BAYESIAN INFERENCE FOR MATCHED CASE-CONTROL STUDIES

By MALAY GHOSH
University of Florida, USA
and
MING-HUI CHEN
University of Connecticut, USA

SUMMARY. The paper introduces general Bayesian inference procedures for the analysis of matched case-control data. The general results hold for a wide class of likelihoods and priors. A real life example is considered to illustrate the methodology. Also, for the logistic regression model, we have shown how Bayesian inference based on the unconditional likelihood can differ substantially from one based on the conditional likelihood.

1. Introduction

Case-control studies occupy a central place in epidemiological research. The basic idea here is the comparison of a group (the “cases”) having the outcome of interest (normally some disease) with respect to a control group (those without disease) regarding one or more characteristics (exposure to certain external events). One of the early examples, which sparked a lot of debate in the fifties between Fisher and Cornfield, is the comparison of persons affected with lung cancer (cases) with persons not affected (controls) in order to find out the effect of smoking (exposure) on lung cancer. Over the years, other important real life examples have emerged, and there is a continuing development of new statistical methodology to address these problems. A very succinct description of many of the important statistical tools for solving case-control problems is given in Breslow and Day (1980). A more recent review article on this subject is due to Breslow (1996) which includes many of the important contributions in this area with an extensive reference list.

Paper received November 2001; revised August 2002.


Keywords and phrases. Likelihood; conditional; unconditional; logit link; probit link; complementary log link; gall-bladder disease; hypertension.
The traditional approach towards the analysis of case-control studies is based on grouping or cross-classification of the data. This grouping is usually done on the basis of several confounding variables (such as age, sex, etc.). However, if cases and controls are selected in advance, and stratification takes place only at the analysis stage, it is possible to have a gross imbalance between cases and controls. For example, in the Ille-et-Velain study of oesophageal cancer discussed in Chapter IV of Breslow and Day (1980), there is one stratum with 1 case and 115 controls. To avoid such inefficiencies and wastage of data, often stratification is introduced at the design stage, allowing a constant ratio of controls to cases in each stratum. Thus each case is individually matched to one or more controls chosen to have similar values for some of the important confounding variables. A second advantage of matching at the design stage is that some of the exact inference procedures which might not be feasible under general stratification, seem possible for matched data.

Although there are numerous statistical papers for matched case-control studies, most of this research has been and continues to be frequentist. Bayesian methods addressing these problems seem to be sparse, and indeed there is a long-awaited need for the development of a unified Bayesian methodology for matched case-control problems.

The present paper intends to develop general Bayesian inferential techniques for matched case-control problems in the presence of one or more exposure variables. The model considered turns out to be more general than that of Zelen and Parker (1986) who assumed a constant stratum effect. Also, unlike Diggle, Morris, and Wakefield (2000), our analysis is based on an unconditional likelihood rather than a conditional likelihood after elimination of nuisance parameters. It may be noted also that any methodology based on the conditional likelihood seems applicable only for the logit link. The general Bayesian methodology based on the full likelihood as proposed in this paper works beyond the logit link. For example, our procedure works for the probit and the complementary log links as well as some new skewed links as proposed in Chen, Dey and Shao (1999). We will also find that a Bayesian procedure based on a full likelihood often provides an answer substantially different from the conditional likelihood.

The outline of the remaining sections is as follows. In Section 2, we provide a general Bayesian approach for the analysis of matched case-control data with one exposure variable, Section 3 discusses the Markov chain Monte Carlo (MCMC) approach for the implementation of the Bayesian procedure and illustrates the same with a dataset studying the effect of gall-bladder disease on the risk of endometrial cancer. Section 4 discusses possible exten-
sions of the results of Section 2 in the presence of multiple binary exposure variables, and illustrates the methodology for the same dataset of Section 3 when both gall-bladder disease and hypertension are considered as exposure variables. Finally, Section 5 contains some concluding remarks.

2. Bayesian Inference

In this section, we consider the case of a single binary exposure variable. For the $i$th stratum, let $X_i = 1$ ($0$) if a person is exposed (unexposed); $Y_i = 1$ ($0$) if the person is a case (control). Following Cox (1970), the joint probability function (pf) of $Y_i$ and $X_i$ is bivariate Bernoulli and is given by

$$f(y_i, x_i) = \frac{\exp(\eta_i y_i + \gamma_i x_i + \beta x_i y_i)}{1 + \exp(\eta_i) + \exp(\gamma_i) + \exp(\eta_i + \gamma_i + \beta)}.$$ (2.1)

Then the marginal pf of $Y_i$ is

$$f(y_i) = \frac{\exp(\eta_i y_i)[1 + \exp(\gamma_i + \beta y_i)]}{1 + \exp(\eta_i) + \exp(\gamma_i) + \exp(\eta_i + \gamma_i + \beta)}.$$ (2.2)

Accordingly, the conditional pf $f(x_i|y_i)$ is

$$f(x_i|y_i) = \frac{\exp[(\gamma_i + \beta y_i)x_i]}{1 + \exp(\gamma_i + \beta y_i)}.$$ (2.3)

It is easy to check that

$$\exp(\beta) = \frac{f(1,1)f(0,0)}{f(0,1)f(1,0)},$$

so that $\beta$ can be interpreted as the common log-odds ratio across all the strata. With one exposure variable, our goal is inference for $\beta$.

Case-control studies are retrospective studies with data generated from the pdf’s $f(x_i|y_i)$. Beginning with Anderson (1972), several authors (e.g., Prentice and Pyke, 1979) have demonstrated how a prospective analysis, i.e., those based on $f(y_i|x_i)$’s is possible even in case-control problems. But that requires the marginal pf’s of the $y_i$’s to be free from the parameters, which is not the case here.

Zelen and Parker (1986) considered a special case of (2.2) where the $\gamma_i$ and the $\eta_i$ are the same across all strata. The $\eta_i$ do not affect our inference based on the pf’s given in (2.2). However, we generalize the method of Zelen and Parker (1986) by allowing variable stratum effects $\eta_i$.

Suppose now we consider the situation when each stratum has $n_1$ ($\geq 1$) cases and $n_0$ ($\geq 1$) controls. (In many epidemiological applications, $n_1 = 1$ and $n_0 \geq 1$.) Then the $2 \times 2$ table for $i$th stratum has the form
<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>(n_{i11})</td>
<td>(n_{i10})</td>
<td>(n_i)</td>
</tr>
<tr>
<td>Unexposed</td>
<td>(n_{i01})</td>
<td>(n_{i00})</td>
<td>(n - n_i)</td>
</tr>
<tr>
<td>Total</td>
<td>(n_1)</td>
<td>(n_0)</td>
<td>(n)</td>
</tr>
</tbody>
</table>

We assume that there are \(I\) strata. Let

\[
p_{ijk} = P(X_i = j | Y_i = k).
\]

Then, from (2.3),

\[
p_{i11} = 1 - p_{i01} = \frac{\exp(\gamma_i + \beta)}{1 + \exp(\gamma_i + \beta)}, \quad i = 1, 2, \ldots, I, \tag{2.4}
\]

and

\[
p_{i10} = 1 - p_{i00} = \frac{\exp(\gamma_i)}{1 + \exp(\gamma_i)}, \quad i = 1, 2, \ldots, I. \tag{2.5}
\]

Let \(\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_I)\). Based on the conditional pf’s \(f(x_i|y_i)\), the likelihood function is given by

\[
L(\beta, \gamma) = \prod_{i=1}^{I} \prod_{j=0}^{1} \prod_{k=0}^{1} p_{ijk}^{n_{ijk}}
\]

\[
= \prod_{i=1}^{I} \left[ \left\{ \frac{\exp(\gamma_i + \beta)}{1 + \exp(\gamma_i + \beta)} \right\}^{n_{i11}} \left\{ \frac{1}{1 + \exp(\gamma_i + \beta)} \right\}^{n_{i01}} \times \left\{ \frac{\exp(\gamma_i)}{1 + \exp(\gamma_i)} \right\}^{n_{i10}} \left\{ \frac{1}{1 + \exp(\gamma_i)} \right\}^{n_{i00}} \right]. \tag{2.6}
\]

The above likelihood is based on the assumption of a logistic distribution of the \(f(x_i|y_i)\) as given in (2.3). A more general likelihood is of the form

\[
L(\beta, \gamma) = \prod_{i=1}^{I} \left[ F^{n_{i11}}(\gamma_i + \beta) \bar{F}^{n_{i01}}(\gamma_i + \beta) F^{n_{i10}}(\gamma_i) \bar{F}^{n_{i00}}(\gamma_i) \right], \tag{2.7}
\]

where \(F\) is a distribution function and \(\bar{F} = 1 - F\).

We consider priors of the form

\[
\pi(\beta, \gamma) = h(\beta) \prod_{i=1}^{I} g(\gamma_i), \tag{2.8}
\]

where the pdf \(g\) is proper, but the pdf \(h\) may be proper or improper. The joint posterior of \(\beta\) and \(\gamma\) is denoted by \(\pi(\beta, \gamma|\text{data})\).
The selection of priors of this form possibly deserves a comment. In the extreme situation when one puts a flat prior on both $\beta$ and $\gamma$, the posterior modes become the MLE’s of these parameters. As is well-known (see e.g. Andersen, 1973) this leads to an inconsistent estimator of $\beta$, an example of the Neyman-Scott phenomenon where the number of nuisance parameters grows in direct proportion to the sample size. Moreover, a flat prior on both $\beta$ and $\gamma$ can possibly lead to an improper posterior, especially in situations where each stratum contains one case and one control. The iid assumption for the components of $\gamma$ may not be appropriate under all circumstances, but is often quite satisfactory, such as in multicenter clinical trials, when the different clinics are used as strata.

Let $s_{m,k} =$ # of $i$ for which $n_{i1} = m$ and $n_{i0} = k$. We first prove a result providing easily verifiable sufficient conditions to check the propriety of the posterior $\pi(\beta, \gamma|\text{data})$.

**Theorem 2.1** Suppose (i) $g$ is a proper pdf, (ii) for $\beta > 0$, $F(\gamma + \beta)\bar{F}(\gamma) \leq u_1(\beta)$, where

$$\int_0^\infty [u_1(\beta)]^{\sum_{m=0}^{n_1} \sum_{k=0}^{n_0} s_{m,k} h(\beta)} d\beta < \infty,$$

and (iii) for $\beta < 0$, $F(\gamma + \beta)\bar{F}(\gamma) \leq u_2(\beta)$, where

$$\int_{-\infty}^{0} [u_2(\beta)]^{\sum_{m=1}^{n_1} \sum_{k=0}^{n_0-1} s_{m,k} h(\beta)} d\beta < \infty.$$

Then, the posterior $\pi(\beta, \gamma|\text{data})$ is proper.

**Proof.** Writing $K$ as the proportionality constant for the posterior,

$$\pi(\beta, \gamma|\text{data}) = KL(\beta, \gamma)\pi(\beta, \gamma).$$

Now, integrating with respect to $\gamma$,

$$\pi(\beta|\text{data}) = Kh(\beta) \prod_{m=0}^{n_1} \prod_{k=0}^{n_0} \int_{-\infty}^{\infty} F^m(\gamma + \beta)\bar{F}^{n_1-m}(\gamma + \beta)F^k(\gamma)\bar{F}^{n_0-k}(\gamma)g(\gamma) d\gamma.]^{s_{m,k}} \quad (2.9)$$

Since each integral is less than or equal to $\int_{-\infty}^{\infty} g(\gamma) d\gamma = 1$, by condition (ii),
for $\beta > 0$,
\[
\pi(\beta|\text{data}) 
\leq K_h(\beta) \prod_{m=0}^{n_1-1} \prod_{k=1}^{n_0} \left[ \int_{-\infty}^{\infty} F^m(\gamma + \beta) F^{n_1-m}(\gamma + \beta) F^k(\gamma) F^{n_0-k}(\gamma) g(\gamma) d\gamma \right]^{s_{m,k}} 
\leq K_h(\beta) \prod_{m=0}^{n_1-1} \prod_{k=1}^{n_0} \left[ \int_{-\infty}^{\infty} \bar{F}(\gamma + \beta) F(\gamma) g(\gamma) d\gamma \right]^{s_{m,k}} 
\leq K_h(\beta) [u_1(\beta)]^{\sum_{m=0}^{n_1-1} \sum_{k=1}^{n_0} s_{m,k}}. 
\tag{2.10}
\]
Similarly, for $\beta < 0$, by condition (iii),
\[
\pi(\beta|\text{data}) \leq K_h(\beta) [u_2(\beta)]^{\sum_{m=1}^{n_1} \sum_{k=0}^{n_0-1} s_{m,k}}. 
\tag{2.11}
\]
The theorem follows now from (2.10) and (2.11) utilizing conditions (ii) and (iii). \hfill \Box

We now verify the conditions of Theorem 2.2 first for the three important special cases when $F$ is standard logistic, standard normal and the extreme value distribution functions. In the more familiar terminology, the corresponding links (i.e., $F^{-1}$) are referred to as the logit, probit and complementary log links.

First, when $F$ is standard logistic, for $\beta > 0$,
\[
\bar{F}(\gamma + \beta) F(\gamma) = [1 + \exp(\gamma + \beta)]^{-1} \exp(\gamma) [1 + \exp(\gamma)]^{-1} 
\leq \exp(-\gamma - \beta) \exp(\gamma) = \exp(-\beta), 
\tag{2.12}
\]
while for $\beta < 0$,
\[
F(\gamma + \beta) \bar{F}(\gamma) = \exp(\gamma + \beta) [1 + \exp(\gamma + \beta)]^{-1} [1 + \exp(\gamma)]^{-1} 
\leq \exp(\gamma + \beta) \exp(-\gamma) = \exp(\beta). 
\tag{2.13}
\]
Hence, when $g$ is a proper pdf, writing
\[
u = \min \left( \sum_{m=0}^{n_1-1} \sum_{k=1}^{n_0} s_{m,k}, \sum_{m=1}^{n_1} \sum_{k=0}^{n_0-1} s_{m,k} \right),
\]
conditions (ii) and (iii) of Theorem 2.2 hold if
\[
\int_{-\infty}^{\infty} \exp(-u|\beta|) h(\beta) d\beta < \infty. 
\tag{2.14}
\]
In particular, when \( \sum_{m=0}^{n_1} s_{m,k} \geq 1 \) and \( \sum_{m=1}^{n_0} s_{m,k} \geq 1 \), a simple sufficient condition guaranteeing (??) is \( \int_{-\infty}^{\infty} \exp(-|\beta|) h(\beta) d\beta < \infty \).

Next, for the probit link, writing \( \Phi \) as the standard normal df, \( F \equiv \Phi \).

Then, by the inequality \( \exp(-x^2/2) \leq 2[\exp(x) + \exp(-x)]^{-1} \) for all real \( x \),

one gets for \( \beta > 0 \),

\[
\Phi(\gamma + \beta) \Phi(\gamma) \leq \frac{1}{2\pi} \left[ \int_{-\infty}^{\infty} \exp(-x) dx \right] \left[ \int_{-\infty}^{\infty} \exp(x) dx \right] = \frac{2}{\pi} \exp(-\gamma^2) \exp(\gamma) = \frac{2}{\pi} \exp(-\beta). \quad (2.15)
\]

Again, for \( \beta < 0 \), similarly one has

\[
\Phi(\gamma + \beta) \Phi(\gamma) \leq \frac{1}{2\pi} \exp(\gamma) \exp(-\gamma) = \frac{1}{2\pi} \exp(\beta). \quad (2.16)
\]

Then, by (??) and (??), (ii) and (iii) hold under the same condition as given in (??).

For the complementary log-link, \( F(x) = \exp[-\exp(-x)] \), \( x \) real. This is the so-called extreme-value distribution. In this case

\[
\bar{F}(x) = 1 - \exp[-\exp(-x)] \leq \exp(-x).
\]

Hence, for \( \beta > 0 \), by the inequality, \( \exp(-x) \geq 1 - x \) for all real \( x \),

\[
\bar{F}(\gamma + \beta) \bar{F}(\gamma) \leq \exp(-\gamma - \beta) \exp[-\exp(-\gamma)] \leq \exp(-\gamma - \beta) \exp[-(1 - \gamma)] = \exp(-\beta - 1). \quad (2.17)
\]

Also, for \( \beta < 0 \),

\[
\bar{F}(\gamma + \beta) \bar{F}(\gamma) \leq \exp[-1 + (\gamma + \beta)] \exp(-\gamma) = \exp(\beta - 1). \quad (2.18)
\]

Once again, by (??) and (??), conditions (ii) and (iii) of Theorem ?? hold under (??).

The extreme-value distribution is one of many examples of asymmetric \( F \). Chen, Dey and Shao (1999) have provided a general method for generating a class of asymmetric \( F \)’s. The extreme-value distribution, however, does not belong to this class. According to the formulation of Chen et al., \( F \) is the distribution function of \( \delta Z_1 + Z_2 \), where \( Z_2 \) has a symmetric distribution,
and $Z_1$ has an asymmetric distribution. The parameter $\delta$ can be viewed as the skewness parameter. Thus, if $Z_1$ has distribution function $G$ and $Z_2$ has distribution function $H$, then $F(x)$ can be written as

$$F(x) = \int_0^\infty H(x-z\delta)dG(z).$$

For simplicity, we consider only the case when $H$ is logistic. Suppose $\int_0^\infty \exp(tz)dG(z) < \infty$ for all real $t$. Since $\bar{F}(x) = \int_0^\infty \bar{H}(x-z\delta)dG(z)$, where $\bar{H} \equiv 1 - H$, for a logistic $H$, for $\beta > 0$,

$$\bar{F}(\gamma + \beta) F(\gamma) \leq \int_0^\infty \exp(-\gamma - \beta + z\delta)dG(z) \int_0^\infty \exp(\gamma - z\delta)dG(z) \leq \exp(-\beta) \int_0^\infty \exp(z\delta)dG(z) \int_0^\infty \exp(-z\delta)dG(z),$$

(2.19)

and the multiplier of $\exp(-\beta)$ is finite by our assumption. Under the same assumption, for $\beta < 0$, $F(\gamma + \beta) \bar{F}(\gamma)$ is less than or equal to a constant multiple of $\exp(\beta)$. Accordingly, conditions (ii) and (iii) of Theorem 1 hold under (??) as in the case of logistic $F$. With the same assumption on $G$, it can be shown that when $H$ is the standard normal df, the conditions (ii) and (iii) of Theorem ?? hold under (??).

**Remark 2.1.** When there is 1 case and 1 control per stratum, that is, $n_0 = n_1 = 1$, Ghosh, Chen, Ghosh, and Agresti (2000) proved the propriety of $\pi(\beta, \gamma|\text{data})$ for logit, probit and complementary log links when $s_1, 0 \geq 1$, $s_0, 1 \geq 1$, $g(\gamma)$ is normal and $h(\beta) \propto 1$. Even in these special cases, as verified already, we have a more general class of priors under which $\pi(\beta, \gamma|\text{data})$ is proper. Also, the present work generalizes the work of Altham (1971).

Next we investigate how the Bayesian methodology can be used when one starts with a conditional likelihood involving $\beta$ only rather than the unconditional likelihood for the logistic model. Since $n_{i11}$ and $n_{i10}$ constitute the minimal sufficient statistic for $(\beta, \gamma_i)$ in the $i$th stratum, we begin with the joint pf

$$P(n_{i11} = m, n_{i10} = k) = \binom{n_1}{m} \binom{n_0}{k} \frac{\exp[\gamma_i(m + k)] \exp(m^2)}{[1 + \exp(\gamma_i + \beta)]^{n_1}[1 + \exp(\gamma_i)]^{n_0}},$$

(2.20)

for $m = 0, 1, \ldots, n_1$, and $k = 0, 1, \ldots, n_0$. Thus, $n_{i11} + n_{i10}$ is sufficient for the nuisance parameter $\gamma_i$, and one works with the conditional pf of $n_{i11}$
given \( n_{i11} + n_{i10} \). After some calculations, from (2.21), it follows that

\[
P(n_{i11} = m | n_{i11} + n_{i10} = t_i) = \frac{\binom{n_1}{m} \binom{n_0}{t_i-m} \exp(m \beta)}{\sum_{x=\max(0,t_i-n_0)}^{\min(n_1,t_i)} \binom{n_i}{x} \binom{n_0}{t_i-x} \exp(x \beta)},
\]

for \( m = \max(0, t_i - n_0), \ldots, \min(n_1, t_i) \), which is the well-known non-central hypergeometric pf. Consequently the conditional likelihood given \( t = (t_1, t_2, \ldots, t_I) \) is

\[
L_C(\beta) = \prod_{i=1}^{I} \frac{\binom{n_1}{n_{i11}} \binom{n_0}{t_i-n_{i11}} \exp(n_{i11} \beta)}{\sum_{x=\max(0,t_i-n_0)}^{\min(n_1,t_i)} \binom{n_i}{x} \binom{n_0}{t_i-x} \exp(x \beta)}.
\]

If now one assigns a prior \( \pi(\beta) \) to \( \beta \), the posterior

\[
\pi_C(\beta | \text{data}) \propto L_C(\beta) \pi(\beta).
\]

Bayesian approaches based on conditional likelihoods have been considered in Liao (1999) for combining multiple \( 2 \times 2 \) tables, and in Diggle, Morris, and Wakefield (2000) for the analysis of case-control data arising in spatial epidemiology.

The expression given in (2.22) simplifies considerably when \( n_1 = 1 \), that is each stratum consists exactly of 1 case. Then,

\[
L_C(\beta) = \prod_{i=1}^{I} \frac{\binom{n_0}{t_i-n_{i11}} \exp(n_{i11} \beta)}{\sum_{x=\max(0,t_i-n_0)}^{\min(n_1,t_i)} \binom{n_i}{x} \binom{n_0}{t_i-x} \exp(x \beta)}.
\]

If now \( s_{0,k} \) denotes the number of strata for which \( n_{i11} = 0 \) and \( n_{i10} = k \), while \( s_{1,k-1} \) denotes the number of strata for which \( n_{i11} = 1 \) and \( n_{i10} = k-1 \) \((k = 1, 2, \ldots, n_0)\), then \( L_C(\beta) \) given in (2.22) simplifies to

\[
L_C(\beta) = \prod_{k=1}^{\min(n_0, n_1)} \left[ \frac{k \exp(\beta)}{k \exp(\beta) + n_0 - k + 1} \right]^{s_{1,k-1}} \left[ \frac{n_0 - k + 1}{k \exp(\beta) + n_0 - k + 1} \right]^{s_{0,k}},
\]

an expression given in (5.15) of Breslow and Day (1980) with different notations.

**Remark 2.2.** Although the conditional likelihood \( L_C(\beta) \) given in (2.22) does not involve any nuisance parameters, and simplifies the Bayesian analysis considerably, observations from those strata \( i \) for which \( t_i = 0 \), i.e. \( n_{i11} = n_{i10} = 0 \) and those for which \( t_i = n_0 + n_1 \), i.e., \( n_{i11} = n_1 \) and \( n_{i10} = n_0 \) do not contribute at all to this conditional likelihood. This can potentially
provide substantially different inference from one based on a full likelihood. This will be evidenced from the following theorem and its corollary. The theorem is proved for a general distribution function $F$.

**Theorem 2.2** Consider the likelihood given in (??), and the prior given in (??) where $g$ is a proper pdf. Then $\pi(\beta|\text{data})$ as given in (??) is non-decreasing in each $s_{m,k}$ for all real $\beta$.

**Proof.** Since each integral in the right hand side of (??) is less than or equal to $\int_{-\infty}^{\infty} g(\gamma) d\gamma = 1$, the result follows immediately from (2.9). $\square$

Theorem ?? leads to the following corollary.

**Corollary 2.1** Suppose that the conditions of Theorem ?? hold. Then

(i) $P(|\beta| > t|\text{data})$ is non-increasing in each $s_{m,k}$ for all $t > 0$, and

(ii) $E[|\beta| | \text{data})$ is non-increasing in each $s_{m,k}$.

**Proof.** (i) $P(|\beta| > t|\text{data}) = \int_t^{\infty} \pi(\beta|\text{data}) d\beta + \int_{-\infty}^{-t} \pi(\beta|\text{data}) d\beta = \int_{t}^{\infty} [\pi(\beta|\text{data}) + \pi(-\beta|\text{data})] d\beta$. Since $\pi(\beta|\text{data})$ is non-increasing in each $s_{m,k}$ for all real $\beta$, the result follows.

(ii) The proof is immediate when one writes $E[|\beta| | \text{data}) = \int_0^{\infty} P(|\beta| > t|\text{data}) dt$ and applies (i). $\square$

**Remark 2.3.** It follows from Theorem ?? and its corollary that the effect of $s_{0,0}$ and $s_{n1,n0}$ can be significant on the posterior of $\beta$ when one or both of these is much larger compared to other $s_{m,k}$’s. On the other hand, as pointed out already, neither $s_{0,0}$ nor $s_{n1,n0}$ appears in $L_C(\beta)$ so that any Bayesian inference based on $L_C(\beta)$ will ignore $s_{0,0}$ and $s_{n1,n0}$. Indeed, since $s_{m,k}$ ($m = 0, 1, \ldots, n1; k = 0, 1, \ldots, n0$) is jointly sufficient for the posterior of $\beta$ based on the full likelihood, discarding $s_{0,0}$ and $s_{n1,n0}$ can lead to potential inefficiency. Particularly, when $n_1 = n_0 = 1$, and $s_{0,0}$ and/or $s_{1,1}$ is substantially larger than $s_{1,0}$ and $s_{0,1}$, then the loss of efficiency could be significant. An example to this effect appears in Ghosh et al., (2000). The work involves the analysis of some sociological data, where $s_{1,1} = 292$, $s_{0,0} = 9$, $s_{1,0} = 25$ and $s_{0,1} = 14$. In this case, a Bayesian analysis based on a full likelihood showed nearly 20% reduction in the posterior mean of $|\beta|$ when compared to the conditional MLE.

Finally, the conditional likelihood approach does not work except for logistic $F$. Any inference based on a full likelihood goes beyond that. It may be added here that the standard frequentist method in such cases is to
find the marginal likelihood of $\beta$ which needs assigning some distribution on the $\gamma_i$, and then integrating them out. Clearly, this is a pseudo-Bayesian approach, and instead, we prefer to carry out a full Bayesian analysis.

It is possible to extend the class of priors proposed in (??), where instead of iid $\gamma_i$’s, one considers exchangeable $\gamma_i$’s. From De Finetti’s representation theorem, it is then possible to express the joint pdf of the $\gamma_i$’s as

$$g(\gamma) = \int_{-\infty}^{\infty} \left[ \prod_{i=1}^{f} h_\eta(\gamma_i) \right] w(\eta) d\eta,$$

where $h_\eta$ and $w$ are proper pdf’s. Now the posterior of $\beta$ is

$$\pi(\beta|\text{data}) = Kh(\beta)\int_{-\infty}^{\infty} \prod_{m=0}^{n_1} \prod_{k=1}^{n_0} \left\{ \int_{-\infty}^{\infty} F^m(\gamma + \beta) F^{n_1-m}(\gamma + \beta) \times F^k(\gamma) F^{n_0-k}(\gamma) h_\eta(\gamma) d\gamma \right\} s_{m,k} w(\eta) d\eta. \tag{2.25}$$

Since $h_\eta$ and $w$ are proper pdf’s, one gets a result similar to that of Theorem ?? under conditions (ii) and (iii).

A very important application of the above result is the situation when $\gamma$ has a multivariate $t$-distribution with, say, $\nu$ degrees of freedom. In this case, conditional on some precision parameter $r$, the $\gamma_i$ are i.i.d $N(0, r^{-1})$. Also, $r$ is assumed to have the gamma pdf

$$\pi(r) \propto \exp\left( -\frac{1}{2} \nu r \right) r^{\frac{1}{2} \nu - 1}. \tag{2.26}$$

This is an example of a prior for $\gamma$ which is a scale-mixture of normals. More generally, one can assign any specified pdf $w(r)$ for $r$, generating in this way a class of scale-mixed normal priors for $\gamma$ (see, for example, Chen and Dey 1998).

Diggle, Morris, and Wakefield (2000) have considered both frequentist and Bayesian analysis of matched case-control data arising in spatial epidemiology. They have argued quite reasonably that in practice, it is not always possible to assign an iid or even exchangeable prior for $\gamma$. Our objective is not to refute this statement, but to point out that when the iid or the exchangeability assumption of the $\gamma_i$’s is reasonable, it is possible to carry out a Bayesian analysis based on a full likelihood which can yield significantly different answer from one based on a conditional likelihood. This
is illustrated in the next section with a real dataset. This section also discuss a general MCMC approach towards implementation of the Bayesian procedure.

3. Numerical Study and An Example

Implementation of the Bayesian procedure described in Section 2 is greatly facilitated by the MCMC numerical integration technique. This requires generating samples from the full conditionals \( \pi(\beta | \gamma, \text{data}) \) and \( \pi(\gamma_i | \beta, \gamma_j \, (j \neq i), \text{data}) \) \((i = 1, 2, \ldots, I)\).

For a specified distribution function \( F \), and the prior \( \pi(\beta, \gamma) \) given in (3.2), these full conditionals are given by

\[
\pi(\gamma_i | \beta, \gamma_j \, (j \neq i), \text{data}) \propto F_{n11}(\gamma_i + \beta) \bar{F}_{n01}(\gamma_i + \beta) F_{n10}(\gamma_i) \bar{F}_{n00}(\gamma_i) g(\gamma_i),
\]

which does not depend on the other \( \gamma_j \). Also,

\[
\pi(\beta | \gamma, \text{data}) \propto I \prod_{i=1}^{I} [F_{n11}(\gamma_i + \beta) \bar{F}_{n01}(\gamma_i + \beta)] h(\beta),
\]

When \( \gamma \) has joint pdf

\[
g(\gamma) \propto \int_{-\infty}^{\infty} \prod_{i=1}^{I} h_{\eta}(\gamma_i) w(\eta) d\eta,
\]

then it is convenient to generate samples from the conditionals

\[
\pi(\gamma_i | \beta, \gamma_j \, (j \neq i), \eta, \text{data}) \propto F_{n11}(\gamma_i + \beta) \bar{F}_{n01}(\gamma_i + \beta) F_{n10}(\gamma_i) \bar{F}_{n00}(\gamma_i) g(\gamma),
\]

\[
\pi(\beta | \gamma, \eta, \text{data}) \propto I \prod_{i=1}^{I} [F_{n11}(\gamma_i + \beta) \bar{F}_{n01}(\gamma_i + \beta)] h(\beta),
\]

and

\[
\pi(\eta | \beta, \gamma, \text{data}) \propto \prod_{i=1}^{I} h_{\eta}(\gamma_i) w(\eta).
\]

It may be noted that the last conditional pdf does not depend on the data.

We consider now an example discussed extensively in Breslow and Day (1980). We focus on the effect of gall-bladder disease on the risk of endometrial cancer. The data which appear in Mack et al. (1976), and also in
Breslow and Day (1980) consist of 63 cases of endometrial cancer occurring among retirement community in Los Angeles, California between 1971 and 1975. Each case was matched to four control women who were alive and living in the community at the time the case was diagnosed, who were born within one year of the case, who had the same marital status, and who had entered the community at approximately the same time.

We use this example mainly to illustrate the results of the previous section, showing in particular that Bayesian inference based on unconditional likelihood can often differ substantially from the one based on conditional likelihood. Here, the “cases” refer to those having endometrial cancer, which “controls” are the ones not having the same. The exposure variable is binary taking values 1 and 0 depending on whether or not the woman had gall-bladder disease. For comparison purposes, we consider only the logit link. Once again $h(\beta) \propto 1$, while several $t$ priors with different degrees of freedom and different scale parameters are chosen as $g$. We tried also some normal priors in this example, but the $t$ priors turn out to be more robust than the normal priors. The generic version of such a $t$ prior is given by

$$g(\gamma) \propto [1 + \gamma^2/(\nu \sigma^2)]^{-\nu+1}.$$ 

For the gall-bladder disease, the $s(m, k)$ values are given in the following table.

<table>
<thead>
<tr>
<th>$m$</th>
<th>$s_{m,0}$</th>
<th>$s_{m,1}$</th>
<th>$s_{m,2}$</th>
<th>$s_{m,3}$</th>
<th>$s_{m,4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>32</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

In this example, the conditional MLE of $\beta$ is given by $\hat{\beta}_{\text{MLE}} = 1.306$, while its Mantel-Haenszel estimate is given by $\hat{\beta}_{\text{MH}} = 1.369$. The posterior mean of $\beta$ based on the conditional likelihood and the flat prior $h(\beta) \propto 1$ turns out to be 1.311 with the associated standard error 0.377. The closeness of the conditional MLE (which is the posterior mode under the said prior) and the posterior mean of $\beta$ is not surprising, since this posterior is near symmetric. However, this analysis ignores $s_{0,0}$ which in this case is more than 50% of the total number of 63 strata.

In contrast, the results of the Bayesian analysis for $\beta$ based on the full likelihood, as provided in the following two tables, make full use of $s_{0,0}$. The two tables correspond to two different values of $\sigma$, the scale parameter.
in the $t$-prior, while each table contains several values of $\nu$, the degrees of freedom. The L measures are also reported in the last column of these two tables. L measure is a Bayesian criterion for model assessment, which is defined as the sum of the posterior predictive variances plus $k$ times the sum of the squared biases, where $0 \leq k \leq 1$ is a weight. A small value of L measure indicates a better model. In our calculation, we used $k = 0.5$ as recommended by Ibrahim, Chen, and Sinha (2001). The detailed formulation and computation of L measure can be found in Gelfand and Ghosh (1998) and Ibrahim, Chen, and Sinha (2001).

Table 2. The posterior means, standard deviations and HPD regions for $\beta$ for the Gall-Bladder disease when $\sigma^2 = 25$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>posterior mean</th>
<th>posterior s.d.</th>
<th>95% HPD interval</th>
<th>L Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.474</td>
<td>0.443</td>
<td>(0.609,2.334)</td>
<td>50.06</td>
</tr>
<tr>
<td>3</td>
<td>1.390</td>
<td>0.435</td>
<td>(0.528,2.238)</td>
<td>51.71</td>
</tr>
<tr>
<td>10</td>
<td>1.351</td>
<td>0.425</td>
<td>(0.517,2.186)</td>
<td>52.55</td>
</tr>
<tr>
<td>20</td>
<td>1.341</td>
<td>0.425</td>
<td>(0.532,2.190)</td>
<td>52.75</td>
</tr>
</tbody>
</table>

Table 3. The posterior means, standard deviations and HPD regions for $\beta$ for the Gall-Bladder disease when $\sigma^2 = 50$

<table>
<thead>
<tr>
<th>$\nu$</th>
<th>posterior mean</th>
<th>posterior s.d.</th>
<th>95% HPD interval</th>
<th>L Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.684</td>
<td>0.462</td>
<td>(0.779,2.590)</td>
<td>46.87</td>
</tr>
<tr>
<td>3</td>
<td>1.623</td>
<td>0.458</td>
<td>(0.716,2.512)</td>
<td>47.84</td>
</tr>
<tr>
<td>10</td>
<td>1.597</td>
<td>0.456</td>
<td>(0.744,2.536)</td>
<td>48.34</td>
</tr>
<tr>
<td>20</td>
<td>1.589</td>
<td>0.449</td>
<td>(0.701,2.459)</td>
<td>48.42</td>
</tr>
</tbody>
</table>

An important point to note here that the posterior mean of $\beta$ based on a full likelihood may be greater than the one based on conditional likelihood. Also, it appears from the above two tables that there is a fair amount of robustness in the Bayesian procedure for fixed $\sigma$ under different choices of $\nu$. The results are also somewhat robust against the different choices of $\sigma$. The important point to note here is that even for moderately large values of $\sigma^2$, the posterior means of $\beta$ based on the full likelihood are greater than the posterior mean based on the conditional likelihood. This is because the conditional likelihood ignores $s_{0,0}$, and, therefore overestimates $\beta$, the difference between the logits of the exposure probabilities for the cases and the controls. In addition, based on the L measure values, the models with large values of $\sigma^2$ and small values of $\nu$ fit the data better.

In the following section, we consider Bayesian inference in the presence of multiple exposure variables.
4. Multiple Binary Exposure Variables

In this section, we consider some extensions of the methods of Section 2 when there are multiple binary exposure variables. For simplicity of notations, we consider the case of only two binary variables, although the results can be extended similarly for more than two binary variables. In this case for the \( i \)th stratum we consider the triplet \((Y_i, X_{i1}, X_{i2})\), where \( Y_i = 1 \) (0) if the person is a case (control); \( X_{i1} = 1 \) (0) if the person is subject to (not subject to) exposure 1; and \( X_{i2} = 1 \) (0) if the person is subject to (not subject to) exposure 2.

For simplicity, we consider only the logistic regression model. Following Cox (1970) and Zhao and Prentice (1990), we begin with the quadratic exponential model, where for the \( i \)th stratum,

\[
f(y_i, x_{i1}, x_{i2}) = D_i^{-1} \exp(\eta_i y_i + \gamma_{i1} x_{i1} + \gamma_{i2} x_{i2} + \beta_{11} x_{i1} y_i + \beta_{22} x_{i2} y_i + \beta_{12} x_{i1} x_{i2}),
\]

where

\[
D_i = 1 + \exp(\eta_i) + \exp(\gamma_{i1}) + \exp(\gamma_{i2}) + \exp(\eta_i + \gamma_{i1} + \beta_{11}) + \exp(\eta_i + \gamma_{i2} + \beta_{22}) + \exp(\gamma_{i1} + \gamma_{i2} + \beta_{12}) + \exp(\eta_i + \gamma_{i1} + \gamma_{i2} + \beta_{11} + \beta_{22} + \beta_{12}).
\]

Then \( Y_i \) has the marginal pf

\[
f(y_i) = D_i^{-1} \exp(\eta_i y_i) [1 + \exp(\gamma_{i1} + \beta_{11} y_i) + \exp(\gamma_{i2} + \beta_{22} y_i) + \exp(\gamma_{i1} + \gamma_{i2} + (\beta_{11} + \beta_{22}) y_i + \beta_{12})].
\]

Accordingly, the joint conditional pf of \( X_{i1} \) and \( X_{i2} \) given \( Y_i = y_i \) is

\[
f(x_{i1}, x_{i2}|y_i) = \frac{\exp(\gamma_{i1} x_{i1} + \gamma_{i2} x_{i2} + \beta_{11} x_{i1} y_i + \beta_{22} x_{i2} y_i + \beta_{12} x_{i1} x_{i2})}{1 + \exp(\gamma_{i1} + \beta_{11} y_i) + \exp(\gamma_{i2} + \beta_{22} y_i) + \exp(\gamma_{i1} + \gamma_{i2} + (\beta_{11} + \beta_{22}) y_i + \beta_{12})}.
\]

Here

\[
\exp(\beta_{11}) = \frac{f(1, 0|1)f(0, 0|0)}{f(1, 0|0)f(0, 0|1)} \quad \text{and} \quad \exp(\beta_{22}) = \frac{f(0, 1|1)f(0, 0|0)}{f(0, 1|0)f(0, 0|1)}
\]

for all the strata, while

\[
\exp(\beta_{12}) = \frac{f(1, 1|1)f(0, 0|1)}{f(1, 0|1)f(0, 1|1)} = \frac{f(1, 1|0)f(0, 0|0)}{f(1, 0|0)f(0, 1|0)}
\]
for all the strata. Thus \( \exp(\beta_{12}) \) is the common odds ratio between the two exposure variables conditional on either a case or a control. It reflects the partial association between the two exposure variables rather than a direct association.

Suppose again we consider the situation where each stratum has \( n_1 (\geq 1) \) cases and \( n_0 (\geq 1) \) controls. Then the \( 4 \times 2 \) table for the \( i \)th stratum takes the form

<table>
<thead>
<tr>
<th></th>
<th>Case</th>
<th>Control</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed/Exposed</td>
<td>( n_{i11} )</td>
<td>( n_{i10} )</td>
<td>( t_{i1} )</td>
</tr>
<tr>
<td>Exposed/Unexposed</td>
<td>( n_{i01} )</td>
<td>( n_{i00} )</td>
<td>( t_{i0} )</td>
</tr>
<tr>
<td>Unexposed/Exposed</td>
<td>( n_{011} )</td>
<td>( n_{010} )</td>
<td>( t_{01} )</td>
</tr>
<tr>
<td>Unexposed/Unexposed</td>
<td>( n_{001} )</td>
<td>( n_{000} )</td>
<td>( t_{00} )</td>
</tr>
</tbody>
</table>

Once again, let \( I \) denote the number of strata. Now, write \( \gamma_1 = (\gamma_{11}, \gamma_{21}, \ldots, \gamma_{11}) \) and \( \gamma_2 = (\gamma_{12}, \gamma_{22}, \ldots, \gamma_{12}) \), the likelihood function based on \( f(x_{i1}, x_{i2}|y_i) \) \((i = 1, 2, \ldots, I)\) is given by

\[
L(\beta_1, \beta_2, \beta_{12}, \gamma_1, \gamma_2) = \prod_{i=1}^{I} L_i(\beta_1, \beta_2, \beta_{12}, \gamma_1, \gamma_2),
\]

where

\[
L_i(\beta_1, \beta_2, \beta_{12}, \gamma_1, \gamma_2)
= \exp \left[ (n_{i111} + n_{i101} + n_{i110} + n_{i010})\gamma_{i1} + (n_{i111} + n_{i011} + n_{i110} + n_{i010})\gamma_{i2}
+ \beta_1(n_{i111} + n_{i101}) + \beta_2(n_{i111} + n_{i011}) + \beta_{12}(n_{i111} + n_{i110}) \right]
\times D_{i1}^{-n_1} D_{i0}^{-n_0}, \tag{4.4}
\]

\[
D_{i1} = 1 + \exp(\gamma_{i1} + \beta_1) + \exp(\gamma_{i2} + \beta_2) + \exp(\gamma_{i1} + \gamma_{i2} + \beta_1 + \beta_2 + \beta_{12}),
\]

and

\[
D_{i0} = 1 + \exp(\gamma_{i1}) + \exp(\gamma_{i2}) + \exp(\gamma_{i1} + \gamma_{i2} + \beta_{12}).
\]

We consider priors of the form

\[
\pi(\beta_1, \beta_2, \beta_{12}, \gamma_1, \gamma_2) \propto h(\beta_1, \beta_2, \beta_{12}) \prod_{i=1}^{I} g(\gamma_{i1}, \gamma_{i2}). \tag{4.5}
\]
The joint posterior of $\beta_1, \beta_2, \beta_{12}, \gamma_1$ and $\gamma_2$ is then given by

$$
\pi(\beta_1, \beta_2, \beta_{12}, \gamma_1, \gamma_2 \mid \text{data}) \propto L(\beta_1, \beta_2, \beta_{12}, \gamma_1, \gamma_2) \times \pi(\beta_1, \beta_2, \beta_{12}, \gamma_1, \gamma_2).
$$

(4.6)

The implementation of the Bayesian procedure once again requires MCMC numerical integration. This requires generating samples from the full conditionals:

$$
\pi(\gamma_{i1} \mid \beta_1, \beta_2, \beta_{12}, \gamma_j, \text{data}) \propto D_{-i1}^{-n_1} D_{0}^{-n_0} \exp\{(n_{i111} + n_{i101} + n_{i110} + n_{i100})\gamma_{i1}\} g(\gamma_{i1}, \gamma_{i2}),
$$

(4.7)

$$
\pi(\gamma_{i2} \mid \beta_1, \beta_2, \beta_{12}, \gamma_j, \text{data}) \propto D_{-i1}^{-n_1} D_{0}^{-n_0} \exp\{(n_{i111} + n_{i011} + n_{i110} + n_{i010})\gamma_{i2}\} g(\gamma_{i1}, \gamma_{i2}),
$$

(4.8)

$$
\pi(\beta_1 \mid \beta_2, \beta_{12}, \gamma_1, \gamma_2, \text{data}) \propto \prod_{i=1}^{l} (D_{-i1}^{-n_1} D_{0}^{-n_0}) \exp\left\{ \beta_1 \sum_{i=1}^{l} (n_{i111} + n_{i101}) \right\},
$$

(4.9)

$$
\pi(\beta_2 \mid \beta_1, \beta_{12}, \gamma_1, \gamma_2, \text{data}) \propto \prod_{i=1}^{l} (D_{-i1}^{-n_1} D_{0}^{-n_0}) \exp\left\{ \beta_2 \sum_{i=1}^{l} (n_{i111} + n_{i011}) \right\},
$$

(4.10)

and

$$
\pi(\beta_{12} \mid \beta_1, \beta_2, \gamma_1, \gamma_2, \text{data}) \propto \prod_{i=1}^{l} (D_{-i1}^{-n_1} D_{0}^{-n_0}) \exp\left\{ \beta_{12} \sum_{i=1}^{l} (n_{i111} + n_{i110}) \right\}.
$$

(4.11)

As before, we observe that when one is interested in inference for $(\beta_1, \beta_2, \beta_{12})$, one approach to eliminate the nuisance parameters $\gamma_i$ ($i = 1, 2, \ldots, I$) is to work with the conditional likelihood of $(\beta_1, \beta_2, \beta_{12})$, where conditioning is based on the sufficient statistics $(n_{i111} + n_{i101} + n_{i110} + n_{i100}, n_{i111} + n_{i011} + n_{i110} + n_{i010})$ for $\gamma_i$ ($i = 1, 2, \ldots, I$). To this end, using the symbol $\binom{n}{a,b,c} = n!/[a!b!c!(n-a-b-c)!]$, where $a \geq 0$, $b \geq 0$, $c \geq 0$, and $a + b + c \leq n$, and $(\binom{n}{a,b,c}) = 0$, otherwise, one gets

$$
P(n_{i111} = a_{i1}, n_{i101} = b_{i1}, n_{i110} = c_{i1}, n_{i110} = a_{i0}, n_{i100} = b_{i0}, n_{i010} = c_{i0})
$$

$$
= \prod_{i=1}^{l} (D_{-i1}^{-n_1} D_{0}^{-n_0}) \binom{n_1}{a_{i1}, b_{i1}, c_{i1}} \binom{n_0}{a_{i0}, b_{i0}, c_{i0}}
$$

$$
\times \exp[a_{i1}(\gamma_{i1} + \gamma_{i2} + \beta_1 + \beta_2 + \beta_{12}) + b_{i1}(\gamma_{i1} + \beta_1) + c_{i1}(\gamma_{i2} + \beta_2)]
$$

$$
\times \exp[a_{i0}(\gamma_{i1} + \gamma_{i2} + \beta_{12}) + b_{i0}\gamma_{i1} + c_{i0}\gamma_{i2}].
$$

(4.12)
Hence, one has the conditional pf

\[
P\left( n_{111} = a_{11}, n_{101} = b_{11}, n_{011} = c_{11}, n_{110} = a_{00} \mid n_{111} + n_{101} + n_{011} + n_{110} = w_{11},
\right.
\]

\[
\left. n_{111} + n_{101} + n_{110} + n_{010} = w_{12} \right) = \left( \sum_{i=1}^{n_1} \frac{n_0}{n_0, w_{11} - a_{11} - b_{11} - a_{00}, w_{12} - a_{11} - c_{11} - a_{00}} \right) e^{[\beta_1(a_{11} + b_{11}) + \beta_2(c_{11} + a_{11}) + \beta_{12}(a_{11} + a_{00})]} e^{[\beta_1(x_1 + y_1) + \beta_2(x_1 + z_1) + \beta_{12}(x_1 + z_0)]},
\]

(4.13)

where \( U \) consists of all permissible values of \((x_1, y_1, z_0)\). When \( n_1 = 1 \) (as in the case of endometrial cancer data), (4.13) simplifies considerably. Then, for each stratum \( i \), \((n_{i111}, n_{i101}, n_{i011}, n_{i001})\) can assume only one of the four possible values: \((1, 0, 0, 0)\), \((0, 1, 0, 0)\), \((0, 0, 1, 0)\), and \((0, 0, 0, 1)\). Thus in this case, the conditional likelihood is given by

\[
L_C(\beta_1, \beta_2, \beta_{12}) = \prod_{i=1}^{I} \left( \frac{n_0}{n_0, w_{11} - a_{11} - b_{11} - a_{00}, w_{12} - a_{11} - c_{11} - a_{00}} \right) e^{[\beta_1(a_{11} + b_{11}) + \beta_2(c_{11} + a_{11}) + \beta_{12}(a_{11} + a_{00})]} e^{[\beta_1(x_1 + y_1) + \beta_2(x_1 + z_1) + \beta_{12}(x_1 + z_0)]}.
\]

(4.14)

Assume that \( b(\beta_1, \beta_2, \beta_{12}) \propto 1 \). Then, the posterior of \((\beta_1, \beta_2, \beta_{12})\) based on the conditional likelihood is

\[
\pi_C(\beta_1, \beta_2, \beta_{12} | \text{data}) \propto L_C(\beta_1, \beta_2, \beta_{12}).
\]

(4.15)

It can be shown that each of \( \pi_C(\beta_1 | \beta_2, \beta_{12}, \text{data}) \), \( \pi_C(\beta_2 | \beta_1, \beta_{12}, \text{data}) \) and \( \pi_C(\beta_{12} | \beta_1, \beta_2, \text{data}) \) is log-concave. Thus, it is convenient to generate samples from each of these conditionals using the adaptive rejection/acceptance algorithm of Gilks and Wild (1992).

We refer back to the same endometrial cancer data, but consider this time two exposure variables—gall-bladder disease and hypertension. We compute posterior means and HPD intervals for each of \( \beta_1 \), \( \beta_2 \) and \( \beta_{12} \) based on the conditional likelihood as well as the full likelihood. When working with the full likelihood, we consider several iid bivariate t-priors for the \((\gamma_{11}, \gamma_{12})\). The results are summarized in Tables 4 and 5.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE posterior mean</th>
<th>posterior s.d.</th>
<th>95% HPD interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_1 )</td>
<td>1.302</td>
<td>1.313</td>
<td>0.380</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>0.401</td>
<td>0.401</td>
<td>0.316</td>
</tr>
<tr>
<td>( \beta_{12} )</td>
<td>0.097</td>
<td>0.099</td>
<td>0.442</td>
</tr>
</tbody>
</table>
Bayesian inference for matched case-control studies

Table 5. The posterior estimates of \((\beta_1, \beta_2, \beta_{12})\) for Gall-Bladder disease and Hypertension based on the full Likelihood

<table>
<thead>
<tr>
<th>((\sigma_1^2, \sigma_2^2))</th>
<th>(\nu)</th>
<th>Parameter posterior mean posterior s.d. 95% HPD interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>(50,50)</td>
<td>1</td>
<td>(\beta_1) 1.585 0.463 (0.664, 2.473)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_2) 0.635 0.380 (−0.091, 1.406)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_{12}) −0.556 0.470 (−1.478, 0.372)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(\beta_1) 1.638 0.468 (0.731, 2.570)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_2) 0.627 0.380 (−0.117, 1.378)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_{12}) −0.453 0.485 (−1.400, 0.503)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>(\beta_1) 1.661 0.470 (0.726, 2.569)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_2) 0.624 0.381 (−0.125, 1.370)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_{12}) −0.419 0.486 (−1.384, 0.522)</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>(\beta_1) 1.660 0.467 (0.741, 2.569)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_2) 0.621 0.380 (−0.117, 1.372)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_{12}) −0.407 0.485 (−1.349, 0.549)</td>
</tr>
<tr>
<td>(100,100)</td>
<td>1</td>
<td>(\beta_1) 1.770 0.483 (0.822, 2.715)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_2) 0.644 0.386 (−0.083, 1.426)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_{12}) −0.375 0.491 (−1.351, 0.578)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(\beta_1) 1.824 0.488 (0.858, 2.775)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_2) 0.639 0.391 (−0.128, 1.407)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_{12}) −0.289 0.502 (−1.268, 0.692)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>(\beta_1) 1.837 0.490 (0.887, 2.800)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_2) 0.631 0.386 (−0.126, 1.391)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_{12}) −0.254 0.506 (−1.249, 0.723)</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>(\beta_1) 1.838 0.490 (0.876, 2.795)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_2) 0.629 0.388 (−0.131, 1.394)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_{12}) −0.249 0.504 (−1.211, 0.772)</td>
</tr>
<tr>
<td>(100,50)</td>
<td>1</td>
<td>(\beta_1) 1.774 0.485 (0.828, 2.724)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_2) 0.613 0.381 (−0.118, 1.373)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_{12}) −0.400 0.485 (−1.348, 0.543)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(\beta_1) 1.824 0.490 (0.872, 2.783)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_2) 0.604 0.382 (−0.140, 1.361)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_{12}) −0.310 0.502 (−1.290, 0.677)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>(\beta_1) 1.834 0.490 (0.858, 2.785)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_2) 0.604 0.385 (−0.151, 1.350)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_{12}) −0.274 0.501 (−1.225, 0.745)</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>(\beta_1) 1.841 0.489 (0.920, 2.844)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_2) 0.601 0.382 (−0.179, 1.318)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(\beta_{12}) −0.265 0.501 (−1.241, 0.725)</td>
</tr>
</tbody>
</table>

The Bayesian analysis with the conditional likelihood is based on the prior \(h(\beta_1, \beta_2, \beta_{12}) \propto 1\). Thus it is not surprising to find the closeness of the conditional MLE’s (which are also the posterior modes) and the posterior means. However, this need not be so when we work with the full likelihood.

Once again, we consider
\[ h(\beta_1, \beta_2, \beta_{12}) \propto 1, \]
but we assume in addition that
\[ g(\gamma_{i1}, \gamma_{i2}) \propto \left[ 1 + \gamma_{i1}^2 / (\nu \sigma_{i1}^2) + \gamma_{i2}^2 / (\nu \sigma_{i2}^2) \right]^{-\frac{\nu+1}{2}}, \]
where \( \sigma_{i1}^2, \sigma_{i2}^2, \) and \( \nu \) are prespecified hyperparameters.

The Bayesian analysis based on the full likelihood shows that while the posterior means and posterior s.d.’s are fairly robust with respect to the choice of the degrees of freedom, and are somewhat robust against the choice of the scale parameters. Moreover, while the Bayesian analysis based on the conditional likelihood suggests essentially no partial association between the two exposure variables, Bayesian analysis based on the full likelihood hints at mild partial negative association.

5. Concluding Remarks

The paper considers Bayesian inference for matched case-control studies. General Bayesian inferential procedures are discussed not necessarily confined to the logit link. For the particular case of logit link, it is shown that Bayesian procedures based on conditional and unconditional likelihoods can often significantly differ. Finally, we consider multiple exposure variables, and once again bring out the difference in the Bayesian analysis based on conditional and unconditional likelihoods for the logistic model. It is possible (though necessarily more computer-intensive) to extend the results to more than two exposure variables, for example use of estrogen in addition to gall-bladder disease and hypertension. The analysis can be carried out in a similar fashion, and we may still anticipate different results based on conditional and unconditional likelihoods.

References


MALAY GHOSH
UNIVERSITY OF FLORIDA
223 GRIFFIN-FLOYD HALL
GAINESVILLE
FL 32611-8545, USA
E-mail: ghosh@mailhost.stat.uf.edu

MING-HUI CHEN
DEPARTMENT OF STATISTICS
UNIVERSITY OF CONNECTICUT
215, GLENBROOK ROAD, U-4120
STORRS, CT 06269-4120, USA
E-mail: mhchen@stat.uconn.edu