Estimation of the false negative fraction of a diagnostic kit through Bayesian regression model averaging

Ranyimbo, A O; Held, L
Estimation of the false negative fraction of a diagnostic kit through Bayesian regression model averaging

Abstract

In modelling we usually endeavour to find a single 'best' model that explains the relationship between independent and dependent variables. Selection of a single model fails to take into account the prior uncertainty in the model space. The Bayesian model averaging (BMA) approach tackles this problem by considering the set of all possible models. We apply BMA approach to the estimation of the false negative fraction (FNF) in a particular case of a two-stage multiple screening test for bowel cancer. We find that after taking model uncertainty into consideration the estimate of the FNF obtained is largely dependent on the covariance structure of the priors. Results obtained when the Zellner g-prior for the prior variance is used is largely influenced by the magnitude of g.
Estimation of the false negative fraction of a diagnostic kit through Bayesian regression model averaging

A. O. Ranyimbo and L. Held*;†

Department of Statistics, Ludwig-Maximilians-Universität München, Ludwigstr. 33, 80539 München, Germany

SUMMARY

In modelling we usually endeavour to find a single ‘best’ model that explains the relationship between independent and dependent variables. Selection of a single model fails to take into account the prior uncertainty in the model space. The Bayesian model averaging (BMA) approach tackles this problem by considering the set of all possible models. We apply BMA approach to the estimation of the false negative fraction (FNF) in a particular case of a two-stage multiple screening test for bowel cancer. We find that after taking model uncertainty into consideration the estimate of the FNF obtained is largely dependent on the covariance structure of the priors. Results obtained when the Zellner g-prior for the prior variance is used is largely influenced by the magnitude of g. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: Bayesian model averaging; false negative fraction; screening selection

1. INTRODUCTION

In attempting to find the relationship between a dependent variable and explanatory variables we endeavour to elicit the single model that ‘best’ explains the relationship. A typical example is in linear modelling where a single model which includes the important covariates is finally selected. The selection of such a single model certainly ignores the prior uncertainty about the covariates to include [1]. Such an approach may lead to overconfident [2] inference since certain independent variables may be reflected as highly significant when in reality they are not. It may also be the case that certain explanatory variables are depicted as non-significant when the converse is actually the case. As a consequence the interpretation derived from such results would be erroneous. Kass and Raftery [3] give an example (see reanalysis of the Educational Transition data) where ignoring uncertainty does lead to conclusions that may not be holding. The incorporation of model uncertainty in the Bayesian framework is quite straightforward. This is achieved by considering hierarchical set-up with one additional
hierarchy for the models. Using Bayesian model averaging (BMA) we can take the model uncertainty into account so that predictions and inferences are based on a set of models and each model contributing proportionally to the support it receives from the observed data [1]. In this article we apply the concept of BMA in the estimation of the false negative fraction (FNF) in the case of a two-stage multiple screening test for bowel cancer [4].

In assessing the performance of a diagnostic kit we usually use sensitivity and specificity as the accuracy measures. By definition sensitivity is the proportion of the true-diseased who are correctly identified by the test while specificity is the proportion of the non-diseased population identified as disease-free by the test. When all the study subjects undergo a gold standard procedure, that is all are verified, the estimation of sensitivity and specificity is straightforward. Alternative summary measures are the false negative fraction (FNF), which is \(1 - \text{sensitivity}\), and the false positive fraction (FPF), which is defined as \(1 - \text{specificity}\).

Lloyd and Frommer [4] describe a regression-based approach of the estimation of the FNF when multiple negatives are unverified. Starting with the set of basis functions \(\{1/k, 1/k^2, x/k, x/k^2, x/k^3\}\), where \(x\) is the number of positives out of a total of \(k\) diagnostic tests, they model the probability that a study subject with a history of \(x\) positives in the previous \(k\) tests gives a negative result in the next test. The selection of their final single best model is based on the Akaike information criterion (AIC). However, the selection of a single model fails to take into account the model uncertainty. The question that we would like to address, therefore, is how we can estimate the FNF after accounting for model uncertainty.

Supposing that \(y\) is a vector of observations and \(\Gamma = \{M_1, \ldots, M_K\}\) is the model space, then under BMA the posterior density of \(\Delta\), the quantity of interest such as relative risk or a future observable, is

\[
P(\Delta | y) = \sum_{j=1}^{K} P(\Delta | M_j, y) P(M_j | y) \tag{1}
\]

This is a weighted average of the model-specific posterior distributions where the weights are the individual posterior model probabilities. Madigan and Raftery [5] have shown that averaging over all models provides better average predictive ability, as measured by a logarithmic scoring rule, than using any single model \(M_j \in \Gamma\). Min and Zellner [6] have also shown that such weighted averaging leads to minimization of the expected predictive squared error loss when the set of models considered is exhaustive. The posterior model probability for model \(M_j\) is

\[
P(M_j | y) = \frac{P(y | M_j)P(M_j)}{\sum_{j=1}^{K} P(y | M_j)P(M_j)} \tag{2}
\]

where

\[
P(y | M_j) = \int P(y | \omega_j, M_j)P(\omega_j | M_j) \, d\omega_j \tag{3}
\]

is the marginal distribution of \(y\) conditioned on model \(M_j\) after integration of the model-specific parameters \(\omega_j\), \(P(M_j)\) is the prior probability for model \(M_j\), \(P(y | \omega_j, M_j)\) is the likelihood and \(P(\omega_j | M_j)\) is the prior distribution of \(\omega_j\) assuming model \(M_j\).

Two immediate difficulties that arise are that the number of models \((K)\) may be very large thus making the computation of the denominator in (2) insurmountable and secondly the
integral in (3) may not be analytically evaluated. The former problem has been addressed by the use of MCMC algorithm whereby sampling is done from a Markov chain on the model space. The Markov chain has the posterior model distribution as its stationary distribution. The model composition, MC^3, of Madigan and York [7] is an example of such a stochastic search approach. The Occam’s window approach of Madigan and Raftery [5] is another alternative to addressing the problem of large number of models to be considered. It sieves the models to be finally considered by omitting those models which have relatively lower posterior model probabilities. When the integral in (3) is analytically intractable the Laplace method can be used as an approximation [3]. Similarly, the reversible jump approach of Green [8], which is a generalization of the Metropolis–Hastings algorithm to parameter spaces of varying dimension, could be used to jointly sample from the parameter and model spaces.

In order to put the above formulation in context, suppose that the diagnostic kit as in Reference [4] is applied to a study subject T times. Let x denote the number of positive tests out of a total of k previous tests and \( f_i(x, k), \ldots, f_p(x, k) \) be basis functions dependent on x and k. Then the probability, denoted \( p_{sk} \), that an individual tests negative in the next test given that the individual had x positive tests out of k previous tests is given by the following logistic model:

\[
\log \left( \frac{p_{sk}}{1 - p_{sk}} \right) = \beta_0 + \beta_1 f_1(x, k) + \cdots + \beta_p f_p(x, k)
\]

In this case model uncertainty is equivalent to uncertainty as to which basis functions to include in the model. We reflect this uncertainty by introducing a vector \( \gamma = (\gamma_1, \ldots, \gamma_p) \) so that \( \gamma_\i = 1 \) with probability 0.5 if the i\textsuperscript{th} basis function is in the model and \( \gamma_\i = 0 \) otherwise. We then have,

\[
\log \left( \frac{p_{sk\gamma}}{1 - p_{sk\gamma}} \right) = \beta_0 + \gamma_1 \beta_1 f_1(x, k) + \cdots + \gamma_p \beta_p f_p(x, k)
\]

where \( p_{sk\gamma} \) is the probability of having a negative result in the next test having had x positive tests out of k previous tests for the model with indicator vector \( \gamma \). Each logistic regression model is defined by a unique \( \gamma \). The basis functions that we consider are \( f_1 = 1/k \), \( f_2 = x/k \), \( f_3 = x/k^2 \) and \( f_4 = x/k^3 \). We use the notation \( \beta_\gamma \) to denote the vector of parameters \( \beta \) for the corresponding indicator vector \( \gamma \). The parameter vector \( \beta_\gamma \) is assumed to be having a \( N(0, \nu) \) prior distribution. We assume a prior mean of zero to reflect the uncertainty on the sign of the parameters.

The model-specific FNF \( \delta_\gamma \) as in Reference [4] is given by the expression

\[
\delta_\gamma = \frac{p_{11\gamma}}{1 - p_{01\gamma} + p_{11\gamma}} \prod_{z=1}^{T-1} p_{0z\gamma}
\]

The quantity of interest, in this case FNF, will be the weighted average of all the \( \delta_\gamma \).

This article is organized as follows: Section 2 outlines probit regression model using auxiliary variable. In Section 3 we give an overview of the application of BMA to selection of basis functions. In Section 4 we describe the results obtained when BMA is applied to the bowel cancer data found in Reference [4]. In Section 5 we give our research findings and a brief discussion.


2. PROBIT REGRESSION WITH AUXILIARY VARIABLE

In order to have a probit model we use the standard Gaussian link in the above formulation of the logistic regression model. The probit regression model would then take the form

\[
\Phi^{-1}(p_{xy}) = \beta_0 + \gamma_1 \beta_1 f_1(x, k) + \cdots + \gamma_p \beta_p f_p(x, k)
\]

where \( \Phi \) is the standard Gaussian cdf. The FNF can then be estimated as a weighted average of all the model-specific FNFs, \( \delta_q \), as in the logistic regression case.

Consider the following Bayesian binary regression model:

\[
y_i \sim \text{Bernoulli}(h^{-1}(\eta_i)) \\
\eta_i = x_i \beta \\
\beta \sim \pi(\beta)
\]  

(4)

where \( y_i \in \{0, 1\}, \ i = 1, \ldots, n \), is a binary response variable with a corresponding covariate vector \( x_i = (x_{i1}, \ldots, x_{ip}) \), \( h \) is a link function and \( \eta_i \) is the linear predictor. The unknown regression parameters \( \beta \in R^p \) are assumed \textit{a priori} to have distribution \( \pi(\beta) \).

Albert and Chib [9] have shown that when \( h \) is the standard Gaussian cumulative distribution function (this case corresponds to the probit link), the introduction of \( n \) independent auxiliary variables \( z_1, \ldots, z_n \) leads to the representation of the model in (4) to be

\[
y_i = \begin{cases} 
1 & \text{if } z_i > 0 \\
0 & \text{otherwise}
\end{cases} \\
z_i = x_i \beta + \epsilon_i \\
\epsilon_i \sim N(0, 1) \\
\beta \sim \pi(\beta)
\]

Since \( \beta \) has a multivariate normal prior, \( \pi(\beta) = N(0, \nu) \), the full conditional distribution of \( \beta \) given \( z \) is then normal,

\[
\beta | z \sim N(B, V) \\
B = Vx'z \\
V = (v^{-1} + x'x)^{-1}
\]

(5)

where \( x = (x'_1, x'_2, \ldots, x'_n)' \). The full conditional for each \( z_i \) is truncated normal,

\[
z_i | \beta, x_i, y_i \propto \begin{cases} 
N(x_i \beta, 1)I(z_i > 0) & \text{if } y_i = 1 \\
N(x_i \beta, 1)I(z_i \leq 0) & \text{otherwise}
\end{cases}
\]

where \( I(.) \) is the indicator function. Sampling from the truncated normal can then be done as discussed in Reference [10]. Alternatively Gibbs sampling can be used in updating \( z \) from its marginal distribution having integrated over \( \beta \) [11]. The procedure discussed above is quite
common in Bayesian statistics and similar work include George and McCulloch [12, 13], Mitchell and Beauchamp [14] and Fernandez et al. [15]. The stochastic search variable selection (SSVS) of George and McCulloch [12] involves the embedding all the models in the full model and excluding those variables whose coefficients are ‘close’ to zero relative to some threshold. Since in SSVS the dimension of the parameter vector remains fixed its implementation proceeds via a stochastic search Gibbs sampler. Smith and Kohn [16] discuss a Bayesian approach for estimating additive regression model semi-parametrically while automatically selecting the significant independent variables.

3. SELECTION OF BASIS FUNCTIONS

In order to approximate (1) we make use of Markov chain Monte Carlo (MCMC) technique. The standard Metropolis–Hastings algorithm is known to be applicable only in the case of a fixed dimension problem. However, in the BMA approach we are dealing with a dimension changing problem since it involves averaging over all the different sets of basis functions. Green [8] discusses the reversible jump MCMC approach which is appropriate when the dimension is either fixed or varying. In the sequel we find the acceptance probability of moving from one set of basis functions to another.

Suppose that \( \gamma = (\gamma_1, \ldots, \gamma_p) \), \( \gamma_i \in \{0, 1\} \), \( i = 1, \ldots, p \), such that \( \gamma_i = 1 \) if the \( i \)-th covariate is present in the model and \( \gamma_i = 0 \) otherwise. A prior on the model space is specified via a prior on the covariate indicator, \( \pi(\gamma) \). Holmes and Held [11] have shown that it is more efficient to jointly update \( \{\gamma, \beta\} \) while conditioning on \( z \). The joint posterior distribution of \( \{\gamma, \beta\} \) is given as

\[
\pi(\gamma, \beta \mid z) = \pi(\gamma \mid z) \pi(\beta \mid \gamma, z)
\]

We jointly update \( \gamma \) and \( \beta \) as

\[
q(\gamma^*, \beta^*) = \pi(\beta^* \mid \gamma^*, z) q(\gamma^*)
\]

where \( \pi(\beta^* \mid \gamma^*, z) \) is the conditional multivariate normal posterior distribution (5), given the covariate set defined by \( \gamma^* \), and \( q(\gamma^*) \) is a proposal density for \( \gamma^* \). Through this formulation the acceptance probability [8] takes the form

\[
\alpha = \min \left\{ 1, \frac{\pi(\gamma^*, \beta^* \mid z)}{\pi(\gamma, \beta \mid z)} \times \frac{\pi(\beta \mid \gamma, z)}{\pi(\beta^* \mid \gamma^*, z)} \times \frac{q(\gamma \mid \gamma^*)}{q(\gamma^* \mid \gamma)} \right\}
\]

We assume independent priors on the set of basis functions and the regression parameters such that \( \pi(\gamma_i = 1) = 0.5 \) for \( i = 1, \ldots, p \) and \( \pi(\beta) = N(0, \Sigma) \), respectively. A basis function is selected at random and a move is proposed by setting \( \gamma_i^* = 1 \) if currently \( \gamma_i = 0 \) or \( \gamma_i^* = 0 \) otherwise. The acceptance probability [11] reduces to (see Appendix A for the derivation),

\[
\alpha = \min \left\{ 1, \frac{|V_{\gamma^*}|^{1/2}}{|V_{\gamma}|^{1/2}} \times \frac{|V_{\gamma^*}|^{1/2}}{|V_{\gamma}|^{1/2}} \times \exp \left( \frac{0.5B_{\gamma^*}'V_{\gamma^*}^{-1}B_{\gamma^*}}{0.5B_{\gamma}'V_{\gamma}^{-1}B_{\gamma}} \right) \right\}
\]

where subscript \( \gamma \) refers to the indicator vector for the current set of basis functions included in the model while \( \gamma^* \) refers to the updated indicator vector after making a move to include
or drop a particular basis function. The posterior model probability is estimated to be the relative number of times a model is accepted in the MCMC algorithm out of the total number of runs after exclusion of the burn-in runs.

4. RESULTS

We apply the BMA approach discussed above to the bowel cancer data found in Table II of Reference [4]. The data are from a voluntary bowel cancer screening program conducted over several years at St Vincent’s Hospital in Sydney, Australia. The number of patients who were screened was 38,000. The screening test in the primary phase involved self-administered testing for blood in stool on six consecutive days using a provided screening kit. About 3000 patients returned a positive test result at least once out of the six screenings. Those who had at least one positive test result had their true disease status verified using physical examination, sigmoidoscopy and colonoscopy. Only 196 of those who were verified were found to be true-diseased cases. In the secondary phase each of the 122 out of the 196 true-diseased cases had a further set of six screenings about one week after the primary screenings.

We considered BMA using the four basis functions mentioned in Section 1 with a probit link. In order to define the prior variance for $\beta$ we use the prior knowledge about the estimates of some of its components and their standard deviations obtained from the model with the minimum AIC value as in Reference [4]. Since the standard deviations in this case were less than 1 and the estimates of the parameters were in the range $(-2.5, 1)$ (with probit link), we adopt a prior variance as $2.5I$ for $\beta$, that is $\beta$ was assumed a priori to be multivariate normal $N(0, 2.5I)$. The range $(-q, q)$, where $q \approx 3.1$ is the 97.5 per cent quantile for $N(0, 2.5)$ distribution, should be vague enough to cover the probable values of the parameters.

An MCMC algorithm with 100,000 runs and a burn-in of 5000 runs was used. Mixing in this case was fairly good (see Figure 1). Figure 1 is a line plot of the every tenth posterior

![Figure 1. Line plot of posterior samples from the false negative fraction.](image-url)
draw of the FNF after the initial burn-in period. The MCMC appears to have converged after
the burn-in period. Figure 2 is a plot of the autocorrelation functions (ACF). The left panel is
the ACF plot for the 100 000 draws of the FNF and indicates that the draws remain correlated
even at a lag of 50. The second panel gives the ACF when the draws of FNF which are at
a lag of 10 are considered. The correlation is now considerably reduced. A histogram of the
posterior draws of the FNF is given in Figure 3.

Based on the AIC Lloyd and Frommer [4] selected the model with basis functions \((x/k, x/k^3)\).
For comparison purpose another MCMC which incorporates the Gamerman’s iterative weighted
least squares algorithm [17] was used in fitting a Bayesian logistic regression model with fixed
basis functions \((x/k, x/k^3)\). A probit model using auxiliary variables was also used to model
the data with the same two basis functions.

Table I contains the maximum likelihood estimates of the FNF from all the \(2^4 = 16\) models
assuming the probit link. The last two columns of this table are the posterior model probabil-
ities (pmp) and clearly depict the uncertainty that is inherent in the model space. The results
indicate that even after taking uncertainty in the models space into account the model with
basis functions \(x/k\) and \(x/k^3\) (minimum AIC model) is still preferred with a posterior proba-
bility value of 0.3925. The top seven models in this table appear to be quite competitive if
one is to consider only the AIC values. However, the estimates of the FNF from these models
are quite varied. The models 9–16 may or may not be better than the BMA model in terms
Figure 3. Histogram of posterior samples from the false negative fraction.

Table I. Maximum likelihood estimates of the false negative fraction under probit models using different basis functions. The last two columns are the posterior model probabilities (pmp). The models have been sorted according to the AIC values.

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>FNF</th>
<th>pmp(^1)</th>
<th>pmp(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(x/k, x/k^3)</td>
<td>861.2423</td>
<td>875.2508</td>
<td>0.2367</td>
<td>0.3925</td>
</tr>
<tr>
<td>2</td>
<td>(x/k, x/k^2)</td>
<td>861.3684</td>
<td>875.3769</td>
<td>0.2353</td>
<td>0.3010</td>
</tr>
<tr>
<td>3</td>
<td>(1/k, x/k, x/k^2)</td>
<td>861.4759</td>
<td>880.1539</td>
<td>0.0536</td>
<td>0.0000</td>
</tr>
<tr>
<td>4</td>
<td>(1/k, x/k, x/k^3)</td>
<td>862.8206</td>
<td>881.4985</td>
<td>0.1631</td>
<td>0.0001</td>
</tr>
<tr>
<td>5</td>
<td>(x/k, x/k^2, x/k^3)</td>
<td>863.2246</td>
<td>881.9026</td>
<td>0.2368</td>
<td>0.0000</td>
</tr>
<tr>
<td>6</td>
<td>(1/k, x/k)</td>
<td>863.2359</td>
<td>877.2444</td>
<td>0.3272</td>
<td>0.1736</td>
</tr>
<tr>
<td>7</td>
<td>(1/k, x/k, x/k^2, x/k^3)</td>
<td>863.3983</td>
<td>886.7458</td>
<td>0.0441</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>(x/k)</td>
<td>870.8532</td>
<td>880.1922</td>
<td>0.1394</td>
<td>0.1328</td>
</tr>
<tr>
<td>9</td>
<td>(1/k, x/k^2, x/k^3)</td>
<td>880.0581</td>
<td>898.7361</td>
<td>0.3919</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>(x/k^2, x/k^3)</td>
<td>893.8957</td>
<td>907.9042</td>
<td>0.0923</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>(1/k, x/k^2)</td>
<td>903.3562</td>
<td>917.3647</td>
<td>0.2462</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>(x/k^2)</td>
<td>966.4923</td>
<td>975.8313</td>
<td>0.0022</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>(1/k, x/k^3)</td>
<td>979.5940</td>
<td>993.6025</td>
<td>0.0053</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>(x/k^3)</td>
<td>980.2814</td>
<td>989.6204</td>
<td>0.0012</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>(1/k)</td>
<td>982.2485</td>
<td>991.5875</td>
<td>0.0007</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>983.5737</td>
<td>988.2432</td>
<td>0.0010</td>
<td>0</td>
</tr>
</tbody>
</table>

FNF: false negative fraction.
BIC: Bayesian information criterion.
AIC: Akaike information criterion.
pmp\(^1\): posterior model probabilities under Bayesian model averaging.
pmp\(^2\): approximate posterior model probabilities obtained from BIC.
of diagnostic performance but they are not of great interest here because their posterior model probabilities are all essentially zero. The ranking of the models on the basis of the Bayesian information criterion (BIC) in this case closely resembles that from AIC. Approximate posterior model probabilities (pmp2) were derived from the BIC values (see Reference [3] for details). Roughly they agree with the pmp1 as far as the ranking of the models is concerned. The slight disparity could be as a result of the error emanating from the poor approximation of Bayes factors from the BIC. The relative error in this case is of order $O(1)$ and as such even for large samples the approximation is far from correct [18]. The approximated posterior mean estimate of the FNF in this case is 22.9 per cent and compares quite well with the value 22.8 per cent obtained from BMA with multivariate normal prior $N(0,2.5I)$ for $\beta$.

Table II gives a summary of the posterior mean and median estimates of the FNF for the three particular cases considered and the corresponding 95 per cent credible intervals. The models referred to in the table are as follows: BMA for the result from Bayesian model averaging, P1 for the probit model with auxiliary variables using fixed basis functions ($x/k, x/k^3$), L1 for the Bayesian logistic model with fixed basis functions ($x/k, x/k^3$) and ML for the logistic model with fixed basis functions ($x/k, x/k^3$). In the last row of this table we report the 95 per cent confidence interval after working on the logit scale of the FNF. The mean and median posterior estimate of the FNF based on the BMA are 0.228 and 0.218, respectively. The value for the median estimate is notably lower relative to the estimates from the logistic and probit models with fixed basis functions. The corresponding 95 per cent credible interval (0.091–0.393) is marginally wider in comparison to that obtained from models with fixed basis functions. This increase in length of credible interval can be attributed to induced variability as a result of uncertainty in the model space. We also find that the mean posterior estimate from model P1 is slightly higher (0.241) than that from a model ML (0.236). The estimate from model L1 is the same as from model ML although the credible interval is narrower.

In Table III we find that the model based on the AIC leads to different conclusion in comparison to the BMA results under the multivariate normal $N(0,2.5I)$ prior for $\beta$. The posterior effect probability, $P(\beta \neq 0 \mid y)$, for the parameters corresponding to the respective basis function are given in the last column of this table. The basis function $x/k$ has a probability one that its effect is not equal zero. Similarly, the basis function $x/k^3$ has a probability of about 0.39 that its effect is different from zero. Both the AIC and the BMA approaches agree on the strong evidence of the effect of the basis function $x/k$ ($p < 0.001$ and $P(\beta \neq 0 \mid y) = 1$).
Table III. Comparison of \( p \)-values for the models 1 and 7 using probit link (see Table I) to the posterior effect probability from the BMA.

<table>
<thead>
<tr>
<th>Basis function</th>
<th>Model 1 ( p )-value</th>
<th>Model 7 ( p )-value</th>
<th>BMA ( P(\beta \neq 0 \mid y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k )</td>
<td>.</td>
<td>0.178</td>
<td>0.174</td>
</tr>
<tr>
<td>( x/k )</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>1.000</td>
</tr>
<tr>
<td>( x/k^2 )</td>
<td></td>
<td>0.233</td>
<td>0.301</td>
</tr>
<tr>
<td>( x/k^3 )</td>
<td>&lt;0.001</td>
<td>0.781</td>
<td>0.393</td>
</tr>
</tbody>
</table>

However, in the AIC model (model 1) we find that the \( p \)-value tends to overstate the evidence of the effect associated with the basis function \( x/k^3 \), \( p < 0.001 \). In contrast the posterior effect probability indicates that there is weak evidence against an effect. There also appears to be mild evidence of the effect associated with basis function \( x/k^2 \).

The ‘full’ model (model 7) is also interesting because it agrees with model 1 on the significance of the basis functions \( x/k \). The results from model 7 also match fairly well those from the BMA in terms of the significance of the effects associated with the different basis functions. However, selecting the ‘full’ model alone would still not account for the inherent uncertainty. In fact, model 7 leads to a FNF value of 4.4 per cent which is very different from the corresponding BMA value of 22.8 per cent. The former value is also at variant with the value 23.6 per cent reported in Lloyd and Frommer [4]. Since model 7 has a posterior probability 0, it is an indication that the data do not support this particular model quite well.

To check how sensitive the estimates of the FNF are with regard to the prior chosen for \( \beta \) when using the BMA approach, we considered four other forms for the prior variance. In the above results the prior distribution of the parameter \( \beta \) was taken to be \( \text{N}(0, 2.5I) \). The choice of an appropriate prior covariance matrix is never straightforward. In the regression framework the choice of independent priors (diagonal covariance matrix) is quite common. However, given the structure of the basis functions, a prior that takes the covariance structure into account may be appropriate. This leads to choosing a prior covariance matrix that is not diagonal. The prior covariance matrix can be written as \( \mathbf{v} = g\mathbf{M} \sigma^2 \) where \( g > 0 \) is a constant to be selected and \( \mathbf{M} \) is a \( p \times p \) matrix (\( p = \text{number of parameters in } \beta \)). If \( \mathbf{M} = (\mathbf{x}'\mathbf{x})^{-1} \) we get the Zellner \( g \)-priors [19]. Fernandez et al. [15] have given ‘benchmark’ prior specifications in linear regression context under model uncertainty based on the Zellner \( g \)-prior. With \( \sigma^2 \) fixed to 1 we consider four particular cases of the Zellner \( g \)-prior: first case is when \( g = n \) (\( n \) is the number of observations) and corresponds to the unit-information prior of Kass and Wasserman [18], second is the case when \( g = p^2 \) (\( p \) is the number of parameters), thirdly we have \( g = (\log(n))^3 \) and finally we have \( g = \log(n) \). The last two are chosen in line with the Hannan–Quinn criterion [20]. Note that all these suggestions have been made for Bayesian linear model, and so are perhaps not directly applicable to the binary regression context.
Table IV. Mean and median posterior estimates of the false negative fraction together with the 95 per cent credible interval under different priors for $\beta$.

<table>
<thead>
<tr>
<th>Prior</th>
<th>Mean</th>
<th>Median</th>
<th>95 per cent Credible interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N(0, 2.5I)</td>
<td>0.228</td>
<td>0.218</td>
<td>0.091–0.393</td>
</tr>
<tr>
<td>$N(0, n(x'x)^{-1})</td>
<td>0.144</td>
<td>0.139</td>
<td>0.069–0.248</td>
</tr>
<tr>
<td>$N(0, p^2(x'x)^{-1})$</td>
<td>0.125</td>
<td>0.112</td>
<td>0.002–0.347</td>
</tr>
<tr>
<td>$N(0, (\log(n))^3(x'x)^{-1})$</td>
<td>0.239</td>
<td>0.232</td>
<td>0.110–0.407</td>
</tr>
<tr>
<td>$N(0, \log(n)(x'x)^{-1})$</td>
<td>0.103</td>
<td>0.096</td>
<td>0.002–0.278</td>
</tr>
</tbody>
</table>

$p = 4$ (number of basis functions), $n = 788$ (total observations).

Table V. The models with highest posterior probability for different prior specification.

<table>
<thead>
<tr>
<th>Prior</th>
<th>Model</th>
<th>pmp</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N(0, 2.5I)$</td>
<td>1</td>
<td>0.3925</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.3010</td>
</tr>
<tr>
<td>$N(0, n(x'x)^{-1})$</td>
<td>8</td>
<td>1.0000</td>
</tr>
<tr>
<td>$N(0, p^2(x'x)^{-1})$</td>
<td>7</td>
<td>0.2958</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.2491</td>
</tr>
<tr>
<td>$N(0, (\log(n))^3(x'x)^{-1})$</td>
<td>1</td>
<td>0.8284</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.0837</td>
</tr>
<tr>
<td>$N(0, \log(n)(x'x)^{-1})$</td>
<td>7</td>
<td>0.3029</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.2291</td>
</tr>
</tbody>
</table>

pmp: posterior model probability, $n = 788$ (total observations), $p = 4$ (number of basis functions).

The MCMC was run again but with the different prior variances as given in the previous paragraph. The results in Table IV show there is some sensitivity with respect to the choice of the prior. Table V lists, for each prior specification, the models frequently accepted in the MCMC algorithm. The priors with $g = (\log(n))^3$ and $g = n$ select the simpler models 1 and 8 with high posterior probability. On the contrary having $g = \log(n)$ and $g = p^2$ leads to models with more basis functions.

5. RESEARCH FINDINGS AND DISCUSSION

In this work we have attempted to use Bayesian model averaging approach in estimating the FNF in a two-phase multiple screening trial. We find that the posterior estimate of the FNF based on the basis-functions $\{1/k, x/k, x/k^2, x/k^3\}$ to be largely dependent on the specification of the prior variance. Selecting smaller prior variance values lead to the choice of simpler models. Large values for prior variance in this case always resulted in the model with only the intercept being selected. When the covariance structure of the predictors is taken into account by using the Zellner $g$-prior the results obtained are dependent largely on the magnitude of $g$ with larger values leading to higher posterior probability for the simpler models.
We have applied the BMA to a relatively small data set with only four basis functions, which leads to a maximum of $2^4$ models. Implementation of the applied MCMC algorithm in the case of a large data set with many subject-specific variables and basis functions presents a practical challenge. By design the proposed MCMC should traverse the whole model space irrespective of the number of models. The less favourable models will have low posterior model probabilities and therefore will not have a considerable influence on the estimation of the quantity of interest. One particular advantage of the proposed approach is the joint update of the $S$ and $R$ vectors. By a joint update we enhance the efficiency of the MCMC algorithm and avoid slow mixing. The sensitivity of the BMA results to the choice of prior is also not quite exhaustive. There is therefore a need for further formal simulation studies with different priors not only for $R$ but also for $\gamma$.

Further, this approach is also possible in the case of logit model [11], but in our application there is no interest in odds ratio interpretation of parameter estimates since all we want is to determine the FNF posterior, which is possible with both link functions. In addition, it would be interesting to compare the BMA and the AIC model in terms of their predictive capabilities.

APPENDIX A: DERIVATION OF FORMULA FOR ACCEPTANCE PROBABILITY $\alpha$

The target (conditional) distribution is

$$
\pi(\gamma, \beta | z, \lambda) = \pi(\gamma | z, \lambda) \times \pi(\beta | \gamma, z, \lambda)
$$

The proposal distribution is

$$
q(\gamma^*, \beta^* | \gamma, \beta) = q(\gamma^*, \beta^* | \gamma) = \pi(\beta^* | \gamma^*, z, \lambda) \times q(\gamma^* | \gamma)
$$

where symbols with * denote the proposed value and symbols without * denote the current value.

The Metropolis–Hastings acceptance probability of this joint proposal for $\beta$ and $\gamma$ is hence

$$
\alpha = \min \left\{ 1, \frac{\pi(\gamma^* | z, \lambda) \times q(\gamma | \gamma^*)}{\pi(\gamma | z, \lambda) \times q(\gamma^* | \gamma)} \right\}
$$

(A1)

and depends neither on $\beta$ nor $\beta^*$.

The interesting term here is $\pi(\gamma | z, \lambda)$. We can rewrite this as

$$
\pi(\gamma | z, \lambda) = \frac{\pi(z | \gamma, \lambda) \pi(\gamma | \lambda)}{\pi(z | \lambda)} = \frac{\pi(z | \gamma, \lambda) \pi(\gamma)}{\pi(z | \lambda)}
$$

Equation (A1) can therefore be written as

$$
\alpha = \min \left\{ 1, \frac{\pi(\gamma^* | z, \lambda) \times q(\gamma | \gamma^*)}{\pi(\gamma | z, \lambda) \times q(\gamma^* | \gamma)} \right\}
$$
\[
\begin{align*}
&= \min \left\{ 1, \frac{\pi(z \mid \gamma^*, \lambda)}{\pi(z \mid \gamma, \lambda)} \times \frac{\pi(\gamma^*)}{\pi(\gamma)} \times \frac{q(\gamma \mid \gamma^*)}{q(\gamma^*)} \right\} \\
&= \min \left\{ 1, \frac{\pi(z \mid \gamma^*, \lambda)}{\pi(z \mid \gamma, \lambda)} \times \frac{\pi(\gamma^*)}{\pi(\gamma)} \right\}
\end{align*}
\]

We note that \(\pi(z \mid \gamma^*, \lambda)\) and \(\pi(z \mid \gamma, \lambda)\) are the marginal likelihoods (integrated likelihoods). Denison et al. [21] derive this quantity in the case of an additional unknown variance. Here we have a simpler case where the variance is known to be one and so

\[
\pi(z \mid \gamma, \lambda) = \frac{\pi(z \mid \beta_T) \times \pi(\beta_T)}{\pi(\beta_T \mid z)}
\]

where \(\beta_T\) has a prior multivariate normal \(N(0, \nu_T)\) distribution. That is,

\[
\pi(\beta_T) = (2\pi)^{-\frac{(p/2)}{2}} |\nu_T|^{-1/2} \exp\{-0.5\beta_T'\nu_T^{-1}\beta_T\}
\]

The distribution of \(z \mid \beta_T\) is also multivariate normal \(N(x\beta_T, I)\);

\[
\pi(z \mid \beta_T) = (2\pi)^{-\frac{(n/2)}{2}} \exp\{-0.5(z - x\beta_T)'(z - x\beta_T)\}
\]

The posterior distribution \(\pi(\beta_T \mid z)\), using Bayes theorem, is simply proportional to the likelihood times the prior. That is

\[
\pi(\beta_T \mid z) \propto \pi(z \mid \beta_T)\pi(\beta_T)
\]

\[
\propto \exp\{-0.5(\beta_T'\nu_T^{-1}\beta_T + (z - x\beta_T)'(z - x\beta_T))\}
\]

\[
= \exp\{-0.5(\beta_T'\nu_T^{-1}\beta_T + \beta_T'x'x\beta_T - 2\beta_T'x'z + z'z)\}
\]

\[
\propto \exp\{-0.5(\beta_T'x'x\beta_T - 2\beta_T'x'z)\}
\]

\[
\propto \exp\{-0.5(\beta_T - B_T)'(V_T^{-1}(\beta_T - B_T))\}
\]

where

\[
V_T = (\nu_T^{-1} + x'x)^{-1}
\]

\[
B_T = V_Tx'z
\]

It then follows that

\[
\pi(z \mid \gamma, \lambda) = \frac{(2\pi)^{-\frac{((n+p)/2)}{2}} |\nu_T|^{-1/2} \exp\{-0.5[\beta_T'\nu_T^{-1}\beta_T + (z - x\beta_T)'(z - x\beta_T)]\}}{(2\pi)^{-\frac{(p/2)}{2}} |V_T|^{-1/2} \exp\{-0.5(\beta_T - B_T)'V_T^{-1}(\beta_T - B_T)\}}
\]

\[
= \frac{(2\pi)^{-\frac{(n/2)}{2}} |V_T|^{1/2} \exp\{-0.5[\beta_T'\nu_T^{-1}\beta_T + (z - x\beta_T)'(z - x\beta_T)]\}}{|\nu_T|^{1/2} \exp\{-0.5(\beta_T - B_T)'V_T^{-1}(\beta_T - B_T)\}}
\]
We then have that

\[
\pi(\mathbf{z} | \gamma^*, \lambda) = \frac{\pi(\gamma) V_{\gamma}^{1/2} | \mathbf{v}_{\gamma} |^{1/2} \exp(-0.5(\mathbf{z} - \mathbf{B}_{\gamma} V_{\gamma}^{-1} \mathbf{B}_{\gamma}'))}{\pi(\mathbf{z} | \gamma, \lambda) \frac{\mathbf{v}_{\gamma}^{1/2}}{|\mathbf{v}_{\gamma}^{1/2}|} \exp(-0.5(\mathbf{z} - \mathbf{B}_{\gamma} V_{\gamma}^{-1} \mathbf{B}_{\gamma}'))}
\]

The expression for the acceptance probability is therefore,

\[
z = \min \left\{ 1, \frac{\pi(\mathbf{z} | \gamma^*, \lambda)}{\pi(\mathbf{z} | \gamma, \lambda)} \frac{|\mathbf{v}_{\gamma}^{1/2}|}{|\mathbf{v}_{\gamma^*}^{1/2}|} \frac{\exp(0.5(\mathbf{B}_{\gamma} V_{\gamma}^{-1} \mathbf{B}_{\gamma}'))}{\exp(0.5(\mathbf{B}_{\gamma^*} V_{\gamma^*}^{-1} \mathbf{B}_{\gamma^*}'))} \right\}
\]