Journal of Statistical Planning and Inference 180 (2017) 30–48

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

The value of information for correlated GLMs

Evangelos Evangelou [a,*], Jo Eidsvik [b]

[a] Department of Mathematical Sciences, University of Bath, Bath, UK
[b] Department of Mathematical Sciences, NTNU, Trondheim, Norway

ARTICLE INFO

Article history:
Received 28 June 2015
Received in revised form 30 May 2016
Accepted 23 August 2016
Available online 28 August 2016

Keywords:
Decision analysis
Generalised linear mixed model
Laplace approximation
Sampling design
Value of information

ABSTRACT

We examine the situation where a decision maker is considering investing in a number of projects with uncertain revenues. Before making a decision, the investor has the option to purchase data which carry information about the outcomes from pertinent projects. When these projects are correlated, the data are informative about all the projects. The value of information is the maximum amount the investor would pay to acquire these data.

The problem can be seen from a sampling design perspective where the sampling criterion is the maximisation of the value of information minus the sampling cost. The examples we have in mind are in the spatial setting where the sampling is performed at spatial coordinates or spatial regions.

In this paper we discuss the case where the outcome of each project is modelled by a generalised linear mixed model. When the distribution is non-Gaussian, the value of information does not have a closed form expression. We use the Laplace approximation and matrix approximations to derive an analytical expression to the value of information, and examine its sensitivity under different parameter settings and distributions. In the Gaussian case the proposed technique is exact. Our analytical method is compared against the alternative Monte-Carlo method, and we show similarity of results for various sample sizes of the data. The closed form results are much faster to compute. Model weighting and bootstrap are used to measure the sensitivity of our analysis to model and parameter uncertainty. A general guidance on making decisions using our results is offered.

Application of the method is presented in a spatial decision problem for treating the Bovine Tuberculosis in the United Kingdom, and for rock fall avoidance decisions in a Norwegian mine.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

One goal of statistical modelling and methodology is to provide useful inputs for decision making under uncertainty. The planning and evaluation of various data acquisition schemes for making improved decision is also a field where statistics is expected to contribute. We apply value of information (VOI) analysis to study when a data set is likely to help us make sufficiently better decisions, i.e. whether it is worthwhile acquiring. We also use VOI analysis for the comparison of various possible experiments. The VOI is a monetary amount, which is computed from the statistical model as well as the costs and revenues of the decision situations. A recent review of decision analysis is given in Howard and Abbas (2015).

We consider the situation with dependent projects having uncertain profits. In our applications the projects will be associated with spatial coordinates, and their correlation depends on the distance between projects. Eidsvik et al. (2015)

* Corresponding author.
E-mail address: ee224@bath.ac.uk (E. Evangelou).

http://dx.doi.org/10.1016/j.jspi.2016.08.005
0378-3758/© 2016 Elsevier B.V. All rights reserved.

present a framework for VOI analysis in this spatial context. Our methods also work for other kinds of dependence. We assume that the decision maker freely selects projects with positive expected monetary value. Initially, the investor has prior knowledge about the outcome of the projects, including dependence, and the overall prior value of projects. There is much at stake, and one can purchase some data before making the decisions. With the option to purchase some data, the posterior value of projects can be computed. When the projects are correlated, the data will be informative of the probability distribution of all projects. The VOI is the difference between the expected posterior value averaged over all possible data sets, and the prior value.

A typical example of this situation is presented in Section 6.2. In this example a mine operator is considering adding rock support at selected locations to avoid rock fall. The support will ensure that the rock will not fall but comes with the cost of equipment and labour. Without the rock support, a rock fall will cause loss of revenue. To assess the likelihood of rock fall, the mining operator can collect data at a number of spatial locations. The number of rock joints counted at those locations is a measure of the rock strength and is modelled by a Poisson spatial model. However, the data are not free and different sampling schemes are considered. VOI analysis can be used to compare sampling schemes for different price ranges, and for various statistical models and/or parameter settings. It then forms a solid basis for management who is making information gathering decisions in the light of budgets and resources.

Mathematically speaking, we consider the set $\mathbb{S}$ of spatial projects. The latent variable of interest is denoted x_s, $s \in \mathbb{S}$. We allow for the components of $\boldsymbol{X} = \{x_s, s \in \mathbb{S}\}$ to be correlated and normally distributed. The decision is tied to this variable. For the case where the distribution of $\boldsymbol{X}$ is categorical, see Bhattacharjya et al. (2010). The potential outcomes of experiments are denoted y_s, $s \in \mathbb{S}$. The distribution of y_s is defined to be conditionally independent of the outcomes of the other experiments with mean $g(x_s)$ where $g(\cdot)$ denotes the inverse link function. In the examples discussed in this paper the outcome of each experiment is either binary or a count variable. The generalised linear model (GLM) is used for modelling data of this type where the response y is then assumed to follow a conditional distribution in the form of the exponential family.

Suppose that the cost of making a decision at any site s is C_s, while the revenue is a fixed amount R_s times the expectation of the binary or count variable. When no data are available, the prior value (PV) is

$$\mathrm{PV}(\mathbb{S}) = \sum_{s\in\mathbb{S}} \max\{0, R_s \times \mathsf{E}_x g(x_s) - C_s\}, \tag{1}$$

i.e. a risk-neutral decision maker selects site s if its expected profit is positive, otherwise the decision maker avoids this site. The decision maker is free to select as many sites as are profitable, thus the sum over all sites. Note that in some situations the objective is to maximise the negative loss, rather than the revenues.

Now suppose that there is the potential of obtaining data $\boldsymbol{y}$ corresponding to a collection of spatial experiments S. We assume that the data from each experiment at $s \in S$ consists of the total over m_s replications of the experiment. In the context of exponential families, m_s would denote the number of trials in a binomial experiment or the time length, area or volume for Poisson responses. The resulting data y_s are informative of the latent variable x_s. Under these circumstances the posterior value (PoV) for the experiments $\mathbb{S}$ is

$$\mathrm{PoV}(\mathbb{S}|S) = \mathsf{E}_y \sum_{s\in\mathbb{S}} \max\{0, R_s \times \mathsf{E}_x[g(x_s)|\boldsymbol{y}] - C_s\}. \tag{2}$$

The difference of (2) from (1) is the VOI provided by the experiments S, i.e.

$$\mathrm{VOI}(\mathbb{S}|S) = \mathrm{PoV}(\mathbb{S}|S) - \mathrm{PV}(\mathbb{S}). \tag{3}$$

It can be shown by an application of Jensen's inequality that $\mathrm{VOI}(\mathbb{S}|S) \geq 0\ \forall S$. Thus, there is always the incentive of collecting more data. However, one must weight this information against its cost so accurate calculation of (3) is important for planning purposes. Moreover, when the optimal experiment set S is sought, these calculations need to be quick. From a computational point of view, calculation of (1) is straightforward and in some cases it can be written in closed-form. The calculation of (2) is more difficult due to the intractable conditional expectation inside the maximum, and the outer expectation over the data.

The case where the outcome of each experiment is normally distributed has been studied by Bhattacharjya et al. (2013). The contribution of the current paper is to extend these results to the general exponential family case. We also consider the risk of decisions and suggest methods to account for model and parameter uncertainty. In some sense the context is similar to that of spatial design. This is usually done based on entropy, see e.g. Fuentes et al. (2007), prediction variance, see e.g. Evangelou and Zhu (2012), or prediction error, see e.g. Peyrard et al. (2013). The main difference between these measures of information and VOI analysis is that the latter is based on decision theoretic concepts and directly tied to monetary units. The VOI analysis is commonly done for medicine and health, see e.g. Baio (2012), and in the context of conservation biology, see Moore and McCarthy (2010); Moore and Runge (2012), but this has not been done in the setting with spatial decisions and latent models incorporating dependence and GLM likelihoods. Analytical expressions can also be useful in sequential decision problems (Morgan and Cressie, 1997). The contribution of our paper is to formulate analytical results for the large class of hierarchical GLMs.

The remaining parts of the paper are organised as follows. Section 2 presents some pertinent asymptotic results for the conditional mean and variance of the latent process. These results are used in Section 3 to derive the approximation to the VOI for different models. Section 4 presents methods for dealing with model and parameter uncertainty. In Section 5 we

Download English Version:

https://daneshyari.com/en/article/5129606

Download Persian Version:

https://daneshyari.com/article/5129606

Daneshyari.com