11}!\, n_{10}!\, n_{01}!\, n_{00}!} & [\pi S_1 S_2 + (1 - \pi)(1 - C_1)(1 - C_2)]^{n_{11}} \\ & \times [\pi S_1(1 - S_2) + (1 - \pi)(1 - C_1)C_2]^{n_{10}} \\ & \times [\pi(1 - S_1)S_2 + (1 - \pi)C_1(1 - C_2)]^{n_{01}} \\ & \times [\pi(1 - S_1)(1 - S_2) + (1 - \pi)C_1C_2]^{n_{00}}, \\ & 0 \leq \pi, S_1, S_2, C_1, C_2 \leq 1. \end{aligned} \quad (3.3)$$

A helpful 'trick' to solving this problem is to imagine what would happen *if* we knew the number of truly diseased subjects in each of the four cells $(T_1 = 1, T_2 = 1)$,

| Diseased | | |
|---|---|---|
| | $T_1 = 1$ | $T_1 = 0$ |
| $T_2 = 1$ | $y_{11}$ | $y_{01}$ |
| $T_2 = 0$ | $y_{10}$ | $y_{00}$ |

| Not Diseased | | |
|---|---|---|
| | $T_1 = 1$ | $T_1 = 0$ |
| $T_2 = 1$ | $n_{11} - y_{11}$ | $n_{01} - y_{01}$ |
| $T_2 = 0$ | $n_{10} - y_{10}$ | $n_{00} - y_{00}$ |

Table 3.2: 'Complete' data from two dichotomous tests.

$(T_1 = 1, T_2 = 0)$, $(T_1 = 0, T_2 = 1)$ and $(T_1 = 0, T_2 = 0)$. If this were true, estimating the prevalence and test parameters would be straightforward as (in Example 1.1) when results from a gold standard test are available. This unobservable or latent data will be denoted by $y_{11}$, $y_{10}$, $y_{01}$ and $y_{00}$. We hypothesize that if the latent data were available, then the latent data along with the observed data would constitute the 'complete' data set as presented in Table 3.2.

The likelihood function of the 'complete' data is then given by

$$L_c = \frac{(n_{11}+n_{10}+n_{01}+n_{00})!}{y_{11}!\,(n_{11}-y_{11})!\,y_{10}!\,(n_{10}-y_{10})!\,y_{01}!\,(n_{01}-y_{01})!\,y_{00}!\,(n_{00}-y_{00})!}$$

$$\times [\pi S_1 S_2]^{y_{11}}\,[(1-\pi)(1-C_1)(1-C_2)]^{n_{11}-y_{11}}\,[\pi S_1(1-S_2)]^{y_{10}}$$
$$\times [(1-\pi)(1-C_1)C_2]^{n_{10}-y_{10}}\,[\pi(1-S_1)S_2]^{y_{01}}\,[(1-\pi)C_1(1-C_2)]^{n_{01}-y_{01}}$$
$$\times [\pi(1-S_1)(1-S_2)]^{y_{00}}\,[(1-\pi)C_1C_2]^{n_{00}-y_{00}},$$
$$0 \le \pi, S_1, S_2, C_1, C_2 \le 1. \quad (3.4)$$

The 'complete' and observed likelihoods are related by

$$L_o = \sum_{y_{11},y_{10},y_{01},y_{00}} L_c. \quad (3.5)$$

Equation (3.5) shows that the probability density associated with the conceptual 'complete' data may not be unique. In most cases, given $L_o$, the choice of the $L_c$ is guided by convenience. In the following sections we will show how the 'complete'

| Number of tests | p | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Number of parameters | $2p+1$ | 3 | 5 | 7 | 9 | 11 |
| Number of df | $2^p-1$ | 1 | 3 | 7 | 15 | 31 |

Table 3.3: Number of unknown parameters and available degrees of freedom as a function of the number of tests.

data likelihood can be used to obtain estimates of parameters when data are latent or missing.

### 3.2.2 *Identifiability conditions for the frequentist model*

The number of diagnostic tests used determines the number of cross-classifications into which the observed results are grouped and hence the number of degrees of freedom. In the general situation when we have results from $p$ conditionally *independent* diagnostic tests we have $p$ sensitivities, $p$ specificities and the prevalence that we are interested in estimating, *i.e.* we have a total of $2p+1$ parameters to estimate. Since each test can take two possible values, $T_j = 0$ or 1, $j = 1, \ldots, p$, there are $2^p$ possible cross-classifications of test results and hence $2^p - 1$ degrees of freedom, since the total sample size is fixed. Table 3.3 summarizes the number of unknown parameters and the number of degrees of freedom available as a function of the number of tests. From this table we can see that in the case when there are less than three conditionally independent tests the problem is non-identifiable and therefore a frequentist approach to the solution of such a problem will require applying certain constraints as will be discussed in Section 3.2.4. For a complete description of identifiability conditions in the general case when the latent and manifest variables have more than two exclusive classes see Goodman, 1974.

| | | Stool Examination | | |
|---|---|---|---|---|
| | | + | - | Total |
| Serology Test | + | 38 | 87 | 125 |
| | − | 2 | 35 | 37 |
| Total | | 40 | 122 | 162 |

Table 3.4: Results of tests for Strongyloides infection among a group of Cambodian refugees.

| | Parameter | Median | 95% CI |
|---|---|---|---|
| Stool Examination | $S_1$ | 0.24 | 0.07-0.47 |
| | $C_1$ | 0.95 | 0.89-0.99 |
| Serology Test | $S_2$ | 0.81 | 0.63-0.92 |
| | $C_2$ | 0.72 | 0.31-0.96 |

Table 3.5: Test parameters of stool examination and serology test.

### 3.2.3 *The Strongyloides infection problem*

In this section we outline a diagnostic test problem which could be modeled using latent class analysis. Following this we describe two methods - a frequentist and a Bayesian method - to estimate the parameters in the latent class model. We apply both methods to the problem described here.

Joseph et al., 1995 were interested in studying the prevalence of Strongyloides infection among a group of Cambodian refugees arriving in Montréal, Canada. They had available to them results from two imperfect tests - a serology test and a stool examination - as illustrated in Table 3.4. Assuming the two tests to be conditionally independent, we see from Table 3.3 that this problem is non-identifiable. Consultation with the literature showed that there was no gold standard test available with which to compare these two tests and hence there was great uncertainty about the performance

parameters of the tests, even though they were commonly prescribed. For instance, the specificity of the serology test was found to range anywhere from 35% to 100%! The prior median and 95% CI of the test parameters are summarized in Table 3.5. While the stool examination had a very poor sensitivity it had an extremely high specificity. The serology test on the other hand had a higher sensitivity but a poorer specificity than the stool examination. The study's main objective was to see if any interventions were required in this population when they immigrated to Canada.

### *3.2.4 A frequentist solution for the 2LC model*

Frequentist methods for parameter estimation require that a problem be identifiable in order to obtain a meaningful solution. Therefore, when the degrees of freedom are less than the number of unknown parameters, constraints must be added to the model by assuming some of the parameters to be known constants. For example, when we have two tests we need to fix the values of any two parameters in order to ensure that there are as many unknown parameters ($5 - 2 = 3$) to estimate as there are degrees of freedom (3) available. Different combinations of parameters may be held fixed depending on the context of the problem, as discussed in great detail in the review by Walter and Irwig, 1988. The remaining parameters are then estimated conditional on the values of the constrained parameters.

Estimates of parameters in the latent class model are commonly obtained by the method of maximum likelihood. The EM algorithm, described in the next section, is an iterative method which can be used to obtain maximum likelihood estimates for this type of problem.

*The EM algorithm*

The EM algorithm, popularized by the paper of Dempster et al., 1977, was developed to obtain maximum likelihood estimates in a situation when we have missing or incomplete data. This method can also be applied to great advantage in situations when the data are 'missing' by virtue of their being unobservable or latent, as in the case of the diagnostic testing problem. Detailed descriptions of the algorithm and its properties are given by Dempster et al., 1977 and Wu, 1983. Louis, 1982 presents a method of estimating the covariance matrix of the parameter estimates from an EM algorithm.

Let $n$ denote the observed data and $y$ denote the complete data, with likelihood functions $g(n|\phi)$ and $f(y|\phi)$, respectively. For a given $n$, the purpose of the EM algorithm is to determine the value of $\phi$ which maximizes $g(n|\phi)$ by making use of the complete data density $f(y|\phi)$. Each iteration of the EM algorithm consists of two steps - the 'expectation' or E-step and the 'maximization' or M-step. The algorithm typically proceeds as follows:

1. Given that the estimate of $\phi$ in the $i^{th}$ step is $\phi^{(i)}$, the E-step consists of computing the expectation of the complete data likelihood conditional on $\phi^{(i)}$, with respect to the conditional density of $y$ given $n$ as follows:

$$Q(\phi|\phi^{(i)}) = E_{y|n}(log(f(y|\phi))|n, \phi^{(i)})$$

   When the probability density of the complete data comes from the exponential family of distributions, this step reduces to calculating the expectation of the sufficient statistics for each parameter conditional on the observed incomplete data and $\phi^{(i)}$.

2. In the M-step the new estimate of $\phi$, $\phi^{(i+1)}$, is obtained by maximizing the expectation computed in the E-step.

The E and M steps are repeated until some pre-determined convergence criterion such as $|\phi^{(i+1)} - \phi^{(i)}| < \epsilon$ is met, for some small $\epsilon$. The main advantage of the EM algorithm is its simplicity. It does not rely on the calculation and inversion of the information matrix. Another feature of the EM algorithm is that the likelihood function never decreases at successive iterations. The convergence of the EM algorithm is linearly related to the amount of missing information in the data. While convergence is reached quickly in problems where the likelihood functions has a single mode, there is a danger that the algorithm could get stuck at local maxima or saddle points. This can happen when there is more than one mode, for example when we have a large sample size resulting in individual points being associated with a large weight.

*The EM Algorithm Applied to the Diagnostic Testing Problem*

We now illustrate the application of the EM algorithm to the two-test diagnostic problem. As noted earlier this non-identifiable problem can be solved using a frequentist approach by constraining some two parameters to be fixed. For the Strongyloides data problem let us arbitrarily hold the two specificities constant. We then have

$$n = (n_{11}, n_{10}, n_{01}, n_{00}), \text{ and}$$
$$\phi = (\pi, S_1, S_2)$$

with $C_1$ and $C_2$ held constant. The likelihood function in equation (3.4) can be rewritten as

$$\begin{aligned} & L_c(\pi, S_1, S_2 | C_1, C_2, n, y) \\ &= \frac{(n_{11}+n_{10}+n_{01}+n_{00})!}{y_{11}!(n_{11}-y_{11})!\, y_{10}!(n_{10}-y_{10})!\, y_{01}!(n_{01}-y_{01})!\, y_{00}!(n_{00}-y_{00})!} \\ & \times \pi^{y_{11}+y_{10}+y_{01}+y_{00}}(1-\pi)^{N-(y_{11}+y_{10}+y_{01}+y_{00})} S_1^{y_{11}+y_{10}}(1-S_1)^{y_{01}+y_{00}} \\ & \times S_2^{y_{11}+y_{01}}(1-S_2)^{y_{10}+y_{00}} C_1^{(n_{01}-y_{01})+(n_{00}-y_{00})}(1-C_1)^{(n_{11}-y_{11})+(n_{10}-y_{10})} \\ & \times C_2^{(n_{10}-y_{10})+(n_{00}-y_{00})}(1-C_2)^{(n_{11}-y_{11})+(n_{01}-y_{01})} \end{aligned}$$

$$\propto \pi^{y_{11}+y_{10}+y_{01}+y_{00}}(1-\pi)^{N-(y_{11}+y_{10}+y_{01}+y_{00})}S_1^{y_{11}+y_{10}}(1-S_1)^{y_{01}+y_{00}}$$
$$S_2^{y_{11}+y_{01}}(1-S_2)^{y_{10}+y_{00}}. \tag{3.6}$$

Let $N_D = y_{11} + y_{10} + y_{01} + y_{00}$, $y_{1\cdot} = y_{11} + y_{10}$ and $y_{\cdot 1} = y_{11} + y_{01}$. Therefore $N_D$ represents the total number of truly diseased subjects, $y_{1\cdot}$ represents the number of subjects who obtain a positive result on the first test and $y_{\cdot 1}$ represents the number of subjects who obtain a positive result on the second test. equation (3.6) can now be rewritten as

$$L_c(\pi, S_1, S_2|C_1, C_2, y) \propto \pi^{N_D}(1-\pi)^{N-N_D}S_1^{y_{1\cdot}}(1-S_1)^{N_D-y_{1\cdot}}S_2^{y_{\cdot 1}}(1-S_2)^{N_D-y_{\cdot 1}}, \tag{3.7}$$

where $N = n_{11}+n_{10}+n_{01}+n_{00}$. The factorization theorem for the exponential family of distributions states that a necessary and sufficient condition for a statistic, say $t(x)$, to be sufficient for the distribution of a variable $x$, is that there exist non-negative functions, $\alpha_\theta(.)$ and $\beta(.)$ such that the density of $x$, $f(x|\theta)$, satisfies

$$f(x|\theta) = \alpha_\theta(t(x))\beta(x).$$

By the factorization theorem, it is easy to see that $N_D, y_{1\cdot}$ and $y_{\cdot 1}$ are the sufficient statistics for $\phi$. The steps of the EM algorithm for the diagnostic problem can therefore be summarized as follows:

1. Since the multinomial likelihood comes from the family of exponential distributions, the E-step involves the computation of the expectations of the three sufficient statistics conditional on the observed data and the current estimate of $\phi$. Therefore,

$$E(N_D|n_{11}, n_{10}, n_{00}, n_{01}, \phi^{(i)})$$
$$= E(y_{11} + y_{10} + y_{01} + y_{00}|n_{11}, n_{10}, n_{00}, n_{01}, \phi^{(i)}), \tag{3.8}$$

$$E(y_{1\cdot}|n_{11}, n_{10}, n_{00}, n_{01}, \phi^{(i)})$$
$$= E(y11 + y10|n_{11}, n_{10}, n_{00}, n_{01}, \phi^{(i)}), \text{ and} \quad (3.9)$$
$$E(y_{\cdot 1}|n_{11}, n_{10}, n_{00}, n_{01}, \phi^{(i)})$$
$$= E(y01 + y00|n_{11}, n_{10}, n_{00}, n_{01}, \phi^{(i)}). \quad (3.10)$$

The expectations in equations (3.8), (3.9) and (3.10) can be obtained using the expectations of $y_{11}, y_{10}, y_{01}, y_{00}$, which are as follows. From equation (3.6) we can deduce that

$$E\{y_{11}|n_{11}, n_{10}, n_{00}, n_{01}, \phi^{(i)}\}$$
$$= n_{11}\frac{\pi_1^{(i)}S_1^{(i)}S_2^{(i)}}{\pi_1^{(i)}S_1^{(i)}S_2^{(i)} + (1-\pi_1^{(i)})(1-C_1^{(i)})(1-C_2^{(i)})},$$

$$E\{y_{10}|n_{11}, n_{10}, n_{00}, n_{01}, \phi^{(i)}\}$$
$$= n_{10}\frac{\pi_1^{(i)}S_1^{(i)}(1-S_2^{(i)})}{\pi_1^{(i)}S_1^{(i)}(1-S_2^{(i)}) + (1-\pi_1^{(i)})(1-C_1^{(i)})C_2^{(i)}},$$

$$E\{y_{01}|n_{11}, n_{10}, n_{00}, n_{01}, \phi^{(i)}\}$$
$$= n_{01}\frac{\pi_1^{(i)}(1-S_1^{(i)})S_2^{(i)}}{\pi_1^{(i)}(1-S_1^{(i)})S_2^{(i)} + (1-\pi_1^{(i)})C_1^{(i)}(1-C_2^{(i)})},$$

$$E\{y_{00}|n_{11}, n_{10}, n_{00}, n_{01}, \phi^{(i)}\}$$
$$= n_{00}\frac{\pi_1^{(i)}(1-S_1^{(i)})(1-S_2^{(i)})}{\pi_1^{(i)}(1-S_1^{(i)})(1-S_2^{(i)}) + (1-\pi_1^{(i)})C_1^{(i)}C_2^{(i)}}.$$

2. For a member of the exponential family of distributions, maximizing $Q(\phi|\phi^{(i)})$ is equivalent to solving $E(t(x)|\phi) = t$. In the M-step the parameter estimates $\phi^{(i+1)}$ are found to be

$$\pi^{(i+1)} = \frac{N_D}{N},$$

| | $C_1 = 0.95,\ C_1 = 0.72$ | $C_1 = 0.89,\ C_1 = 0.31$ |
|---|---|---|
| $\pi$ | 0.521 | 0.253 |
| $S_1$ | 0.543 | 0.520 |
| $S_2$ | 0.547 | 0.479 |

Table 3.6: Estimates of $\pi, S_1$ and $S_2$ obtained using a frequentist solution to the $2LC$ model at two different points on the $(C_1, C_2)$ plane.

$$S_1^{(i+1)} = \frac{y_{1\cdot}}{N_D},$$
$$S_2^{(i+1)} = \frac{y_{\cdot 1}}{N_D}.$$

The E and M steps are repeated till some convergence criterion such as

$$|\pi^{(i+1)} - \pi^{(i)}| + |S_1^{(i+1)} - S_1^{(i)}| + |S_2^{(i+1)} - S_2^{(i)}| < 0.000001$$

is met. The EM algorithm converges fairly quickly for this problem (within about 50 iterations) but this may not always be the case. Two sets of point estimates of the prevalence and sensitivities obtained by constraining the two specificities to be at their medians, and to be at the lower endpoint of their 95% CI, are given in Table 3.6. A sensitivity analysis, that is results from repeated runs of the above algorithm for different values of $C_1$ and $C_2$, could be used to get an idea of the range of the prevalence and sensitivities. The prevalence estimate almost doubles in value as the specificities change across their range of plausible values.

As mentioned earlier, the major advantage of using this method is the ease of implementation of the EM algorithm. Unlike a Bayesian method there is no need to determine prior distributions for the parameters, although one does have to exactly specify $C_1$ and $C_2$, or two other parameters in order to draw inferences. This ease comes at the cost, however, of having to determine which parameters must be constrained and what values the constrained parameters must take. From the results in Table 3.6 we can see that by altering the values of the specificities the estimate of the

prevalence changed drastically. This is particularly worrisome because there is much uncertainty about the values of the specificities as discussed earlier, and with this method we are not able to simultaneously account for the variability of the different parameters. A sensitivity analysis, while giving an idea of the point estimates at several points in the $(C_1, C_2)$ plane, still does not provide the complete picture.

### 3.2.5 *A Bayesian solution for the 2LC model*

As seen in the previous section, when we have a non-identifiable model the classical frequentist method cannot be used to obtain estimates of *all* unknown parameters without constraining some of them to be fixed constants. This is a rather unrealistic requirement since none of the parameters are truly known and hence the division into constrained and unconstrained parameters is often quite arbitrary. Further, the uncertainty in the value of the constrained parameters is not accounted for while estimating the variability of the unconstrained parameters, for example in terms of their confidence intervals.

*The Bayesian approach to a non-identifiable problem*

Under non-identifiability, a Bayesian approach can be used to obtain closed form, interpretable point and interval estimates for each of the unknown parameters (Neath and Samaniego, 1997). The basic idea behind the Bayesian approach is to use the available information about each parameter summarized in the form of a prior distribution, and thus eliminate the need for constraining them. By doing so the uncertainty in our knowledge of the parameter is accounted for. Prior distribution functions can be determined in consultation with the available literature and expert opinion. This approach is numerically equivalent to the frequentist approach when a degenerate prior distribution is used over the constrained parameters, which concentrates all the probability mass at a single point. The problem of formulating a suitable prior

distribution or deriving the posterior density are not worsened by virtue of the non-identifiability of a problem, at least in this case since all parameters have an easily understood 'natural' interpretation.

Roughly speaking, in order to obtain a useful solution using this approach, fairly strong priors would be needed on at least as many parameters as would be constrained when using the frequentist approach. Since in this situation the prior distribution tends to have a strong influence on the posterior, it is important to interpret the results of such an analysis carefully. Neath and Samaniego, 1997 determine conditions to identify the subset of prior densities that would ensure the posterior estimate is an improvement over the prior guess. They considered the non-identifiable problem involving results from a single dichotomous test, although by treating only point estimates they did not consider interval estimation.

*A Bayesian approach to the diagnostic testing problem*

In this section we describe a Bayesian solution to the $2LC$ problem which was presented in the paper by Joseph et al., 1995. For the Bayesian approach the observed data and the parameters to be estimated would be

$$n = (n_{11}, n_{10}, n_{01}, n_{00}), \text{ and}$$
$$\phi = (\pi, S_1, S_2, C_1, C_2),$$

respectively. Let $y_{11}, y_{10}, y_{01}$ and $y_{00}$ be defined as in Table 3.2

As explained in Chapter 2, the joint posterior distribution is proportional to the product of the likelihood function and the prior distributions. Denoting the prior density corresponding to a parameter $\theta$ as $p(\theta|.)$, and assuming all parameters in $\phi$ to be independent of each other, the joint posterior distribution can be written as

$$p(\phi|n) \propto \frac{(n_{11}+n_{10}+n_{01}+n_{00})!}{y_{11}!\,(n_{11}-y_{11})!\,y_{10}!\,(n_{10}-y_{10})!\,y_{01}!\,(n_{01}-y_{01})!\,y_{00}!\,(n_{00}-y_{00})!}$$

$$\times [\pi S_1 S_2]^{y_{11}} [(1-\pi)(1-C_1)(1-C_2)]^{n_{11}-y_{11}} [\pi S_1(1-S_2)]^{y_{10}}$$
$$\times [(1-\pi)(1-C_1)C_2]^{n_{10}-y_{10}} [\pi(1-S_1)S_2]^{y_{01}} [(1-\pi)C_1(1-C_2)]^{n_{01}-y_{01}}$$
$$\times [\pi(1-S_1)(1-S_2)]^{y_{00}} [(1-\pi)C_1C_2]^{n_{00}-y_{00}}$$
$$\times p_\pi(\pi)p_{S_1}(S_1)p_{S_2}(S_2)p_{C_1}(C_1)p_{C_2}(C_2). \quad (3.11)$$

A convenient choice of prior distribution for all the parameters of interest would be from the $Beta(\alpha, \beta)$ family of distributions, since this family is constrained to the range $0-1$, which matches the range of all parameters in $\phi$. Further, the *Beta* distribution is conjugate to a *Binomial* likelihood function (see Section 2.1.2), thus simplifying the derivation of the full conditional distributions of the Gibbs sampler that will be employed below. A random variable, $X$, has a $Beta(\alpha, \beta)$ distribution if its probability density function is of the form

$$f(X=x) = \begin{cases} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 \le x \le 1 \;\; \alpha, \beta > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Another advantage of the *Beta* distribution is that several density shapes can be obtained by varying $\alpha$ and $\beta$. For example, a non-informative or uniform prior is obtained by setting $\alpha = \beta = 1$, a symmetric prior by setting $\alpha = \beta = m(>1)$, a right-skewed prior can be obtained by setting $\alpha > \beta$, etc.

The joint posterior distribution can now be rewritten as

$$p(\phi|n) \propto [\pi S_1 S_2]^{y_{11}} [(1-\pi)(1-C_1)(1-C_2)]^{n_{11}-y_{11}} [\pi S_1(1-S_2)]^{y_{10}}$$
$$\times [(1-\pi)(1-C_1)C_2]^{n_{10}-y_{10}} [\pi(1-S_1)S_2]^{y_{01}} [(1-\pi)C_1(1-C_2)]^{n_{01}-y_{01}}$$
$$\times [\pi(1-S_1)(1-S_2)]^{y_{00}} [(1-\pi)C_1C_2]^{n_{00}-y_{00}}$$
$$\times \pi^{\alpha_\pi-1}(1-\pi)^{\beta_\pi-1} S_1^{\alpha_{S_1}-1}(1-S_1)^{\beta_{S_1}-1} S_2^{\alpha_{S_2}-1}(1-S_2)^{\beta_{S_2}-1}$$
$$\times C_1^{\alpha_{C_1}-1}(1-C_1)^{\beta_{C_1}-1} C_2^{\alpha_{C_2}-1}(1-C_2)^{\beta_{C_2}-1}.$$

$$
\begin{aligned}
&= \pi^{\alpha_\pi+y_{11}+y_{10}+y_{01}+y_{00}-1}(1-\pi)^{\beta_\pi+N-(y_{11}+y_{10}+y_{01}+y_{00})-1} \\
&\times S_1^{\alpha_{S_1}+y_{11}+y_{10}-1}(1-S_1)^{\beta_{S_1}+y_{01}+y_{00}-1} S_2^{\alpha_{S_2}+y_{11}+y_{10}-1}(1-S_2)^{\beta_{S_1}+y_{10}+y_{00}-1} \\
&\times C_1^{\alpha_{C_1}+(n_{01}-y_{01})+(n_{00}-y_{00})-1}(1-C_1)^{\beta_{C_1}+(n_{11}-y_{11})+(n_{10}-y_{10})-1} \\
&\times C_2^{\alpha_{C_2}+(n_{10}-y_{10})+(n_{00}-y_{00})-1}(1-C_2)^{\beta_{C_1}+(n_{11}-y_{11})+(n_{01}-y_{01})-1}. \qquad (3.12)
\end{aligned}
$$

The latent data are not observed, preventing direct use of the posterior distribution. Therefore, samples from the marginal posterior distributions of the prevalence and test parameters are drawn using the Gibbs sampler described in Section 2.2.2. The basic idea behind this method is that if the latent data are known, then the marginal full conditional distributions of the prevalence and test parameters are known. Conversely, conditional on the exact values of the prevalence, test parameters and the observed test results, we can derive the full conditional distributions of $y_{11}$, $y_{10}$, $y_{01}$ and $y_{00}$. By alternating between these two steps a sample from the marginal posterior distributions of each parameter is obtained.

From equation 3.12 it is easy to see that the full-conditional distributions, namely the distributions of each parameter conditional on the others are as follows

$$
\begin{aligned}
y_{11}|n_{11},\pi,S_1,C_1,S_2,C_2 &\sim Binomial(n_{11}, \frac{\pi S_1 S_2}{\pi S_1 S_2 + (1-\pi)\pi(1-C_1)(1-C_2)}), \\
y_{10}|n_{10},\pi,S_1,C_1,S_2,C_2 &\sim Binomial(n_{10}, \frac{\pi S_1 (1-S_2)}{\pi S_1 (1-S_2) + (1-\pi)\pi(1-C_1)C_2}), \\
y_{01}|n_{01},\pi,S_1,C_1,S_2,C_2 &\sim Binomial(n_{01}, \frac{\pi (1-S_1) S_2}{\pi (1-S_1) S_2 + (1-\pi)\pi C_1(1-C_2)}), \\
y_{00}|n_{00},\pi,S_1,C_1,S_2,C_2 &\sim Binomial(n_{00}, \frac{\pi (1-S_1)(1-S_2)}{\pi (1-S_1)(1-S_2) + (1-\pi)\pi C_1 C_2}),
\end{aligned}
$$

$$
\begin{aligned}
&\pi\ |n_{11},n_{10},n_{01},n_{00},y_{11},y_{10},y_{01},y_{00},\alpha_\pi,\beta_\pi \\
&\sim Beta(\alpha_\pi + y_{11} + y_{10} + y_{01} + y_{00}, \beta_\pi + N - (y_{11} + y_{10} + y_{01} + y_{00})),
\end{aligned}
$$

$$
\begin{aligned}
&S_1|y_{11},y_{10},y_{01},y_{00},\alpha_{S_1},\beta_{S_1} \\
&\sim Beta(\alpha_{S_1} + y_{11} + y_{10}, \beta_{S_1} + y_{01} + y_{00}),
\end{aligned}
$$

| | Parameter | Median | 95% PI | $\alpha$ | $\beta$ |
|---|---|---|---|---|---|
| Stool | $S_1$ | 0.24 | 0.07-0.47 | 4.44 | 13.31 |
| Examination | $C_1$ | 0.95 | 0.89-0.99 | 71.25 | 3.75 |
| Serology | $S_2$ | 0.81 | 0.63-0.92 | 21.96 | 5.49 |
| Test | $C_2$ | 0.72 | 0.31-0.96 | 4.1 | 1.76 |

Table 3.7: Prior distribution parameters corresponding to sensitivities and specificities of the stool examination and serology test.

$$
\begin{aligned}
&C_1|n_{11}, n_{10}, n_{01}, n_{00}, y_{11}, y_{10}, y_{01}, y_{00}, \alpha_{C_1}, \beta_{C_1} \\
&\quad \sim Beta(\alpha_{C_1} + (n_{01} - y_{01}) + (n_{00} - y_{00}), \\
&\quad \beta_{C_1} + (n_{11} - y_{11}) + (n_{10} - y_{10})), \\
&S_2|y_{11}, y_{10}, y_{01}, y_{00}, \alpha_{S_2}, \beta_{S_2} \\
&\quad \sim Beta(\alpha_{S_2} + y_{11} + y_{10}, \beta_{S_2} + y_{10} + y_{00}), \\
&C_2|n_{11}, n_{10}, n_{01}, n_{00}, y_{11}, y_{10}, y_{01}, y_{00}, \alpha_{C_2}, \beta_{C_2} \\
&\quad \sim Beta(\alpha_{C_2} + (n_{10} - y_{10}) + (n_{00} - y_{00}), \\
&\quad \beta_{C_1} + (n_{11} - y_{11}) + (n_{01} - y_{01})).
\end{aligned}
\tag{3.13}
$$

In the case of the Strongyloides data, since no information was available about the prevalence of this disease in the population under study, a $Beta(1, 1)$ diffuse prior distribution was used for the prevalence. The parameters of the *Beta* prior for the sensitivities and specificities of the two tests were obtained by equating the center of the range (*i.e.* the 95% probability interval) to the mean of the *Beta* distribution, which is given by $\frac{\alpha}{\alpha+\beta}$ and, one quarter of the range to the standard deviation of the *Beta* distribution, which is given by $\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}$. These parameters are presented in Table 3.7. The estimated posterior medians and 95% posterior probability intervals for the prevalence and test parameters are summarized in Table 3.8. Here

| | | Median | 95% PI |
|---|---|---|---|
| | Prevalence | 0.76 | 0.52-91 |
| Stool Examination | $S_1$ | 0.31 | 0.22-0.44 |
| | $C_1$ | 0.96 | 0.91-0.99 |
| Serology Test | $S_2$ | 0.89 | 0.80-0.95 |
| | $C_2$ | 0.67 | 0.36-0.95 |

Table 3.8: Posterior Medians and 95% posterior probability intervals of the prevalence and test parameters obtained using a Bayesian solution to the $2LC$ model.

and elsewhere PI denotes 'probability interval'. Figure 3.1 illustrates the prior and posterior density function for the prevalence.

The advantage of this method over the frequentist method is that it avoids the unrealistic constraining of unknown parameters and takes into account their variability. In the words of Neath and Samaniego "To the classical statistician the estimation of a non-identifiable problem is, purely and simply, an ill-posed problem ... (for which) the only alternative is to ... solve a related problem which is identifiable. ... Bayesian methods can provide point estimates of a non-identifiable parameter that are unambiguous and unique (relative to a given prior)." The Gibbs sampler is fairly simple to implement and was found to converge quickly. One drawback of this method is that, like the frequentist approach described earlier, it assumes tests to be independent of each other conditional on the latent disease status. This assumption rarely holds in practice as the two tests could in fact be the same test conducted at two different points in time, or the two tests could be based on the same underlying phenomenon. That is, there may be a variable besides the disease status which causes the test results to be related. The reasons for this are described in greater detail in Section 5.1. In the next section we present some of the methods developed to date to model this dependence between tests.

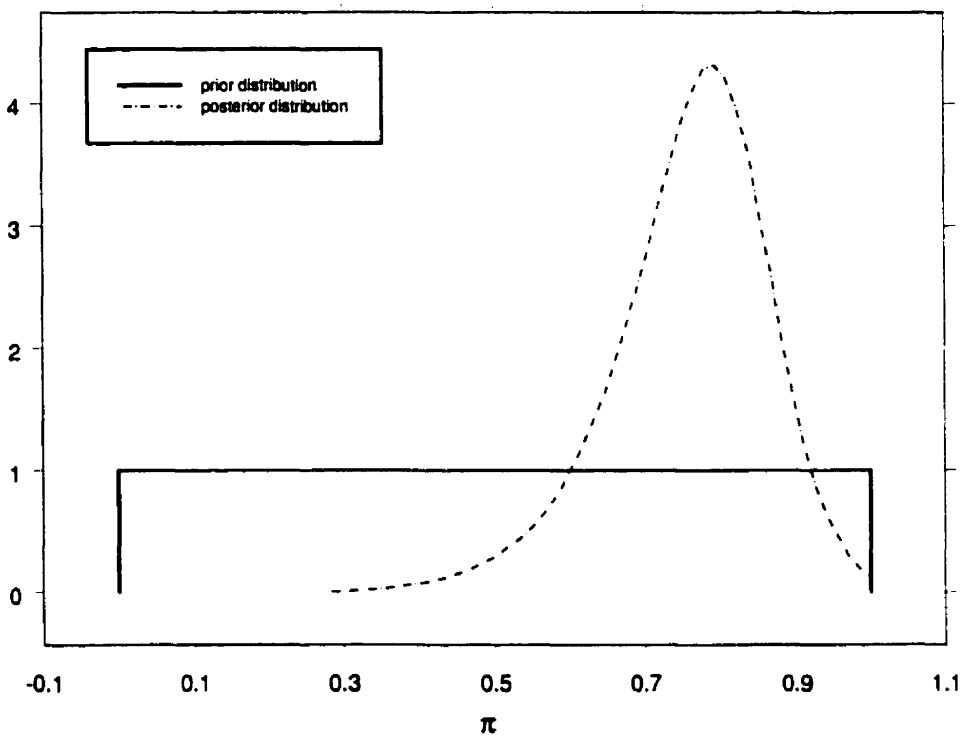


Figure 3.1: Prior distribution of the prevalence overlaid by the posterior distribution obtained using the Gibbs sampler.

## 3.3 *Modeling the conditional dependence between diagnostic tests*

The possible inadequacy of the assumption of conditional independence between multiple tests/raters has been recognized by several researchers and consequently many methods have been suggested to model the data more realistically. While earlier methods tried to describe the agreement between tests using measures such as the kappa statistic (Fleiss, 1971, Landis and Koch, 1977), later methods sought to model this agreement and correct for its influence on estimates of the test parameters and

prevalence. In this section we discuss some of the methods that were developed for the situation when there was no gold standard among the test/rater results.

### 3.3.1 *Increasing the number of latent classes*

When the conditional independence model is inappropriate for the observed data, the proportion of subjects for whom all tests give identical results are underestimated. One way to account for this is to increase the number of latent classes in the latent variable under study such that there is an exclusive latent class corresponding to the unequivocally diseased and another corresponding to the unequivocally non-diseased subjects for whom all test results are in agreement. This method has been described by several authors including Hagenaars, 1988, Espeland and Handelman, 1989, Uebersax and Grove, 1990 and Formann, 1994.

To illustrate the application of this method we use the example from the paper by Espeland and Handelman, 1989, where they analyzed the results from five diagnosticians on 3,869 radiographic images of dental caries. The images were classified by each diagnostician as sound (0) or carious (1). The authors begin by fitting the classic two-latent class model (Section 3.2.1) reformulated as a log-linear model. Let $m_{dpqrst}, d, p, q, r, s, t = 0, 1$ denote the frequency for the cross-classifications by the five dentists in the latent class $D = d$. Then, the equivalent of equation 3.2 in log-linear form is

$$log(m_{dpqrst}) = u_0(d) + u_1(d,p) + u_2(d,q) + u_3(d,r) + u_4(d,s) + u_5(d,t),$$
$$d, p, q, r, s, t = 0, 1. \tag{3.14}$$

where $u_0(.)$ references the latent variable and $u_1(.,.), \ldots, u_5(.,.)$ reference individual diagnosticians. This model can be solved using the EM-algorithm by assuming the $m_{dpqrst}$ to be known at one step and computing the maximum likelihood estimates of the $u_j(.)'s$ in the next step. The value of the likelihood ratio chi-square statistic

corresponding to the model in equation (3.14) was $G^2 = 129.84$, $df = 20$, indicating that the data did not fit the model adequately.

The authors hypothesize that some images are unequivocally classified as sound or carious by all the dentists but this unqualified agreement is underestimated by the $2LC$ model. More generally, a look at the residuals from the $2LC$ model, *i.e.* the difference in the observed and estimated values of the cell means, will suggest which cross-classifications have been under- or overestimated. The problem at hand can be modeled by adding two more latent classes $D = 2$ and 3 corresponding to unequivocally sound and unequivocally carious teeth. Thus all subjects in class $D = 2$ will receive a 'sound' diagnosis and all subjects in class $D = 3$ will receive a 'carious' diagnosis from all five dentists. The log-linear model for these two additional cells has the following parameterization

$$\begin{aligned} log(m_{200000}) &= u_0(2), \\ log(m_{311111}) &= u_0(3). \end{aligned} \tag{3.15}$$

The value of the $G^2$ statistic for the new model was found to be 53.08, $df = 18$, indicating a marked improvement in the fit of the model to the data.

The prevalence and test parameters can be estimated using the estimated cell means as follows:

$$\pi = P(D = 1 \text{ or } D = 3) = m_{311111} + \sum_{pqrst=0,1} m_{1pqrst},$$

Sensitivity of Dentist 1

$$= P(T_1 = 1 | D = 1 \text{ or } D = 3) = \frac{m_{311111} + \sum_{qrst=0,1} m_{11qrst}}{m_{311111} + \sum_{pqrst=0,1} m_{1pqrst}},$$

*and*

Specificity of Dentist 1

$$= P(T_1 = 0 | D = 0 \text{ or } D = 2) = \frac{m_{200000} + \sum_{qrst=0,1} m_{00qrst}}{m_{200000} + \sum_{pqrst=0,1} m_{0pqrst}}.$$

The sensitivities and specificities of the other dentists can be calculated similarly.

Lu, 1968 suggests the further addition of a latent class corresponding to 'undiagnosable' cases for whom there is no right or wrong result and for whom diagnosticians are providing no more than a random guess. Thus each one of the $2^5$ possible cross-classifications is equally probable for subjects who lie in this latent class, which we shall denote as $D = 4$. A log-linear model for this fifth latent class would be of the form

$$log(m_{4pqrst}) = u_0(4), \quad p, q, r, s, t = 0, 1. \tag{3.16}$$

This model only marginally improves the likelihood chi-square ratio statistic to $G^2 = 50.57$, $df = 17$.

Addition of latent classes, however, accentuates the problem of non-identifiability. This model is identifiable only when $(2^p - 1) \geq (2p + 1 + m)$, where $m$ is the number of latent classes added to the $2LC$ model. Therefore, in order to have sufficient degrees of freedom to estimate all parameters when a single latent class is added, a minimum of four tests would be required. The other problem with this method is that it is not clear how many additional latent classes need to be added to improve the fit of the model. In the dental caries example, Espeland and Handelman found that the addition of two latent classes was not sufficient to explain the relations in the data and they had to extend their model in other ways as will be described in the next section. Clearly, with the addition of a sufficient number of latent classes, the problem will become saturated resulting in $G^2 = 0$ but this may not necessarily yield a meaningful model. Finally, it is not clear whether each additional latent class will be substantively meaningful. For instance in the case when some subjects are deemed 'undiagnosable' it is unclear how the prevalence is to be estimated.

### 3.3.2 *Addition of interaction terms*

In this section we describe yet another extension of the $2LC$ model used in the paper by Espeland and Handelman, 1989. When the addition of latent classes did not sufficiently improve the fit of the model, the authors added interaction terms between pairs of dentists who showed a tendency to agree with each other. Such a pair of dentists (or tests) among whom there is agreement can be identified by collapsing the $2^5$ cross-classifications into marginal tables for each pair of dentists and looking at the residuals obtained from the four-class latent model of the previous section. From such a table, the authors found a significant agreement between the raters 3 and 4 and hence modified the earlier model to include an interaction term as follows

$$\begin{aligned} log(m_{dpqrst}) &= u_0(d) + u_1(d,p) + u_2(d,q) + u_3(d,r) \\ &\quad + u_4(d,s) + u_5(d,t) + u_{34}(d,r,s), \\ &\quad d,p,q,r,s,t = 0,1, \\ log(m_{200000}) &= u_0(2), \\ log(m_{311111}) &= u_0(3). \end{aligned} \tag{3.17}$$

This model reduces the likelihood ratio chi-square statistic to $G^2 = 25.57$, $df = 16$ suggesting a significant improvement in the fit as compared to the earlier model. The prevalence and test parameters can be estimated from the estimated 'complete' cross-classification table as described in equation (3.16).

### 3.3.3 *Marginal models*

In a recent paper, Yang and Becker, 1997 used latent class marginal models to adjust for the correlation between diagnostic tests. In this model the sensitivities and specificities are simple functions of the model parameters, unlike the methods described in the previous sections. The authors describe the particular situation when results

from four dichotomous tests are available and only associations between pairs of tests are considered. Three and four-factor associations are constrained to be absent. For a general description of marginal models see Becker, 1994.

Let the four tests be denoted by $T_j = t$, where $j = 1,2,3,4$ and $t = 0,1$ and disease status be denoted by $D = d$, $d = 0,1$. The marginal model for an observed $2^4$ contingency table involves four univariate logits, six log-odds ratios for bivariate associations, four tri-variate associations and one four-factor association. The model is formally expressed as follows:

1. The univariate logits corresponding to $T_1$ are

$$\ln\left(\frac{\pi_{0d}^{T_1|D}}{\pi_{1d}^{T_1|D}}\right) = \alpha_d^{T_1}, \ d = 0,1,$$

where $\pi_{td}^{T_j|D} = P(T_j = t|D = d)$.

2. The log-odds ratios for the association between $T_1$ and $T_2$ are

$$\ln\left(\frac{\pi_{00d}^{T_1T_2|D}\pi_{11d}^{T_1T_2|D}}{\pi_{01d}^{T_1T_2|D}\pi_{10d}^{T_1T_2|D}}\right) = \psi_d^{T_1T_2}, \ d = 0,1.$$

The univariate logits for the other tests $\alpha_d^{T_2}$, $\alpha_d^{T_3}$, $\alpha_d^{T_4}$ and the log-odds ratios $\psi_d^{T_1T_3}$, $\psi_d^{T_1T_4}$, $\psi_d^{T_2T_3}$, $\psi_d^{T_2T_4}$, $\psi_d^{T_3T_4}$ for the other bivariate associations follow similarly. At this point we note that if all univariate and bivariate associations were present there would be $2\times4+2\times6 = 20$ parameters to estimate which would exceed the degrees freedom available, namely, $2^4 - 1 = 15$. Hence constraints must be applied on at least 5 of the parameters. In their paper, Yang and Becker, 1997 consider only one bivariate association $\psi_t^{T_3T_4}$ to be non-zero so that the number of unknown parameters is 10 and hence the model is identifiable.

3. Constraints are placed on all three-factor associations,

$$\ln\left(\frac{\pi_{000d}^{T_1T_2T_4|D}\pi_{110d}^{T_1T_2T_4|D}}{\pi_{010d}^{T_1T_2T_4|D}\pi_{100d}^{T_1T_2T_4|D}}\right) - \ln\left(\frac{\pi_{001d}^{T_1T_2T_4|D}\pi_{111d}^{T_1T_2T_4|D}}{\pi_{011d}^{T_1T_2T_4|D}\pi_{101d}^{T_1T_2T_4|D}}\right) = 0,$$

$$\ln\left(\frac{\pi_{000d}^{T_1T_3T_4|D}\pi_{110d}^{T_1T_3T_4|D}}{\pi_{010d}^{T_1T_3T_4|D}\pi_{100d}^{T_1T_3T_4|D}}\right) - \ln\left(\frac{\pi_{001d}^{T_1T_3T_4|D}\pi_{111d}^{T_1T_3T_4|D}}{\pi_{011d}^{T_1T_3T_4|D}\pi_{101d}^{T_1T_3T_4|D}}\right) = 0,\ d = 0, 1,$$

$$\ln\left(\frac{\pi_{000d}^{T_2T_3T_4|D}\pi_{110d}^{T_2T_3T_4|D}}{\pi_{010d}^{T_2T_3T_4|D}\pi_{100d}^{T_2T_3T_4|D}}\right) - \ln\left(\frac{\pi_{001d}^{T_2T_3T_4|D}\pi_{111d}^{T_2T_3T_4|D}}{\pi_{011d}^{T_2T_3T_4|D}\pi_{101d}^{T_2T_3T_4|D}}\right) = 0,$$

and, the single four factor-association

$$\ln\left(\frac{\pi_{0000d}^{T_1T_2T_3T_4|D}\pi_{1100d}^{T_1T_2T_3T_4|D}}{\pi_{0100d}^{T_1T_2T_3T_4|D}\pi_{1000d}^{T_1T_2T_3T_4|D}}\right) - \ln\left(\frac{\pi_{0010d}^{T_1T_2T_3T_4|D}\pi_{1110d}^{T_1T_2T_3T_4|D}}{\pi_{0110d}^{T_1T_2T_3T_4|D}\pi_{1010d}^{T_1T_2T_3T_4|D}}\right),$$

$$-\ln\left(\frac{\pi_{0001d}^{T_1T_2T_3T_4|D}\pi_{1101d}^{T_1T_2T_3T_4|D}}{\pi_{0101d}^{T_1T_2T_3T_4|D}\pi_{1001d}^{T_1T_2T_3T_4|D}}\right) + \ln\left(\frac{\pi_{0011d}^{T_1T_2T_3T_4|D}\pi_{1111d}^{T_1T_2T_3T_4|D}}{\pi_{0111d}^{T_1T_2T_3T_4|D}\pi_{1011d}^{T_1T_2T_3T_4|D}}\right) = 0,\ d = 0, 1.$$

to ensure that they are absent.

Maximum likelihood estimates of the parameters in the marginal latent class model can be obtained using the EM algorithm. One of the benefits of modeling the marginal distributions is that the sensitivity and specificity can be expressed as functions of the parameters for the univariate marginal logits. For example, to obtain the sensitivity of test $T_1$ we proceed as follows:

$$\ln\left(\frac{\pi_{01}^{T_1|D}}{\pi_{11}^{T_1|D}}\right) = \alpha_1^{T_1},$$

$$\Rightarrow \exp(\alpha_1^{T_1}) = \frac{\pi_{01}^{T_1|D}}{\pi_{11}^{T_1|D}} = \frac{1 - S_1}{S_1},$$

$$\Rightarrow S_1 = \frac{1}{1 + \exp(\alpha_1^{T_1})}.$$

Similarly for the specificity:

$$\ln\left(\frac{\pi_{00}^{T_1|D}}{\pi_{10}^{T_1|D}}\right) = \alpha_0^{T_1},$$

$$\Rightarrow \exp(\alpha_0^{T_l}) = \frac{\pi_{00}^{T_l|D}}{\pi_{10}^{T_l|D}} = \frac{C_l}{1-C_l},$$

$$\Rightarrow C_l = \frac{\exp(\alpha_0^{T_l})}{1+\exp(\alpha_0^{T_l})}.$$

The sensitivities and specificities for the remaining tests are obtained similarly.

The main drawback to this method is that of non-identifiability in the absence of a sufficient number of tests. In fact, while the degrees of freedom available from $p$ tests remains the same, *i.e.* $2^p - 1$ (see Table 3.3), the number of parameters to be estimated increases to $2 \times ({}^pC_1 + {}^pC_2 + \ldots + {}^pC_n) + 1 = 2(2^p - 1) + 1$. So the number of parameters to be estimated always exceeds the degrees of freedom available, meaning that constraints would always have to be placed on a subset of the parameters to obtain an identifiable model. The authors cite the parameterization of the model in terms of test parameters as the model's main advantage. However, when weighed against the complexity of the model it is questionable whether it is worth the effort.

### 3.3.4 *A random effects model*

Random effects models are often used to model similarity within groups, as for example, the similarity among serial observations on the same unit in a repeated measures analysis, or the similarity within clusters in a two-stage sampling design. This is done by ... (see Kutner et al., 1996 for an introduction to random effects models) Qu et al., 1996, use such an approach to model the correlation between multiple tests via their sensitivities and specificities. The similarity between the test results is hypothesized to arise due to a latent, subject-specific characteristic which is different from the disease status. Some examples of such subject-specific variables which affect test performance are given in Chapter 5 where we propose a Bayesian solution to the random effects model.

The sensitivities and specificities are taken to be probit functions of an unobserved, continuous variable, say $I$, which is assumed to have a $N(0,1)$ distribution. Probit functions are convenient because they take values between 0 and 1. Let $t_{jk} = 0, 1$ denote the test result of the $k^{th}$ subject, $k = 1, \ldots, N$, on the $j^{th}$ test, $j = 1, \ldots, p$, and $t_k = (t_{1k}, , t_{2k}, \ldots, t_{pk})$ denote the vector of test results for the $k^{th}$ subject on each test. The probability that the $k^{th}$ subject has a positive test result on the $j^{th}$ test is given by

$$P(t_{jk} = 1|D = d, I = i) = \Phi(a_{jd} + b_{jd}i), \quad d = 0, 1, \ I \sim N(0,1) \tag{3.18}$$

where $\Phi$ represents the cumulative distribution function of the $N(0,1)$ distribution and $(a_{jd}, b_{jd})$, $j = 1, \ldots, p$, $d = 0, 1$ are real constants.

The disease status $D$ and the latent variable $I$ are assumed to be independent of each other. The sensitivity of the $j$th test is then given by

$$\begin{aligned} S_j &= P(t_{jk} = 1|D = 1) \\ &= \int_{-\infty}^{\infty} \Phi(a_{j1} + b_{j1}i) d\Phi(i) \\ &= \int_{-\infty}^{\frac{a_{j1}}{1+b_{j1}^2}} \Phi(x) d\Phi(x) \quad \text{(using a } \tan^{-1}(\tfrac{a_{j1}}{1+b_{j1}^2}) \text{ anticlockwise rotation)} \\ &= \Phi(\frac{a_{j1}}{1 + b_{j1}^2}) \end{aligned}$$

Similarly, the specificity of the $j$th test is given by

$$\begin{aligned} C_j &= P(t_{jk} = 0|D = 0) \\ &= 1 - \Phi(\frac{a_{j0}}{1 + b_{j0}^2}) \\ &= \Phi(-\frac{a_{j0}}{1 + b_{j0}^2}) \text{ using the result } \ \Phi(-x) = 1 - \Phi(x) \end{aligned}$$

The results of different tests are taken to be independent of each other conditional on the disease status $D$ and the variable $I$. This means that within each disease class the tests are independent of each other conditional on the variable $I$. Qu et al., 1996

call this model the $2LCR$ model which is short for the 'Two Latent Class Random Effects Model'. The likelihood function of the observed data is given by

$$\begin{aligned} L &\propto \prod_{k=1}^{N} \sum_{d=0}^{1} P(D=d)\, P(t_k|D=d, I=i) \\ &= \prod_{k=1}^{N} (\pi \prod_{j=1}^{p} \Phi(a_{j1} + b_{j1} i_k)^{t_{jk}} (1 - \Phi(a_{j1} + b_{j1} i_k))^{(1-t_{jk})} \\ &\quad + (1-\pi) \prod_{j=1}^{p} \Phi(a_{j0} + b_{j0} i_k)^{(1-t_{jk})} (1 - \Phi(a_{j0} + b_{j0} i_k))^{t_{jk}}), \end{aligned} \tag{3.19}$$

The estimates of the parameters $\pi_d$ and $(a_{jd}, b_{jd})$, $j = 1, \ldots, p$, $d = 0, 1$ are obtained using the EM algorithm.

Apart from providing an elegant way to model the dependence between two or more tests simultaneously, this method is also more substantively meaningful. Similarity between test results often arises due to a factor, such as severity of disease, which is independent of disease status. A more severe case of the disease is likely to be detected by all tests as positive resulting in an agreement between them. The model is also flexible and the $b_{jd}$ parameters can be set so that the dependence is between a subset of tests, of the same or varying strength between different pairs of tests.

Once again, the drawback of this model is that a frequentist approach to its solution requires a minimum number of tests. Corresponding to each sensitivity and specificity there is a pair of $(a_{jd}, b_{jd})$'s to be estimated. Hence the total number of parameters is given by

$$2 \times (2 \times p) + 1 = (4 \times p) + 1 \tag{3.20}$$

This means that in order for the problem to be identifiable we require

$$(4 \times p) + 1 \leq 2^p - 1 \Rightarrow p \geq 5 \tag{3.21}$$

A special case of the model, which the authors term the $2LCR1$ model, occurs when the variance components are all equal, *i.e.* $b_{jd} = b_d$, $j = 1, \ldots, p$, $d = 0, 1$. This

means we have only two $b_{jd}$ values now, one among the diseased subjects and one among the non-diseased subjects. This would mean that the effect of the variable $I$ on the performance of each test is the same. In this situation we need

$$(2 \times p) + 2 + 1 \leq 2^p - 1 \Rightarrow p \geq 4 \tag{3.22}$$

resulting in a less stringent constraint of a minimum of four tests.

## 3.4 *Other Bayesian methods for the analysis of diagnostic tests*

Several researchers have found Bayesian analysis advantageous for developing methodology for the analysis of diagnostic tests. We briefly mention some of this work in this section. Gastwirth et al., 1991, used Bayesian inference to obtain large-sample maximum likelihood estimates of the accuracy of screening tests used to detect HIV antibodies in donated blood. More recently, Neath and Samaniego, 1997 looked at the problem of estimating the prevalence of HIV based on results from a single test. Their model is set up to estimate the proportion of truly diseased subjects among those who test positive by applying a $Dirichlet(\alpha_1, \alpha_2, \alpha_3)$ prior distribution over the pair of probabilities $(p_1, p_2)$ of true-positive and false-positive results, as follows

$$f(p_1, p_2) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} p_3^{\alpha_3 - 1},$$
$$p_1 \geq 0, p_2 \geq 0, p_1 + p_2 \leq 1, p_3 = 1 - p_1 - p_2.$$

If $X \sim Binomial(n, p_1 + p_2)$ denotes the number of subjects who tested positive, then the posterior distribution of $(p_1, p_2)$ is given by

$$f(p_1, p_2 | X = x) = \frac{\sum_{k=0}^{x} {}^xC_k \, p_1^{\alpha 1 + k - 1} p_1^{\alpha_2 + x - k - 1} p_3^{\alpha_3 + n - x - 1}}{\sum_{l=0}^{x} {}^xC_l \, \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)}}.$$

The authors use this example to provide a template for assessing the efficacy of Bayesian updating in non-identifiable problems. They conclude that Bayesian analy-

sis is worthwhile in the vast majority of cases because the posterior estimate of $(p_1, p_2)$ is usually closer to the true value than the prior estimate.

Matchar et al., 1990, proposed a Bayesian approach to address the problem of evaluating test performance when some patient's remain undiagnosed by a gold standard test. Estimating test parameters by ignoring these subjects could lead to work-up or verification bias, as discussed by Begg, 1987. The authors used a joint prior distribution over the sensitivity and specificity and employed a Monte Carlo Markov Chain method to sample from the posterior distributions of the prevalence and test parameters. It was found that taking into account the undiagnosed subjects markedly affects the estimates of the prevalence and test parameters.

Peng and Hall, 1996, propose a Bayesian solution to regression models of ordered ordinal response data from radiological tests. While most methods for the analysis of diagnostic tests postulate that test scores from diseased and non-diseased subjects as following a binormal distribution, as depicted in Figure 1.1, in practice test scores are usually ordinal. The approach suggested by Peng and Hall, 1996, overcomes this problem by imputing the unobserved continuous observations from the latent binormal distributions using data augmentation. They postulate that if there are $J$ possible outcomes on a test which are determined by the latent cut-off values $\theta_1, \ldots, \theta_{J-1}$ and $I$ possible covariate levels, then the probability of a subject at the $i^{th}$ covariate level lying in any one of the first $j$ ordered categories of the test is given by

$$\gamma_j(\chi_i) = \phi(\frac{\theta_j - \alpha^T \chi_i}{exp(\beta^T \chi_i)}), \quad j = 1, \ldots, J, i = 1, \ldots, I$$

where $\chi_i$ is the covariate vector corresponding to the $i^{th}$ level. Given the cut-off values, the unobserved test results are imputed using the constraint $\theta_{j-1} \leq z_{ij} \leq \theta_j$. The papers by Gatsonis, 1995 and Ishwaran and Gatsonis, 1997 extend this method to the case when we have correlated ordinal data from multiple measurements on each subject.

Joseph and Gyorkos, 1996, propose a Bayesian method for calculating point and interval estimates of likelihood ratios in the absence of a gold standard diagnostic test. They observed that their results were numerically similar to those obtained by the standard frequentist approach in the presence of a gold standard test, but typically provide larger interval estimates reflecting the inherent uncertainty when a gold standard is not present. This method was an improvement over earlier frequentist methods which required that the data under study be normally distributed.

## 3.5 Summary

In this chapter we have discussed various methods for the analysis of diagnostic tests, focusing mostly on those that have provided the background for the methods developed in this thesis. The $2LC$ model, while providing an elegant way to analyze results, from independent diagnostic tests, cannot be solved directly using a frequentist approach when we have less than 3 tests, *i.e.* when we have a non-identifiable problem. Another drawback of the $2LC$ model is its assumption of conditional independence between the tests which may not always be satisfied. Similar comments, however, apply to the $2LCR$ model when there are less than four or five tests. Bayesian approaches can be used in such situations to provide estimates of the disease prevalence and test parameters without imposing any unrealistic constraints, as Joseph et al., 1995 showed for the $2LC$ model.

Although there is an enormous literature on statistical methods for diagnostic test data, there are no frequentist solutions that directly address the problem of estimating parameters in the presence of three or less correlated tests. To our knowledge there is no literature discussing a Bayesian solution to this problem, even for identifiable cases. In the next chapter we describe the first of the two methods proposed in this thesis for modeling the conditional dependence between two imperfect tests.

# 4

# Modeling Conditional Dependence Using Fixed Effects

In the preceding chapters we have seen that the assumption of conditional independence between diagnostic tests is often made to simplify the statistical analysis of test results, even though it may be of questionable validity. We have also seen that frequentist approaches which address this problem would require a minimum of four tests to estimate all parameters without imposing unrealistic constraints on the parameter values. In this chapter we present the first approach developed in this thesis for modeling the conditional dependence between tests. We begin with a look at the effect of correlation on test results and how this can be modeled. The next section describes the implementation of a Bayesian fixed effects model to estimate the prevalence and diagnostic test parameters of two tests, while adjusting for the correlation between them. This is followed by a simulated example to illustrate the method. We will apply this method to real data in Chapter 6.

## *4.1 Modeling the correlation between a pair of tests*

To demonstrate the effect on test results when tests are correlated, we use an example involving two hypothetical tests $T_1$ and $T_2$. In the extreme situation when these tests are perfectly positively correlated *i.e.* when they have a correlation of +1, their combined results would appear as in Table 4.1. Since two perfectly correlated tests

| | $T_1+$ | $T_1-$ |
|---|---|---|
| $T_2+$ | $N_{11}$ | 0 |
| $T_2-$ | 0 | $N_{00}$ |

Table 4.1: Results from two tests with a correlation of +1.

would give the same result for every subject, the two cells where the test results are in conflict have a frequency of $N_{10} = N_{01} = 0$. In this case the second test adds no information once the results of the first test are known.

Of course, it would be unlikely to encounter such an extreme correlation in practice. Nevertheless, this example serves to illustrate that when two tests are positively correlated, there is a greater tendency for their results to agree. The frequencies of the $(T_1 = 1, T_2 = 1)$ and $(T_1 = 1, T_2 = 0)$ cells are increased, and the frequencies of the $(T_1 = 1, T_2 = 0)$ and $(T_1 = 0, T_2 = 1)$ cells are decreased compared to the case of conditionally independent tests. Therefore, while modeling this situation, an increased probability must be assigned to the diagonal cells at the expense of the off-diagonal cells

The conditional dependence between two tests can be measured using the covariance or correlation between the two tests within each disease class. This 'conditional covariance' can be expressed in terms of the sensitivities and specificities of the two tests involved. Let $T_1$ and $T_2$ be two tests such that:

$$T_1 = \begin{cases} 1 & \text{if the test is positive,} \\ 0 & \text{otherwise,} \end{cases}$$

and

$$T_2 = \begin{cases} 1 & \text{if the test is positive,} \\ 0 & \text{otherwise.} \end{cases}$$

Using the notation first presented in Chapter 1, the covariance between the tests

among the diseased subjects, denoted here by *covp*, can be derived as follows (Vacek, 1985):

$$\begin{aligned} covp = cov(T_1, T_2|D=1) &= E(T_1T_2|D=1) - E(T_1|D=1)E(T_2|D=1), \\ &= \sum_{t_1,t_2=0}^{1} t_1 t_2 \, P(T_1 = t_1, T_2 = t_2|D=1) \\ &\quad - \sum_{t_1=0}^{1} t_1 \, P(T_1 = t_1|D=1) \sum_{t_2=0}^{1} t_2 \, P(T_2 = t_2|D=1), \\ &= P(T_1 = 1, T_2 = 1|D=1) \\ &\quad - P(T_1 = 1|D=1)P(T_2 = 1|D=1), \\ &= P(T_1 = 1, T_2 = 1|D=1) - S_1S_2. \end{aligned} \tag{4.1}$$

Similarly among the non-diseased subjects

$$\begin{aligned} covn = cov(T_1, T_2|D=0) &= E(T_1T_2|D=0) - E(T_1|D=0)E(T_2|D=0), \\ &= P(T_1 = 0, T_2 = 0|D=0) - C_1C_2. \end{aligned}$$

Equation (4.1) can be re-written as

$$\begin{aligned} P(T_1 = 1, T_2 = 1|D=1) &= S_1S_2 + covp, \\ &= S_1S_2, \qquad \text{when } covp = 0. \end{aligned} \tag{4.2}$$

Therefore, the probability of observing $(T_1 = 1, T_2 = 1)$ when two tests are correlated, is increased by a factor of *covp* as compared to the case when the tests are conditionally independent. Similarly,

$$\begin{aligned} P(T_1 = 1, T_2 = 0|D=1) &= S_1(1 - S_2) - covp, \\ &= S_1(1 - S_2), \qquad \text{when } covp = 0, \end{aligned} \tag{4.3}$$

so that the probability of observing $(T_1 = 1, T_2 = 0)$ when two tests are correlated is *covp* less than in the case when the tests are conditionally independent. The

probabilities of observing the remaining combinations of test results are

$$\begin{aligned}
P(T_1 = 0, T_2 = 1|D = 1) &= (1 - S_1)S_2 - covp, \\
P(T_1 = 0, T_2 = 0|D = 1) &= (1 - S_1)(1 - S_2) + covp, \\
P(T_1 = 1, T_2 = 1|D = 0) &= (1 - C_1)(1 - C_2) + covn, \\
P(T_1 = 1, T_2 = 0|D = 0) &= (1 - C_1)C_2 - covn, \\
P(T_1 = 0, T_2 = 1|D = 0) &= C_1(1 - C_2) - covn, \\
P(T_1 = 0, T_2 = 0|D = 0) &= C_1C_2 + covn.
\end{aligned} \tag{4.4}$$

In the remainder of the thesis, we take *covp* and *covn* to be positive as this is the case that arises most frequently in practice. Analogous results to those presented above can be derived for negative correlations.

From the above equations we see that a constant covariance between a pair of tests causes the probability of each combination of test results to be altered by a fixed value. This suggests that the dependence between a pair of tests could be modeled as a fixed effect due to their covariance on their combined results. As discussed in the previous chapter, in order to obtain an identifiable solution using this model we require that there be at least as many degrees of freedom as the number of unknown parameters.

$$\begin{aligned}
&\text{number of unknown parameters} \leq \text{number of degrees of freedom}, \\
&\Rightarrow (2 \times p) + (2 \times {}^pC_2) + 1 \leq 2^p - 1, \\
&\Rightarrow p \geq 5.
\end{aligned}$$

Hence a minimum of 5 tests would be required to obtain a solution for this model using a frequentist approach.

To address the problem of non-identifiability when we have less than 5 tests, we propose to extend the Bayesian approach used by Joseph et al., 1995, in the case when tests are conditionally independent. A problem with using the covariance to model conditional dependence is that, since it is defined only for pairs of tests it is

not possible to model the simultaneous dependence between three or more tests. The Bayesian method developed in the following section pertains to the situation when there are only two tests, although it could be extended to situations when there are more than two tests and there is a dependence between different pairs of tests. For a discussion of the effect of conditional dependence between pairs of tests, in the situation when there are three or four tests, on the estimates of the prevalence and test parameters see Torrance-Rynard and Walter, 1997.

## 4.2 A Bayesian fixed effects model

As outlined in Figure 1.2, the parameters of interest in the general situation, when the conditional dependence between tests is taken into account, are the prevalence, sensitivities and specificities of the tests and the covariances between them. As in the case of the method developed by Joseph et al., 1995, discussed in the previous chapter, by assigning suitable prior distributions to these parameters, we can do away with the need for 5 tests and still be able to obtain a solution for all unknown parameters simultaneously. The parameters of the prior distributions can be determined in consultation with the literature, be based on expert opinion, or some combination of these sources. This is an important step, since under non-identifiability, the influence of the prior distributions is considerable even as the sample size increases.

### 4.2.1 Notation

Let us assume we have results from two dichotomous tests $T_1$ and $T_2$. We will use the same notation for the prevalence, sensitivities, specificities and covariance as introduced in the first chapter. The number of subjects who fall into the cross-classification $(T_1 = i, T_2 = j)$ is denoted by $N_{ij}$, $i, j = 0, 1$ and the total number of subjects in the study is denoted by $N$. The true number of diseased subjects for each

| | D=1 | | D=0 | |
|---|---|---|---|---|
| | $T_1$=1 | $T_1$=0 | $T_1$=1 | $T_1$=0 |
| $T_2$=1 | $Y_{11}$ | $Y_{01}$ | $N_{11} - Y_{11}$ | $N_{01} - Y_{01}$ |
| $T_2$=0 | $Y_{10}$ | $Y_{00}$ | $N_{10} - Y_{10}$ | $N_{00} - Y_{00}$ |

Table 4.2: Cross-classification of observed and latent data from two tests.

combination of test results will be denoted by $Y_{ij}$, $i, j = 0, 1$. The $Y_{ij}$'s are latent values which are not observed. The observed and latent data can be summarized as in Table 4.2.

### 4.2.2 *The model*

Using equations (4.2), (4.3) and (4.4), we can write the likelihood function of the observed and latent data in the spirit of equation (3.4) as

$$\begin{aligned} L \propto\ & (\pi(S_1 S_2 + covp))^{Y_{11}} (\pi(S_1(1 - S_2) - covp))^{Y_{10}} \\ & \times (\pi((1 - S_1)S_2 - covp))^{Y_{01}} (\pi((1 - S_1)(1 - S_2) + covp))^{Y_{00}} \\ & \times ((1 - \pi)((1 - C_1)(1 - C_2) + covn))^{N_{11} - Y_{11}} \\ & \times ((1 - \pi)((1 - C_1)C_2 - covn))^{N_{10} - Y_{10}} \\ & \times ((1 - \pi)(C_1(1 - C_2) - covn))^{N_{01} - Y_{01}} ((1 - \pi)(C_1 C_2 + covn))^{N_{00} - Y_{00}}. \end{aligned} \tag{4.5}$$

As described earlier in Chapter 2, perhaps the most important practical aspect of developing a Bayesian solution for a non-identifiable problem is the proper elicitation of prior distributions for the model parameters. As explained there, we will use prior distributions that are members of standard distributional families, whose parameters are fixed such that they reflect the experimenter's beliefs prior to observing any data. The forms of the prior densities that we used for each parameter of interest

| | $T_1$=1 | $T_1$=0 | Total |
|---|---|---|---|
| $T_2$=1 | $(1-C_1)(1-C_2)+covn$ | $C_1(1-C_2)-covn$ | $1-C_2$ |
| $T_2$=0 | $(1-C_1)C_2-covn$ | $C_1C_2+covn$ | $C_2$ |
| Total | $1-C_1$ | $C_1$ | 1 |

Table 4.3: Probabilities of obtaining each possible combination of test results among the non-diseased subjects.

are described below. The choice of these functions is not unique and they may be replaced by other suitable densities.

1. The prevalence is assumed to follow a beta prior distribution with parameters $\alpha_\pi$ and $\beta_\pi$, *i.e.* $\pi \sim Beta(\alpha_\pi, \beta_\pi)$. The *Beta* distribution is chosen since it is defined over the entire range of possible values, $(0,1)$, of the prevalence. It is a versatile distribution and can be set to be diffuse, symmetric or skewed by a suitable choice of parameter values. Further, since the *Beta* prior distribution is conjugate to the *Binomial* likelihood (see Section 2.1.2, we will see that it simplifies the calculation of the full conditional density for $\pi$ when using the Gibbs sampler.

2. The sensitivities and specificities are also assumed to have *Beta* prior densities such that $S_j \sim Beta(\alpha_{S_j}, \beta_{S_j})$ , $j = 1,2$ and $C_j \sim Beta(\alpha_{C_j}, \beta_{C_j})$ , $j = 1,2$. Again, the range of these distributions match those for $S_j$ and $C_j$.

3. The feasible range of the covariance is determined by the sensitivities among the diseased subjects and the specificities among the non-diseased subjects. This can be verified from Table 4.3, which depicts the situation among the non-diseased subjects. Clearly,

$$C_1C_2 + covn \leq min(C_1, C_2),$$

where $min(a, b)$ is the minimum of $a$ and $b$. Similarly,

$$\begin{aligned} C_2 - (C_1C_2 + covn) &\leq 1 - C_1 \\ \Rightarrow\ covn &\geq C_1 + C_2 - C_1C_2 - 1 \\ &= (C_1 - 1)(1 - C_2). \end{aligned} \tag{4.6}$$

However, as mentioned in the previous section, we are only interested in the situation when the two tests are positively correlated, requiring that the lower bound of *covn* is 0. Since the expression in equation (4.6) is always negative, the lower bound of *covn* was fixed at 0. Therefore the upper and lower bounds for *covn* are given by

$$0 \leq covn \leq min(C_1, C_2) - C_1C_2. \tag{4.7}$$

Analogously, the bounds for *covp* are

$$0 \leq covp \leq min(S_1, S_2) - S_1S_2. \tag{4.8}$$

The *generalized Beta* distribution is suitable for the covariance parameters, since it can be defined over a range determined by their lower and upper bounds. A variable is said to have a *generalized Beta*$(\alpha, \beta)$ distribution when its density function is of the following form (Johnson et al., 1994):

$$f(x) = \frac{1}{B(\alpha, \beta)} \frac{(y-a)^{\alpha-1}(b-y)^{\beta-1}}{(b-a)^{\alpha+\beta-1}}, \quad a \leq y \leq b,\ \alpha > 0,\ \beta > 0.$$

This distribution has the same properties as the *Beta* distribution in that it is flexible and can take on various shapes by an appropriate selection of $(\alpha, \beta)$. The notation for the prior distribution parameters of the covariances is as follows:

$$covp \sim GenBeta(\alpha_{covp}, \beta_{covp}), \qquad 0 \leq covp \leq u_p,$$
$$\text{where } u_p = min(S_1, S_2) - S_1S_2.$$

$$\text{and,}\quad covn \sim GenBeta(\alpha_{covn}, \beta_{covn}), \qquad 0 \leq covn \leq u_n,$$
$$\text{where } u_n = min(C_1, C_2) - C_1C2.$$

### 4.2.3 *The Gibbs sampler algorithm*

When the likelihood function in equation (4.5) is combined with the above prior distributions, we obtain the following expression for the joint posterior distribution of the parameters:

$$
\begin{aligned}
& p(\pi, S_1, C_1, S_2, C_2, covp, covn, Y_{11}, Y_{10}, Y_{01}, Y_{00} | N_{11}, N_{10}, N_{01}, N_{00}) \\
& \propto (\pi(S_1 S_2 + covp))^{Y_{11}} (\pi(S_1(1-S_2) - covp))^{Y_{10}} \\
& \quad \times (\pi((1-S_1)S_2 - covp))^{Y_{01}} (\pi((1-S_1)(1-S_2) + covp))^{Y_{00}} \\
& \quad \times ((1-\pi)((1-C_1)(1-C_2) + covn))^{N_{11}-Y_{11}} \\
& \quad \times ((1-\pi)((1-C_1)C_2 - covn))^{N_{10}-Y_{10}} \\
& \quad \times ((1-\pi)(C_1(1-C_2) - covn))^{N_{01}-Y_{01}} ((1-\pi)(C_1 C_2 + covn))^{N_{00}-Y_{00}} \\
& \quad \times \pi^{\alpha_\pi - 1}(1-\pi)^{\beta_\pi - 1} S_1^{\alpha_{S_1} - 1}(1-S_1)^{\beta_{S_1} - 1} S_2^{\alpha_{S_2} - 1}(1-S_2)^{\beta_{S_2} - 1} \\
& \quad \times C_1^{\alpha_{C_1} - 1}(1-C_1)^{\beta_{C_1} - 1} C_2^{\alpha_{C_2} - 1}(1-C_2)^{\beta_{C_2} - 1} \\
& \quad \times covp^{\alpha_{covp} - 1}(u_p - covp)^{\beta_{covp} - 1} covn^{\alpha_{covn} - 1}(u_n - covn)^{\beta_{covn} - 1}.
\end{aligned}
$$

Due to the complexity of this expression, it is not possible to obtain the marginal distribution for the parameters analytically. Since we are interested in the marginal posterior densities of all parameters, we can use a Gibbs sampler algorithm as outlined in the following three steps:

1. Arbitrary starting values are chosen for each parameter as follows:

$$
\begin{aligned}
\pi = \pi^{(1)} & \text{ such that } 0 \le \pi^{(1)} \le 1, \\
S_j = S_j^{(1)} & \text{ such that } 0 \le S_j^{(1)} \le 1, \quad j = 1, 2, \\
C_j = C_j^{(1)} & \text{ such that } 0 \le C_j^{(1)} \le 1, \quad j = 1, 2, \\
covp = covp^{(1)} & \text{ such that } 0 \le covp^{(1)} \le min(S_1^{(1)}, S_2^{(1)}) - S_1^{(1)} S_2^{(1)}, \\
covn = covn^{(1)} & \text{ such that } 0 \le covn^{(1)} \le min(C_1^{(1)}, C_2^{(1)}) - C_1^{(1)} C_2^{(1)}, \text{ and} \\
y_{ij} = y_{ij}^{(1)} & \text{ such that } 0 \le y_{ij}^{(1)} \le N_{ij}, \quad i, j = 0, 1.
\end{aligned}
$$

2. At each iteration of the Gibbs sampler, a single value is sampled in turn from the full conditional distribution of each parameter. The full conditional distribution is obtained by selecting the terms containing the parameter of interest, from the product of the likelihood and the prior densities, and normalizing this function. For our problem, the full conditional distributions are given below. Note that while in theory we are conditioning on all parameters except the one whose distribution is being derived, in practice some parameters do not appear in all equations because of simplifications that arise.

   The full conditional distribution of the prevalence, $\pi$, given the other variables and latent data is

$$\begin{aligned} p(\pi \mid Y_{11}, Y_{10}, Y_{01}, Y_{00}) &\propto \pi^{\alpha_\pi + Y_{11} + Y_{10} + Y_{01} + Y_{00} - 1} \\ &\quad (1-\pi)^{\beta_\pi + N_{11} + N_{10} + N_{01} + N_{00} - (Y_{11} + Y_{10} + Y_{01} + Y_{00}) - 1}, \\ \Rightarrow \pi \mid Y_{11}, Y_{10}, Y_{01}, Y_{00} &\sim Beta(\alpha_\pi + Y_{11} + Y_{10} + Y_{01} + Y_{00}, \beta_\pi + N - (Y_{11} + Y_{10} + Y_{01} + Y_{00})). \end{aligned} \tag{4.9}$$

   Therefore, at each iteration of the Gibbs sampler a single value of $\pi$ is sampled from the *Beta* distribution in equation (4.9). The full conditional distributions for the other variables, namely the sensitivities, specificities and the covariance parameters were not of any standard form, so a SIR algorithm (see Chapter 2) was used to sample from these distributions. The full conditional distributions of the sensitivities are given (up to a constant of integration) by

$$\begin{aligned} &p(S_j | S_{3-j}, covp, Y_{11}, Y_{10}, Y_{01}, Y_{00}) \\ &\quad \propto (\pi(S_1 S_2 + covp))^{Y_{11}} (\pi(S_1(1-S_2) - covp))^{Y_{10}} \\ &\quad\quad \times (\pi((1-S_1)S_2 - covp))^{Y_{01}} \end{aligned}$$

$$\times (\pi((1-S_1)(1-S_2)+covp))^{Y_{00}} S_j^{\alpha_{S_j}-1}(1-S_j)^{\beta_{S_j}-1},$$

where j=1,2.

The full conditional distributions of the specificities are similarly given by

$$\begin{aligned} p(C_j|C_{3-j}, covn, Y_{11}, Y_{10}, Y_{01}, Y_{00}) \\ \propto ((1-\pi)((1-C_1)(1-C_2)+covn))^{N_{11}-Y_{11}} \\ \times ((1-\pi)((1-C_1)C_2-covn))^{N_{10}-Y_{10}} \\ \times ((1-\pi)(C_1(1-C_2)-covn))^{N_{01}-Y_{01}} \\ \times ((1-\pi)(C_1C_2+covn))^{N_{00}-Y_{00}} \\ \times C_j^{\alpha_{C_j}-1}(1-C_j)^{\beta_{C_j}-1}, \qquad \text{where j=1,2.} \end{aligned}$$

The full conditional distributions for the covariances are given by

$$\begin{aligned} p(covp|S_1, S_2, Y_{11}, Y_{10}, Y_{01}, Y_{00}) \\ \propto (\pi(S_1S_2+covp))^{Y_{11}} \\ \times (\pi(S_1(1-S_2)-covp))^{Y_{10}} \\ \times (\pi((1-S_1)S_2-covp))^{Y_{01}} \\ \times (\pi((1-S_1)(1-S_2)+covp))^{Y_{00}} \\ \times (covp-l_p)^{\alpha_{covp}-1}(u_p-covp)^{\beta_{covp}-1}, \end{aligned}$$

where $l_p = 0$ and $u_p = min(S_1, S_2) - S_1S_2$, and,

$$\begin{aligned} p(covn|C_1, C_2, Y_{11}, Y_{10}, Y_{01}, Y_{00}) \\ \propto ((1-\pi)((1-C_1)(1-C_2)+covn))^{N_{11}-Y_{11}} \\ \times ((1-\pi)((1-C_1)C_2-covn))^{N_{10}-Y_{10}} \\ \times ((1-\pi)(C_1(1-C_2)-covn))^{N_{01}-Y_{01}} \\ \times ((1-\pi)(C_1C_2+covn))^{N_{00}-Y_{00}} \\ \times (covp-l_n)^{\alpha_{covp}-1}(u_n-covp)^{\beta_{covp}-1}, \end{aligned}$$

where $l_n = 0$ and $u_n = min(C_1, C_2) - C_1C_2$.

The latent variables $Y_{ij}$ have *Binomial* full conditional distributions as follows:

$$Y_{11} \mid \pi, S_1, S_2, C_1, C_2, covp, covn \sim Bin(N_{11}, p_{11}),$$

$$\text{where } p_{11} = \frac{\pi(S_1S_2 + covp)}{\pi(S_1S_2 + covp) + (1-\pi)((1-C_1)(1-C_2) + covn)},$$

$$Y_{10} \mid \pi, S_1, S_2, C_1, C_2, covp, covn \sim Bin(N_{10}, p_{10}),$$

$$\text{where } p_{10} = \frac{\pi(S_1(1-S_2) - covp)}{\pi(S_1(1-S_2) - covp) + (1-\pi)((1-C_1)C_2 - covn)},$$

$$Y_{01} \mid \pi, S_1, S_2, C_1, C_2, covp, covn \sim Bin(N_{01}, p_{01}),$$

$$\text{where } p_{01} = \frac{\pi((1-S_1)S_2 - covp)}{\pi((1-S_1)S_2 - covp) + (1-\pi)(C_1(1-C_2) - covn)},$$

$$Y_{00} \mid \pi, S_1, S_2, C_1, C_2, covp, covn \sim Bin(N_{00}, p_{00}),$$

$$\text{where } p_{00} = \frac{\pi((1-S_1)(1-S_2) + covp)}{\pi((1-S_1)(1-S_2) + covp) + (1-\pi)(C_1C_2 + covn)}.$$

3. Step 2 is repeated a large number of times to obtain a sufficiently large sample from the full conditional distribution of each parameter. The resulting samples are approximate random samples from the marginal posterior density of each parameter, as discussed in Chapter 2.

In the following section we illustrate the application of this Gibbs sampler algorithm to a problem involving simulated data.

| Parameter | True value |
|---|---|
| $\pi$ | 0.73 |
| $S_1$ | 0.50 |
| $S_2$ | 0.80 |
| $covp$ | 0.07 |
| $C_1$ | 0.90 |
| $C_2$ | 0.70 |
| $covn$ | 0.05 |

Table 4.4: 'True' prevalence and test parameters.

## 4.3 A simulated example

In order to illustrate the performance of the Bayesian fixed effects method developed in the previous section, we simulated a hypothetical problem involving the results of two tests, neither of which was a gold standard. The parameters of the two tests were set up such that they had complementary characteristics *i.e.* the sensitivity of one test was very poor but it had a high specificity, while the other test had a reasonably high sensitivity but a worse specificity than the first test. Thus, it may be expected that the combined result of the two tests will provide more accurate results than either test alone.

### 4.3.1 Simulating the 'observed' data

The true values of the prevalence and test parameters were set to be as in Table 4.4. The tests were designed to be conditionally dependent both among the diseased and non-diseased subjects. The range of the covariance among the diseased and non-diseased subjects is, of course, limited by the values of the sensitivities and specificities

| | | Test1 | | |
|---|---|---|---|---|
| | | Positive | Negative | Total |
| Test2 | Positive | 73 | 60 | 133 |
| | Negative | 5 | 62 | 67 |
| | Total | 78 | 122 | 200 |

Table 4.5: Simulated cross-classification of results from two correlated tests.

as defined in equations (4.7) and (4.8), so that

$$\begin{aligned} 0 \leq covp &\leq min(S_1, S_2) - S_1 S_2 \\ &= min(0.5, 0.8) - (0.5)(0.8) = 0.1, \\ \text{and, } 0 \leq covp &\leq min(C_1, C_2) - C_1 C_2, \\ &= min(0.9, 0.70) - (0.9)(0.70) = 0.07. \end{aligned}$$

For this example we set $covp = 0.07$ and $covn = 0.05$, which lie within these admissible ranges.

The test results were generated by calculating the expected frequency of each cross-classification based on the equations (4.2), (4.3) and (4.4), using the values in Table 4.4. The 'observed' data would then appear as in Table 4.5. The S-plus program used to obtain these results is listed in the Appendix. If the tests were conditionally independent, *i.e.* the covariance between them was 0 in each disease category, we would observe the test results presented in Table 4.6. Note that the frequency of the $N_{00}$ and $N_{11}$ cells is less than in the correlated case, while the margins remain fixed.

| | | Test1 | | |
|---|---|---|---|---|
| | | Positive | Negative | Total |
| Test2 | Positive | 60 | 73 | 133 |
| | Negative | 18 | 49 | 67 |
| | Total | 78 | 122 | 200 |

Table 4.6: Simulated cross-classification of results from two independent tests.

### 4.3.2 *Determining the parameters of the prior distributions*

The prior means and 95% prior probability intervals of the test parameters, or in other words, the marginal prior information was set to be as in Table 4.7. Here and elsewhere, the abbreviation PI stands for probability interval. The corresponding *Beta* distribution parameters for the sensitivities and specificities were obtained by first determining the ratio of $\alpha : \beta$ by solving the equation $\frac{\alpha}{\alpha+\beta} = Mean$, and then finding the *Beta* distribution whose 95% probability interval matches that given in Table 4.7. For the two covariance parameters we used the same procedure, replacing the expression for the mean by $\frac{\alpha}{\alpha+\beta} \times (l-u) + u$, where $l$ and $u$ are the lower and upper bounds of the covariances given in Table 4.7, and then finding the *Generalized Beta* distribution whose 95% probability interval matches these values.

### 4.3.3 *Results*

We tried to see if we could re-capture the true prevalence when using a non-informative prior distribution over the prevalence along with the informative priors for the test parameters described above. This situation could arise, for instance, when two common tests are used to assess the prevalence of a disease in a population in which the disease distribution is unknown. The C++ program used to carry out the Gibbs sampler is listed in Appendix B.1.2. This program takes about 2 minutes to complete

| Parameter | Mean | 95% PI | $\alpha$ | $\beta$ |
|---|---|---|---|---|
| $S_1$ | 0.50 | 0.4-0.6 | 34 | 34 |
| $S_2$ | 0.80 | 0.7-0.9 | 90 | 10 |
| $Covp$ | 0.07 | 0.0-0.1 | 1.167 | 0.5 |
| $C_1$ | 0.90 | 0.85-0.95 | 32 | 8 |
| $C_2$ | 0.70 | 0.6-0.8 | 42 | 18 |
| $Covn$ | 0.05 | 0.0-0.07 | 2.5 | 1 |

Table 4.7: Prior means and 95% prior probability intervals of the test parameters and corresponding *Beta* distribution parameters for the two hypothetical tests.

20,000 iterations of the Gibbs sampler on a Pentium class computer. The posterior medians and 95% posterior probability intervals of the prevalence and test parameters thus obtained are listed in Table 4.8. The corresponding results using the Bayesian conditional independence model are presented in Table 4.9.

Figure 4.1 is a plot of the non-informative prior and the two posterior distributions which would be obtained when using the Bayesian fixed effects model and the Bayesian conditional independence model. We find that the prevalence would be underestimated if the correlation between the two tests was ignored. Figure 4.2 is a plot of the prior and the two posterior distributions for the sensitivity of the first test, obtained when using the Bayesian fixed effects model and the Bayesian conditional independence model respectively. Ignoring the correlation would result in an overestimate of the sensitivity. This is to be expected, since obtaining the results in Table 4.5 when using two independent tests would require higher values of the sensitivities and specificities for both tests. The same effect was observed on the remaining test parameters.

The CODA software package was used to obtain diagnostic statistics for the Gibbs sampler. The value of Gelman and Rubin's $\hat{R}$ statistic for each parameter was found

| Variable | Median | 95% PI |
|---|---|---|
| $\pi$ | 0.7341 | 0.5662 - 0.8984 |
| $S_1$ | 0.4995 | 0.4177 - 0.5814 |
| $C_1$ | 0.8932 | 0.8281 - 0.9422 |
| $pvp1$ | 0.9291 | 0.8435 - 0.9794 |
| $pvn1$ | 0.3920 | 0.1573 - 0.6033 |
| $S_2$ | 0.7902 | 0.7143 - 0.8652 |
| $C_2$ | 0.6966 | 0.5729 - 0.8019 |
| $pvp2$ | 0.8812 | 0.7431 - 0.9614 |
| $pvn2$ | 0.5531 | 0.2481 - 0.7550 |
| $covp$ | 0.0809 | 0.0478 - 0.0989 |
| $covn$ | 0.0379 | 0.0116 - 0.0639 |

Table 4.8: Posterior medians and 95% posterior probability intervals of the prevalence and test parameters obtained using the fixed effects model.

| Variable | Median | 95% PI |
|---|---|---|
| $\pi$ | 0.6137 | 0.5010 - 0.7163 |
| $S_1$ | 0.5680 | 0.4869 - 0.6510 |
| $C_1$ | 0.9242 | 0.8717 - 0.9605 |
| $pvp1$ | 0.8892 | 0.8128 - 0.9427 |
| $pvn1$ | 0.6649 | 0.5789 - 0.7483 |
| $S_2$ | 0.8965 | 0.8260 - 0.9484 |
| $C_2$ | 0.7287 | 0.6255 - 0.8228 |
| $pvp2$ | 0.7975 | 0.6935 - 0.8778 |
| $pvn2$ | 0.8559 | 0.7486 - 0.9309 |

Table 4.9: Posterior medians and 95% posterior probability intervals of the prevalence and test parameters obtained using the conditional independence model.

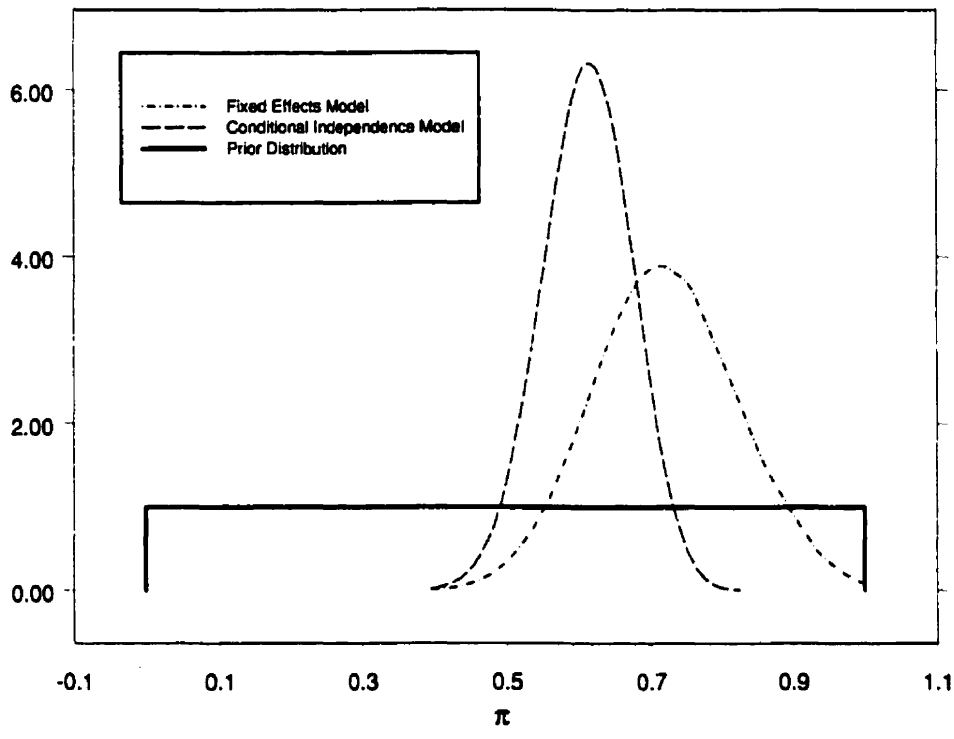


Figure 4.1: Prior distribution of $\pi$ overlaid by posterior distributions obtained using the fixed effects and conditional independence models.

to be close to one as shown in Table 4.10. Figure 4.3 shows the overlaid trace plots for the prevalence obtained from five different runs of the Gibbs sampler for the fixed effects model are indistinguishable after about 100 iterations, corroborating the evidence from the $\hat{R}$ statistic that the sequences have converged. Only the first 500 iterations are included here for clarity. Similar plots were obtained for the remaining parameters.

Raftery and Lewis' method was used to determine the minimum number of iterations required to estimate the 0.025 quantile with 95% probability with an accuracy of +/- 0.005. The results are presented in Table 4.11. The value of the $N_{min}$ statistic

| Iterations used for diagnostic = 2450:4899<br>Thinning interval = 1<br>Sample size per chain = 4899 | | |
|---|---|---|
| Variable | Point est. of $\hat{R}$ | 97.5% quantile |
| prev | 1.01 | 1.02 |
| sens1 | 1.00 | 1.00 |
| spec1 | 1.00 | 1.00 |
| pvp1 | 1.01 | 1.02 |
| pvn1 | 1.01 | 1.02 |
| sens2 | 1.00 | 1.00 |
| spec2 | 1.00 | 1.00 |
| pvp2 | 1.01 | 1.02 |
| pvn2 | 1.01 | 1.02 |
| covn | 1.00 | 1.00 |
| covp | 1.00 | 1.00 |

Table 4.10: Gelman and Rubin 50% and 97.5% shrink factors.

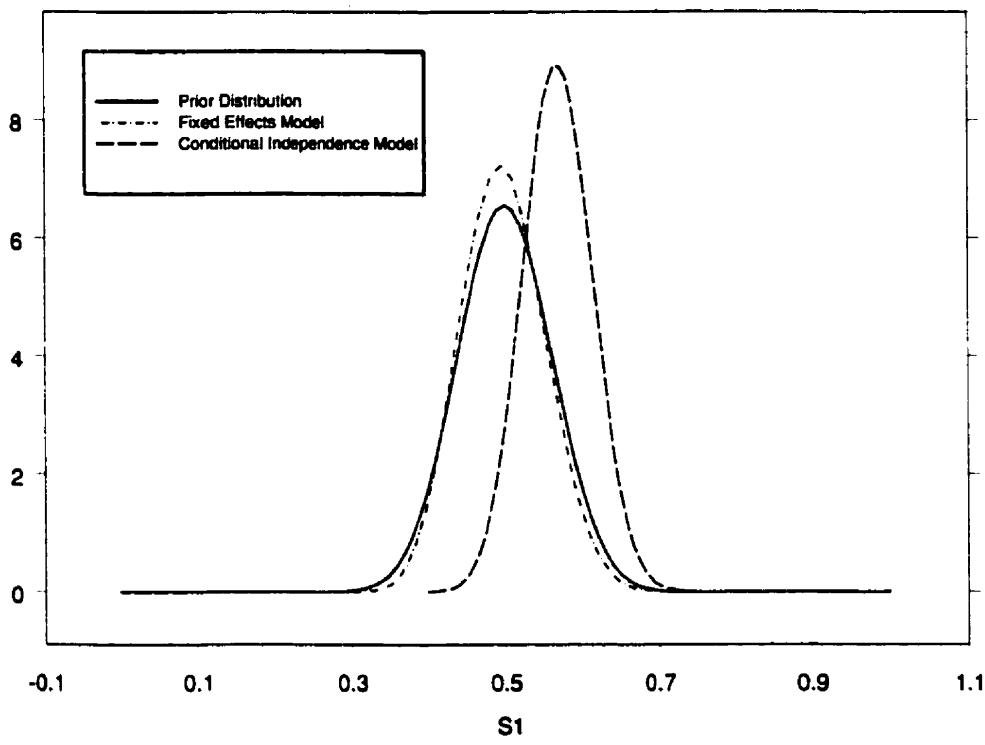


Figure 4.2: Prior distribution of $S_1$ overlaid by posterior distributions obtained using the fixed effects and conditional independence models.

was fairly high for some of the parameters resulting in a dependency factor greater than 1. This suggests a high correlation between successive iterations long after the sequence has converged. This is not surprising since in our non-identifiable problem the values of certain parameters are largely determined by the values of other parameters, and vice versa, thereby producing autocorrelations. This can be best seen by looking at the full conditional distributions of Section 4.2.3. While a high degree of autocorrelation suggests that the chain moves slowly through the range of the parameter, the accuracy of the parameter estimates becomes sufficiently high if a large enough number of iterations are run. Nevertheless, careful attention to Gibbs

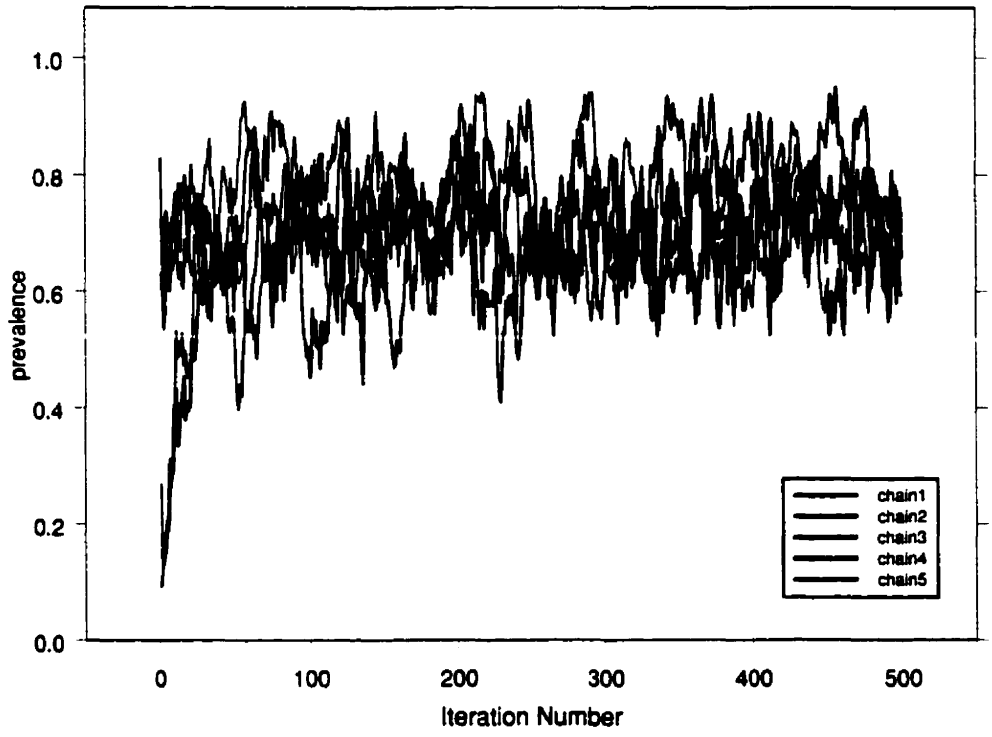


Figure 4.3: Overlaid trace plots of the prevalence from 5 different chains of the Gibbs sampler.

sampler diagnostics is required whenever this method is used.

## 4.4 Summary

In this chapter we have demonstrated that a Bayesian approach can be used to obtain a solution for the non-identifiable problem that arises when we have two conditionally dependent tests in the absence of a gold-standard. This fixed effects model further has the advantage that cross-classification probabilities are modeled in terms of test parameters and the covariance, allowing for a simple interpretation of the effect of conditional dependence on test results. The most challenging aspect of using this

| Iterations used = 101:4999<br>Thinning interval = 1<br>Sample size per chain = 4899 | | | | | |
|---|---|---|---|---|---|
| Quantile = 0.025<br>Accuracy = +/- 0.005<br>Probability = 0.95 | | | | | |
| Variable | Thin (k) | Burn-in (M) | Total (N) | Lower bound (Nmin) | Dependence factor (I) |
| prev | 3 | 24 | 21972 | 3746 | 5.87 |
| sens1 | 1 | 3 | 4206 | 3746 | 1.12 |
| spec1 | 2 | 8 | 8670 | 3746 | 2.31 |
| pvp1 | 2 | 16 | 15186 | 3746 | 4.05 |
| pvn1 | 2 | 16 | 18232 | 3746 | 4.87 |
| sens2 | 2 | 8 | 8670 | 3746 | 2.31 |
| spec2 | 2 | 10 | 13936 | 3746 | 3.72 |
| pvp2 | 3 | 30 | 26958 | 3746 | 7.2 |
| pvn2 | 2 | 18 | 19830 | 3746 | 5.29 |
| covn | 1 | 3 | 4198 | 3746 | 1.12 |
| covp | 1 | 2 | 3837 | 3746 | 1.02 |

Table 4.11: Raftery and Lewis convergence diagnostic.

method is to determine prior distributions which accurately represent the available information on the prevalence and the test parameters. This method is easy to implement using computational techniques such as a Gibbs sampler and a SIR. While our simulation showed the method to work well, we chose prior densities for four of the parameters which, while relatively wide, were centered on the 'true' values. Other prior distributions, of course, may not work as well, and in practice we will never be certain that our prior distribution 'covers' the true parameter values. In this sense, the methodology developed here may be viewed as a 'mapping' from a given set of prior distributions to the corresponding set of posterior distributions. Therefore, the posterior density can always be interpreted as a coherent updating of the prior distribution upon seeing the data, but any extrapolation to the 'truth' involves a leap of faith. A possible drawback of this method is its limitation to modeling the dependence between pairs of tests only. The next chapter presents a method free of this restriction.

# 5

# Modeling Conditional Dependence Using Random Effects

In this chapter, we present an approach to modeling the conditional dependence between multiple tests using random effects. The sensitivities and specificities of the tests are modeled as functions of a latent, subject-specific random variable. Applying the same latent value within each patient across all tests induces a dependence between the tests, without explicit reference to a covariance parameter. The chapter commences with some examples of such subject-specific variables. This is followed by the description of a Bayesian random effects approach to modeling this situation. The next section is devoted to a simulated example used to illustrate this method, and the last section summarizes our findings. The model is applied to real data in Chapter 6, where we compare the results to a fixed effects model.

## *5.1 Background*

In the classical diagnostic testing model, the sensitivity and specificity of a test are usually assumed to remain constant over all individuals to whom the test is applied. In practice however, test performance often varies between subjects for a variety of reasons. The source of this variation could be due to random or systematic errors of the kind discussed in Section 1.1. However, apart from factors related to the test or its laboratory analysis, there is often one or more covariates inherent to the subject that

accounts for this disparity. Three typical situations where this occurs are described below:

1. *Fecal Occult Blood Test*: This test, which is commonly used to screen for colorectal cancer, diagnoses a patient as positive when it detects traces of blood in the patient's stool. This is because malignant polyps are known to bleed intermittently, causing traces of blood to be present in the stool. As a corollary, a fecal occult blood test cannot detect polyps that do not bleed. Therefore, the bleeding biology of colorectal cancer ultimately determines the upper limit of screening efficacy when using this test. In subjects who have 'small' polyps, however, the test has a very poor sensitivity since smaller polyps do not bleed at all in certain subjects, for a reason not yet determined by medical research (Ransohoff and Lang, 1997).

2. *Stool Examination*: In a stool examination for an infectious or parasitic disease, a positive test means that the parasite of interest was directly observed under a microscope to be in the subject's stool specimen. The sensitivity of such a test is thus dependent on the ease of detecting the parasite in the stool. In a severely diseased case there is a larger concentration of parasites, making it easier to detect, and resulting in a more sensitive test. By the same token, the test has high specificity among subjects who are disease free since the absence of the parasite increases the likelihood of a negative test. Nevertheless a false positive test usually occurs when a different parasite is wrongly identified as the one of interest, so that subjects carrying other parasites have decreased specificity.

3. *Diagnosis of group A streptococcal infection*: A common problem faced by doctors in primary care is treating a sore throat. If the sore throat is diagnosed to be due to streptococcal infection the patient should be given antibiotics, but not if it is diagnosed to be due to a viral infection. While it is possible to find

this information using laboratory tests which take over 24 hours, the physician may want to prescribe the course of treatment immediately. To address this problem several predictor variables for streptococcal infection have been identified such as fever, no cough, and tonsillar swelling. However, the validity of these markers varies with age. An accurate scoring system therefore assigns a point for the presence of each of these symptoms/signs but modifies the score based on age (eg., 1 extra point for 3-14 year olds, 0 for 15-44 year olds, and -1 for age $\geq$ 45 years, McIsaac et al., 1998).

In the first example, the 'covariate' which determines whether or not the polyp bleeds is unknown and hence cannot be measured. The second example illustrates the type of covariate that most typically affects test performance. Subjects whose stool sample has a larger concentration of parasites, *i.e.* those with a more severe case of infection, are more likely to be detected. This is because the observed value of the separator variable has a much greater likelihood of satisfying the positivity criterion. The covariate of interest could be termed 'severity of illness', but it cannot be easily quantified, and is usually unknown at the time that the test must be interpreted. Finally, in the third example, the covariate is clearly defined and can also be measured.

The general situation, encompassing all three of the above cases, can be conceptualized as one where the performance of a subject on a test is a function of a continuous random variable, which we will term the 'intensity'. This 'intensity' can be thought of as a summary measure of the severity of illness / ease of detection along with any other covariates which affect a subject's performance on a test. The sensitivity and specificity of a test for *each* subject are functions of this underlying intensity, of the form $f(I)$, where $f$ is a continuous, monotonically increasing function taking values between 0 and 1. The higher the intensity of a subject, the greater the value of the sensitivity and the specificity of every test for that subject.

In the situation when we have two correlated tests, illustrated in Section 4.1, subjects who have a higher 'intensity' will tend to be correctly detected by both tests and will therefore fall into the $(T_1 = 1, T_2 = 1)$ or $(T_1 = 0, T_2 = 0)$ cells where the tests are in agreement. Conversely, subjects who fall into the $(T_1 = 1, T_2 = 0)$ or $(T_1 = 0, T_2 = 1)$ cells, where test results are not consistent, tend to have a lower intensity. Thus a dependence is induced between tests via the test parameters without explicitly using a parameter for the covariance. The range of values, as well as the 'meaning' of the intensity may be different among the diseased and non-diseased subjects. In Example 2 above, it may be that 'intensity' indeed measures severity of disease among diseased subjects, but among non-diseased subjects higher intensity may correspond to an absence of other parasites that could lead to a false-positive diagnosis.

A possible approach to represent the above situation is by way of a random effects model since $I$ is usually latent. While it is unknown what distribution the values of $I$ take, in the Bayesian random effects model described below $I$ is taken to be a random variable following a $N(\mu, \sigma^2)$ distribution. Without loss of generality we will use a $N(0, 1)$ distribution. Densities other than the *Normal* could be used, although we do not investigate them here.

## 5.2 A Bayesian random effects model

In this section we propose a Bayesian solution to the 'Two Latent Class Random Effects (2LCR) Model' discussed in the paper by Qu et al., 1996 which was described earlier in Section 3.3.4. The main advantage of this model over the earlier approach, is that it provides a solution even in the situation when we have a non-identifiable model, *i.e.* when the number of tests is less than 4 or 5. By placing a prior distribution over all unknown parameters, it is possible to obtain a joint posterior distribution over

these same parameters, even though the problem is non-identifiable.

### 5.2.1 Notation

We consider the general scenario with $p$ tests and $N$ subjects. The test result for the $k^{th}$ subject on the $j^{th}$ test is denoted by $t_{jk} = 1$ or 0 for a positive or a negative result, respectively. The vector of results for the $k^{th}$ subject on each of the $p$ tests is denoted by $t_k = (t_{1k}, \ldots, t_{pk})$. The true disease status of the $k^{th}$ subject is denoted by $D = d_k$ $d_k = 0, 1$. The 'intensity' of the $k^{th}$ subject is denoted by $i_k$.

### 5.2.2 The model

As defined earlier, the probability that the $k^{th}$ subject has a positive result on the $j^{th}$ test is given by

$$P(t_{jk} = 1|D = d_k, I = i_k) = \Phi(a_{jd} + b_{jd}i_k), \quad d_k = 0, 1, \ I \sim N(0, 1),$$

where $\Phi$ represents the cumulative distribution function of the $N(0, 1)$ distribution and $(a_{jd}, b_{jd})$, $j = 1, \ldots, p$, $d = 0, 1$ are real, unknown parameters. It follows that the probability that the $k^{th}$ subject has a negative result on the $j^{th}$ test is given by

$$P(t_{jk} = 0|D = d_k, I = i_k) = 1 - \Phi(a_{jd} + b_{jd}i_k), \quad d_k = 0, 1, \ I \sim N(0, 1).$$

The disease status $D$ and the latent variable $I$ are assumed to be independent of each other. The mean sensitivity of the $j^{th}$ test over all subjects is then given by integrating the expression for the sensitivity of the $k^{th}$ subject with respect to $I$, as follows:

$$\begin{aligned} S_j &= P(t_{jk} = 1|D = 1) = \int_{-\infty}^{\infty} S_{jk}\, d\Phi(i_k) \\ &= \int_{-\infty}^{\infty} \Phi(a_{j1} + b_{j1}i_k) d\Phi(i_k) \end{aligned}$$

$$= \int_{-\infty}^{\frac{a_{j1}}{1+b_{j1}^2}} \Phi(x)d\Phi(x) \quad \text{(using a } \tan^{-1}(\tfrac{a_{j1}}{1+b_{j1}^2}) \text{ anticlockwise rotation)}$$
$$= \Phi(\frac{a_{j1}}{1+b_{j1}^2}).$$

Similarly, the specificity of the $j$th test is given by

$$\begin{aligned} C_j &= P(t_{jk} = 0|D = 0) \\ &= 1 - \Phi(\frac{a_{j0}}{1+b_{j0}^2}) \\ &= \Phi(-\frac{a_{j0}}{1+b_{j0}^2}) \text{ using the result } \Phi(-x) = 1 - \Phi(x). \end{aligned}$$

The sensitivity and specificity of each individual subject can be thought of as being shifted from the mean by an amount determined by the magnitude of the subject's 'intensity' $i_k$.

The results of different tests are taken to be independent of each other conditional on the disease status $D$ and the latent variable $I$. This means that within each disease class the test results are independent of each other conditional on the variable $I$. Therefore the likelihood for the $k^{th}$ subject given $i_k$ is

$$\begin{aligned} P(t_{1k}, \ldots, t_{pk}|\psi, i_k) &= \pi \prod_{j=1}^{p} S_{jk}^{t_{jk}}(1 - S_{jk})^{(1-t_{jk})} + (1-\pi) \prod_{j=1}^{p} C_{jk}^{(1-t_{jk})}(1 - C_{jk})^{t_{jk}} \\ &= \pi \prod_{j=1}^{p} \Phi(a_{j1} + b_{j1}i_k)^{t_{jk}}(1 - \Phi(a_{j1} + b_{j1}i_k))^{(1-t_{jk})} \\ &\quad + (1-\pi) \prod_{j=1}^{p} \Phi(a_{j0} + b_{j0}i_k)^{(1-t_{jk})}(1 - \Phi(a_{j0} + b_{j0}i_k))^{t_{jk}}, \end{aligned} \tag{5.1}$$

where $\psi$ is the vector of parameters to be estimated, so that $\psi = (\pi, (a_{jd}, b_{jd})\ j = 1, \ldots, p,\ d = 0, 1)$. The 'complete' likelihood for the $k^{th}$ subject, given the latent data $i_k$ and $d_k$, is given by

$$P(t_{1k}, \ldots, t_{pk}|\psi, i_k, d_k) = (\pi \prod_{j=1}^{p} \Phi(a_{j1} + b_{j1}i_k)^{t_{jk}}(1 - \Phi(a_{j1} + b_{j1}i_k))^{(1-t_{jk})})^{d_k}$$

$$\times ((1-\pi)\prod_{j=1}^{p}\Phi(a_{j0}+b_{j0}i_k)^{(1-t_{jk})}$$
$$(1-\Phi(a_{j0}+b_{j0}i_k))^{t_{jk}})^{(1-d_k)}.$$

Clearly, the case when $b_{jd}=0, j=1,\ldots,p,\ d=0,1$ corresponds to the Bayesian conditional independence model. It follows that the likelihood function of the observed and latent data for all subjects is given by

$$\begin{aligned}L &\propto \prod_{k=1}^{N} P(t_{1k},\ldots,t_{pk}|\psi,i_k,d_k)\\ &= \prod_{k=1}^{N}(\pi\prod_{j=1}^{p}\Phi(a_{j1}+b_{j1}i_k)^{t_{jk}}(1-\Phi(a_{j1}+b_{j1}i_k))^{(1-t_{jk})})^{d_k}\\ &\quad\times((1-\pi)\prod_{j=1}^{p}\Phi(a_{j0}+b_{j0}i_k)^{(1-t_{jk})}(1-\Phi(a_{j0}+b_{j0}i_k))^{t_{jk}})^{(1-d_k)}.\end{aligned}$$

Let $p(\theta)$ denote the prior distribution of a parameter $\theta$. The expression for the joint posterior distribution of $\psi$ is obtained by multiplying the likelihood above by the prior distribution for each parameter as follows:

$$\begin{aligned}&p(\psi \mid t_{1k},\ldots,t_{pk},i_k,d_k,\ k=1,\ldots,N)\\ &\propto \prod_{k=1}^{N}(\pi\prod_{j=1}^{p}\Phi(a_{j1}+b_{j1}i_k)^{t_{jk}}(1-\Phi(a_{j1}+b_{j1}i_k))^{(1-t_{jk})})^{d_k}\\ &\quad\times((1-\pi)\prod_{j=1}^{p}\Phi(a_{j0}+b_{j0}i_k)^{(1-t_{jk})}(1-\Phi(a_{j0}+b_{j0}i_k))^{t_{jk}})^{(1-d_k)}\\ &\quad\times p(\pi)\prod_{d=0}^{1}\prod_{j=1}^{p}p(a_{jd},b_{jd}).\end{aligned} \tag{5.2}$$

As explained in Chapter 2, the elicitation of the parameters for the prior distributions is perhaps the most important step in implementing this method. These parameters are obtained in consultation with results from earlier studies and expert opinion. We provide an example of this in Chapter 6. When the prior information can be expressed using the prior densities given below, the expressions for the full conditional distribution of each parameter, which will be used in the Gibbs sampler algorithm, are simplified.

1. A $Beta(\alpha_\pi, \beta_\pi)$ prior distribution can be used for the prevalence, $\pi$, since it is conjugate to the binomial distribution. This distribution can take on a variety of shapes over the range $(0, 1)$ by appropriate adjustment of the parameters $(\alpha_\pi, \beta_\pi)$.

2. A bivariate normal prior distribution, $N_2(\begin{pmatrix} a_{JD} \\ b_{JD} \end{pmatrix}, \Sigma_{JD})$, over the parameter pairs $(a_{jd}, b_{jd})$ can be used. The vector $\begin{pmatrix} a_{JD} \\ b_{JD} \end{pmatrix}$ denotes the bivariate mean and $\Sigma_{JD}$ is the $2 \times 2$ covariance matrix.

   This prior distribution also facilitates sampling from the full conditional distribution using a method developed by Albert and Chib, 1993, which is described in the Appendix.

### 5.2.3 *Implementing the Gibbs sampler*

Assuming the joint prior distribution is the product of the individual densities discussed above, the product of the likelihood function in equation (5.2) and the joint prior distribution is:

$$
\begin{aligned}
& p(\psi \mid t_{1k}, \ldots, t_{pk}, i_k, d_k,\ k = 1, \ldots, N) \\
& \propto \pi^{\sum_{k=1}^{N} d_k + \alpha_\pi - 1} (1-\pi)^{\sum_{k=1}^{N} (1-d_k) + \beta_\pi - 1} \\
& \quad \times \prod_{k=1}^{N} \Big( \prod_{j=1}^{p} \Phi(a_{j1} + b_{j1} i_k)^{t_{jk}} (1 - \Phi(a_{j1} + b_{j1} i_k))^{(1-t_{jk})} \Big)^{d_k} d\Phi(a_{j1}, b_{j1}) \\
& \quad \times \Big( \prod_{j=1}^{p} \Phi(a_{j0} + b_{j0} i_k)^{(1-t_{jk})} (1 - \Phi(a_{j0} + b_{j0} i_k))^{t_{jk}} \Big)^{(1-d_k)} d\Phi(a_{j0}, b_{j0}),
\end{aligned}
$$

where $d\Phi(a_{jd}, b_{jd})$ is the bivariate normal density

$$
f(\tilde{x} | \tilde{X}, \Sigma) = \frac{exp(-(\tilde{x} - \tilde{X})^T \Sigma^{-1} (\tilde{x} - \tilde{X})/2)}{2\pi |\Sigma|},
$$

$$\text{with } \tilde{x} = \begin{pmatrix} a_{jd} \\ b_{jd} \end{pmatrix} \text{ and } \tilde{X} = \begin{pmatrix} a_{JD} \\ b_{JD} \end{pmatrix}.$$

As in the case of the fixed effects model, we use a Gibbs sampler algorithm to sample from the full conditional distributions of the parameters in $\psi$ as follows:

1. Assign random starting values to each of the parameters and latent variables. Possible starting values are, for example,

$$\begin{aligned} \pi &= \pi^{(1)},\ 0 \leq \pi^{(1)} \leq 1, \\ (a_{jd}, b_{jd}) &= (a_{jd}^{(1)}, b_{jd}^{(1)}),\ j = 1, \ldots, p,\ d = 0, 1, \\ &\qquad \text{such that } \begin{pmatrix} a_{jd}^{(1)} \\ b_{jd}^{(1)} \end{pmatrix} \sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right), \\ d_k &= d_k^{(1)},\ k = 1, \ldots, N, \\ &\qquad \text{such that } d_k^{(1)} = 0 \text{ or } 1, \\ i_k &= i_k^{(1)},\ k = 1, \ldots, N, \\ &\qquad \text{such that } i_k^{(1)} \sim N(0, 1). \end{aligned}$$

   When running a Gibbs sampler, however, it is usual to run it several times with different starting values that are overdispersed with respect to the range of the expected posterior densities, in order to verify convergence.

2. Draw a single value from the full conditional distribution of each parameter, *i.e.* the distribution of each parameter conditional on the most recently updated values of the other parameters. This distribution is obtained by isolating from the product of the likelihood and the prior, the product of the terms which contain the parameter of interest, and normalizing it, whenever possible. If normalization is difficult, a SIR or other algorithm may be used as needed.

(a) For the prevalence, the full conditional distribution is given by

$$p(\pi|d_1,\ldots,d_N) \propto \pi^{\sum_{k=1}^N d_k+\alpha_\pi-1}(1-\pi)^{N-\sum_{k=1}^N d_k+\beta_\pi-1},$$
$$\Rightarrow \pi|d_1,\ldots,d_N \sim Beta(\sum_{k=1}^N d_k+\alpha_\pi, N-\sum_{k=1}^N d_k+\beta_\pi). \quad (5.3)$$

Thus at each iteration $\pi$ is set equal to a random value drawn from $Beta(\sum_{k=1}^N d_k+\alpha_\pi, N-\sum_{k=1}^N d_k+\beta_\pi)$.

(b) The full conditional distribution for each $d_k$ is given by

$$\begin{aligned} &p(d_k|t_{1k},\ldots,t_{pk},\psi,i_k) \\ &\quad\propto (\pi\prod_{j=1}^p \phi(a_{j1}+b_{j1}i_k)^{t_{jk}}(1-\phi(a_{j1}+b_{j1}i_k))^{(1-t_{jk})})^{d_k} \\ &\qquad\times((1-\pi)\prod_{j=1}^p \phi(a_{j0}+b_{j0}i_k)^{(1-t_{jk})}(1-\phi(a_{j0}+b_{j0}i_k))^{t_{jk}})^{(1-d_k)}, \end{aligned}$$

$$\Rightarrow d_k|t_{1k},\ldots,t_{pk},\psi,i_k \sim Bernoulli(p_k), \quad (5.4)$$

where

$$p_k = \frac{\pi\prod_{j=1}^p \Phi(a_{j1}+b_{j1}i_k)^{t_{jk}}(1-\Phi(a_{j1}+b_{j1}i_k))^{(1-t_{jk})}}{Q},$$

$$\begin{aligned} \text{and } Q = &\pi\prod_{j=1}^p \Phi(a_{j1}+b_{j1}i_k)^{t_{jk}}(1-\Phi(a_{j1}+b_{j1}i_k))^{(1-t_{jk})} \\ &+(1-\pi)\prod_{j=1}^p \phi(a_{j0}+b_{j0}i_k)^{(1-t_{jk})}(1-\phi(a_{j0}+b_{j0}i_k))^{t_{jk}}. \end{aligned}$$

Thus, at every iteration we draw a value for each $d_k$, $k=1\ldots,N$ from a $Bernoulli(p_k)$, $k=1\ldots,N$ distribution.

(c) The full conditional distribution for each $(a_{j1},b_{j1})$ is given by

$$\begin{aligned} &p(a_{j1},b_{j1}|t_{j1},\ldots,t_{jN},d_1,\ldots,d_N) \\ &\qquad\propto \prod_{k=1}^N \Phi(a_{j1}+b_{j1}i_k)^{d_k t_{jk}} \\ &\qquad\quad\times(1-\Phi(a_{j1}+b_{j1}i_k))^{d_k(1-t_{jk})} d\Phi(a_{j1},b_{j1}). \end{aligned} \quad (5.5)$$

Similarly, the full conditional distribution for each $(a_{j0}, b_{j0})$ is given by

$$p(a_{j0}, b_{j0}|t_{11}, \ldots, t_{jN}, d_1, \ldots, d_N) \propto \prod_{k=1}^{N} \Phi(a_{j0} + b_{j0} i_k)^{(1-d_k)(1-t_{jk})} \times (1 - \Phi(a_{j0} + b_{j0} i_k))^{(1-d_k) t_{jk}} d\Phi(a_{j0}, b_{j0}). \tag{5.6}$$

The full conditional distributions for the $(a_{jd}, b_{jd})$ pairs are not of a standard distributional form but they can be sampled from using approximate methods like a SIR algorithm (described in section 2.4.1). Albert and Chib, 1993 have developed a method which is of particular interest in these situations which is described the Appendix. We used this method here. Metropolis sampling (Hastings, 1970) could also be used.

(d) The full conditional distribution of the intensity, of the $k^{th}$ subject, $i_k$, $k = 1, \ldots, N$, is given by

$$p(i_k|t_{1k}, \ldots, t_{pk}, (a_{jd}, b_{jd}),\ j = 1, \ldots, p,\ d = 0, 1) \tag{5.7}$$

$$\propto \prod_{j=1}^{p} \Phi(a_{j1} + b_{j1} i_k)^{d_k t_{jk}} (1 - \Phi(a_{j1} + b_{j1} i_k))^{d_k (1-t_{jk})} \times \prod_{j=1}^{p} \Phi(a_{j0} + b_{j0} i_k)^{(1-d_k)(1-t_{jk})} (1 - \Phi(a_{j0} + b_{j0} i_k))^{(1-d_k) t_{jk}} \tag{5.8}$$

Since equation (5.8) is not reducible to the form of any standard density function the $i_k$ values are sampled at each iteration using a SIR algorithm.

3. Step two is repeated many times to obtain a sufficiently large sample from the full conditional distribution of each parameter. The resulting samples are approximate random samples from the marginal posterior densities of each parameter.

The next section applies this method to a simulated data set.

## 5.3 A simulated example

### 5.3.1 Simulating the 'observed' data

As in the case of the fixed effects model, we used simulated results from two hypothetical, non-gold standard tests to examine whether adjusting for the conditional dependence between the two tests, by way of the random effects model, affects the inference about the prevalence and test parameters. The 'true' values of the prevalence and test parameters along with the corresponding values of the $(a_{jd}, b_{jd})$ parameters for the sensitivities and specificities are listed in Table 5.1. Even though there is no explicit parameter for the covariance, the value of the $(a_{jd}, b_{jd})$ pairs was determined such that the tests are conditionally dependent, as will be explained later in this section. It must be noted that the parameters $S_1, S_2, C_1$ and $C_2$ do not have their usual meanings, since the test properties are different for each subject depending on $i_k$. Therefore, the values in Table 5.1 represent mean values averaged over all possible $i_k \sim N(0,1)$. As was explained in Chapter 4, the range of values for the covariances is determined by equation (4.7). As in the previous chapter we set the two covariances to be $covp = 0.07$ and $covn = 0.05$.

The probability of falling into each possible cross-classification $(T_1 = i, T_2 = j)$ $i, j = 0, 1$ for each of the $N = 200$ subjects is calculated using equation (5.1). The overall 'observed' cross-classification is obtained by summing the probabilities for each classification over all subjects. For example, the number of subjects in the cell $(T_1 = 1, T_2 = 1)$ is given by

$$\begin{aligned}
&\text{Number of subjects in cross-classification } (T_1 = 1, T_2 = 1),\\
&= \sum_{k=1}^{N} P(t_k = (1,1)),\\
&= \sum_{k=1}^{N} (\pi \prod_{j=1}^{2} \Phi(a_{j1} + b_{j1} i_k) + (1-\pi) \prod_{j=1}^{2} (1 - \Phi(a_{j0} + b_{j0} i_k))).
\end{aligned}$$

| Parameter | True value | $a_{jd}$ | $b_{jd}$ |
|---|---|---|---|
| $\pi$ | 0.73 | | |
| $covp$ | 0.07 | | |
| $covn$ | 0.05 | | |
| $S_1$ | 0.50 | 0 | 1.2731 |
| $S_2$ | 0.80 | 1.3625 | 1.2731 |
| $C_1$ | 0.90 | 2.2536 | 1.4465 |
| $C_2$ | 0.70 | 0.9221 | 1.4465 |

Table 5.1: 'True' prevalence and test parameters with corresponding $(a_{jd}, b_{jd})$ values.

| | | Test1 | | |
|---|---|---|---|---|
| | | Positive | Negative | Total |
| Test2 | Positive | 75 | 58 | 133 |
| | Negative | 5 | 62 | 67 |
| | Total | 80 | 120 | 200 |

Table 5.2: Simulated cross-classification of results from two correlated tests.

where the $I = i_k$ values for each subject are drawn randomly from a $N(0, 1)$ distribution for each subject. The program for implementing this algorithm is listed in the Appendix. It is of interest to note that the data set in Table 4.1 which was simulated using a fixed effects model can also be observed under the random effects model.

### 5.3.2 *Determining the parameters of the prior distributions*

Once again we assumed that there was no prior information available about the prevalence and used a $Beta(1, 1)$ prior distribution for it. The prior means and 95% prior probability intervals for the test parameters was set to be as in Table 5.3.

| Parameter | Mean | 95% PI |
|---|---|---|
| *covp* | 0.07 | 0.01-0.09 |
| *covn* | 0.05 | 0.01-0.07 |
| $S_1$ | 0.50 | 0.4-0.6 |
| $C_1$ | 0.90 | 0.85-0.95 |
| $S_2$ | 0.80 | 0.7-0.9 |
| $C_2$ | 0.70 | 0.6-0.8 |

Table 5.3: Prior means and 95% prior probability intervals of the test parameters of the two hypothetical tests.

Determining the $(a_{JD}, b_{JD}), J = 1, 2, D = 0, 1$ values is a difficult exercise. Here we use the $2LCR1$ model which is a particular case of the $2LCR$ model when $b_{jd} = b_d, \ j = 1, 2, d = 0, 1$. This means that a change in the value of $i_k$ will cause the sensitivities and specificities of all tests for the $k^{th}$ subject to change by the same amount on the probit scale. We use the $2LCR1$ model in part because the available information about the mean values of the test parameters is not sufficient to uniquely determine the $(a_{JD}, b_{JD}), \ J = 1, 2, \ D = 0, 1$ parameters in the $2LCR$ model using the method described below. This problem does not arise when we have more than two tests. However, the restriction of the $2LCR1$ model does not substantially affect the estimates of the sensitivities and specificities since the effect of the $b_{jd}$ parameters on their mean values is small, as will be shown later in this section. Further, this model has the advantage of being easier to interpret since it has fewer parameters. One method to elicit the means of the prior distributions of the $(a_{j1}, b_{j1}), \ j = 1, 2$ parameters in the $2LCR1$ model from knowledge about the mean values of the sensitivities in the population, is by solving the following 3 equations for $(a_{11}, a_{21}, b_1)$:

$$\Phi\left(\frac{a_{11}}{\sqrt{1 + b_1^2}}\right) = S_1, \tag{5.9}$$

$$\Phi\left(\frac{a_{2l}}{\sqrt{1+b_l^2}}\right) = S_2, \tag{5.10}$$

$$\int_{-\infty}^{\infty} \Phi(a_{1l} + b_l i_k)\, \Phi(a_{2l} + b_l i_k) d\Phi(i_k) - S_1 S_2 = covp. \tag{5.11}$$

The first two equations relate the information about the mean sensitivities in Table 5.3 to the expression for the mean sensitivity of each test in terms of the $(a_{jd}, b_{jd})$'s. The last equation relates the mean covariance to the expression for the covariance in terms of the sensitivities.

These equations can be solved using a bisection algorithm as follows:

1. Transform the equations (5.9) and (5.10) such that $a_{Jl} = \Phi^{-1}(S_J)\sqrt{1+b_l^2}$, $J = 1, 2$.

2. Substitute the expressions for $a_{Jl}$, $J = 1, 2$ in terms of $b_l$ in equation (5.11) to obtain:

$$\int_{-\infty}^{\infty} \prod_{J=1}^{2} \Phi(\Phi^{-1}(S_J)\sqrt{1+b_l^2} + b_l i_k) d\Phi(i_k) - S_1 S_2 = covp$$

$$\Rightarrow \int_{-\infty}^{\infty} \prod_{J=1}^{2} \Phi(\Phi^{-1}(S_J)\sqrt{1+b_l^2} + b_l i_k) d\Phi(i_k)$$

$$-S_1 S_2 - covp = 0$$

3. Let $f(b_l) = \int_{-\infty}^{\infty} \prod_{J=1}^{p} \Phi(\Phi^{-1}(S_J)\sqrt{1+b_l^2} + b_l i_k) d\Phi(i_k) - S_1 S_2 - covp$. The solution for $b_l$ must satisfy $f(b_l) = 0$. We start by fixing the lower $(l)$ and upper $(u)$ bounds between which the solution must lie. The idea is that if $b_1$ is truly bounded by $l$ and $u$ then $f(l)f(u) < 0$ since $f(b_1) = 0$. A reasonable starting value for $b_l$ is $x = (l+u)/2$. If $f(x)f(l) < 0$ then the two are of opposite signs and the solution must lie between $l$ and $x$, so the upper bound is changed to $u = x$. If on the other hand, $f(x)f(l) > 0$ then $f$ takes the same sign at both $l$ and $u$ and therefore the lower bound is altered to $l = x$.

4. Step 3 is repeated till $f(x)$ is smaller than a predetermined value $\epsilon$.

The parameters for the prior distributions of the specificities can be calculated similarly. It must be noted that while eliciting the prior distributions for these parameters we have considered what is known about their mean values, which are in fact averaged over the individual-specific properties in the population under study. The degree to which the properties vary over the population is controlled mostly by the $b_d$ values.

The solution to the equations (5.9), (5.10) and (5.11) is $(a_{11} = 0, a_{21} = 1.362, b_1 = 1.273)$. To determine the approximate prior standard deviations of $a_{11}$, $a_{21}$ and $b_1$ we used contour plots of $S_l$ on the $(a_{11}, b_1)$ plane and $S_2$ on the $(a_{21}, b_1)$ plane, respectively, which were constructed using the following steps:

1. We first generated sequences of points lying between (mean-1, mean+1) for each of the three parameters. Then, a 3-D grid of $(a_{11}, a_{21}, b_1)$ values was created using all possible combinations of values from the three sequences.

2. At each point on this 3-D grid we calculated the values of $S_1 = \Phi(\frac{a_{11}}{\sqrt{1+b_1^2}})$, $S_2 = \Phi(\frac{a_{21}}{\sqrt{1+b_1^2}})$ and $covp = \int_{-\infty}^{\infty} \Phi(a_{1l} + b_l i_k)\, \Phi(a_{2l} + b_l i_k) d\Phi(i_k) - S_1 S_2$.

3. We then plotted $S_1$ values on a 2-D plot of $b_1$ vs. $a_{11}$, $S_2$ values on a 2-D plot of $b_1$ vs. $a_{21}$ and *covp* values on both the $(a_{11}, b_1)$ and $(a_{21}, b_1)$ planes.

From Figure 5.1, which is the contour plot of $S_1$ on the $(a_{11}, b_1)$ plane, we can see that as $S_1$ ranges from 0.4 to 0.6 (its 95% prior probability interval), $a_{11}$ ranges approximately from -0.270369 to 0.270369. The standard deviation of $a_{11}$ was taken to be a quarter of this range namely $sd(a_{11}) = \frac{0.270369-(-0.270369)}{4} = 0.1351845$. The range of $b_1$ is less obvious, since the same value of $b_1$ could correspond to the entire range of values of $S_1$. Therefore we set the standard deviation to a relatively large

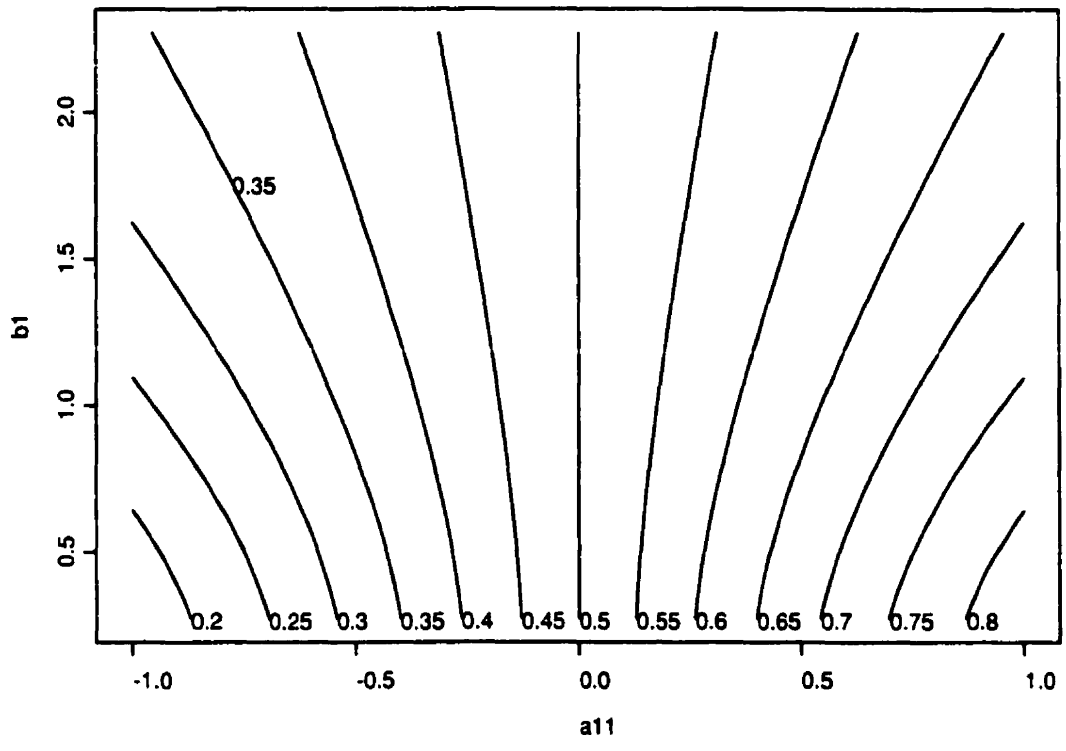


Figure 5.1: Contour plot of $S_1$ on the $(a_{11}, b_1)$ plane.

value. We can deduce from this that the value of $S_1$ is mainly determined by $a_{11}$, while the value of $b_1$ has a greater bearing on the value of the covariance between the tests. Intuitively, this could have been expected since the covariance enters into the model via the 'intensity' $i_k$, which is multiplied by $b_1$.

The standard deviations of the remaining parameters were deduced similarly. The values of the means and standard deviations of the prior distributions of the $(a_{jd}, b_{jd})$ parameter pairs are summarized in Table 5.4. In order to validate our method of prior elicitation we generated a random sample from the prior distributions of the $(a_{jd}, b_{jd})$ pairs using the parameter values in Table 5.4 and calculated the mean sensitivities

| | Mean | S.D.* |
|---|---|---|
| $a_{11}$ | 0 | 0.1332 |
| $a_{21}$ | 1.3625 | 0.2135 |
| $b_1$ | 1.2731 | 0.5 |

| | Mean | S.D.* |
|---|---|---|
| $a_{10}$ | 2.2536 | 0.0801 |
| $a_{20}$ | 0.9221 | 0.2579 |
| $b_0$ | 1.4465 | 0.5 |

* S.D. = Standard Deviation

Table 5.4: Prior means and standard deviations for parameters determining sensitivity and specificity in the random effects model.

and specificities. The means and 95% probability intervals of these samples were found to be close to the corresponding values for the sensitivities and specificities presented earlier in Table 5.3.

### 5.3.3 *Results*

A C++ program, which is listed in the Appendix, was used to implement the Gibbs sampler for the random effects model. The values of posterior medians and 95% posterior probability intervals thus obtained are listed in Table 5.5. The Gibbs sampler takes about one hour to complete 20000 iterations. It is slowed down chiefly by the fact that there are several parameter values, such as the individual sensitivities and specificities of each subject and their intensities, that need to be calculated and stored. The results that would have been obtained had we ignored the conditional dependence and used the Bayes conditional independence model are presented in Table 5.6. Figure 5.2 is a plot of the prior distribution of the prevalence overlaid by the posterior distributions obtained when the conditional dependence is taken into account and when it is ignored. Figure 5.3 is a similar plot for the sensitivity of Test 2. Once again we see that the prevalence has been underestimated when the conditional dependence is ignored, whereas the sensitivity of Test 2 is overestimated. The same is true for the remaining sensitivities and specificities. A representative sample of the posterior medians and 95% posterior probability intervals of the intensity for 16

| Variable | Median | 95% PI |
|---|---|---|
| $\pi$ | 0.7405 | 0.5872 - 0.9107 |
| $a_{11}$ | 0.0032 | -0.7002 - 0.7116 |
| $a_{21}$ | 1.4411 | 0.9152 - 2.0906 |
| $b_1$ | 1.4826 | 0.4386 - 2.3731 |
| $a_{10}$ | 2.1914 | 1.5050 - 2.8794 |
| $a_{20}$ | 0.9016 | 0.0139 - 1.7298 |
| $b_0$ | 1.4543 | 0.1488 - 2.2381 |
| $S_1$ | 0.5006 | 0.3330 - 0.6711 |
| $C_1$ | 0.9243 | 0.7971 - 0.9910 |
| $pvp1$ | 0.9530 | 0.8329 - 0.9964 |
| $pvn1$ | 0.3888 | 0.1524 - 0.6051 |
| $S_2$ | 0.7873 | 0.6771 - 0.9318 |
| $C_2$ | 0.7203 | 0.5079 - 0.8968 |
| $pvp2$ | 0.8929 | 0.7811 - 0.9780 |
| $pvn2$ | 0.5348 | 0.1817 - 0.8709 |

Table 5.5: Posterior medians and 95% posterior probability intervals of the prevalence and test parameters obtained using the random effects model.

| Variable | Median | 95% PI |
|---|---|---|
| $\pi$ | 0.6113 | 0.5037 - 0.7159 |
| $S_1$ | 0.5778 | 0.4960 - 0.6612 |
| $C_1$ | 0.8723 | 0.8710 - 0.9597 |
| $pvp1$ | 0.9130 | 0.8173 - 0.9438 |
| $pvn1$ | 0.6686 | 0.5857 - 0.7478 |
| $S_2$ | 0.8962 | 0.8262 - 0.9472 |
| $C_2$ | 0.7300 | 0.6319 - 0.8206 |
| $pvp2$ | 0.7979 | 0.6967 - 0.8767 |
| $pvn2$ | 0.8570 | 0.7452 - 0.9282 |

Table 5.6: Posterior medians and 95% posterior probability intervals of the prevalence and test parameters obtained using the conditional independence model.

subjects falling into each of the four classifications is presented in Table 5.7. We can see a clear distinction in the distribution of the intensity across the four cells with the subjects falling in the $(T_1 = 1, T_2 = 1)$ having the highest intensity and those falling in the $(T_1 = 0, T_2 = 0)$ having the lowest intensity.

The overlaid trace plots of the posterior distribution of the prevalence obtained from 5 different runs of the Gibbs sampler starting from overdispersed initial values are presented in Figure 5.4. We present only the first 500 observations for the sake of clarity. Clearly all sequences reach convergence fairly quickly. This is confirmed by Gelman and Rubin's statistic for all the parameters which is displayed in Table 5.9. The high value of Raftery and Lewis' $N_{min}$ diagnostic, in Table 5.8, indicates the presence of autocorrelation between successive observations. This is to be expected since our parameters are intimately related. For example, decreasing the sensitivity of a test will result in an increase in the prevalence, since it is presumed that there are more false negative subjects and so on.

| Cross classification | Median | 95% PI |
|---|---|---|
| $T_1 = 1, T_2 = 1$ | 0.6004 | -1.6256 - 2.2733 |
| | 0.6127 | -1.4234 - 2.0749 |
| | 0.5868 | -1.221 - 2.2423 |
| | 0.5582 | -1.4935 - 2.2745 |
| $T_1 = 1, T_2 = 0$ | -0.3990 | -1.8241 - 0.7308 |
| | -0.4630 | -1.7565 - 0.8066 |
| | -0.4109 | -1.7688 - 0.7321 |
| | -0.4481 | -1.6904 - 0.6503 |
| $T_1 = 0, T_2 = 1$ | -0.3265 | -1.5983 - 0.9799 |
| | -0.3350 | -1.6545 - 0.9131 |
| | -0.3792 | -1.6687 - 0.9445 |
| | -0.3340 | -1.66057 - 0.8924 |
| $T_1 = 0, T_2 = 0$ | -0.4424 | -2.225 - 1.8914 |
| | -0.5232 | -2.2791 - 1.7015 |
| | -0.2846 | -2.2674 - 1.8679 |
| | -0.4189 | -2.2854 - 1.7990 |

Table 5.7: Posterior medians and 95% posterior probability intervals of the marginal posterior distributions of a sample of the $i_k$ values.

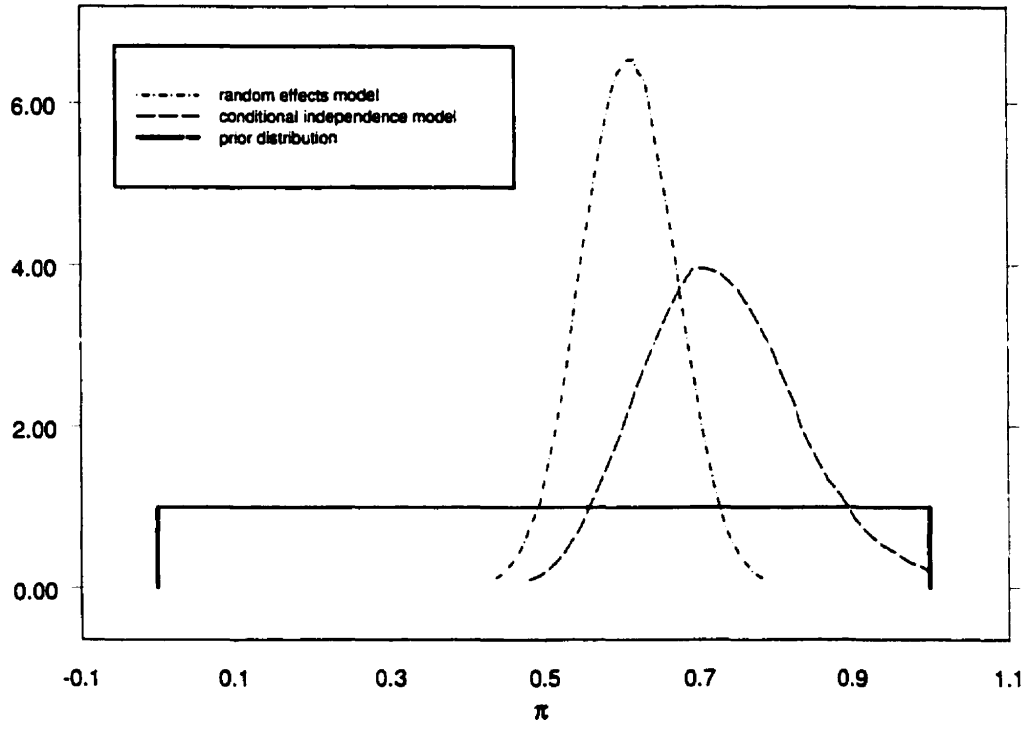


Figure 5.2: Prior distribution of $\pi$ overlaid by posterior distributions obtained using the random effects and conditional independence models.

## 5.4 Summary

In this chapter we have presented a Bayesian method to model the conditional dependence between tests by allowing for individual variability in test performance. This method accounts for the simultaneous dependence between three or more tests. The main advantage of the Bayesian approach is that it provides a solution even in the situation when we have non-identifiable parameters due to there being less than 4 or 5 tests. While no closed-form solution exists, this method can be implemented using a Gibbs sampler. With 4 or more tests, the Bayesian approach can still be useful if there is good prior information on one or more of the parameters which will lead

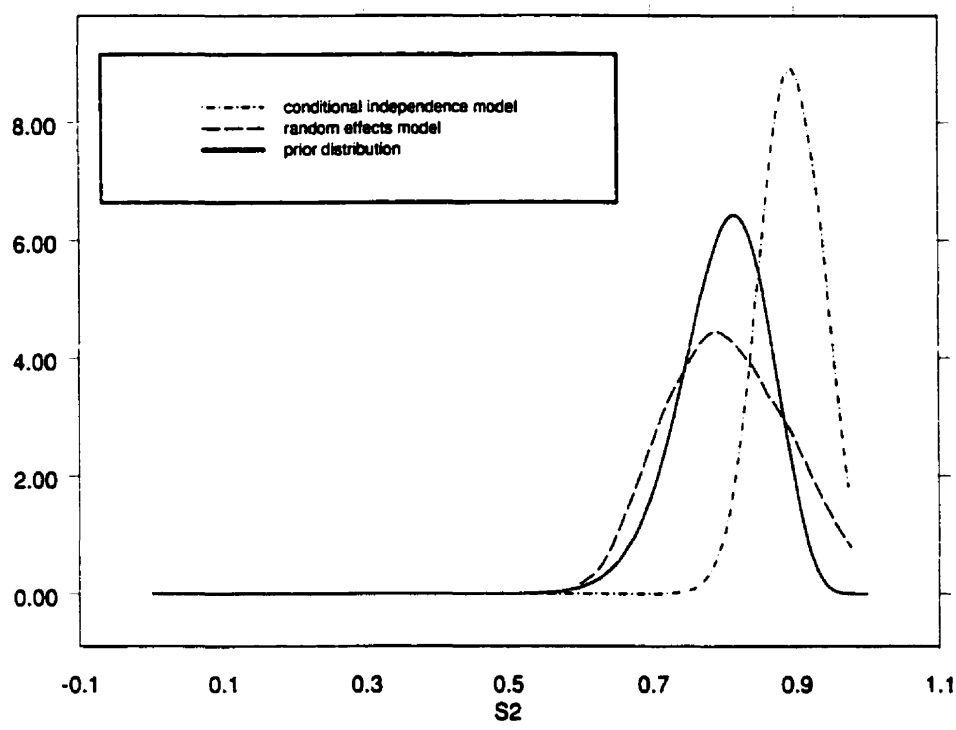


Figure 5.3: Prior distribution of $S_2$ overlaid by posterior distributions obtained using the random effects and conditional independence models.

to narrowed posterior distributions. A minor drawback of this method is that it is computationally expensive, since the Gibbs sampler takes almost 10 times as much time to complete a given number of iterations compared to the fixed effects model.

In the next chapter we compare the performance of both the fixed and random effects models in adjusting for the conditional dependence between two tests using a 'real-life' problem.

| Iterations used for diagnostic = 2500:4999<br>Thinning interval = 1<br>Sample size per chain = 4999 | | |
|---|---|---|
| Variable | Point est. of $\hat{R}$ | 97.5% quantile |
| $\pi$ | 1.01 | 1.02 |
| $a_{11}$ | 1.00 | 1.01 |
| $a_{21}$ | 1.01 | 1.02 |
| $b_1$ | 1.01 | 1.02 |
| $a_{10}$ | 1.00 | 1.02 |
| $a_{20}$ | 1.00 | 1.03 |
| $b_0$ | 1.01 | 1.03 |
| $S_1$ | 1.01 | 1.02 |
| $C_1$ | 1.00 | 1.00 |
| PV+ | 1.01 | 1.02 |
| PV- | 1.01 | 1.02 |
| $S_2$ | 1.00 | 1.00 |
| $C_2$ | 1.00 | 1.00 |
| PV+ | 1.01 | 1.03 |
| PV- | 1.00 | 1.00 |

Table 5.8: Gelman and Rubin 50% and 97.5% shrink factors.

| Iterations used = 1:1999<br>Thinning interval = 1<br>Sample size per chain = 1999 | | | | | |
|---|---|---|---|---|---|
| Quantile = 0.025<br>Accuracy = +/- 0.01<br>Probability = 0.95 | | | | | |
| Variable | Thin (k) | Burn-in (M) | Total (N) | Lower bound (Nmin) | Dependence factor (I) |
| $\pi$ | 1 | 7 | 1959 | 937 | 2.09 |
| $a_{11}$ | 1 | 2 | 930 | 937 | 0.993 |
| $a_{21}$ | 1 | 5 | 1322 | 937 | 1.41 |
| $b_1$ | 1 | 34 | 9251 | 937 | 9.87 |
| $a_{10}$ | 1 | 4 | 1216 | 937 | 1.3 |
| $a_{20}$ | 1 | 8 | 2185 | 937 | 2.33 |
| $b_0$ | 1 | 8 | 2222 | 937 | 2.37 |
| $S_1$ | 1 | 3 | 1011 | 937 | 1.08 |
| $C_1$ | 1 | 8 | 2053 | 937 | 2.19 |
| pvp1 | 1 | 10 | 2497 | 937 | 2.66 |
| pvn1 | 1 | 7 | 1959 | 937 | 6.18 |
| $S_2$ | 1 | 8 | 2053 | 937 | 2.19 |
| $C_2$ | 1 | 7 | 2078 | 937 | 2.22 |
| pvp2 | 2 | 14 | 4394 | 937 | 4.69 |
| pvn2 | 2 | 22 | 5356 | 937 | 5.72 |

Table 5.9: Raftery and Lewis convergence diagnostic.

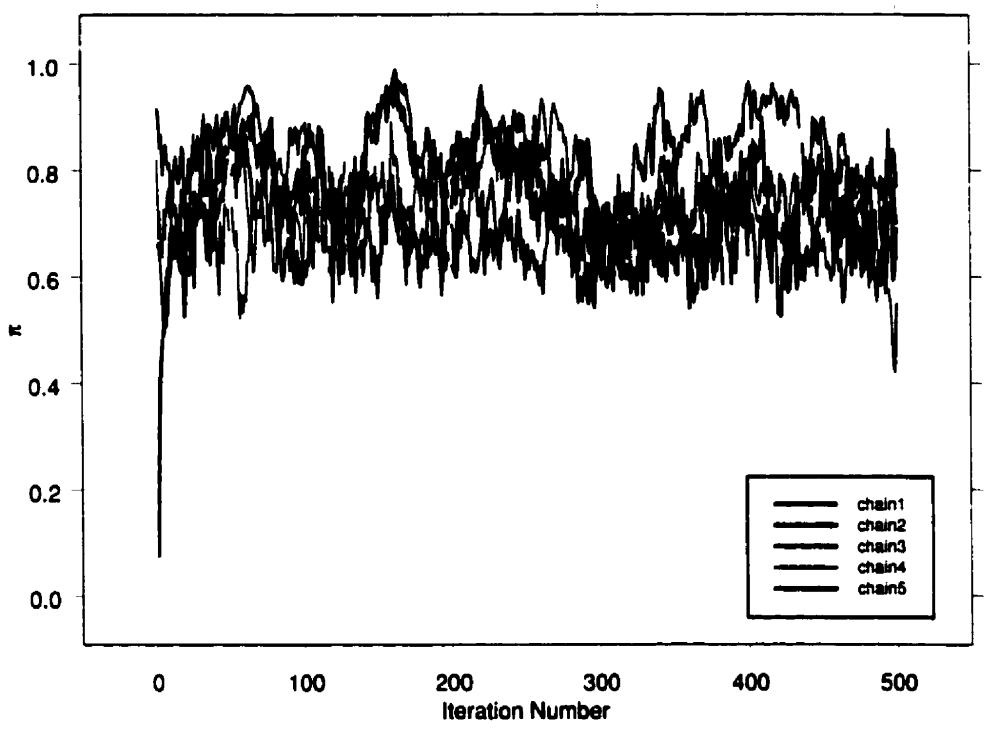


Figure 5.4: Overlaid trace plots of the prevalence from 5 different chains of the Gibbs sampler.

# 6

# THE STRONGYLOIDES INFECTION PROBLEM REVISITED

In this chapter we apply the methods developed in Chapters 4 and 5 to the Strongyloides infection problem which was introduced in Section 3.2.3. This data set was obtained as part of a study conducted among a group of Cambodian refugees in Montréal, Canada. The purpose of this chapter is to illustrate the practical aspects of using the methods described earlier, to account for the conditional dependence between two imperfect tests.

The cross-classification of the results from the stool examination and the serology test are repeated in Table 6.1 for convenience.

| | | Stool Examination | | |
|---|---|---|---|---|
| | | + | - | Total |
| Serology Test | + | 38 | 87 | 125 |
| | − | 2 | 35 | 37 |
| Total | | 40 | 122 | 162 |

Table 6.1: Results of tests for Strongyloides infection among a group of Cambodian refugees.

| | Parameter | 95% PI |
|---|---|---|
| Stool Examination | Sensitivity | 0.05-0.45 |
| | Specificity | 0.90-1.00 |
| Serology Test | Sensitivity | 0.65-0.95 |
| | Specificity | 0.35-1.00 |

Table 6.2: 95% Prior probability intervals for sensitivity and specificity of the stool examination and the serology test.

## 6.1 *Elicitation of the prior distributions*

Both the stool examination and serology test are commonly used diagnostic tools in infectious disease practice. Since a positive result on the stool examination requires that the parasite actually be detected in the stool specimen, this test tends to underestimate the population prevalence. The serology test, on the other hand, is expected to overestimate the prevalence due to cross-reactivity or persistence of reactivity following the parasite cure. The lack of gold standard tests for most parasitic infections, however, means that the parameters for these tests are not known with a high accuracy. In consultation with the faculty from the McGill Centre for Tropical Disease, Joseph et al., 1995, determined equal tailed 95% prior probability intervals for the sensitivities and specificities of the two tests as presented in Table 6.2. These were determined from information documented in previous studies and clinical opinion (Gam et al., 1987, Genta, 1988, Nutman et al., 1987, Genta, 1989, Caroll et al., 1981, Bailey, 1989, Pelletier et al., 1988, Douce et al., 1987).

Since very little was known a priori about the prevalence of Strongyloides infection in a Cambodian population, a diffuse or non-informative prior was used over this parameter.

For this example, we retain the prior distributions used by Joseph et al., 1995,

| | Parameter | $\alpha$ | $\beta$ | Mean* | S.D.** |
|---|---|---|---|---|---|
| Stool Examination | Sensitivity | 4.44 | 13.31 | 0.25 | 0.10 |
| | Specificity | 71.25 | 3.75 | 0.95 | 0.025 |
| Serology Test | Sensitivity | 21.96 | 5.49 | 0.80 | 0.075 |
| | Specificity | 4.1 | 1.76 | 0.70 | 0.1625 |

* Mean $= \frac{\alpha}{\alpha+\beta}$, ** S.D.$=\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}$

Table 6.3: Prior distribution parameters for sensitivities and specificities in the fixed effects model.

and explain in the following two sections how the corresponding prior distribution parameters are elicited for the fixed and the random effects models. In doing so, it is important to note that the parameters for the sensitivity and specificity given in Table 6.2 represent marginal prior information, as the tests are now correlated. In addition, the random effects model allows for subject-to-subject variations in the test properties, depending on the 'intensity'. In this case, the values given in Table 6.2 represent marginal prior information for the **mean** over all subjects in the population. Subject-specific sensitivities and specificities vary about this mean, as discussed below.

### *Elicitation of prior distribution parameters for the fixed effects model*

The parameters for the $Beta(\alpha, \beta)$ prior distributions of the sensitivities and specificities were determined by solving the two equations which match the center of the parameter range to its mean, $\frac{\alpha}{\alpha+\beta}$, and a quarter of the 95% prior probability interval to its standard deviation, $\sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}}$. These two equations determine $\alpha$ and $\beta$ uniquely. The $(\alpha, \beta)$ values for the prior distributions of the sensitivities and specificities are presented in Table 6.3. Since the prior distribution for the prevalence is diffuse the corresponding *Beta* distribution parameters are $(\alpha_\pi = 1, \beta_\pi = 1)$.

As discussed in Chapter 3, in order to obtain a meaningful solution for a non-identifiable problem, we need to have informative prior distributions on at least as many parameters as would need to be constrained in a frequentist approach. For the fixed effects model this means we must have informative distributions on at least

$$(2 \times 2 + 2 \times {}^2C_2 + 1) - (2^2 - 1) = 4 \text{ parameters.}$$

In this particular example, we were able to determine informative prior distributions for the sensitivities and specificities. Since we did not have exact information about the covariance between tests, it was decided to use diffuse generalized beta prior distributions over the two covariance parameters, *i.e.* prior distributions which assign equal weight to all values in the admissible range of the two covariances as follows:

$$\begin{aligned} covp &\sim GenBeta(1,1), \quad 0 \leq covp \leq min(S_1, S_2) - S_1 S_2, \\ \text{and,} \quad covn &\sim GenBeta(1,1), \quad 0 \leq covn \leq min(C_1, C_2) - C_1 C_2. \end{aligned}$$

A *Generalized Beta* distribution is simply a standard *Beta* distribution as discussed in Chapter 4, which has been stretched or compressed and then translated to accommodate a wider or a narrower range than $(0, 1)$.

## *Elicitation of prior distribution parameters for the random effects model*

In the case of the random effects model, we can use the bisection algorithm described in Section 5.3.2 to obtain the mean values of the prior distribution parameters for the sensitivities and specificities. We denote the Stool Examination as Test 1 and the Serology Test as Test 2. Unlike for the fixed effects model, the values for the covariances among the diseased and non-diseased subjects must be specified in order to uniquely determine the $(a_{JD}, b_D)$, $J = 1, 2$, $D = 0, 1$ parameter pairs. The mean sensitivities and specificities were taken to be equal to the middle value of their ranges in the Table 6.2. Using the expression derived in equation (4.8), and the mean values

of the sensitivities, we can estimate the range of values in which the mean covariance among the diseased subjects lies as

$$\begin{aligned} 0 \leq covp &\leq min(S_1, S_2) - S_1 S_2 \\ &= min(0.25, 0.8) - (0.25)(0.8) = 0.05. \end{aligned} \tag{6.1}$$

For purposes of estimating the prior densities of $a_{11}, a_{21}$ and $b_1$, we arbitrarily fixed $covp = 0.025$ since this value lies in the middle of the range in equation (6.1). We discuss later that this choice has little effect on the final prior parameter values. The mean values of the sensitivities from Table 6.2 together with this value of *covp* were used to solve for the mean values of $(a_{11}, a_{21}, b_1)$ as follows

$$S_1 = \Phi\left(\frac{a_{11}}{\sqrt{1 + b_1^2}}\right) = 0.25,$$

$$S_2 = \Phi\left(\frac{a_{21}}{\sqrt{1 + b_1^2}}\right) = 0.8,$$

$$\int_{-\infty}^{\infty} \Phi(a_{11} + b_1 i_k)\, \Phi(a_{21} + b_1 i_k) d\Phi(i_k) - (0.25)(0.8) = 0.025. \tag{6.2}$$

The possible range of values for the mean covariance among the non-diseased subjects, *covn*, was determined using the equation (4.7) and the two mean specificities such that

$$\begin{aligned} 0 \leq covp &\leq min(C_1, C_2) - C_1 C_2, \\ &= min(0.95, 0.70) - (0.95)(0.70) = 0.035. \end{aligned} \tag{6.3}$$

Once again we arbitrarily set $covn = 0.0175$ which is the mid-point of the range in equation (6.3). The mean values of the specificities in Table 6.2 together with this value of *covn* were used to solve the equations involving $(a_{10}, a_{20}, b_0)$ as follows:

$$C_1 = \Phi\left(\frac{a_{10}}{\sqrt{1 + b_0^2}}\right) = 0.95,$$

$$C_2 = \Phi\left(\frac{a_{20}}{\sqrt{1 + b_0^2}}\right) = 0.70,$$

$$\int_{-\infty}^{\infty} \Phi(a_{10} + b_0 i_k)\, \Phi(a_{20} + b_0 i_k) d\Phi(i_k) - (0.95)(0.7) = 0.0175.$$

It must be stressed that the values we have selected for the covariance parameters are by no means unique. In the absence of any information about the covariance between the tests, it seems sensible to use the mid-point of the range as the prior mean, so that the prior distribution can easily cover the feasible range. Another approach may be to first run the fixed effects model and then use the mean values of the posterior distributions for the covariance parameters obtained there.

The solution to the equations in (6.2) is $(a_{11} = -0.856, a_{21} = 1.068, b_1 = 0.782)$. To determine the approximate prior standard deviations for $a_{11}$, $a_{21}$ and $b_1$ we used contour plots of $S_1$ on the $(a_{11}, b_1)$ plane and $S_2$ on the $(a_{21}, b_1)$ plane as illustrated in Figures 6.1 and 6.2, respectively. From Figure 6.1, we can see that as $S_1$ ranges from 0.05 to 0.45 (its 95% prior probability interval), $a_{11}$ ranges approximately from -2.08 to 0.16. The standard deviation of $a_{11}$ was taken to be a quarter of this range namely $sd(a_{11}) = \frac{0.16-(-2.08)}{4} = 0.56$. The range of $b_1$ is less obvious, since the same value of $b_1$ could correspond to the entire range of values of $S_1$. We can deduce from this that the value of $S_1$ is mainly determined by $a_{11}$, while the value of $b_1$ has a greater bearing on the value of the covariance between the tests.

Keeping in mind that we have no prior information on *covp*, and that its 'mean' value was arbitrarily selected, it was thought prudent to use a wide prior distribution for $b_1$, *ie* one with a high standard deviation. Such a prior distribution would assign similar probabilities to a sufficiently large range of values of $b_1$ corresponding to a wide range of values of the covariance. Similar to our comment in our discussion of the prior distributions for the fixed effects model, we should be able to attain reasonable results here, since we have fairly strong priors on the four $a_{jd}$, $j = 1, 2$, $d = 0, 1$ parameters. Of course, if in other applications there is better information on $b_1$, this will further sharpen the posterior inferences.

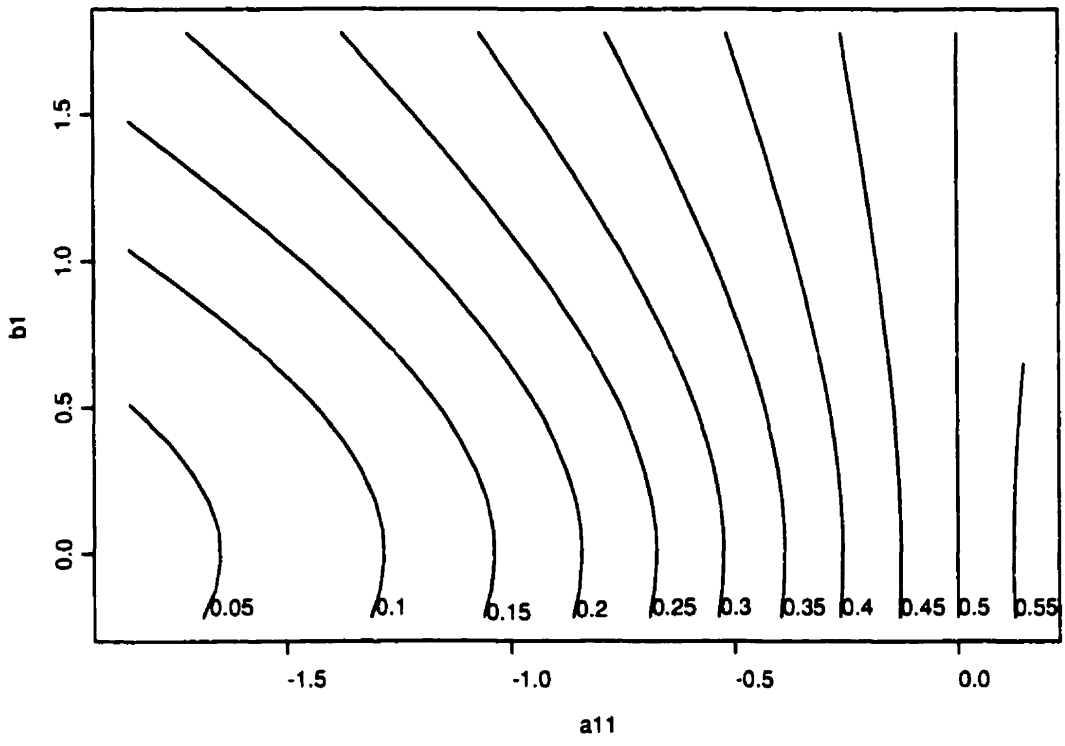


Figure 6.1: Contour plot of $S_1$ on the $(a_{11}, b_1)$ plane.

The range of values of the $a_{21}$ can similarly be deduced from Figure 6.2 to be 0.48 to 1.68. Once again we notice that the value of $S_2$ is mainly determined by $a_{21}$ and not $b_1$. The standard deviations for $(a_{10}, a_{20}, b_0)$ were determined in a similar fashion.

The values of the means and standard deviations of the prior distributions of the $(a_{jd}, b_{jd})$ parameter pairs are summarized in Table 6.4. In order to validate our method of prior elicitation we generated a random sample of 1000 observations from the prior distributions of the $(a_{jd}, b_{jd})$ pairs using the parameter values in Table 6.4 and calculated the mean sensitivities and specificities. The medians and 95% probability intervals of these samples, which are presented in Table 6.5, were found

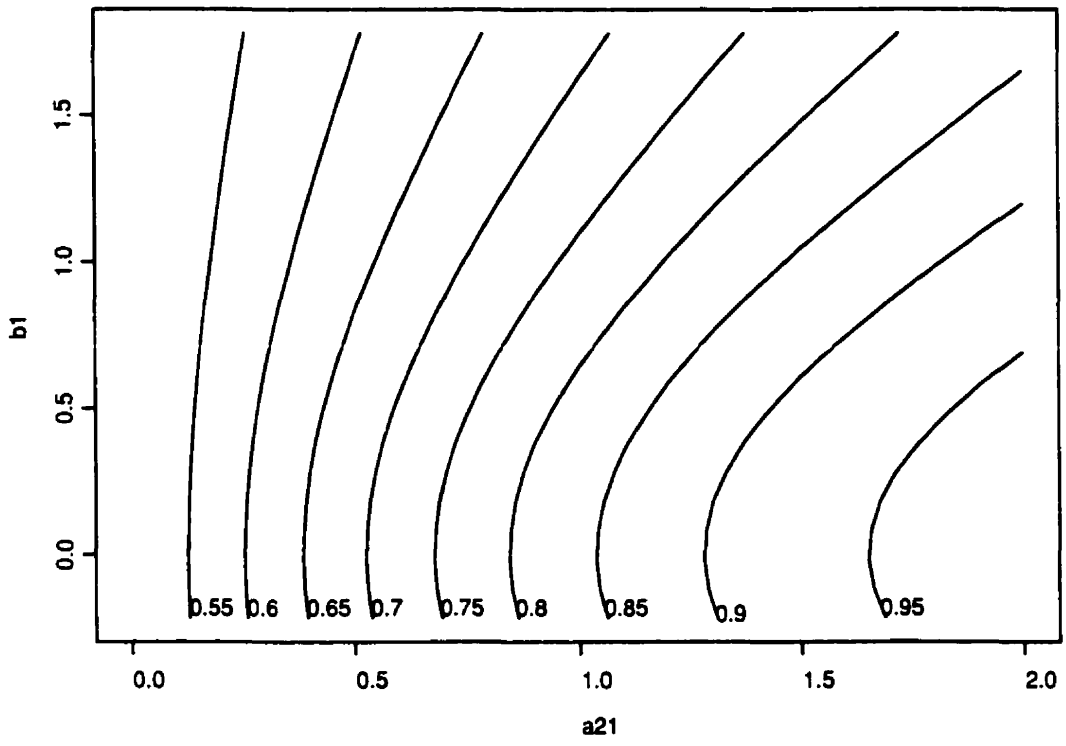


Figure 6.2: Contour plot of $S_2$ on the $(a_{21}, b_1)$ plane.

to be very close to the desired values in Table 6.2. Therefore, we conclude that our somewhat ad hoc method of determining the prior parameters for this problem has worked well.

## 6.2 Results

The results obtained by Joseph et al., 1995, when applying the Bayesian conditional independence model to the Strongyloides infection problem are repeated Table 6.6 to facilitate comparison with the results obtained from the fixed and random effects models.

| | Mean | S.D.* |
|---|---|---|
| $a_{11}$ | -0.856 | 0.562 |
| $a_{21}$ | 1.069 | 0.302 |
| $b_1$ | 0.782 | 1.000 |

| | Mean | S.D.* |
|---|---|---|
| $a_{10}$ | 3.108 | 0.427 |
| $a_{20}$ | 0.983 | 0.693 |
| $b_0$ | 1.584 | 1.000 |

* S.D. = Standard Deviation

Table 6.4: Prior mean and standard deviation for parameters determining sensitivity and specificity in the random effects model.

| | Parameter | Median | 95% PI |
|---|---|---|---|
| Stool Examination | Sensitivity | 0.252 | 0.046-0.493 |
| | Specificity | 0.947 | 0.802-0.995 |
| Serology Test | Sensitivity | 0.766 | 0.602-0.920 |
| | Specificity | 0.684 | 0.417-0.974 |

Table 6.5: Prior medians and 95% prior probability intervals for sensitivity and specificity of the stool examination and the serology test calculated using estimated $(a_{jd}, b_d)$ parameters.

| | | Median | 95% PI |
|---|---|---|---|
| | Prevalence | 0.76 | 0.52-0.91 |
| Stool Examination | Sensitivity | 0.31 | 0.22-0.44 |
| | Specificity | 0.96 | 0.91-0.99 |
| | PV+ | 0.98 | 0.88-1.00 |
| | PV- | 0.30 | 0.11-0.63 |
| Serology Test | Sensitivity | 0.89 | 0.80-0.95 |
| | Specificity | 0.67 | 0.36-0.95 |
| | PV+ | 0.90 | 0.62-1.00 |
| | PV- | 0.70 | 0.28-0.92 |

Table 6.6: Posterior medians and 95% posterior probability intervals of the prevalence and test parameters obtained using the conditional independence model.

### *Results from the fixed effects model*

Using the prior distributions determined in the previous section, we implemented the Gibbs sampler for the Bayesian fixed effects model described in Section 4.2.3. The posterior medians and 95% posterior probability intervals for the prevalence and test parameters thus obtained are presented in Table 6.7. As was noticed in the example using simulated data in Chapter 4, the median prevalence obtained using the fixed effects model is greater than that obtained using the conditional independence model. The 95% posterior probability interval, however, is not very different and does not give any clear indication of a shift in the value of the prevalence due to accounting for the dependence between tests.

To determine the degree to which the two posterior densities differ, we calculated the probability $P(\pi_{BCI} < \pi_{FE})$ where $BCI$ refers to the Bayesian Conditional Independence model, and $FE$ refers to the Fixed Effects model. This was done by sampling with replacement 10,000 pairs of values, one from each of the two posterior

| | Variable | Median | 95% PI |
|---|---|---|---|
| | $\pi$ | 0.8549 | 0.5461 - 0.9903 |
| | *covp* | 0.0276 | 0.0063 - 0.0531 |
| | *covn* | 0.01878 | 0.0026 - 0.0575 |
| Stool Examination | Sensitivity | 0.2749 | 0.1993 - 0.3914 |
| | Specificity | 0.9353 | 0.8640 - 0.9785 |
| | PV+ | 0.9824 | 0.8189 - 0.9979 |
| | PV- | 0.1801 | 0.0123 - 0.5460 |
| Serology Test | Sensitivity | 0.8305 | 0.7391 - 0.9247 |
| | Specificity | 0.6776 | 0.3013 - 0.9369 |
| | PV+ | 0.9479 | 0.6243 - 0.9979 |
| | PV- | 0.4073 | 0.0302 - 0.7857 |

Table 6.7: Posterior medians and 95% posterior probability intervals of the marginal posterior distributions of the prevalence and test parameters obtained using the fixed effects model.

distributions of $\pi$, and then calculating the proportion of times $\pi_{BCI}$ was less than $\pi_{FE}$. It was estimated that $P(\pi_{BCI} < \pi_{FE})$=0.715, indicating that the prevalence was more likely to be greater when the conditional dependence between the tests was taken into account than when it was ignored. If the two distributions were identical, we would have $P(\pi_{BCI} < \pi_{FE})$=0.5, while if the distributions were non-overlapping we would have $P(\pi_{BCI} < \pi_{FE})$=0 or 1, indicating certainty of a difference. Our result is intermediate to these extremes. The shift in the prevalence when accounting for conditional dependence is clearly seen in Figure 6.3, which is a plot of the posterior densities obtained from the three different methods.

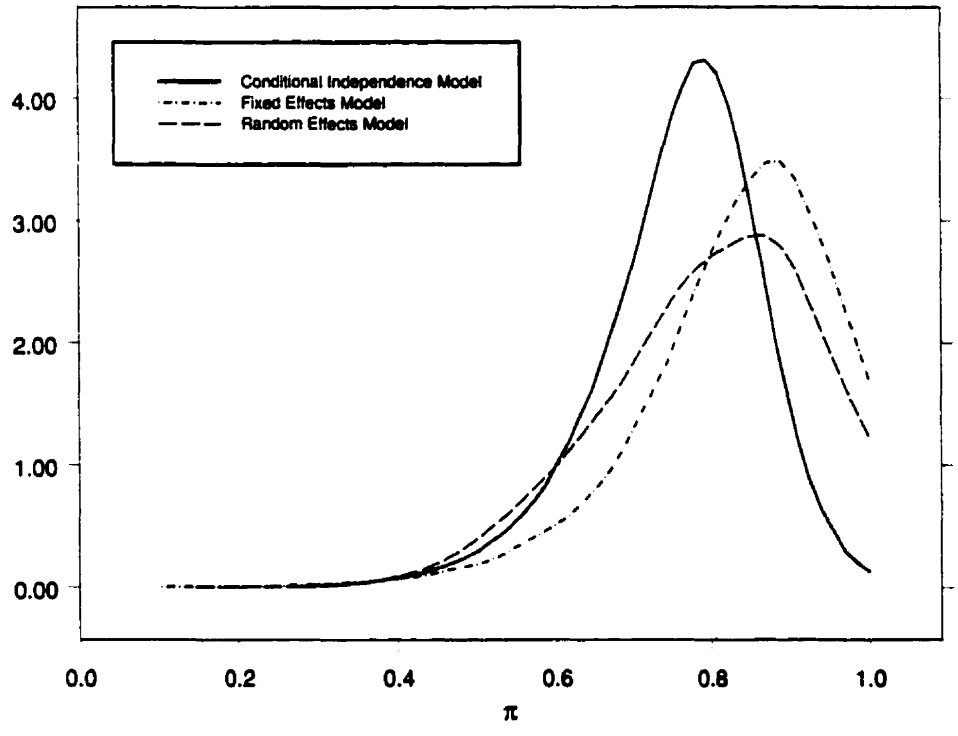


Figure 6.3: Posterior distributions of the prevalence obtained using the three models.

The posterior medians of the sensitivities and specificities are lower than those obtained using the conditional independence model. The 95% posterior probability intervals of these parameters are, however, wider due to the addition of the covariance parameters, and due to the non-informative prior distributions over them. Our overall conclusion, then, is that if we take into account the possibility of correlation, this may have a substantive effect on the posterior estimates of the prevalence and test properties. Unless something is known a priori about the degree of correlation, however, we also add 'noise' to the modeling process, especially in the non-identifiable case of two tests.

The diagnostic parameters obtained from the methods due to Gelman and Rubin, 1992, and Raftery and Lewis, 1992 are presented at the end of the chapter in Tables 6.12 and 6.13, respectively. The value of the $\hat{R}$ statistic remained close to 1 for all parameters when comparing five different runs with over-dispersed starting points, indicating that the Gibbs sampler had likely converged. The same inference can be drawn from the low value of the burn-in iterations suggested by the method of Raftery and Lewis. However, the high value of the dependency factor, $I$, is indicative of high-autocorrelation between successive values sampled. Therefore, we ran a large number of iterations (20,000), in order to obtain accurate inferences from the Gibbs sampler. Since the run-time was only about 2 minutes, it was not thought necessary to seek a reparameterization or other method to reduce the autocorrelations.

### *Results from the random effects model*

The posterior medians and 95% posterior probability intervals obtained using the random effects model are presented in Table 6.8. The median prevalence is greater than that obtained with the conditional independence model, though not as great as that obtained with the fixed effects model. However, the 95% posterior probability interval is similar to that obtained with the other two models. We found that

| | Variable | Median | 95% PI |
|---|---|---|---|
| | $\pi$ | 0.8215 | 0.5276 - 0.9869 |
| | $a_{11}$ | 3.0407 | 1.9849 - 4.1331 |
| | $a_{21}$ | 0.2658 | -1.0566 - 2.1746 |
| | $b_1$ | 0.9530 | -0.2378 - 2.9828 |
| | $a_{10}$ | -0.8456 | -2.1389 - 0.4655 |
| | $a_{20}$ | 1.3896 | 0.8457 - 2.0987 |
| | $b_0$ | 1.3018 | 0.5660 - 2.2795 |
| Stool Examination | Sensitivity | 0.2761 | 0.0761 - 0.6132 |
| | Specificity | 0.9810 | 0.8148 - 0.9997 |
| | PV+ | 0.9886 | 0.8405 - 0.9998 |
| | PV- | 0.2367 | 0.0172 - 0.5935 |
| Serology Test | Sensitivity | 0.8038 | 0.6632 - 0.9203 |
| | Specificity | 0.6622 | 0.1830 - 0.9230 |
| | PV+ | 0.9093 | 0.5194 - 0.9972 |
| | PV- | 0.3553 | 0.0340 - 0.7466 |

Table 6.8: Posterior medians and 95% posterior probability intervals of the prevalence and test parameters obtained using the random effects model.

$P(\pi_{BCI} < \pi_{RE}) = 0.634$, where $RE$ denotes the Random Effects model. This indicates that the prevalence estimate obtained using the random effects model is greater than that obtained using the conditional independence model, in the sense that the posterior density is somewhat shifted to the right, as seen in Figure 6.3.

The median sensitivities were somewhat lower than those obtained using the conditional independence model, while the median specificities remained about the same. Their 95% posterior probability intervals are even wider than those obtained with the fixed effects model. This is due to the additional uncertainty added on by the latent

'intensity' variable.

The values of the Gelman and Rubin, and Raftery and Lewis diagnostic statistics revealed that while the Gibbs' sampler converges fairly quickly, there was again a high degree of autocorrelation between successive observations. These results are presented in Tables 6.14 and 6.15 at the end of the chapter.

### *Results using priors with smaller variance*

A plausible reason why we do not see a more substantial change from the results obtained with the conditional independence model is because of the lack of strong prior information on the covariance parameters. In fact, the prior distributions for the sensitivities and specificities are also very wide, which further compounds the problem. In order to demonstrate that adjusting for the dependence between tests, when it exists, could create an important difference in the results, we reduced the variability in the parameters by halving the standard deviations of the informative prior distributions. The results for the conditional independence model, the fixed effects model and the random effects models are presented in Tables 6.9, 6.10 and 6.11, respectively.

We see that the median prevalence obtained from the fixed and random effects models is greater than that obtained when assuming conditional independence. The 95% probability intervals are now tighter making it possible to distinguish the two situations when the conditional dependence is taken into account and when it is ignored. This result is illustrated in Figure 6.4.

| | | Median | 95% PI |
|---|---|---|---|
| | Prevalence | 0.8294 | 0.6636-0.9646 |
| Stool Examination | Sensitivity | 0.2858 | 0.2120-0.3769 |
| | Specificity | 0.9529 | 0.9255-0.9727 |
| | PV+ | 0.7368 | 0.6112-0.8406 |
| | PV- | 0.7428 | 0.6763-0.8103 |
| Serology Test | Sensitivity | 0.8352 | 0.7698-0.8889 |
| | Specificity | 0.6938 | 0.5162-0.8399 |
| | PV+ | 0.9232 | 0.8321-0.9676 |
| | PV- | 0.4974 | 0.2928-0.6678 |

Table 6.9: Posterior medians and 95% posterior probability intervals of the prevalence and test parameters obtained using the using the conditional independence model when prior distributions have a reduced variance.

## 6.3 A brief note on model selection

Given that we have described three competing models, each of which provide somewhat different inferences, it is natural to ask which model is best supported by the data. If necessary, this can be accomplished using Bayes Factors (Kass and Raftery, 1995). Given data $D$ and two models, say $M_1$ and $M_2$, the Bayes Factor of Model 2 compared to Model 1 is defined as

$$B_{21} = \frac{p(D|M_2)}{p(D|M_1)}. \tag{6.4}$$

The Bayes Factor, therefore, provides the ratio of the probability of observing the data under Model 2 compared to the probability of obtaining the data under Model 1. Intuitively, if the data are more likely under Model 2, then $B_{21} > 1$, and $M_2$ is preferred. If $B_{21} < 1$ then $M_1$ is preferred, and if $B_{21} = 1$ then the data do not

| | Variable | Median | 95% PI |
|---|---|---|---|
| | $\pi$ | 0.8974 | 0.7211 - 0.9929 |
| | *covp* | 0.0331 | 0.0081 - 0.0578 |
| | *covn* | 0.0187 | 0.0009 - 0.0579 |
| Stool Examination | Sensitivity | 0.2690 | 0.1948 - 0.3547 |
| | Specificity | 0.9433 | 0.8884 - 0.9680 |
| | PV+ | 0.9767 | 0.9130 - 0.9984 |
| | PV- | 0.1280 | 0.0089 - 0.3457 |
| Serology Test | Sensitivity | 0.8089 | 0.7431 - 0.8701 |
| | Specificity | 0.6966 | 0.5137 - 0.8437 |
| | PV+ | 0.9614 | 0.8393 - 0.9975 |
| | PV- | 0.2943 | 0.0221 - 0.6184 |

Table 6.10: Posterior medians and 95% posterior probability intervals of the prevalence and test parameters obtained using the fixed effects model when prior distributions have a reduced variance.

| | Variable | Median | 95% PI |
|---|---|---|---|
| | $\pi$ | 0.9029 | 0.7427 - 0.9916 |
| | $a_{11}$ | 3.087 | 2.3549 - 3.8333 |
| | $a_{21}$ | 0.8814 | -0.1262 - 1.9143 |
| | $b_1$ | 1.5218 | 0.5917 - 2.5301 |
| | $a_{10}$ | -0.8432 | -1.7827 - 0.0682 |
| | $a_{20}$ | 1.2448 | 0.8609 - 1.6952 |
| | $b_0$ | 0.9781 | 0.4748 - 1.5239 |
| Stool Examination | Sensitivity | 0.2772 | 0.0928 - 0.5212 |
| | Specificity | 0.9546 | 0.8490 - 0.9972 |
| | PV+ | 0.9850 | 0.8813 - 0.9995 |
| | PV- | 0.1241 | 0.0111 - 0.3296 |
| Serology Test | Sensitivity | 0.8102 | 0.7291 - 0.8982 |
| | Specificity | 0.6832 | 0.4729 - 0.8824 |
| | PV+ | 0.9632 | 0.8564 - 0.9974 |
| | PV- | 0.2749 | 0.0204 - 0.6544 |

Table 6.11: Posterior medians and 95% probability intervals of the prevalence and test parameters obtained using the random effects model when prior distributions have a reduced variance.

distinguish between these models.

In order the calculate Bayes Factors, one needs to calculate terms of the form $p(D|M)$. In general, suppose that the model $M$ contains the vector of unknown parameters $\theta$. We can write

$$p(D|M) = \int p(D|\theta, M)p(\theta|M)d\theta. \tag{6.5}$$

In other words, $p(D|M)$ can usually be expressed as the integral of the product

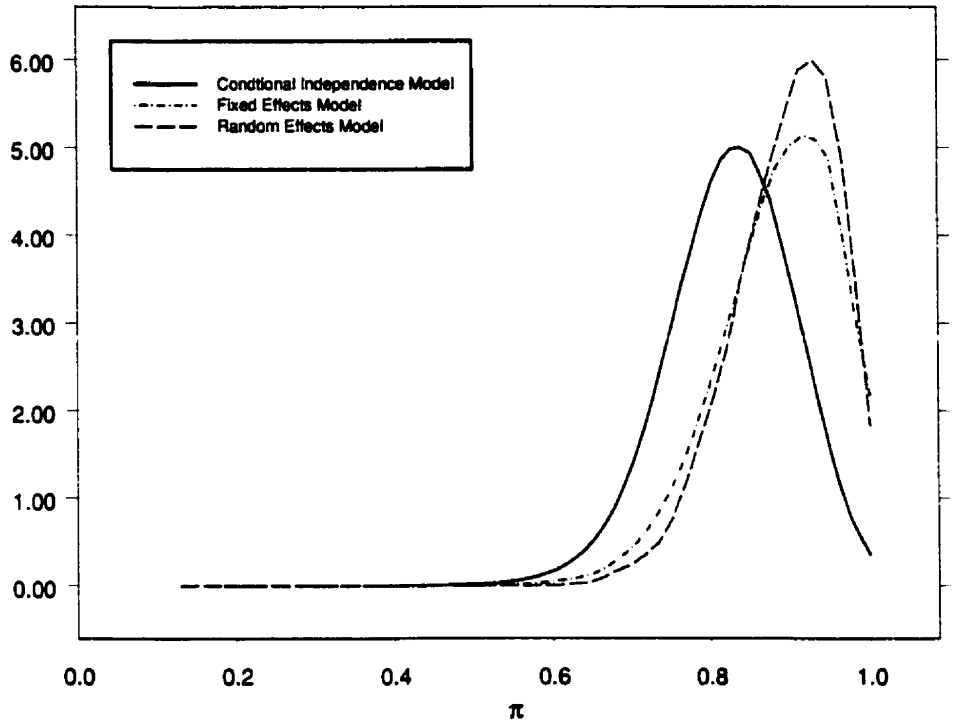


Figure 6.4: Posterior distributions of the prevalence obtained using the three models when prior distributions have a reduced variance.

of the prior distribution of $\theta$ times the likelihood function of the data given $\theta$, with respect to $\theta$.

In our problem, however, equation (6.5) appears intractable even for the simplest case of the conditional independence model, so that we cannot directly apply equation (6.4) to calculate the desired Bayes Factors. Several authors (for example, see Chib, 1995) have discussed approximating Bayes Factors from the output of a Gibbs sampler. The idea is as follows:

For any given model, from Bayes Theorem, we can write:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{m(D)}, \tag{6.6}$$

where $m(D) = \int p(D|\theta)p(\theta)d\theta$ is the marginal distribution of the data D. Rearranging terms, we can write (6.6) as

$$m(D) = \frac{p(D|\theta)p(\theta)}{p(\theta|D)}. \tag{6.7}$$

Since (6.7) must hold for all $\theta$, this equation shows that one can estimate $m(D)$ provided that each of $p(D|\theta)$, $p(\theta)$, and $p(\theta|D)$ can be estimated for at least one value of $\theta$. Often a good choice is the posterior mean of $\theta$, since it is usually a point of high density which induces stability into the estimator, and since it is easily estimated from the Gibbs sampler output.

While this is sufficient for many problems, a further complication that applies to the models discussed in this thesis is that while the likelihood and prior densities are 'fully available' in the sense that the normalizing constants are known, this is not the case for the posterior density. Chib, 1995 discusses how to estimate the normalizing constant in such situations by using a 'Rao-Blackwell' (see Gelfand and Smith, 1990) mixture estimate of the posterior density. Here one takes the full conditional distribution from which each unknown parameter was sampled during the running of the Gibbs sampler, and takes an average of the result over all iterations. This additional step would allow $m(D)$ to be calculated for the conditional independence model, where all full conditional distributions are fully specified, including normalizing constants. However, both models developed in this thesis for correlated had full conditional distributions with unknown normalizing constants, wherein the SIR algorithm was employed, so that the methods of Chib, 1995 could not be employed here.

Other methods have appeared for calculating Bayes Factors from the output of a Gibbs sampler. For example, Carlin and Polson, 1991 suggested running all competing

models simultaneously, and adding a parameter as a model indicator. This method, however, requires a carefully selected tuning parameter which essentially balances the probabilities that each model is selected for the next iteration, so that proper mixing of the Gibbs sampler occurs. Newton and Raftery, 1994 showed that the marginal density could be estimated by a harmonic mean involving only the likelihood function given the Gibbs sampler output, but this estimate has been criticized as being unstable and therefore unreliable as input into a Bayes Factor equation.

Given the above considerations, we decided that it was not worthwhile here to compare models via Bayes Factors, as the posterior inferences were not sufficiently dissimilar to warrant the considerable effort required, and since the only applicable method is known to be unreliable. Further research is clearly required in this area.

## 6.4 Summary

In this chapter we have seen that adjusting for the conditional dependence between diagnostic tests can result in substantial changes to the posterior densities of the prevalence and test parameters. However, the magnitudes of these changes are dependent on informative content of the prior distributions used, and on the data themselves, particularly since we have only two tests.

| Iterations used for diagnostic = 2500:4999<br>Thinning interval = 1<br>Sample size per chain = 4999 | | | |
|---|---|---|---|
| | Variable | Point est. of $\hat{R}$ | 97.5% quantile |
| | $\pi$ | 1.01 | 1.02 |
| | *covp* | 1.00 | 1.00 |
| | *covn* | 1.00 | 1.00 |
| Stool Examination | Sensitivity | 1.01 | 1.02 |
| | Specificity | 1.00 | 1.00 |
| | PV+ | 1.01 | 1.02 |
| | PV- | 1.01 | 1.02 |
| Serology Test | Sensitivity | 1.00 | 1.00 |
| | Specificity | 1.00 | 1.00 |
| | PV+ | 1.01 | 1.03 |
| | PV- | 1.00 | 1.00 |

Table 6.12: Gelman and Rubin 50% and 97.5% shrink factors for the fixed effects model.

| Iterations used = 1:4999<br>Thinning interval = 1<br>Sample size per chain = 4999 | | | | | | |
|---|---|---|---|---|---|---|
| Quantile = 0.025<br>Accuracy = +/- 0.005<br>Probability = 0.95 | | | | | | |
| | Variable | Thin (k) | Burn-in (M) | Total (N) | Lower bound (Nmin) | Dependence factor (I) |
| | $\pi$ | 2 | 32 | 33192 | 3746 | 8.86 |
| | *covp* | 1 | 3 | 4448 | 3746 | 1.19 |
| | *covn* | 1 | 4 | 4636 | 3746 | 1.24 |
| Stool | Sensitivity | 2 | 10 | 10670 | 3746 | 2.85 |
| Examination | Specificity | 1 | 4 | 5165 | 3746 | 1.38 |
| | PV+ | 2 | 20 | 21728 | 3746 | 5.8 |
| | PV- | 2 | 16 | 17508 | 3746 | 4.67 |
| Serology | Sensitivity | 1 | 5 | 6020 | 3746 | 1.61 |
| Test | Specificity | 2 | 10 | 10670 | 3746 | 2.85 |
| | PV+ | 1 | 26 | 26888 | 3746 | 7.18 |
| | PV- | 1 | 12 | 12871 | 3746 | 3.44 |

Table 6.13: Raftery and Lewis convergence diagnostic for the fixed effects model.

| Iterations used for diagnostic = 2500:4999<br>Thinning interval = 1<br>Sample size per chain = 4999 | | | |
|---|---|---|---|
| | Variable | Point est. of $\hat{R}$ | 97.5% quantile |
| | $\pi$ | 1.01 | 1.02 |
| | $a_{11}$ | 1.00 | 1.01 |
| | $a_{21}$ | 1.01 | 1.02 |
| | $b_1$ | 1.01 | 1.02 |
| | $a_{10}$ | 1.00 | 1.02 |
| | $a_{20}$ | 1.00 | 1.03 |
| | $b_0$ | 1.01 | 1.03 |
| Stool Examination | Sensitivity | 1.01 | 1.02 |
| | Specificity | 1.00 | 1.00 |
| | PV+ | 1.01 | 1.02 |
| | PV- | 1.01 | 1.02 |
| Serology Test | Sensitivity | 1.00 | 1.00 |
| | Specificity | 1.00 | 1.00 |
| | PV+ | 1.01 | 1.03 |
| | PV- | 1.00 | 1.00 |

Table 6.14: Gelman and Rubin 50% and 97.5% shrink factors for the random effects model.

| Iterations used = 1:4999<br>Thinning interval = 1<br>Sample size per chain = 4999 | | | | | | |
|---|---|---|---|---|---|---|
| Quantile = 0.025<br>Accuracy = +/- 0.005<br>Probability = 0.95 | | | | | | |
| | Variable | Thin (k) | Burn-in (M) | Total (N) | Lower bound (Nmin) | Dependence factor (I) |
| | $\pi$ | 2 | 32 | 33192 | 3746 | 8.86 |
| | $a_{11}$ | 1 | 2 | 3803 | 3746 | 1.02 |
| | $a_{21}$ | 1 | 6 | 6683 | 3746 | 1.77 |
| | $b_1$ | 1 | 10 | 10383 | 3746 | 2.77 |
| | $a_{10}$ | 1 | 4 | 5165 | 3746 | 1.38 |
| | $a_{20}$ | 2 | 8 | 9730 | 3746 | 2.6 |
| | $b_0$ | 1 | 7 | 5165 | 3746 | 1.38 |
| Stool | Sensitivity | 2 | 10 | 10670 | 3746 | 2.85 |
| Examination | Specificity | 1 | 4 | 5165 | 3746 | 1.38 |
| | PV+ | 2 | 20 | 21728 | 3746 | 5.8 |
| | PV- | 2 | 16 | 17508 | 3746 | 4.67 |
| Serology | Sensitivity | 1 | 5 | 6020 | 3746 | 1.61 |
| Test | Specificity | 2 | 10 | 10670 | 3746 | 2.85 |
| | PV+ | 1 | 26 | 26888 | 3746 | 7.18 |
| | PV- | 1 | 12 | 12871 | 3746 | 3.44 |

**Table 6.15: Raftery and Lewis convergence diagnostic for the random effects model.**

# 7

# DISCUSSION

Development of methods for the analysis of results from diagnostic tests is a very active area of biostatistical research. This is not surprising given the bearing the inferences drawn from these methods have on medical decision making. Accurate estimates of the prevalence and test parameters help to improve the organization of health care at both clinical and public health levels.

In this thesis we have addressed the issue of statistical analysis when tests are conditionally dependent. Although there is an enormous literature on statistical methods for diagnostic test data, there are no frequentist solutions that directly address the problem of estimating parameters in the presence of three or less correlated tests. To our knowledge there is also no literature discussing a Bayesian solution to this problem, even for identifiable cases. This thesis has addressed this gap in the literature. The problem with less than 3 tests is non-identifiable since in the absence of a gold standard test we need at least 4 tests to obtain a direct solution using a frequentist approach. Due to time and cost constraints, for example, there are often occasions when we have to make do with the results from less than four tests, and hence there is a need for methods which provide the best possible estimates of the parameters of interest in such situations. Even though the successful application of the methods presented here depend to a very large extent on the prior distributions, this solution is preferable to no solution at all, especially given the frequency with which the problem occurs.

We have demonstrated how a Bayesian approach can be used when we have a non-identifiable problem, since it can provide simultaneous estimates of all important parameters without having to impose unrealistic constraints on the unknown parameters. The Bayesian approach also allows us to utilize valuable prior information to draw inferences from the data at hand. To carry out the Bayesian analyses, we have used computational methods such as the Gibbs sampler and the Sampling-Importance Resampling (SIR) algorithms.

We have discussed two methods for evaluating the possible effects that correlation between tests may have on their results. The first method postulates that the conditional dependence between two tests has a fixed effect on their joint probability. This can be modeled by way of the covariance between the tests among the diseased and non-diseased populations. In Chapter 4 we showed that the two covariance parameters can be expressed in terms of the sensitivities and specificities, thus making this model easy to interpret. In a simulated example we found that this method gives more accurate estimates of the posterior prevalence and test parameters, than would be obtained by assuming conditional independence between the tests.

The second method we propose allows for variation in the performance of individual subjects on each diagnostic test. This variation is incorporated by way of a latent 'intensity' variable which is independent of the disease status and could be taken to mean 'severity of disease' or 'ease of detection'. The 'intensity' which follows a $N(0,1)$ distribution is modeled as a random effect which induces correlation between the tests. The results of applying this model to a simulated data set showed that it also provided more accurate estimates of the prevalence and test parameters than when ignoring the dependence between tests.

In Chapter 6, we applied both methods to the Strongyloides infection data set used earlier in the paper by Joseph et al., 1995. Here we found that in the absence of strong prior distributions, the improvement in the results obtained by adjusting for

conditional dependence, though evident, is not substantial. Unless good information is available about the degree of correlation between tests or about the test properties themselves, there will always be some uncertainty about whether correlation exists, and about how important it is to correct for possible correlations. This problem is especially acute when there is no gold standard, and when the number of available tests is less than 4, which is usually the case. When results from more than 4 tests are available, one may apply the recently developed method of Qu et al., 1996. Even then the Bayesian methods presented here may be useful in allowing for more accurate estimation via prior information. Nevertheless, when there are less than 4 tests, checking for the effect of correlation is important.

Although in all the examples we have used there was no information about the prevalence and therefore we used a diffuse prior distribution over this parameter, the methods developed here can also be used in other situations. For example, when evaluating the accuracy of a new test by comparing it to a gold standard test whose properties are well known, we could use diffuse distributions over its sensitivity and specificity. We could also use these methods in the situation when we have informative prior distributions over the prevalence and sensitivities and specificities of the two tests and we are interested to know about the covariance between the two tests among diseased and non-diseased subjects.

While concluding this thesis we feel that the following will be important for future research in this area:

1. The methodology developed here may be viewed as a 'mapping' from a given set of prior distributions to the corresponding set of posterior distributions. Therefore, the posterior density can always be interpreted as a coherent updating of the prior distribution upon seeing the data, but any extrapolation to the 'truth' involves a leap of faith. Thus the accurate elicitation of prior distributions is

very important. Though there is much literature on the general problem of elicitation, elicitation of prior distributions for the diagnostic testing problem remains to be addressed. Not all tests perform uniformly well across different populations, and this is difficult to quantify.

2. It would be of interest to compare performance of the models developed here with other models which can be used for adjusting for the conditional dependence between tests, such as a logistic regression model or the ordinal regression models used for the analysis of parametric ROC curves. These models bring their own problems in non-identifiable situations. For example, in a logistic regression the binary outcome of 'disease' or 'free of disease' is latent, so that one would need a good method for eliciting prior information on the regression parameters relating test results to disease status.

3. In both the models developed here we assumed that the observed data were collected from a random sample of the population. Another area worthy of interest would be the extension of these models to the situation when test results are obtained from a non-random population, such as might be observed in a clinic-based study. In such a situation we would need to adjust for the 'work-up' bias that might occur, when the prevalence and test parameters are estimated based only on the results of the subjects studied.

Although much remains to be done, the work presented here shows that the diagnostic testing problem for correlated data is manageable from a Bayesian point-of-view even under non-identifiability, modulo a careful treatment of the prior information.

# APPENDIX A

## A.1 The Albert and Chib method

In this section we describe one of the methods presented in the paper titled 'Bayesian Analysis of Binary and Polychotomous Data' by Albert and Chib, 1993, in which the authors propose Bayesian approaches to modeling categorical response data using a Gibbs sampler.

Let $Y_1, \ldots, Y_N$ be $N$ independent, observed Bernoulli variables with probability of success $p_k = P(Y_k = 1)$, $k = 1, \ldots, N$. The probability $p_k$ is a function of $n$ known covariates such that $p_k = H(\mathbf{x}_k^T\beta)$. The vector of covariates is denoted by $\mathbf{x}_k = (x_{k1}, \ldots, x_{kn})$ and $\beta$ is an $n \times 1$ vector of unknown parameters. $H$ is a known cumulative distribution function. In the case when $H$ is the normal distribution we have the probit model, $p_k = \Phi(\mathbf{x}_k^T\beta)$. The likelihood function of the observed data is then given by

$$L = \prod_{k=1}^{N} \Phi(\mathbf{x}_k^T\beta)^{y_k}(1 - \Phi(\mathbf{x}_k^T\beta))^{(1-y_k)}.$$

Let $\pi(\beta)$ be the joint prior density of the parameters of interest, $\beta$. Then the posterior distribution of $\beta$ would be given by

$$\pi(\beta|y_1, \ldots, y_N) = \frac{\pi(\beta)\prod_{i=1}^{N} \Phi(\mathbf{x}_k^T\beta)^{y_k}(1 - \Phi(\mathbf{x}_k^T\beta))^{(1-y_k)}}{\int \pi(\beta)\prod_{i=1}^{N} \Phi(\mathbf{x}_k^T\beta)^{y_k}(1 - \Phi(\mathbf{x}_k^T\beta))^{(1-y_k)}d\beta}. \qquad \text{(A.1)}$$

In the particular case when $\pi(\beta)$ is the normal distribution $N_k(\tilde{\beta}, \Sigma)$, (A.1) reduces

to

$$\pi(\beta|y_1, \ldots, y_N) = \frac{exp(-\frac{(\beta-\tilde{\beta})^T\Sigma^{-1}(\beta-\tilde{\beta})}{2}) \prod_{i=1}^{N} \Phi(\mathbf{x}_k^T\beta)^{y_k}(1-\Phi(\mathbf{x}_k^T\beta))^{(1-y_k)}}{\int exp(-\frac{(\beta-\tilde{\beta})^T\Sigma^{-1}(\beta-\tilde{\beta})}{2}) \prod_{i=1}^{N} \Phi(\mathbf{x}_k^T\beta)^{y_k}(1-\Phi(\mathbf{x}_k^T\beta))^{(1-y_k)} d\beta}. \quad (A.2)$$

Since the expression in the denominator of A.2 is very difficult to solve analytically, it is not straightforward to sample from this distribution. However, the structure of the problem is simplified by introducing $N$ independent, latent variables $Z_1, \ldots, Z_N$, where $Z_k \sim N(\mathbf{x}_k^T\beta, 1)$. If the $Z_k$ were known this would correspond to the standard normal linear model, for which the solution is available. However, since they are not known, we can link the $Z_k$ to $\mathbf{x}_k^T\beta$ via the known $y_k$. Define $y_k$ such that $y_k = 1$ if $Z_k > 0$ and $y_k = 0$ if $Z_k \leq 0$. The $y_k$ then have a Bernoulli distribution with probability $\Phi(\mathbf{x}_k^T\beta)$.

$$\begin{aligned} P(y_k = 1) &= P(Z_k > 0), \\ &= 1 - P(Z_k \leq 0), \\ &= 1 - P(\frac{Z_k - \mathbf{x}_k^T\beta}{1} \leq \frac{0 - \mathbf{x}_k^T\beta}{1}), \\ &= 1 - \Phi(-\mathbf{x}_k^T\beta) = \Phi(\mathbf{x}_k^T\beta). \end{aligned}$$

Treating the $Z_k$'s as the augmented data, the posterior distribution for $\beta$ can be obtained using the Gibbs sampler. The joint posterior distribution of the unobservables $\beta$ and $Z_k$'s is given by

$$\pi(\beta, \mathbf{Z}|\mathbf{Y}) \propto \pi(\beta) \prod_{k=1}^{N} (1(Z_k > 0)1(Y_k = 1) + 1(Z_k \leq 0)1(Y_k = 0))\Phi(Z_i; \mathbf{x}_k^T\beta, 1). \quad (A.3)$$

The function $1(X \in A)$ is the indicator function that is equal to 1 if the random variable $X$ is contained in set $A$. From (A.3) we can see that the posterior distribution of $\beta$ conditional on $\mathbf{Z}$ is given by

$$\pi(\beta|\mathbf{Z}, \mathbf{y}) \propto \pi(\beta) \prod_{k=1}^{N} \Phi(Z_i; \mathbf{x}_k^T\beta, 1).$$

When the prior distribution of $\beta$ is $N_k(\beta, \Sigma)$, then the posterior distribution of $\beta$ is given by

$$\beta|\mathbf{Z}, \mathbf{y} \sim N_k((\Sigma^{-1} + \mathbf{X}^T\mathbf{X})^{-1}(\Sigma^{-1}\beta + \mathbf{X}^T Z), (\Sigma^{-1} + \mathbf{X}^T\mathbf{X})^{-1}). \quad \text{(A.4)}$$

The posterior distribution of $\mathbf{Z}$ conditional on $\beta$ also has a simple form,

$$\begin{aligned} Z_i|y, \beta &\sim N(x_k^T\beta, 1) \text{ truncated at the left by 0, if } y_k = 1, \text{ and} \\ Z_i|y, \beta &\sim N(x_k^T\beta, 1) \text{ truncated at the right by 0, if } y_k = 0. \end{aligned} \quad \text{(A.5)}$$

Using the Gibbs sampler algorithm one can alternately sample from (A.4) and (A.5) to obtain a sample from the posterior distribution of $\beta$. This method can be applied to obtain a sample from the posterior distribution of the $(a_{jd}, b_{jd})$'s in the random effects model developed in Chapter 5. The form of their posterior distributions can be seen from (5.5) and (5.6) to be similar to that in (A.2). The $y_k$'s representing the true disease status are Bernoulli random variables and the $i$ values are the covariates. Though in truth, neither of these variables is observed, at each iteration of the Gibbs sampler (described in section 5.2.3) they take on specific values.

# APPENDIX B

## *B.1 Programs used for the Bayesian fixed effects model in Section 4.3*

### *B.1.1 S-Plus program used to calculate cross-classification of test results in Tables 4.5 and 4.6*

```
simulate<-function(x) {

# setting test parameters to the mean value of their prior
# distributions
s1<-0.5
s2<-0.8
c1<-0.9
c2<-0.7
covp<-0.046
covn<-0.023
prev<-0.73

n<-200

d<-rep(1,4)
i<-rep(1,4)

# calculating the number of subjects in each cross-classification
# when the tests are dependent
d[1]<-n*(prev*(s1*s2+covp)+(1-prev)*((1-c1)*(1-c2)+covn))
d[2]<-n*(prev*(s1*(1-s2)-covp)+(1-prev)*((1-c1)*c2-covn))
d[3]<-n*(prev*((1-s1)*s2-covp)+(1-prev)*(c1*(1-c2)-covn))
d[4]<-n*(prev*((1-s1)*(1-s2)+covp)+(1-prev)*(c1*c2+covn))

# calculating the number of subjects in each cross-classification
```

```
# when the tests are dependent
i[1]<-n*(prev*(s1*s2)+(1-prev)*((1-c1)*(1-c2)))
i[2]<-n*(prev*(s1*(1-s2))+(1-prev)*((1-c1)*c2))
i[3]<-n*(prev*((1-s1)*s2)+(1-prev)*(c1*(1-c2)))
i[4]<-n*(prev*((1-s1)*(1-s2))+(1-prev)*(c1*c2))

return(d,i)
}
```

### *B.1.2 C++ program to implement the Gibbs sampler for the Bayesian fixed effects model*

For the sake of brevity only the functions related to the Gibbs sampler are included here and standard functions, such as those used to sample random variables, are omitted.

```
#include <iostream.h>
#include <stdlib.h>
#include <math.h>
#include <fstream.h>
#include <time.h>
#include "matrix.h"
#include "random.h"
#include "sir.h"

// global constants
int n11=68;
int n10=10;
int n01=65;
int n00=57;
int sum=200;

// declaring the variables used in the program
int y1,y2,y3,y4;
double *prev, *sens1, *spec1, *sens2, *spec2, p11, p10, p01, p00;
double *covp, *covn, *pvp1, *pvn1, *pvp2, *pvn2;
int size=5000,int i;

// output file
ofstream fout("d:/cpp/fixed");
```

```
// body of main program
void main(void) {

//initiating a random seed
srand( (unsigned)time( NULL ) );
long seed=rand();
long *idum=&seed;

// declaring prior distributions
double alphaprev = 1;
double betaprev = 1;

double alphasens1 = 34;
double betasens1 = 34;
double alphaspec1 = 90;
double betaspec1 = 10;
double alphasens2 = 32;
double betasens2 = 8;
double alphaspec2 = 42;
double betaspec2 = 18;
double alphacovp = 23;
double betacovp = 477;
double alphacovn = 23;
double betacovn = 477;

// allocating memory for pointers
prev_samp = (double *)malloc(size*sizeof(double));
sens1 = (double *)malloc(size*sizeof(double));
sens2 = (double *)malloc(size*sizeof(double));
spec1 = (double *)malloc(size*sizeof(double));
spec2 = (double *)malloc(size*sizeof(double));
pvp1 = (double *)malloc(size*sizeof(double));
pvn1 = (double *)malloc(size*sizeof(double));
pvp2 = (double *)malloc(size*sizeof(double));
pvn2 = (double *)malloc(size*sizeof(double));
covp = (double *)malloc(size*sizeof(double));
covn = (double *)malloc(size*sizeof(double));

// starting values
double prev_start=runif(idum);
```

```
double sens1_start=runif(idum);
double spec1_start=runif(idum);
double sens2_start=runif(idum);
double spec2_start=runif(idum);

// determining the lower and upper bounds for the
// starting values of the covariance parameters

double ubp=__min(sens1_start,sens2_start)
-sens1_start*sens2_start;
double lbp=0;
double ubn=__min(spec1_start,spec2_start)
-spec1_start*spec2_start;
double lbn=0;

// starting values of the covariance parameters
double covp_start=runif(idum)*(ubp-lbp)+lbp;
double covn_start=runif(idum)*(ubn-lbn)+lbn;

// initializing all array elements to 1
for (i=0; i<size; i++)
{
*(prev+i) = 1.0;
*(sens1+i) = 1.0;
*(spec1+i) = 1.0;
*(sens2+i) = 1.0;
*(spec2+i) = 1.0;
*(covp+i) = 1.0;
*(covn+i) = 1.0;
}

// setting the first entry in each array to the starting value
*prev = prev_start;
*sens1 = sens1_start;
*spec1 = spec1_start;
*sens2 = sens2_start;
*spec2 = spec2_start;
*covp = covp_start;
*covn = covn_start;

// the 5000 iterations of the Gibbs sampler begin
```

```
for (i=1; i<size; i++) {

*(pvp1+i-1) = (*(prev+i-1) * *(sens1+i-1))
/ (*(prev+i-1) * *(sens1+i-1) + (1-*(prev+i-1)) * (1-*(spec1+i-1)));
*(pvn1+i-1) = (1-*(prev+i-1))* *(spec1+i-1)
/ (*(prev+i-1)*(1-*(sens1+i-1)) + (1-*(prev+i-1))* *(spec1+i-1));
*(pvp2+i-1) = (*(prev+i-1) * *(sens2+i-1))
/ (*(prev+i-1) * *(sens2+i-1) + (1-*(prev+i-1)) * (1-*(spec2+i-1)));
*(pvn2+i-1) = (1-*(prev+i-1))* *(spec2+i-1)
/ (*(prev+i-1)*(1-*(sens2+i-1)) + (1-*(prev+i-1))* *(spec2+i-1));

fout << i << "\t" << *(prev+i-1) << "\t" << *(sens1+i-1) << "\t"
     << *(spec1+i-1) << "\t" << *(pvp1+i-1) << "\t" << *(pvn1+i-1)
     << "\t" << *(sens2+i-1) << "\t"<< *(spec2+i-1) << "\t"
     << *(pvp2+i-1) << "\t" << *(pvn2+i-1) << "\t"
     << *(covp+i-1) << "\t" << *(covn+i-1) << "\n" ;
fout.flush();

// calculating the value of a true positive in
// each cross-classification of the two tests
p11 = (*(prev+i-1)* (*(sens1+i-1)* *(sens2+i-1)+*(covp+i-1)))/
(*(prev+i-1)* (*(sens1+i-1)* *(sens2+i-1)+*(covp+i-1))
    +(1-*(prev+i-1))*((1- *(spec1+i-1))*(1- *(spec2+i-1))
                      +*(covn+i-1)));
y1 = rbin(n11,p11,idum);

p10 = (*(prev+i-1)*(*(sens1+i-1)*(1- *(sens2+i-1))-*(covp+i-1)))/
(*(prev+i-1)* (*(sens1+i-1)*(1- *(sens2+i-1))-*(covp+i-1))
  +(1-*(prev+i-1))*((1- *(spec1+i-1))* *(spec2+i-1)
                    -*(covn+i-1)));
y2 = rbin(n10,p10,idum);

p01 = (*(prev+i-1)*((1- *(sens1+i-1))* *(sens2+i-1)- *(covp+i-1)))/
(*(prev+i-1)*((1- *(sens1+i-1))* *(sens2+i-1)- *(covp+i-1))
  +(1-*(prev+i-1))* (*(spec1+i-1)*(1- *(spec2+i-1))
                     - *(covn+i-1)));
y3 = rbin(n01,p01,idum);

p00 = (*(prev+i-1)*((1- *(sens1+i-1))*(1- *(sens2+i-1))+*(covp+i-1)))/
  (*(prev+i-1)*((1- *(sens1+i-1))*(1- *(sens2+i-1))+*(covp+i-1))
  +(1-*(prev+i-1))* (*(spec1+i-1)* *(spec2+i-1)+*(covn+i-1)));
```

```
y4 = rbin(n00,p00,idum);

// drawing the estimated prevalence from a beta distribution
        *(prev+i)=rbeta(y1+y2+y3+y4+alphaprev,
sum-(y1+y2+y3+y4)+betaprev,idum);

// updating the sensitivities and specificities
sir_sens1(alphasens1,betasens1,i,
sens1,sens2,covp,y1,y2,y3,y4,idum);
sir_spec1(alphaspec1,betaspec1,i,spec1,spec2,covn,
n11,n10,n01,n00,y1,y2,y3,y4,idum);
sir_sens2(alphasens2,betasens2,i,
sens1,sens2,covp,y1,y2,y3,y4,idum);
sir_spec2(alphaspec2,betaspec2,i,spec1,spec2,covn,
n11,n10,n01,n00,y1,y2,y3,y4,idum);

// updating the covariances
sir_covp(alphacovp,betacovp,covp,i,
sens1,sens2,y1,y2,y3,y4,idum);
sir_covn(alphacovn,betacovn,covn,i,spec1,spec2,
n11,n10,n01,n00,y1,y2,y3,y4,idum);
}
}

\\ SIR for updating the sensitivity of the first test
void sir_sens1(double alphasens1,double betasens1, int ii,
double *sens1, double *sens2, double *covp,
int y1, int y2, int y3, int y4, long *idum)
{
double p,cusum[50],w[50],k[50],g[50],lb,ub;
cusum[0]=0; cusum[49]=1;

lb=*(covp+ii-1)/(1- *(sens2+ii-1));
ub=1 - *(covp+ii-1)/ *(sens2+ii-1);

for (int i=1; i<49; i++)
{
g[i]=runif(idum);
k[i]=(ub-lb)*g[i]+lb;
w[i]=pow(k[i] * *(sens2+ii-1) + *(covp+ii-1),y1)
*pow(k[i]*(1- *(sens2+ii-1)) - *(covp+ii-1),y2)
```

```
*pow((1-k[i]) * *(sens2+ii-1) - *(covp+ii-1),y3)
*pow((1-k[i]) * (1 - *(sens2+ii-1)) + *(covp+ii-1),y4)
*pow(k[i],(alphasens1-1))*pow((1-k[i]),(betasens1-1));

cusum[i]=cusum[i-1]+w[i];
}
p=runif(idum);
for (i=1;i<49;i++) {
cusum[i] /= cusum[48];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
*(sens1+ii)=k[i];
break;
}
}
}

\\ SIR for updating the specificity of the first test
void sir_spec1(double alphaspec1, double betaspec1, int ii,
 double *spec1, double *spec2, double *covn,
 int n1, int n2, int n3, int n4,
 int y1, int y2, int y3, int y4, long *idum)
{
double p,cusum[50],w[50],k[50],g[50],lb,ub;
cusum[0]=0; cusum[49]=1;

lb=*(covn+ii-1)/(1- *(spec2+ii-1));
ub=1 - *(covn+ii-1)/ *(spec2+ii-1);

for (int i=1; i<49; i++)
{
g[i]=runif(idum);
k[i]=(ub-lb)*g[i]+lb;
w[i]=pow(k[i] * *(spec2+ii-1) + *(covn+ii-1),(n4-y4))
*pow(k[i] * (1- *(spec2+ii-1)) - *(covn+ii-1),(n3-y3))
*pow((1-k[i]) * *(spec2+ii-1) - *(covn+ii-1),(n2-y2))
* pow((1-k[i]) * (1 - *(spec2+ii-1)) + *(covn+ii-1),(n1-y1))
*pow(k[i],(alphaspec1-1))*pow((1-k[i]),(betaspec1-1));
cusum[i]=cusum[i-1]+w[i];
}

p=runif(idum);
```

```
for (i=1;i<49;i++) {
cusum[i] /= cusum[48];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
*(spec1+ii)=k[i];
break;
}
}
}

\\ SIR for updating the sensitivity of the second test
void sir_sens2(double alphasens2, double betasens2, int ii,
double *sens1, double *sens2, double *covp,
int y1, int y2, int y3, int y4, long *idum)
{
double p,cusum[50],w[50],k[50],g[50],lb,ub;
cusum[0]=0; cusum[49]=1;

lb=*(covp+ii-1)/(1- *(sens1+ii-1));
ub=1 - *(covp+ii-1)/ *(sens1+ii-1);

for (int i=1; i<49; i++)
{
g[i]=runif(idum);
k[i]=(ub-lb)*g[i]+lb;
w[i]=pow(k[i] * *(sens1+ii) + *(covp+ii-1),y1)
     *pow(k[i] * (1- *(sens1+ii)) - *(covp+ii-1),y3)
     *pow((1-k[i]) * *(sens1+ii) - *(covp+ii-1),y2)
     *pow((1-k[i]) * (1 - *(sens1+ii)) + *(covp+ii-1),y4)
     *pow(k[i],(alphasens2-1))*pow((1-k[i]),(betasens2-1));
cusum[i]=cusum[i-1]+w[i];
}

p=runif(idum);
for (i=1;i<49;i++) {
cusum[i] /= cusum[48];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
*(sens2+ii)=k[i];
break;
}
}
}
```

```
\\ SIR for updating the specificity of the second test
void sir_spec2(double alphaspec2, double betaspec2, int ii,
double *spec1, double *spec2, double *covn,
int n1, int n2, int n3, int n4,
int y1, int y2, int y3, int y4, long *idum)
{
double p,cusum[50],w[50],k[50],g[50],lb,ub;
cusum[0]=0; cusum[49]=1;

lb=*(covn+ii-1)/(1- *(spec1+ii-1));
ub=1 - *(covn+ii-1)/ *(spec1+ii-1);

for (int i=1; i<49; i++)
{
g[i]=runif(idum);
k[i]=(ub-lb)*g[i]+lb;
w[i]=pow(k[i] * *(spec1+ii) + *(covn+ii-1),(n4-y4))
     *pow(k[i] * (1- *(spec1+ii)) - *(covn+ii-1),(n2-y2))
     *pow((1-k[i]) * *(spec1+ii) - *(covn+ii-1),(n3-y3))
     *pow((1-k[i]) * (1 - *(spec1+ii)) + *(covn+ii-1),(n1-y1))
     *pow(k[i],(alphaspec2-1))*pow((1-k[i]),(betaspec2-1));
cusum[i]=cusum[i-1]+w[i];
}

p=runif(idum);
for (i=1;i<49;i++) {
cusum[i] /= cusum[48];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
*(spec2+ii)=k[i];
break;
}
}
}

\\ SIR for updating the covariance among the diseased subjects
void sir_covp(double alphacovp, double betacovp, double *covp, int ii,
double *sens1, double *sens2,
int y1, int y2, int y3, int y4, long *idum)
{
double p,cusum[50],w[50],k[50],g[50],lb,ub;
```

```
cusum[0]=0; cusum[49]=1;

lb=0;
ub=__min(__min(*(sens1+ii),
*(sens2+ii))- *(sens1+ii) * *(sens2+ii),0.1);

for (int i=1; i<49; i++)
{
g[i]=runif(idum);
k[i]=g[i]*(ub-lb)+lb;
w[i]=pow(*(sens1+ii) * *(sens2+ii) + k[i],y1)
*pow(*(sens1+ii) * (1- *(sens2+ii)) - k[i],y2)
*pow((1-*(sens1+ii)) * *(sens2+ii) - k[i],y3)
*pow((1-*(sens1+ii)) * (1 - *(sens2+ii)) + k[i],y4)
*pow(k[i],(alphacovp-1))*pow((1-k[i]),(betacovp-1));
cusum[i]=cusum[i-1]+w[i];
}

p=runif(idum);
for (i=1;i<49;i++) {
cusum[i] /= cusum[48];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
*(covp+ii)=k[i];
break;
}
}
}

\\ SIR for updating the covariance among the non-diseased subjects
void sir_covn(double alphacovn, double betacovn, double *covn,
int ii, double *spec1, double *spec2,
int n1, int n2, int n3, int n4,
int y1, int y2, int y3, int y4, long *idum)
{
double p,cusum[50],w[50],k[50],g[50],lb,ub;
cusum[0]=0; cusum[49]=1;

lb=0;
ub=__min(__min(*(spec1+ii),
*(spec2+ii))- *(spec1+ii) * *(spec2+ii),0.07);
```

```
for (int i=1; i<49; i++)
{
g[i]=runif(idum);
k[i]=g[i]*(ub-lb)+lb;
w[i]=pow(*(spec1+ii) * *(spec2+ii) + k[i],(n4-y4))
     *pow(*(spec1+ii) * (1- *(spec2+ii)) - k[i],(n3-y3))
                  *pow((1-*(spec1+ii)) * *(spec2+ii) - k[i],(n2-y2))
     *pow((1-*(spec1+ii)) * (1 - *(spec2+ii)) + k[i],(n1-y1))
     *pow(k[i],(alphacovn-1))*pow((1-k[i]),(betacovn-1));

cusum[i]=cusum[i-i]+w[i];
}

p=runif(idum);
for (i=1;i<49;i++) {
cusum[i] /= cusum[48];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
*(covn+ii)=k[i];
break;
}
}
}
```

# APPENDIX C

## *C.1 Programs used for the Bayesian random effects model in Section 5.3*

### *C.1.1 S-Plus program used to calculate cross-classification of test results in Table 5.2*

```
sim6<-function(x) {

a11<-0
a21<-1.362548
b1<-1.273193

a10<-2.253658
a20<-0.9221768
b0<-1.446533

t<-rnorm(200,0,1)

prev<-0.73

s1<-pnorm(a11+b1*t)
c1<-pnorm(a10+b0*t)
s2<-pnorm(a21+b1*t)
c2<-pnorm(a20+b0*t)

p11<-prev*(s1*s2)+(1-prev)*((1-c1)*(1-c2))
p10<-prev*(s1*(1-s2))+(1-prev)*((1-c1)*c2)
p01<-prev*((1-s1)*s2)+(1-prev)*(c1*(1-c2))
p00<-prev*((1-s1)*(1-s2))+(1-prev)*(c1*c2)

n11<-sum(p11)
n10<-sum(p10)
n01<-sum(p01)
```

```
n00<-sum(p00)

return(n11,n10,n01,n00)

}
```

### C.1.2 C++ program used to implement the Gibbs sampler for the Bayesian random effects model

For the sake of brevity only the functions related to the Gibbs sampler are included here and standard functions, such as those used to sample random variables, are omitted.

```
#include <iostream.h>
#include <stdlib.h>
#include <math.h>
#include <fstream.h>
#include <time.h>
#include "matrix.h"
#include "random.h"
#include "sir6.h"

// prototypes of functions appearing in the body of the program
int sumarr(int arr[], int n1, int n2);
void sir(int *y_samp, int ii, double *ab1_samp, double *ab0_samp,
double *t, long *idum);

// global variables
int n11=75;
int n10=5;
int n01=58;
int n00=62;
int sum=200;
int mu=0;
int sigma=1;
double *T;
double mean=0;

// declaring the variables used in the program
int *y_samp;
```

```
double *prev_samp, *ab1_samp, *ab0_samp;
double *sens1, *spec1, *sens2, *spec2, p11, p10, p01, p00;
double *sens1_samp, *sens2_samp, *spec1_samp, *spec2_samp;
double *pvp1, *pvn1, *pvp2, *pvn2;
double *sens1_check1, sens1_check2;

// output file
ofstream fout("d:/main/cpp/thesis/results/chap51");

// main body of program
void main (void)
{
//details of iterations
int size=5000;
int i,j;
srand( (unsigned)time( NULL ) );
long seed=rand();
long *idum=&seed;

// declaring prior distributions
double alphaprev = 1;
double betaprev = 1;

double a11, a10, a21, a20, b1, b0;
double a11Bstar, a10Bstar, a21Bstar, a20Bstar, b1Bstar, b0Bstar;

a11 = 0; a10 = 2.253658; a21 = 1.362548; a20 = 0.9221768;
b1 = 1.273193; b0 = 0.91;

b1Bstar = 0.5;
b0Bstar = 0.5;
a11Bstar = 0.133265;
a10Bstar = 0.12849975;
a21Bstar = 0.175;
a20Bstar = 0.19059425;

// allocating memory for pointers
y_samp = (int *)malloc(sum*sizeof(int));
prev_samp = (double *)malloc(size*sizeof(double));
sens1 = (double *)malloc(size*sizeof(double));
sens2 = (double *)malloc(size*sizeof(double));
```

```
spec1 = (double *)malloc(size*sizeof(double));
spec2 = (double *)malloc(size*sizeof(double));
sens1_samp = (double *)malloc(size*sum*sizeof(double));
spec1_samp = (double *)malloc(size*sum*sizeof(double));
sens2_samp = (double *)malloc(size*sum*sizeof(double));
spec2_samp = (double *)malloc(size*sum*sizeof(double));
sens1_check1 = (double *)malloc(sum*sizeof(double));
pvp1 = (double *)malloc(size*sizeof(double));
pvn1 = (double *)malloc(size*sizeof(double));
pvp2 = (double *)malloc(size*sizeof(double));
pvn2 = (double *)malloc(size*sizeof(double));
T = (double *)malloc(2*sum*size*sizeof(double));
ab1_samp = (double *)malloc(3*size*sizeof(double));
ab0_samp = (double *)malloc(3*size*sizeof(double));

// setting the first entry in each array to a random starting value
*(prev_samp)=ran2(idum);
*(ab1_samp)=ran2(idum);
*(ab1_samp+1)=ran2(idum);
*(ab1_samp+2)=ran2(idum);
*(ab0_samp)=ran2(idum);
*(ab0_samp+1)=ran2(idum);
*(ab0_samp+2)=ran2(idum);


// initializing all array elements to 1
for (i=1; i<size; i++)
{
*(prev_samp+i) = 1.0;
*(ab1_samp+3*i) = 1.0;
*(ab1_samp+3*i+1) = 1.0;
*(ab1_samp+3*i+2) = 1.0;
*(ab0_samp+3*i) = 1.0;
*(ab0_samp+3*i+1) = 1.0;
*(ab0_samp+3*i+2) = 1.0;
}

for (i=0; i<size; i++)
{
*(sens1+i) = 1.0;
*(spec1+i) = 1.0;
```

```
*(sens2+i) = 1.0;
*(spec2+i) = 1.0;
*(pvp1+i)=1.0;
*(pvn1+i)=1.0;
*(pvp2+i)=1.0;
*(pvn2+i)=1.0;
}

for (i=0; i<size; i++) {
for (j=0; j<sum; j++) {
*(sens1_samp+i*sum+j) = 1.0;
*(spec1_samp+i*sum+j) = 1.0;
*(sens2_samp+i*sum+j) = 1.0;
*(spec2_samp+i*sum+j) = 1.0;
}
}


// drawing random N(0,1) values for the initial t values
for (j=0; j<sum; j++) {
*(y_samp+j) = 1;
*(T+j*2) = 1;
*(T+j*2+1) = gasdev(0,1,idum);
}

// the 5000 iterations begin
    for (i=1; i<size; i++)
{
*(sens1+i-1) = pnorm(*(ab1_samp+3*(i-1))/sqrt(1
+*(ab1_samp+3*(i-1)+2)*(ab1_samp+3*(i-1)+2)));
*(spec1+i-1) = pnorm(*(ab0_samp+3*(i-1))/sqrt(1
+*(ab0_samp+3*(i-1)+2)*(ab0_samp+3*(i-1)+2)));
*(sens2+i-1) = pnorm(*(ab1_samp+3*(i-1)+1)/sqrt(1
+*(ab1_samp+3*(i-1)+2)*(ab1_samp+3*(i-1)+2)));
*(spec2+i-1) = pnorm(*(ab0_samp+3*(i-1)+1)/sqrt(1
+*(ab0_samp+3*(i-1)+2)*(ab0_samp+3*(i-1)+2)));

/// calculating the predictive value positive and negative
*(pvp1+i-1) = (*(prev_samp+i-1)* *(sens1+i-1))/
(*(prev_samp+i-1)* *(sens1+i-1)
+ (1-*(prev_samp+i-1)) * (1-*(spec1+i-1)));
```

```
*(pvn1+i-1) = ((1-*(prev_samp+i-1))* *(spec1+i-1))/
(*(prev_samp+i-1)*(1-*(sens1+i-1))
+(1-*(prev_samp+i-1))* *(spec1+i-1));
*(pvp2+i-1) = (*(prev_samp+i-1)* *(sens2+i-1))/
(*(prev_samp+i-1)* *(sens2+i-1)
+ (1-*(prev_samp+i-1))* (1-*(spec2+i-1)));
*(pvn2+i-1) = ((1-*(prev_samp+i-1))* *(spec2+i-1))/
(*(prev_samp+i-1)*(1-*(sens2+i-1))
+ (1-*(prev_samp+i-1))* *(spec2+i-1));

fout << i << "\t" << *(prev_samp+i-1) << "\t" << *(sens1+i-1)
<< "\t" << *(spec1+i-1) << "\t" << *(pvp1+i-1)
<< "\t" << *(pvn1+i-1) << "\t" << *(sens2+i-1)
<< "\t"<< *(spec2+i-1) << "\t" << *(pvp2+i-1)
<< "\t" << *(pvn2+i-1) << "\t" << *(ab0_samp+3*(i-1))
<< "\t" << *(ab0_samp+3*(i-1)+1)
<< "\t" << *(ab0_samp+3*(i-1)+2)
<< "\t" << *(ab1_samp+3*(i-1))
<< "\t" << *(ab1_samp+3*(i-1)+1)
<< "\t" << *(ab1_samp+3*(i-1)+2) << "\t";
fout.flush();

// calculating the S and C's for each subject as a function of t
for (j=0; j<sum; j++)
{
fout << *(T+(i-1)*2*sum+j*2+1) << "\t" ;
*(sens1_samp+(i-1)*sum+j) = pnorm(*(ab1_samp+3*(i-1))
+*(ab1_samp+3*(i-1)+2)*(T+(i-1)*2*sum+j*2+1));
*(spec1_samp+(i-1)*sum+j) = pnorm(*(ab0_samp+3*(i-1))
+*(ab0_samp+3*(i-1)+2)*(T+(i-1)*2*sum+j*2+1));
*(sens2_samp+(i-1)*sum+j) = pnorm(*(ab1_samp+3*(i-1)+1)
+*(ab1_samp+3*(i-1)+2)*(T+(i-1)*2*sum+j*2+1));
*(spec2_samp+(i-1)*sum+j) = pnorm(*(ab0_samp+3*(i-1)+1)
+*(ab0_samp+3*(i-1)+2)*(T+(i-1)*2*sum+j*2+1));
}

fout << "\n";

// calculating the value of a true positive in each of the four
// segments n11,n10,n01,n00
for (j=0; j<n11; j++)
```

```
{
p11 = (*(prev_samp+i-1)* *(sens1_samp+(i-1)*sum+j)*
*(sens2_samp+(i-1)*sum+j))/
(*(prev_samp+i-1)* *(sens1_samp+(i-1)*sum+j)*
*(sens2_samp+(i-1)*sum+j)
+(1-*(prev_samp+i-1))*(1- *(spec1_samp+(i-1)*sum+j))
*(1- *(spec2_samp+(i-1)*sum+j)));
    *(y_samp+j) = rbern(p11,idum);
}

for (j=n11; j<(n11+n10); j++)
{
p10 = (*(prev_samp+i-1)* *(sens1_samp+(i-1)*sum+j)
*(1- *(sens2_samp+(i-1)*sum+j)))/
(*(prev_samp+i-1)* *(sens1_samp+(i-1)*sum+j)
*(1- *(sens2_samp+(i-1)*sum+j))
+(1-*(prev_samp+i-1))*(1- *(spec1_samp+(i-1)*sum+j))
* *(spec2_samp+(i-1)*sum+j));
*(y_samp+j) = rbern(p10,idum);
}

for (j=(n11+n10); j<(n11+n10+n01); j++)
{
p01 = (*(prev_samp+i-1)*(1- *(sens1_samp+(i-1)*sum+j))*
*(sens2_samp+(i-1)*sum+j))/
(*(prev_samp+i-1)*(1- *(sens1_samp+(i-1)*sum+j))*
*(sens2_samp+(i-1)*sum+j)
+(1-*(prev_samp+i-1))* *(spec1_samp+(i-1)*sum+j)*
(1- *(spec2_samp+(i-1)*sum+j)));
*(y_samp+j) = rbern(p01,idum);
}

for (j=(n11+n10+n01); j<sum; j++)
{
p00 = (*(prev_samp+i-1)*(1- *(sens1_samp+(i-1)*sum+j))*
(1- *(sens2_samp+(i-1)*sum+j)))/
  (*(prev_samp+i-1)*(1- *(sens1_samp+(i-1)*sum+j))*
(1- *(sens2_samp+(i-1)*sum+j))
  +(1-*(prev_samp+i-1))* *(spec1_samp+(i-1)*sum+j)*
*(spec2_samp+(i-1)*sum+j));
*(y_samp+j) = rbern(p00,idum);
```

```
}

// estimated number of true positives in each segment
        int y1 = sumarr(y_samp,0,n11);
int y2 = sumarr(y_samp,n11,n11+n10);
int y3 = sumarr(y_samp,n11+n10,n11+n10+n01);
int y4 = sumarr(y_samp,n11+n10+n01,sum);

// sirs to pick b values
sir_b0(y_samp,i,T,b0,b0Bstar,ab0_samp,idum,n11,n10,n01,n00);
sir_a10(y_samp,i,T,a10,a10Bstar,ab0_samp,idum,n11,n10,n01,n00);
sir_a20(y_samp,i,T,a20,a20Bstar,ab0_samp,idum,n11,n10,n01,n00);

sir_b1(y_samp,i,T,b1,b1Bstar,ab1_samp,idum,n11,n10,n01,n00);
sir_a11(y_samp,i,T,a11,a11Bstar,ab1_samp,idum,n11,n10,n01,n00);
sir_a21(y_samp,i,T,a21,a21Bstar,ab1_samp,idum,n11,n10,n01,n00);
//sirs to pick a values

// drawing the estimated prevalence from a beta distribution
        *(prev_samp+i)=rbeta(y1+y2+y3+y4+alphaprev,
sum-(y1+y2+y3+y4)+betaprev,idum);

// drawing the updated t's using a sir function
sir(y_samp,i,ab1_samp,ab0_samp,T,idum);
}

}

// sumarr called to sum the elements of an array
int sumarr(int *arr, int n1, int n2)
{
int total = 0;
for (int i=n1;i<n2;i++)
total = total + *(arr+i);
return total;
}

// SIR to pick t values
void sir(int *y_samp, int ii, double *ab1_samp, double *ab0_samp,
double *t, long *idum)
{
```

```
int i,j;
double k[25],w[25],cusum[25];
for (j=0; j<sum; j++) {
double p=ran2(idum);
cusum[0] = 0; cusum[24]=1;
if (y_samp[j]==1) {
for (i=1; i<24; i++) {
k[i]=gasdev(mu,1,idum);
if (j<n11) w[i]=pnorm(*(ab1_samp+3*ii)+*(ab1_samp+3*ii+2)*k[i])*
pnorm(*(ab1_samp+3*ii+1)+*(ab1_samp+3*ii+2)*k[i])
*exp(-k[i]*mu+(mu*mu)/2);
if ((j>=n11) && (j<(n11+n10))) w[i]=pnorm(*(ab1_samp+3*ii)+
*(ab1_samp+3*ii+2)*k[i])*pnorm(-(*(ab1_samp+3*ii+1)+
*(ab1_samp+3*ii+2)*k[i]))*exp(-k[i]*mu+(mu*mu)/2);
if ((j>=(n11+n10)) && (j<(n11+n10+n01)))
w[i]=pnorm(-(*(ab1_samp+3*ii)+
*(ab1_samp+3*ii+2)*k[i]))*pnorm(*(ab1_samp+3*ii+1)+
*(ab1_samp+3*ii+2)*k[i])*exp(-k[i]*mu+(mu*mu)/2);
if (j>=(n11+n10+n01)) w[i]=pnorm(-(*(ab1_samp+3*ii)+
*(ab1_samp+3*ii+2)*k[i]))*pnorm(-(*(ab1_samp+3*ii+1)+
*(ab1_samp+3*ii+2)*k[i]))*exp(-k[i]*mu+(mu*mu)/2);
cusum[i]=cusum[i-1]+w[i];
}
}
else if (y_samp[j]==0) {
for (i=1; i<24; i++) {
k[i]=gasdev(-mu,1,idum);
if (j<n11) w[i]=pnorm(-(*(ab0_samp+3*ii)+
*(ab0_samp+3*ii+2)*k[i]))*pnorm(-(*(ab0_samp+3*ii+1)+
*(ab0_samp+3*ii+2)*k[i]))*exp(+k[i]*mu+(mu*mu)/2);
if ((j>=n11) && (j<(n11+n10))) w[i]=pnorm(-(*(ab0_samp+3*ii)+
*(ab0_samp+3*ii+2)*k[i]))*pnorm((*(ab0_samp+3*ii+1)+
*(ab0_samp+3*ii+2)*k[i]))*exp(+k[i]*mu+(mu*mu)/2);
if ((j>=(n11+n10)) && (j<(n11+n10+n01))) w[i]=pnorm(*(ab0_samp+3*ii)+
*(ab0_samp+3*ii+2)*k[i])*pnorm(-(*(ab0_samp+3*ii+1)+
*(ab0_samp+3*ii+2)*k[i]))*exp(+k[i]*mu+(mu*mu)/2);
if (j>=(n11+n10+n01)) w[i]=pnorm(*(ab0_samp+3*ii)+
*(ab0_samp+3*ii+2)*k[i])*pnorm(*(ab0_samp+2*ii+1)+
*(ab0_samp+2*ii+2)*k[i])*exp(+k[i]*mu+(mu*mu)/2);
cusum[i]=cusum[i-1]+w[i];
}
```

```
}
for (i=1; i<25; i++) {
if (i<24) cusum[i] /= cusum[23];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
*(t+ii*2*sum+j*2+1)=k[i];
break;
}
}
}
}
// SIR for picking b values when bi1 = b1, ie S's share same b
void sir_b1(int *y, int ii, double *T, double b1, double sd,
double *ab1_samp, long *idum, int n11, int n10, int n01, int n00) {
double p,cusum[100],w[100],k[100],wi[200];
int i,j;

p=ran2(idum);
cusum[0]=0; cusum[99]=1;
for (i=1;i<99;i++) {
k[i]=gasdev(b1,sd,idum);
w[i]=0.0;

for (j=0;j<200;j++) {
if (*(y+j)==0) wi[j]=log(1*1);
else {
if (j<n11) {
wi[j]=log(1*pnorm(*(ab1_samp+3*(ii-1))+k[i]**(T+(ii-1)*400+j*2+1))
*pnorm(*(ab1_samp+3*(ii-1)+1)+k[i]**(T+(ii-1)*400+j*2+1)));
}
if (j>=n11 && j<(n11+n10)) {
wi[j]=log(1*pnorm(*(ab1_samp+3*(ii-1))+k[i]**(T+(ii-1)*400+j*2+1))
*(1-pnorm(*(ab1_samp+3*(ii-1)+1)+k[i]**(T+(ii-1)*400+j*2+1))));
}
if (j>=(n11+n10) && j<(n11+n10+n01)) {
wi[j]=log(1*(1-pnorm(*(ab1_samp+3*(ii-1))+k[i]**(T+(ii-1)*400+j*2+1)))
*pnorm(*(ab1_samp+3*(ii-1)+1)+k[i]**(T+(ii-1)*400+j*2+1)));
}
if (j>=n11+n10+n01) {
wi[j]=log(1*(1-pnorm(*(ab1_samp+3*(ii-1))+k[i]**(T+(ii-1)*400+j*2+1)))
*(1-pnorm(*(ab1_samp+3*(ii-1)+1)+k[i]**(T+(ii-1)*400+j*2+1))));
}
```

```
}
w[i] += wi[j];

}
cusum[i]=cusum[i-1]+exp(w[i]);
}

for (i=1;i<100;i++) {
if (i<99) cusum[i] /= cusum[98];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
*(ab1_samp+3*ii+2)=k[i];
break;
}
}
}

// SIR for picking b values when bi0 = b0, ie S's share same b
void sir_b0(int *y, int ii, double *T, double b0, double sd,
double *ab0_samp, long *idum, int n11, int n10, int n01, int n00) {
double p,cusum[100],w[100],k[100],wi[200];
int i,j;

p=ran2(idum);
cusum[0]=0; cusum[99]=1;
for (i=1;i<99;i++) {
k[i]=gasdev(b0,sd,idum);
w[i]=0.0;

for (j=0;j<200;j++) {
if (*(y+j)==1) wi[j]=log(1*1);
else {
if (j<n11) {
wi[j]=log(1*(1-pnorm(*(ab0_samp+3*(ii-1))+k[i]*
*(T+(ii-1)*400+j*2+1)))
*(1-pnorm(*(ab0_samp+3*(ii-1)+1)+k[i]**(T+(ii-1)*400+j*2+1))));
}
if (j>=n11 && j<(n11+n10)) {
wi[j]=log(1*(1-pnorm(*(ab0_samp+3*(ii-1))+k[i]**(T+(ii-1)*400+j*2+1)))
*pnorm(*(ab0_samp+3*(ii-1)+1)+k[i]**(T+(ii-1)*400+j*2+1)));
}
if (j>=(n11+n10) && j<(n11+n10+n01)) {
```

```
wi[j]=log(1*pnorm(*(ab0_samp+3*(ii-1))+k[i]**(T+(ii-1)*400+j*2+1))
*(1-pnorm(*(ab0_samp+3*(ii-1)+1)+k[i]**(T+(ii-1)*400+j*2+1))));
}
if (j>=n11+n10+n01) {
wi[j]=log(1*pnorm(*(ab0_samp+3*(ii-1))+k[i]**(T+(ii-1)*400+j*2+1))
*pnorm(*(ab0_samp+3*(ii-1)+1)+k[i]**(T+(ii-1)*400+j*2+1)));
}
}
w[i] += wi[j];
}
cusum[i]=cusum[i-1]+exp(w[i]);
}
for (i=1;i<100;i++) {
if (i<99) cusum[i] /= cusum[98];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
*(ab0_samp+3*ii+2)=k[i];
break;
}
}
}

// SIR for picking a11 values
void sir_a11(int *y, int ii, double *T, double a11, double sd,
double *ab1_samp, long *idum, int n11, int n10, int n01, int n00) {
double p,cusum[100],w[100],k[100],wi[200];
int i,j;

p=ran2(idum);
cusum[0]=0; cusum[99]=1;
for (i=1;i<99;i++) {
k[i]=gasdev(a11,sd,idum);
w[i]=0;

for (j=0;j<200;j++) {
if (*(y+j)==0) wi[j]=log(1*1);
else {
if (j<(n11+n10)) wi[j]=log(1*pnorm(k[i]+
*(ab1_samp+3*ii+2)**(T+(ii-1)*400+j*2+1)));
else wi[j]=log(1*(1-pnorm(k[i]+*(ab1_samp+3*ii+2)*
*(T+(ii-1)*400+j*2+1))));
}
```

```
}
w[i] += wi[j];
cusum[i]=cusum[i-1]+exp(w[i]);
}
for (i=1;i<100;i++) {
if (i<99) cusum[i] /= cusum[98];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
*(ab1_samp+3*ii)=k[i];
break;
}
}
}

// SIR for picking a10 values
void sir_a10(int *y, int ii, double *T, double a10, double sd,
double *ab0_samp, long *idum, int n11, int n10, int n01, int n00) {
double p,cusum[100],w[100],k[100],wi[200];
int i,j;

p=ran2(idum);
cusum[0]=0; cusum[99]=1;
for (i=1;i<99;i++) {
k[i]=gasdev(a10,sd,idum);
w[i]=0;

for (j=0;j<200;j++) {
if (*(y+j)==1) wi[j]=log(1*1);
else {
if (j<(n11+n10)) wi[j]=log(1*(1-pnorm(k[i]+*(ab0_samp+3*ii+2)*
*(T+(ii-1)*400+j*2+1))));
else wi[j]=log(1*pnorm(k[i]+*(ab0_samp+3*ii+2)*
*(T+(ii-1)*400+j*2+1)));
}
w[i] += wi[j];
}
cusum[i]=cusum[i-1]+exp(w[i]);
}

for (i=1;i<100;i++) {
if (i<99) cusum[i] /= cusum[98];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
```

```
*(ab0_samp+3*ii)=k[i];
break;
}
}
}

// SIR for picking a21 values
void sir_a21(int *y, int ii, double *T, double a21, double sd,
double *ab1_samp, long *idum, int n11, int n10, int n01, int n00) {
double p,cusum[100],w[100],k[100],wi[200];
int i,j;
p=ran2(idum);
cusum[0]=0; cusum[99]=1;
for (i=1;i<99;i++) {
k[i]=gasdev(a21,sd,idum);
w[i]=0;

for (j=0;j<200;j++) {
if (*(y+j)==0) wi[j]=log(1*1);
else {
if (j<n11 || (j>=(n11+n10) && j<(n11+n10+n01)))
wi[j]=log(1*pnorm(k[i]+*(ab1_samp+3*ii+2)*
*(T+(ii-1)*400+j*2+1)));
else wi[j]=log(1*(1-pnorm(k[i]+*(ab1_samp+3*ii+2)*
*(T+(ii-1)*400+j*2+1))));
}
w[i] += wi[j];
}
cusum[i]=cusum[i-1]+exp(w[i]);
}

for (i=1;i<100;i++) {
if (i<99) cusum[i] /= cusum[98];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
*(ab1_samp+3*ii+1)=k[i];
break;
}
}
}

// SIR for picking a20 values
```

```
void sir_a20(int *y, int ii, double *T, double a20, double sd,
double *ab0_samp, long *idum, int n11, int n10, int n01, int n00) {
double p,cusum[100],w[100],k[100],wi[200];
int i,j;

p=ran2(idum);
cusum[0]=0; cusum[99]=1;
for (i=1;i<99;i++) {
k[i]=gasdev(a20,sd,idum);
w[i]=0;

for (j=0;j<200;j++) {
if (*(y+j)==1) wi[j]=log(1*1);
else {
if (j<n11 || (j>=n11+n10 && j<(n11+n10)+n01))
wi[j]=log(1*(1-pnorm(k[i]+*(ab0_samp+3*ii+2)*
*(T+(ii-1)*400+j*2+1))));
else wi[j]=log(1*pnorm(k[i]+*(ab0_samp+3*ii+2)*
*(T+(ii-1)*400+j*2+1)));
}
w[i] += wi[j];
cusum[i]=cusum[i-1]+exp(w[i]);
}

for (i=1;i<100;i++) {
if (i<99) cusum[i] /= cusum[98];
if ((p>cusum[i-1]) && (p<=cusum[i])) {
*(ab0_samp+3*ii+1)=k[i];
break;
}
}
}
```

# REFERENCES


Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88:669–79.

Bailey, J. W. (1989). A serological test for the diagnosis of Strongyloides antibodies in ex-Far East prisoners of war. *Annals of Tropical Medicine and Parasitology*, 83:241–7.

Becker, M. P. (1994). Analysis of cross-classifications of counts using models for marginal distributions: An application to trends in attitudes on legalized abortion. *Sociological Methodology*, 24:229–65.

Begg, C. B. (1987). Biases in the assessment of diagnostic tests. *Statistics in Medicine*, 6:411–32.

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.

Brenner, H. (1996). How independent are multiple 'independent' diagnostic classifications? *Statistics in Medicine*, 15:1377–86.

Carlin, B. D. and Louis, T. A. (1996). *Bayes and Empirical Bayes methods for data analysis*. Chapman & Hall, London.

Carlin, B. D. and Polson, N. (1991). Inference of nonconjugate Bayesian models using Gibbs sampling. *Canadian Journal of Statistics*, 19:399–405.

Caroll, S. M., Karthigasu, K. T., and Grove, D. I. (1981). Serodiagnosis of human strongyloidiasis by an enzyme-linked immunosorbent assay. *Trans R Soc Trop Med Hyg*, 75:706–9.

Chaloner, K. M. (1996). Estimation of prior distributions. In Berry, D. A. and Stangl, D. K., editors, *Bayesian Biostatistics*, chapter 4. Marcel Dekker, New York, first edition.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90:1313–21.

Cowles, M. K. and Carlin, B. P. (1996). Markov Chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91 (434):883–904.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38.

Douce, R. W., Brown, A. F., and Khamboonruang, C. (1987). Seroepidemiology of strongyloidiasis in a Thai village. *International Journal for Parasitology*, 17:1343–8.

Espeland, M. A. and Handelman, S. L. (1989). Using latent class models to characterize and assess relative error in discrete measurements. *Biometrics*, 45:587–99.

Feinstein, A. R. (1977). Clinical biostatistics XXXIX: The haze of Bayes, the aerial palaces of decision analysis. *Clinical Pharmacology and Therapeutics*, 21:482–496.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76:378–82.

Formann, A. K. (1994). Measurement error in caries diagnosis: Some further latent class models. *Biometrics*, 50:865–71.

Fryback, D. G. (1978). Bayes' theorem and conditional nonindependence of data in medical diagnosis. *Computers and Biomedical Research*, 11:423–34.

Gam, A. A., Neva, F. A., and Krotoski, W. A. (1987). Comparative sensitivity and specificity of ELISA and 1HA for serodiagnosis of strongyloidiasis with larval antigens. *American Journal of Tropical Medical Hygiene*, 37:157–61.

Gastwirth, J. L., Johnson, W. O., and Reneau, D. M. (1991). Bayesian analysis of screening data: Application to AIDS in blood donors. *The Canadian Journal of Statistics*, 19:135–50.

Gatsonis, C. A. (1995). Random-effects models for diagnostic accuracy data. *Academic Radiology*, 2 (Supplement 1):S14–21.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman & Hall, first edition.

Gelman, A. and Meng, X. L. (1996). Model checking and model improvement. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in practice*, pages 189–98. Chapman & Hall, London, first edition.

Gelman, A. and Rubin, D. B. (1992). Inferences from iterative simulation and multiple sequences (with discussion). *Statistical Science*, 7:457–511.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–41.

Genta, R. M. (1988). Predictive value of an enzyme-linked immunosorbent assay (ELISA) for the serodiagnosis of Strongyloidiasis. *American Journal of Clinical Pathology*, 89:391–4.

Genta, R. M. (1989). Global prevalence of strongyloidiasis: critical review with epidemiologic insights into the prevention of disseminated disease. *Review of Infectious Disease*, 11:755–67.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors (1996). *Markov Chain Monte Carlo in practice.* Chapman & Hall, London, first edition.

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61 (2):215–231.

Hagenaars, J. A. (1988). Latent structure models with direct effects between indicators. *Sociological Methods & Research*, 16:379–405.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109.

Hilden, J. (1984). Statistical diagnosis based on conditional independence does not require it. *Computers in Biology and Medicine*, 14 (4):429–435.

Ishwaran, H. and Gatsonis, C. (1997). A general class of hierarchical ordinal regression models with applications to correlated ROC analysis. *To be published.*

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous univariate distributions.* Wiley & Sons, New York.

Joseph, L., Gyorkos, T., and Coupal, L. (1995). Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology*, 141(3):263–272.

Joseph, L. and Gyorkos, T. W. (1996). Inferences for likelihood ratios in the absence of a 'gold standard'. *Medical Decision Making*, 16:412–7.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90:773–95.

Kutner, M. H., Nachtscheim, C. J., Wasserman, W., and Neter, J. (1996). *Applied Linear Statistical Models.* Irwin, Chicago, third edition.

Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33:159–74.

Lazarsfeld, P. F. and Henry, N. W. (1968). *Latent Structure Analysis.* Houghton Mifflin, Boston.

Ledley, L. B. and Lusted, R. S. (1959). Reasoning foundation of medical diagnosis: Symbolic logic, probability and value theory aid our understanding of how phycisians reason. *Science*, 130 (9):26–275.

Lincoln, T. L. and Parker, R. D. (1967). Medical diagnosis using Bayes theorem. *Health Services Research*, 2:34–45.

Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 44:226–233.

Lu, K. H. (1968). A critical evaluation of diagnostic errors, true increment and examiner's accuracy in caries experience assessment by a probabilistic model. *Archives of Oral Biology*, 13:1133–47.

Matchar, D. B., Simel, D. L., Geweke, J. F., and Feussner, J. R. (1990). A Bayesian method for evaluating medical test operating characteristics when some patient's conditions fail to be diagnosed by the reference standard. *Medical Decision Making*, 10:102–11.

McCutcheon, A. (1990). *Latent Class Analysis*. Sage Publications, Beverly Hills, California.

McIsaac, D. J., White, D., Tannenbaum, D., and Low, D. E. (1998). A clinical score to reduce unnecessary antibiotic use in patients with sore throat. *Canadian Medical Association Journal*, 158:75–83.

Neath, A. and Samaniego, F. J. (1997). On the efficacy of Bayesian inference for nonidentifiable models. *American Statistician*, 51:225–234.

Newton, M. and Raftery, A. (1994). Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society, Series B*, 56:3–48.

Nutman, T. B., Ottesen, E. A., Ieng, S., Samuels, J., Kimball, E., Lutkoski, M., Zierdt, W. S., Gam, A. A., and Neva, F. A. (1987). Eosinophilia in South East Asian refugees: evaluation at a referral center. *Journal of Infectious Diseases*, 155:309–13.

Pelletier, L. L. J., Baker, C. B., and Gam, A. A. (1988). Diagnosis and evaluation of treatment of chronic strongyloidiasis in ex-prisoners of war. *Journal of Infectious Diseases*, 157:573–6.

Peng, F. and Hall, J. W. (1996). Bayesian analysis of ROC curves using Markov-chain Monte Carlo methods. *Medical Decision Making*, 16:404–11.

Qu, Y., Tan, M., and Kutner, M. H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, 52:797–810.

Raftery, A. E. and Lewis, S. M. (1992). How many iterations in the Gibbs sampler? In Bernardo, J., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics 4*, pages 765–76. Oxford University Press, Oxford, first edition.

Ransohoff, D. F. and Lang, C. A. (1997). Screening for colorectal cancer with the fecal occult blood test: A background paper. *Annals of Internal Medicine*, 126:811–22.

Roberts, G. O. (1996). Model checking and model improvement. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in practice*, pages 45–57. Chapman & Hall, London, first edition.

Rubin, D. (1988). Using the SIR algorithm to simulate posterior distributions. *Bayesian Statistics*, 3:395–402.

Russek, E., Kronmal, R. A., and Fisher, L. D. (1983). The effect of assuming independence in applying Bayes theorem to risk estimation and classification in diagnosis. *Computers and Biomedical Research*, 16:537–552.

Shapiro, S., Goldberg, J. D., and Hutchison, G. B. (1974). Lead time in breast cancer detection and implications for periodicity of screening. *American Journal of Epidemiology*, 100:357–66.

Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears: A sampling-resampling perspective. *American Statistician*, 46:84–88.

Spiegelhalter, D. J., Freedman, L. S., and Parmar, M. K. B. (1994). Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society, Series A*, 157:357–416.

Tanner, M. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82 (398):528–50.

Torrance-Rynard, V. L. and Walter, S. D. (1997). Effects of dependent errors in the assessment of diagnostic test performance. *Statistics in Medicine*, 16:2157–75.

Uebersax, J. S. and Grove, W. M. (1990). Latent class analysis of diagnostic agreement. *Statistics in Medicine*, 9:559–72.

Vacek, P. M. (1985). The effect of conditional dependence on the evaluation of diagnostic tests. *Biometrics*, 41:959–68.

Walter, S. D. and Irwig, L. M. (1988). Estimation of error rates, disease prevalence, and relative risk misclassified data: A review. *Journal of Clinical Epidemiology*, 41:923–937.

Weinstein, M., Fineberg, H. V., and Elstein, A. S. (1980). *Clinical Decision Analysis*. Saunders, Philadelphia.

Wolfson, L. (1995). *Elicitation of prior distributions*. PhD thesis, Carnegie Mellon University.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm. *Annals of Statistics*, 11:95–103.

Yang, I. and Becker, M. P. (1997). Latent variable modeling of diagnostic accuracy. *Biometrics*, 53:948–958.