Adaptive MCMC scheme for Bayesian Variable Selection in Binary and Time-to-Event Endpoints via data augmentation algorithm

Abstract

Advances in technology have enabled thousands of variations in genes to be sequenced simultaneously, providing insights into disease etiology and discovery. Such high-throughput data continue to challenge statisticians and researchers to develop novel techniques. Modern statistical analyses such as regression models have been widely applied to determine which biomarkers are significantly correlated with outcomes of interest. In this paper, we develop an adaptive MCMC algorithm that can be applied to a general class of logistic regression models. The logistic regression model is a fundamental tool used in genetic data analysis to model binary outcomes. The logistic regression model can be used to link a categorical outcome to the linear predictors. The logistic regression model is the basis for many statistical methods used in medical research, including association studies, case-control studies, and survival analysis. The logistic regression model is widely used in the medical field because it allows researchers to quantify the relationship between a binary outcome and one or more predictors. The logistic regression model is also used in the area of cancer research to identify genes that are associated with the development or progression of cancer. The logistic regression model is used to calculate the probability of an event occurring given the values of the predictors. The logistic regression model is used to calculate the probability of an event occurring given the values of the predictors.

Introduction

Griffiths et al. (2014) developed their method in the logistics regression framework. To extend their work further to logistic regression models or accelerated failure time models. The logistic regression model is preferred because the marginal likelihood of the logistic regression model is easier to calculate than the marginal likelihood of another model. The logistic regression model is also easy to interpret. The logistic regression model is used in the area of cancer research to identify genes that are associated with the development or progression of cancer. The logistic regression model is used to calculate the probability of an event occurring given the values of the predictors. The logistic regression model is used to calculate the probability of an event occurring given the values of the predictors. The logistic regression model is used to calculate the probability of an event occurring given the values of the predictors. The logistic regression model is used to calculate the probability of an event occurring given the values of the predictors. The logistic regression model is used to calculate the probability of an event occurring given the values of the predictors.

Logistic Regression model for binary outcome

The logistic regression model can be used to link a categorical outcome to the predictors using a generalized linear model. The logistic regression model is used to calculate the probability of an event occurring given the values of the predictors. The logistic regression model is used to calculate the probability of an event occurring given the values of the predictors. The logistic regression model is used to calculate the probability of an event occurring given the values of the predictors.

Application

The methods were applied to two datasets: a Medulloblastoma datasets for binary outcome, and a breast cancer datasets for survival outcome.

Medulloblastoma data

The Medulloblastoma data has been studied extensively in the literature. Several studies have been conducted to identify genes that are associated with the development or progression of Medulloblastoma. The Medulloblastoma dataset consists of gene expression levels for Medulloblastoma tumors. The gene expression levels for Medulloblastoma tumors were obtained from the Gene Expression Omnibus (GEO) database.

Table 1 – Medulloblastoma data: genes associated with survival

<table>
<thead>
<tr>
<th>Gene</th>
<th>Expression</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEGP1</td>
<td>1.1782</td>
<td>0.118</td>
</tr>
</tbody>
</table>

Breast cancer data

We consider each patient’s failure time as the outcome of interest (as in Sha et al. (2010)). Patient with no clinical information available are considered censored cases. This differs from the data in Veer et al. (2002), where they treated all non-expressor cases as a special group.

The gene expression levels were monitored using the channel arrays with 35,288 probes. The expression levels of genes were compared using the intensity levels with respect to a reference post obtained by combining RNA samples from each tumor.

Two patients had several missing gene expression levels and were removed from further analysis. The gene expression levels were monitored over 10 years. We apply the proposed method to the intensity levels with respect to the reference post obtained by combining RNA samples from each tumor.

Table 2 – Breast cancer data: genes associated with time to metastasis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Expression</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEGP1</td>
<td>1.1782</td>
<td>0.118</td>
</tr>
</tbody>
</table>

Conclusions

We have shown that the method developed by Griffiths et al. (2014) and the time-to-event endpoints framework via data augmentation algorithm. The method is useful for identifying genes that are associated with clinical outcomes with > 95% reliability in a reasonable amount of time.

The algorithm with RLS methods constitutes censored cases. This is differ from van’t Veer et al. (2002), where they treated cases with non-expressor cases as a special group. The method is useful for identifying genes that are associated with clinical outcomes with > 95% reliability in a reasonable amount of time.

Acknowledgements

We thank Stanley Bailey and Mike Wiltshire for their discussions.

Foster presented at International Society for Bayesian Analysis (ISBA) 15th Annual Meeting, 2019, Edinburgh, United Kingdom.

References


Bhattacharya and Wiltshire for their discussions.

Foster presented at International Society for Bayesian Analysis (ISBA) 16th Annual Meeting, 2020, Rio de Janeiro, Brazil.

An adaptive MCMC scheme for Bayesian Variable Selection in Binary and Time-to-Event Endpoints via data augmentation algorithm

Author(s): Kitty Yuen W. Wai, Douglas Robinson and Jim Griffin

Affiliations: *Novartis Pharma AG, Basel; **Novartis Pharmaceuticials Corporation, Cambridge, United States; †School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, United Kingdom.

Website to Download: https://www.novartis.com/uk