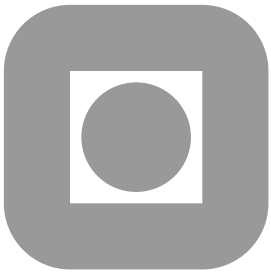
NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET

The Value of Information for Correlated GLMs

by

Evangelos Evangelou and Jo Eidsvik

PREPRINT
STATISTICS NO. 2/2015



NORWEGIAN UNIVERSITY OF SCIENCE AND TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2015/S2-2015.pdf
Jo Eidsvik has homepage: http://www.math.ntnu.no/∼joeid
E-mail: joeid@stat.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

The Value of Information for Correlated GLMs

Evangelos Evangelou[1] and Jo Eidsvik[2]

1. Department of Mathematical Sciences – University of Bath, Bath, UK
2. Department of Mathematical Sciences – NTNU, Trondheim, Norway

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.

In this paper we discuss the case where the outcome of each project is modelled by an exponential family. 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. 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.

Keywords: Decision analysis; Generalised linear mixed model; Laplace approximation; Sampling design; Value of Information.

1 Introduction

In many situations one must make decisions under uncertainty. One goal of statistical modeling and methodology is to help resolve difficult decision problems. 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 whether 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

Date: March 26, 2015

the statistical model as well as the costs and revenues of the decision situations. A recent review of decision analysis and the VOI is provided 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. Our framework also holds 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 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 at all projects. The VOI is the difference between the expected posterior value averaged over all possible data sets, and the prior value.

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. 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 $C_s = C > 0$, while the revenue is a fixed amount $R_s = R$ times the expectation of the binary or count variable. When no data are available, the prior value (PV) is

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

i.e. a risk-neutral decision maker will make the decision at site s if it is expected to have profit, otherwise the decision maker will avoid 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 maximize the negative loss, rather than the revenues.

Now suppose that there is the potential of obtaining data $\boldsymbol{y}$, where each experiment $s \in S$ could be performed multiple times, m_s say. These data are informative of the latent variable x_s and can be used to update our belief about the outcome of experiment s. Under these circumstances, our posterior value (PoV) for the experiments $\mathbb{S}$ given this potential is

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

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 $\text{VOI}(\mathbb{S}|S) \geq 0 \ \forall S$. 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 this paper is to extend these results to the general exponential family case. 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 tied to monetary units. The VOI analysis is commonly done for medical applications, see e.g. Willan and Pinto (2005) on clinical trials and Welton et al. (2008) on medical prioritisation, 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 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 of the paper is 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. In Section 4 we present computational results where we compare the proposed approximation to the Monte-Carlo method and discuss the sensitivity of our approximation to the parameters of the model. In Section 5 we illustrate our method to applications and finally, in Section 6 we present our conclusions. Some technical derivations are given in the Appendix.

2 Some asymptotic results for GLMs

We denote the latent process at S by $\boldsymbol{x} := \{x_1, \ldots, x_n\}$. Let further $\boldsymbol{\mu} := \mathsf{E}_x \, \boldsymbol{x}$ be the mean and $\Sigma := \mathsf{V}_x \, \boldsymbol{x}$ be the covariance matrix of $\boldsymbol{x}$, and write the conditionally independent distribution of $y_{ij}|x_i$ in the form

$$p(y_{ij}|x_i) \propto \exp \int_{y_{ij}}^{g_i} \frac{y_{ij} - u}{\tau^2 v(u)} \, \mathrm{d}u, \ i = 1, \ldots, n, \ j = 1, \ldots, m_i$$

where $g_i := g(x_i)$, τ^2 is called the dispersion parameter and $v(\cdot)$ is the variance function. The case $v(u) = u$ gives the Poisson distribution and the case $v(u) = u(1-u)$ gives the Bernoulli distribution, while $v(u) = 1$ is the normal distribution (McCullagh and Nelder,

1999). Let $\boldsymbol{y} = \{y_{ij},\ i = 1, \ldots, n,\ j = 1, \ldots, m_i\}$. In this section we derive a Gaussian approximation to the distribution $p(x_s|\boldsymbol{y})$ using Laplace's method.

2.1 Laplace approximation

Laplace's method approximates multidimensional integrals of the form

$$I = \int f(\boldsymbol{x}) e^{-h(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x},$$

as $h(\cdot) \to \infty$, around

$$\hat{\boldsymbol{x}} := \underset{\boldsymbol{x}}{\operatorname{argmin}} \, h(\boldsymbol{x}).$$

The first order approximation is

$$I \approx f(\hat{\boldsymbol{x}}) e^{-h(\hat{\boldsymbol{x}})} \left| \frac{1}{2\pi} \hat{H} \right|^{-1/2},$$

where $\hat{H}$ denotes the Hessian matrix of $h(\cdot)$ evaluated at $\hat{\boldsymbol{x}}$.

When the Laplace approximation is applied to ratios of integrals of the form

$$\frac{I_f}{I_1} = \frac{\int f(\boldsymbol{x}) e^{-h(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x}}{\int e^{-h(\boldsymbol{x})} \, \mathrm{d}\boldsymbol{x}},$$

the approximation to first order is

$$\frac{I_f}{I_1} \approx f(\hat{\boldsymbol{x}}). \tag{4}$$

If the dimension of $\boldsymbol{x}$ is fixed, the asymptotic error of (4) is $O(h^{-1})$ as $h(x) \to \infty$. The requirement $h(x) \to \infty$ is equivalent to $m_i \to \infty$ for all i in our setting. The case where $n \to \infty$ has been studied in Shun and McCullagh (1995) and Evangelou et al. (2011) who showed that the approximation error for the geo-spatial case becomes $O(nh^{-1})$ to the first order.

The Laplace approximation is a consequence of the Gaussian approximation to $e^{-h(\boldsymbol{x})}$. In particular, application of second order Taylor expansion to $h(\boldsymbol{x})$ around $\hat{\boldsymbol{x}}$ gives

$$e^{-h(\boldsymbol{x})} = e^{-h(\hat{\boldsymbol{x}})} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \hat{\boldsymbol{x}})^{\intercal} \hat{H} (\boldsymbol{x} - \hat{\boldsymbol{x}}) \right\},$$

so if $e^{-h(\boldsymbol{x})}$ represents a pdf, then it can be approximated by the Gaussian pdf with mean $\hat{\boldsymbol{x}}$ and variance $\hat{H}^{-1}$.

2.2 Gaussian approximation to the conditional distribution of $\boldsymbol{x}|\boldsymbol{y}$

Consider the conditional distribution of $\boldsymbol{x}|\boldsymbol{y}$. This distribution is in general not available in closed-form. A Gaussian approximation to this distribution is derived using

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x}) p(\boldsymbol{x}) = p(\boldsymbol{x}, \boldsymbol{y}),$$

where $p(\boldsymbol{y}|\boldsymbol{x}) = \prod p(y_{ij}|x_i)$ and $p(\boldsymbol{x})$ is the multivariate normal pdf with mean $\boldsymbol{\mu}$ and variance Σ. To that end, let

$$\hat{\boldsymbol{x}} := \underset{\boldsymbol{x}}{\operatorname{argmax}}\, p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}),$$

and $\hat{H} := \Sigma^{-1} + \hat{D}$ denotes the negative Hessian of $\log p(\boldsymbol{y}, \boldsymbol{x})$ with respect to $\boldsymbol{x}$ evaluated at $\hat{\boldsymbol{x}}$. The matrix D is diagonal with ith element $m_i v(g_i)\tau^{-2}$ if a canonical link is used.

Then, an approximation to the mean and variance of $\boldsymbol{x}|\boldsymbol{y}$ is

$$\begin{aligned} \mathsf{E}_x[\boldsymbol{x}|\boldsymbol{y}] &\approx \hat{\boldsymbol{x}} \\ \mathsf{V}_x[\boldsymbol{x}|\boldsymbol{y}] &\approx \hat{H}^{-1}. \end{aligned} \tag{5}$$

This motivates approximation of the conditional distribution of $\boldsymbol{x}|\boldsymbol{y}$ by the normal distribution with mean and variance given by (5), i.e

$$\boldsymbol{x}|\boldsymbol{y} \sim N_n(\hat{\boldsymbol{x}}, \hat{H}^{-1}). \tag{6}$$

Using the result in (6), we can predict x_s at any given experiment s. Let $\boldsymbol{c}_s$ denote the covariance between x_s and $\boldsymbol{x}$. Then,

$$\begin{aligned} \kappa_s &:= \mathsf{E}_x[x_s|\boldsymbol{x}] = \mu_s + \boldsymbol{c}_s^{\mathsf{T}}\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu}), \\ \xi_s^2 &:= \mathsf{V}_x[x_s|\boldsymbol{x}] = \sigma_s^2 - \boldsymbol{c}_s^{\mathsf{T}}\Sigma^{-1}\boldsymbol{c}_s, \\ \nu_s &:= \mathsf{E}_x[x_s|\boldsymbol{y}] \approx \mu_s + \boldsymbol{c}_s^{\mathsf{T}}\Sigma^{-1}(\hat{\boldsymbol{x}} - \boldsymbol{\mu}). \end{aligned} \tag{7}$$

The notation $\approx$ will be used here to denote the first order Laplace approximation.

Since the mean and variance in (5) depend on $\boldsymbol{y}$ only through $\hat{\boldsymbol{x}}$,

$$\begin{aligned} \mathsf{E}_x[\boldsymbol{x}|\hat{\boldsymbol{x}}] &\approx \hat{\boldsymbol{x}} \\ \mathsf{V}_x[\boldsymbol{x}|\hat{\boldsymbol{x}}] &\approx \hat{H}^{-1}. \end{aligned} \tag{8}$$

By an application of the law of iterated expectations on the left and right-hand sides of (8) we have

$$\begin{aligned} \boldsymbol{\mu} = \mathsf{E}_x\, \boldsymbol{x} &= \mathsf{E}_{\hat{x}}[\mathsf{E}_x[\boldsymbol{x}|\hat{\boldsymbol{x}}]] \approx \mathsf{E}_{\hat{x}}[\hat{\boldsymbol{x}}] \\ \Sigma = \mathsf{V}_x\, \boldsymbol{x} &= \mathsf{V}_{\hat{x}}\, \mathsf{E}_x[\boldsymbol{x}|\hat{\boldsymbol{x}}] + \mathsf{E}_{\hat{x}}\, \mathsf{V}_x[\boldsymbol{x}|\hat{\boldsymbol{x}}] \approx \mathsf{V}_{\hat{x}}\, \hat{\boldsymbol{x}} + \mathsf{E}_{\hat{x}}\, \hat{H}^{-1} \\ &\Rightarrow \mathsf{V}_{\hat{x}}\, \hat{\boldsymbol{x}} \approx \Sigma - \mathsf{E}_{\hat{x}}\, \hat{H}^{-1} = \Sigma\, \mathsf{E}_{\hat{x}}(\Sigma + \hat{D}^{-1})^{-1}\Sigma =: \Psi, \end{aligned} \tag{9}$$

where in the last line we used $\hat{H}^{-1} = (\Sigma^{-1} + \hat{D})^{-1} = \Sigma - \Sigma(\Sigma + \hat{D}^{-1})^{-1}\Sigma$. Asymptotically, the distribution of $\hat{\boldsymbol{x}}$ is the n-dimensional multivariate normal with mean $\boldsymbol{\mu}$ and variance Ψ. Note that the elements of $\hat{D}^{-1}$ are negligible for large m_i. In this case, the matrix Ψ is approximated by $\Psi \simeq \Sigma(\Sigma + K)^{-1}\Sigma$ where $K = E_x D^{-1}$. Applying this result to (7), we have

$$\begin{aligned} \mathsf{E}_y[\nu_s] &\approx \mu_s, \\ \mathsf{V}_y[\nu_s] &\approx \boldsymbol{c}_s^{\mathsf{T}}(\Sigma + K)^{-1}\boldsymbol{c}_s =: \chi_s^2. \end{aligned} \tag{10}$$

Equation (10) is the main result of this section and is used for the approximation of VOI as we show next.

3 Approximation to the VOI

In this section we show how the results from section 2 are used to derive the contribution VOI($s|S$) of a single site s to the VOI.

Consider first the expectation $\mathsf{E}_x[g(x_s)|\boldsymbol{y}]$ and define

$$M_g(\kappa_s, \xi_s) = \mathsf{E}_x[g(x_s)|\boldsymbol{x}].$$

Then, by an application of (4),

$$\begin{aligned} \mathsf{E}_x[g(x_s)|\boldsymbol{y}] &= \mathsf{E}_x[\mathsf{E}_x[g(x_s)|\boldsymbol{x}]|\boldsymbol{y}] \\ &= \mathsf{E}_x[M_g(\kappa_s, \xi_s)|\boldsymbol{y}] \\ &\approx M_g(\nu_s, \xi_s). \end{aligned} \tag{11}$$

By combining (1), (2), (3) and (11), we have

$$\begin{aligned} \text{VOI}(s|S) &= \mathsf{E}_y \max\{0, R \times \mathsf{E}_x[g(x_s)|\boldsymbol{y}] - C\} - \max\{0, R \times \mathsf{E}_y[\mathsf{E}_x\, g(x_s)|\boldsymbol{y}] - C\} \\ &\approx \mathsf{E}_y \max\{0, R \times M_g(\nu_s, \xi_s) - C\} - \max\{0, R \times \mathsf{E}_y\, M_g(\nu_s, \xi_s) - C\} \\ &= \mathsf{E}_{\nu_s} \max\{0, R \times M_g(\nu_s, \xi_s) - C\} - \max\{0, R \times \mathsf{E}_{\nu_s}\, M_g(\nu_s, \xi_s) - C\}. \end{aligned} \tag{12}$$

The last expectation is with respect to the distribution of ν_s which from (10) can be taken to be $\nu_s \sim N(\mu_s, \chi_s^2)$. This result can be readily applied to the different distributions considered.

3.1 Normal-identity model

We consider first the case where $y_s|x_s$ is normally distributed with variance τ^2 and $g(x) = x$ so the variance function $v(g) = 1$. Then $K = \text{diag}\{\tau^2/m_s, s \in S\}$ and $M_g(\kappa_s, \xi_s) = \kappa_s$. This gives, for $a = C/R$,

$$\text{VOI}(s|S) = R\chi_s \phi\left(\frac{\mu_s - a}{\chi_s}\right) + R(\mu_s - a)\Phi\left(\frac{\mu_s - a}{\chi_s}\right) - R\max\{0, \mu_s - a\}.$$

Note that in this case the approximation is exact.

Based on the closed form expression one can easily gauge the effect of input parameters on the VOI. For instance, when $\mu_s \to \pm\infty$, the Gaussian density $\phi\left(\frac{\mu_s - a}{\chi_s}\right) \to 0$. The cumulative function $\Phi\left(\frac{\mu_s - a}{\chi_s}\right)$ goes to 0 or 1 in these cases, and the posterior value cancels with the prior value $R\max\{0, \mu_s - a\}$. Thus, the VOI goes towards zero for very low or high values of the prior mean. Data will not help us make better decisions for extreme prior means. For intermediate values of the prior mean parameter the data will likely help us in the decision making and the VOI is positive.

3.2 Poisson-log model

In this case $g(x) = e^x$, and its expectation becomes $M_g(\kappa_s, \xi_s) = \exp(\kappa_s + \frac{1}{2}\xi_s^2)$. For the Poisson model $v(g) = g$, and we get $K = \text{diag}\{\tau^2/m_s \exp(-\mu_s + \frac{1}{2}\sigma_s^2), s \in S\}$. Then, for $a = \log(C/R)$ and using Lemma 1 in the Appendix,

$$\begin{aligned}
\text{VOI}(s|S) &= \mathsf{E}_{\nu_s} \max\left\{0, R \times \exp\left(\nu_s + \frac{1}{2}\xi_s^2\right) - C\right\} - \max\left\{0, R \times \mathsf{E}_{\nu_s} \exp\left(\nu_s + \frac{1}{2}\xi_s^2\right) - C\right\} \\
&= R \exp\left(\mu_s + \frac{1}{2}\xi_s^2 + \frac{1}{2}\chi_s^2\right) \Phi\left(\chi_s + \frac{\mu_s + \frac{1}{2}\xi_s^2 - a}{\chi_s}\right) - Re^a \Phi\left(\frac{\mu_s + \frac{1}{2}\xi_s^2 - a}{\chi_s}\right) \\
&\quad - R \max\left\{0, \exp\left(\mu_s + \frac{1}{2}\xi_s^2 + \frac{1}{2}\chi_s^2\right) - e^a\right\}. \qquad (13)
\end{aligned}$$

The closed form facilitates interpretation. When the prior mean μ_s gets large, the cumulative functions in (13) go to 1. This means the VOI goes to 0. The variance χ_s^2 is influenced by the correlation in the model. We have $\chi_s = 0$ if the outcome at site s is independent of the data. In this case the cumulative functions again go to either 0, 0.5, or 1, depending on whether $\mu_s + \frac{1}{2}\xi_s^2 - a$ is negative, zero or positive, and the VOI contribution at s becomes 0.

3.3 Binomial-logit model

In this case we need $C < R$ otherwise the problem becomes trivial. For $v(g) = g(1-g)$, we have, by an application of Lemma 2 in the Appendix,

$$K = \text{diag}\left\{\frac{\tau^2}{m_s}\left(2 + \exp\left(-\mu_s + \sigma_s^2/2\right) + \exp\left(\mu_s + \sigma_s^2/2\right)\right), s \in S\right\}.$$

The inverse link function is $g(x) = (1+e^{-x})^{-1}$. Then, $M_g(\kappa_s, \xi_s^2) \approx g\left(\kappa_s/\sqrt{1+\alpha^2\xi_s^2}\right)$. This approximation uses the Gaussian approximation to the logistic-normal integral (Demidenko, 2004), derived in the Appendix (section 7.2).

For $a = \log((C/R)/(1-(C/R)))$ we have

$$\begin{aligned}
\text{VOI}(s|S) &= \mathsf{E}_{\nu_s} \max\left\{0, R \times g\left(\nu_s/\sqrt{1+\alpha^2\xi_s^2}\right) - C\right\} - \max\left\{0, R \times \mathsf{E}_{\nu_s}\, g\left(\nu_s/\sqrt{1+\alpha^2\xi_s^2}\right) - C\right\} \\
&= R\Lambda_a\left(\frac{\mu_s}{\sqrt{1+\alpha^2\xi_s^2}}, \frac{\chi_s^2}{1+\alpha^2\xi_s^2}\right) - Rg(a)\Phi\left(\frac{\mu_s - a\sqrt{1+\alpha^2\xi_s^2}}{\chi_s}\right) \\
&\quad - R \max\left\{0, \Lambda(\mu_s, \xi_s^2 + \chi_s^2) - g(a)\right\},
\end{aligned}$$

where $\Lambda(\cdot)$ and $\Lambda_a(\cdot)$ denote the complete and incomplete logistic-normal integrals. See Appendix (section 7.2). As in the other cases at the limits $\mu \to +\infty$ and $\mu \to -\infty$ the functions Λ, Λ_a and Φ tend to 1 or 0 respectively and the VOI tends to 0.

4 Computational Experiments

In this section we compare the approximations to the VOI derived in section 3 against the Monte-Carlo sampling. We also perform a sensitivity analysis of the proposed approximation. The general setup consists of the spatial domain $[0,1]^2$ with the possible experiments consisting of the $n = 25$ pairs $\mathbb{S} = \{s_{ij} = (\frac{i}{4}, \frac{j}{4})\ i,j = 0,1,\ldots,4\}$. We define $S = \{s_{ij} : i,j \text{ odd}\}$ and compute $\text{VOI}(s_{ij}|S)$ with $C_s = C = 0.5$ and $R = 1$.

The latent component $\boldsymbol{x}$ is assumed to have mean at location s_{ij} equal to $\mu_{ij} = -1 + (i+j)/4$ and variance-covariance matrix $\sigma^2 R(\rho)$ where $R(\rho)$ is the matrix whose elements are of the form $\exp(-\rho\|s_{ij} - s_{i'j'}\|)$. Larger values of the parameter ρ decrease the correlation between experiments.

The outcome of each experiment is taken to be from the exponential family. We consider the Gaussian, Poisson and binomial cases, with m replications and dispersion parameter τ^2.

4.1 Comparison with Monte-Carlo

In this section we fix $\sigma^2 = 10$, $\tau^2 = 1$, $\rho = 0.6$ and perform computations for $m = 10^b$, $b = 0,\ldots,4$. The Monte-Carlo method was implemented as follows:

1. Sample N_O times $\boldsymbol{y}_S^{(i)} \sim p(\boldsymbol{y}_S)$ on S. This is done in two steps, first a sample $\boldsymbol{x}_S^{(i)} \sim p(\boldsymbol{x}_S)$ on S is taken and then $\boldsymbol{y}_S^{(i)} \sim p(\boldsymbol{y}_S|\boldsymbol{x}_S^{(i)})$.

2. For $i = 1,\ldots,N_O$

   Compute a Monte-Carlo approximation $A_{\mathbb{S}}^{(i)}$ to the expectation $\mathsf{E}_{\boldsymbol{x}}[g(\boldsymbol{x}_{\mathbb{S}})|\boldsymbol{y}_S^{(i)}]$. This is computed using importance sampling with N_I samples and proposal distribution equal to the Gaussian approximation to $p(\boldsymbol{x}_{\mathbb{S}}|\boldsymbol{y}_S^{(i)})$.

3. Approximate the VOI by

$$\text{VOI}(s|S) \simeq \frac{1}{N_O}\sum_{i=1}^{N_O} \max\left\{0, R \times A_s^{(i)} - C\right\} - \max\left\{0, R \times \left[\frac{1}{N_O}\sum_{i=1}^{N_O} A_s^{(i)}\right] - C\right\}$$

   for $s \in \mathbb{S}$.

For our computations we used $N_O = N_I = 10^4$ samples for all cases.

Figure 1 shows the square-root mean square difference between the analytical approximation to the VOI and the Monte-Carlo approximation for each of the three distributions considered. As the analytical approximation is exact for the Gaussian case, that case indicates the increase in the Monte-Carlo error as m increases. This is due to the larger variance of the simulated $\boldsymbol{y}$, which increases the variability of the Monte-Carlo average. For the Poisson and binomial cases the mean square difference between the two methods drops as m increases which can be explained by the improvement of the analytical approximation for large m.

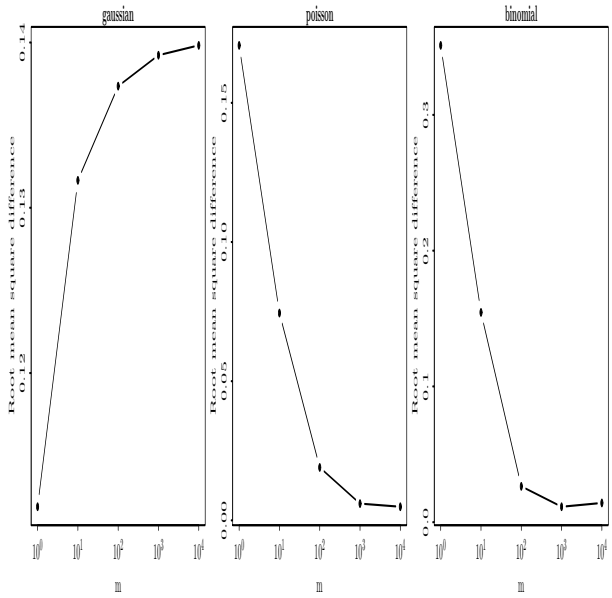


Figure 1: Comparison between the analytical and Monte-Carlo approximations to VOI.

4.2 Sensitivity analysis

In this section we fix $m = 100$ and compute the VOI as a function of the parameters σ^2, τ^2 and ρ. We choose $3 \times 3 \times 3$ combinations with $\sigma^2, \tau^2 \in \{0.1, 1, 10\}$ and $\rho \in \{0.2, 0.6, 1\}$. The analytical approximation to the VOI is computed for each combination and for the three families considered. The results are plotted in Figure 2.

The pattern corresponding to the three distributions is similar. The variance parameter σ^2 has the largest impact and τ^2 the least. The effect of the correlation parameter ρ is more apparent when the σ^2 is large. Also the VOI decreases as the sites become less correlated to each other. For the Poisson distribution we notice a relatively faster decline in VOI when ρ increases.

The case where the range of the mean $\boldsymbol{\mu}$ varies together with the other parameters was also considered but not shown. In this case the results support our interpretation in Section 3 that the effect of the mean is larger for intermediate values, when we are most indifferent and the data can be more helpful.

5 Examples

5.1 Poisson spatio-temporal model for disease pretesting

We consider the bovine tuberculosis (BTB) data collected during the years 1989 to 2002 from farms in Cornwall, UK. The data consist of the locations of infected farms found upon inspection during the fourteen-year period. The data were analysed by Diggle et al. (2005) among others.

To formulate the decision problem, we take the role of the monitoring agency that

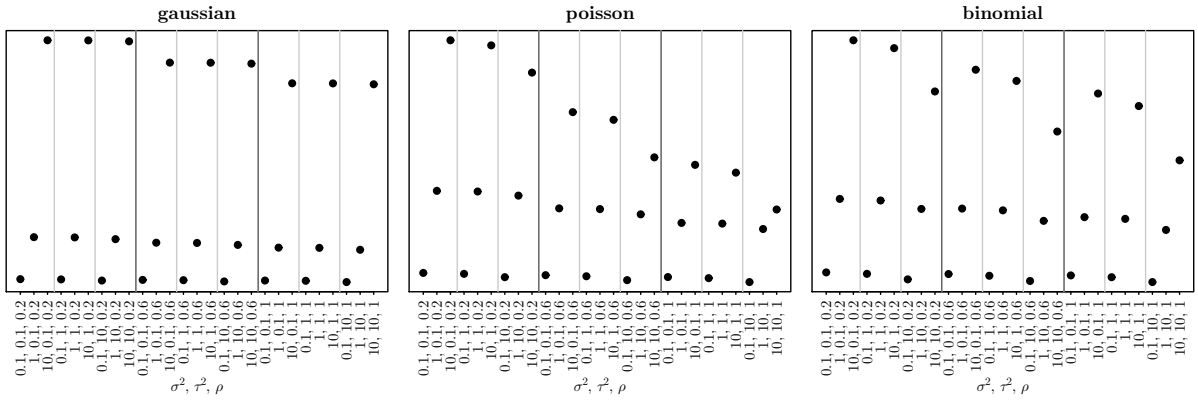


Figure 2: Results from the sensitivity study. For each distribution we plot the approximate VOI against the different parameter values considered.

decides whether to test for the disease or not. To that end, the entire spatial region is split into 90 grid cells with maximum width 8Km and maximum height 8Km as shown in Figure 3. If all cattle within a cell are inspected and all infected farms are eliminated, then that particular cell is considered "treated" for that year. Thus, the reward for treating cell s at time t (number of years since 1988) is $-C_s - R_1 y_{s,t}$, where C_s is the search cost proportional to the area of the cell s, R_1 is the loss occurring when an infected farm is found and therefore eliminated, and $y_{s,t}$ is the number of infected farms at time t in cell s. Alternatively, the agency may decide to "skip" cell s, in which case the reward is $-R_2 y_{s,t}$. We set $R_2 > R_1$ because an infected farm can incur higher losses if it remains undetected. With these rewards, the prior value for treating cell s at time t is

$$\mathrm{PV}_t(s) = \max\left\{-C_s - R_1\,\mathsf{E}_y\, y_{s,t}, -R_2\,\mathsf{E}_y\, y_{s,t}\right\}, \tag{14}$$

i.e. the agency's decision is to treat cell s if its expected loss is less than the expected loss when the cell is skipped.

Let us also suppose that, prior to treatment, the monitoring agency has the option to administer a pretest to a sample of cattle from each farm within a cell. The pretest can be used to gain information, denoted $\boldsymbol{y}$, about the distribution of the disease and help decide which cells to treat. Suppose that the cells $S = \{s_1, \ldots, s_n\}$ have been chosen for the pretest. Then, the posterior value for treating cell s at time t provided by S is

$$\mathrm{PoV}_t(s|S) = \mathsf{E}_{\boldsymbol{y}} \max\left\{-C_s - R_1\,\mathsf{E}_y[y_{s,t}|\boldsymbol{y}], -R_2\,\mathsf{E}_y[y_{s,t}|\boldsymbol{y}]\right\}. \tag{15}$$

By combining (14) and (15), the VOI for treating cell s at time t provided by the pretest S becomes

$$\begin{aligned}
\mathrm{VOI}_t(s|S) &= \mathsf{E}_{\boldsymbol{y}} \max\left\{-C_s - R_1\,\mathsf{E}_y[y_{s,t}|\boldsymbol{y}], -R_2\,\mathsf{E}_y[y_{s,t}|\boldsymbol{y}]\right\} - \max\left\{-C_s - R_1\,\mathsf{E}_y\, y_{s,t}, -R_2\,\mathsf{E}_y\, y_{s,t}\right\} \\
&= \mathsf{E}_{\boldsymbol{y}} \max\left\{-C_s + (R_2 - R_1)\,\mathsf{E}_y[y_{s,t}|\boldsymbol{y}], 0\right\} - R_2\,\mathsf{E}_{\boldsymbol{y}}\,\mathsf{E}_y[y_{s,t}|\boldsymbol{y}] \\
&\qquad - \max\left\{-C_s + (R_2 - R_1)\,\mathsf{E}_y\, y_{s,t}, 0\right\} + R_2\,\mathsf{E}_y\, y_{s,t} \\
&= \mathsf{E}_{\boldsymbol{y}} \max\left\{0, (R_2 - R_1)\,\mathsf{E}_y[y_{s,t}|\boldsymbol{y}] - C_s\right\} - \max\left\{0, (R_2 - R_1)\,\mathsf{E}_y\, y_{s,t} - C_s\right\}.
\end{aligned} \tag{16}$$

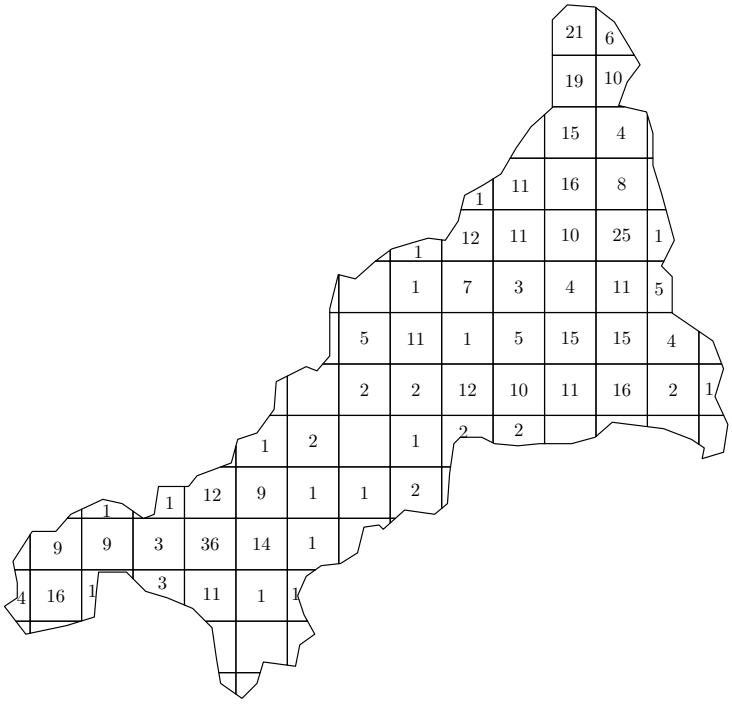


Figure 3: Sampling grid for the BTB example. The numbers show the total number of infected farms in that cell across the years 1989 to 1998; the empty cells correspond to zero counts.

We now specify our modelling framework. Let $x_{s,t}$ denote the logarithmic disease intensity at cell s at time t, $s \in \mathbb{S}$, $t \in \mathbb{T}$. For the purposes of this example, we model $\{x_{s,t}\}$ as a separable spatio-temporal Gaussian process with constant mean, i.e.

$$x_{s,t} = \beta_0 + \eta_s + \epsilon_t,$$

where $\{\eta_s\}$ is a spatial conditional autoregressive process (CAR) on a square lattice (Cressie, 1993, Section 6.3.2) and $\{\epsilon_t\}$ is a temporal CAR process. Specific details on the spatial and temporal model are given in the Appendix. Conditional on $x_{s,t}$, the number of infected farms $y_{s,t}$ within cell s at time t is Poisson distributed with mean $m_s e^{x_{s,t}}$ where m_s denotes the area of cell s divided by 64 in Km^2. The cost C_s for cell s is also set to $C_s = m_s$ while the difference in revenue $R_2 - R_1 = 1$. Then, the VOI from (16) becomes

$$\text{VOI}_t(s|S) = m_s \, \mathsf{E}_{\boldsymbol{y}} \max\{0, \mathsf{E}_x[\exp(x_{s,t})|\boldsymbol{y}] - 1\} - m_s \max\{0, \mathsf{E}_x \exp(x_{s,t}) - 1\}.$$

For any given year t, we assume that all data prior to that year were observed and use them to estimate the parameters of our model by maximum likelihood. Given parameter estimates, the plug-in predictive distribution of $\boldsymbol{x}_t$ is the normal distribution with mean and variance given by (19) in the Appendix.

We assume that the agency is able to pretest a total of $n = 9$ cells (10% of the total). The choice of the pretest cells is done as follows. The posterior value for pretesting each cell is calculated and the cell $S_1 = \{s_1\}$ that corresponds to the highest $\text{VOI}(\mathbb{S}|S_1)$ is selected. The remaining 89 cells are searched again to obtain $S_2 = \{s_1, s_2\}$ which corresponds to the highest $\text{VOI}(\mathbb{S}|S_2)$. This procedure is repeated until we obtain $S_9 = \{s_1, \ldots, s_9\}$. More generally, the choice of the pretest locations can also be seen as a spatial design problem.

After the pretest locations are selected, the pretest is administered and the corresponding cells are observed. The data from the pretest cells are augmented with the existing data and the model is refitted and a new prior value for each cell is computed. Figure 4 shows the treatment scheme before and after the pretest. The middle column of Figure 4 shows that the strategy for identifying pretest cells tends to give tests near borders of the gray-white zones, i.e. zones where we are most indifferent and additional information would assist the decision making. In some cases pretesting increases the number of treated cells and in others it is reduced.

5.2 Poisson for joint counts affecting stability in mining

We consider decisions related to rock support in mining operations, where one would avoid rock fall. The strength of the rock mass depends on a number of attributes such as joint intensity, rock mechanical properties, fluid components, faulting, and so on, see e.g. Nilsen et al. (2003). The joints of the rocks are critical here, and it is the focus of our example from a mine in Norway (Ellefmo and Eidsvik, 2009).

Figure 5 shows the joint frequency data set. Ellefmo and Eidsvik (2009) analysed the data using a Poisson likelihood model and a Gaussian latent log intensity. The

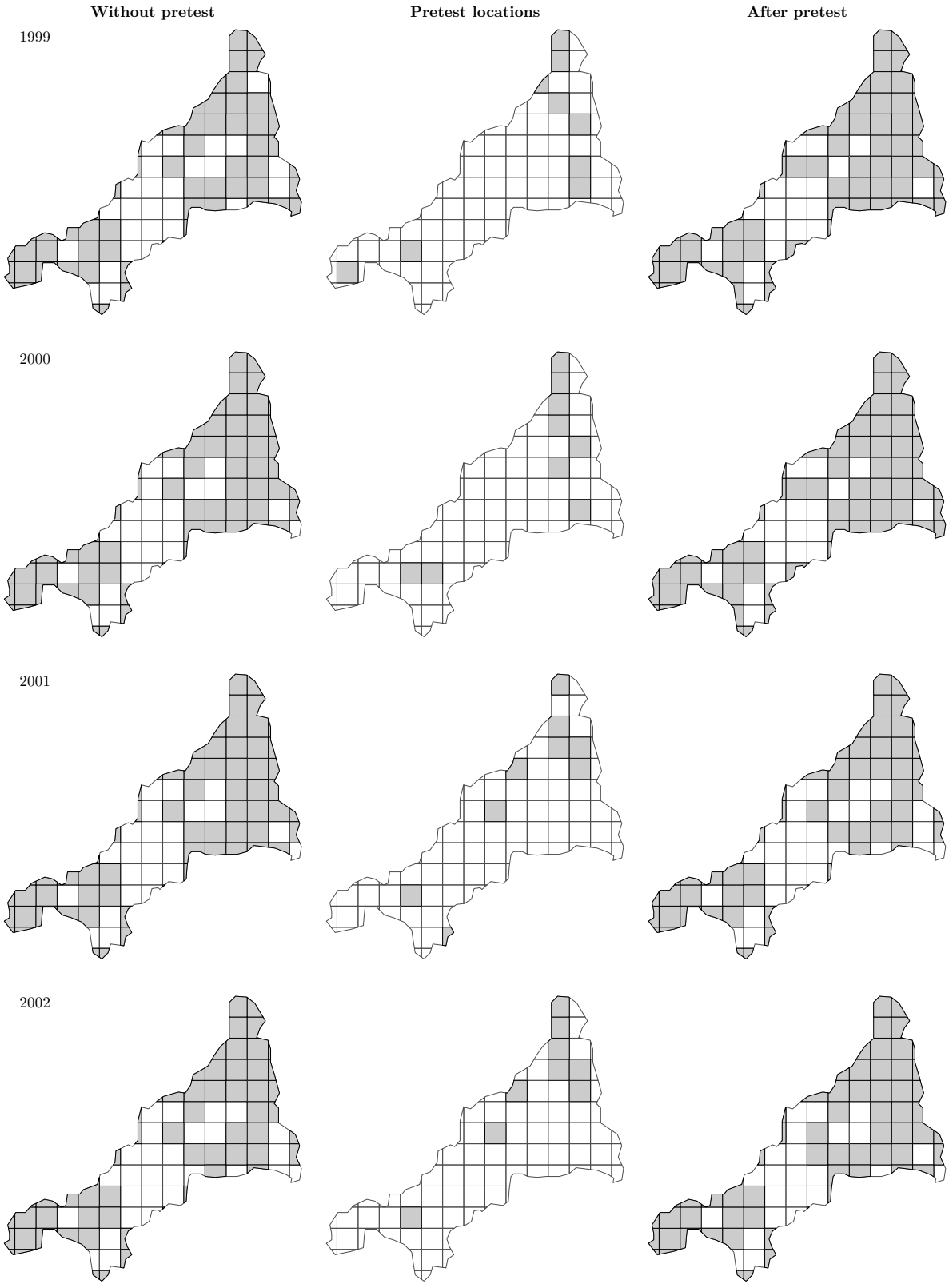


Figure 4: The grey cells indicate for each year: treated cells without pretesting (left column), pretest cells (middle), treated cells after pretesting (right).

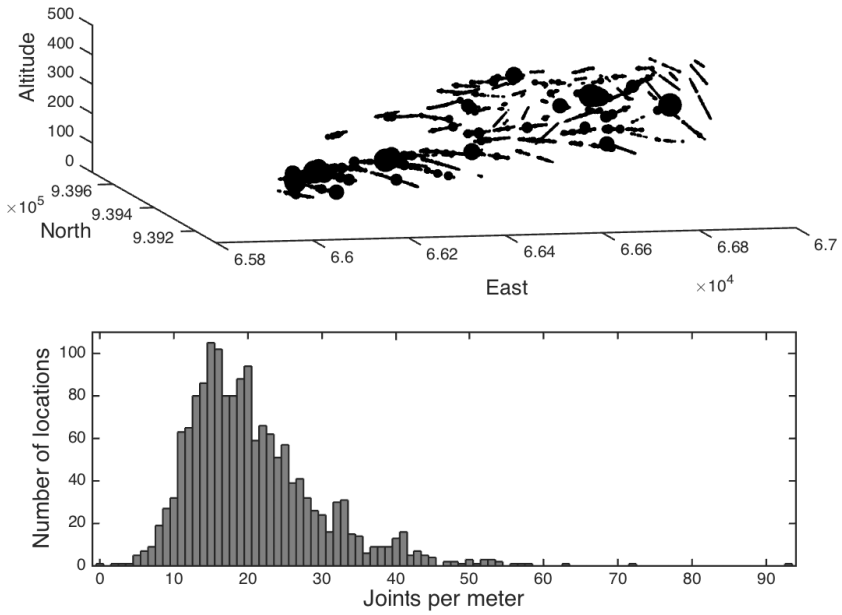


Figure 5: Top: Illustration of a joint frequency count data set. The dots indicate locations of joint counts data. Larger dots mean larger number of joints. The largest count is 93, the smallest is 0. There are 1615 locations in about 100 boreholes. Bottom: Histogram of the joint counts.

authors used a Gaussian model with constant mean and covariance structure defined by a nugget effect plus an exponential anisotropic covariance function. Based on the Laplace approximation parameter values were specified to: mean 1.55, partial sill 0.13, nugget 0.04 and in-strike correlation 300 metres. The range perpendicular to the ore strike was set to a quarter of the in-strike correlation length, i.e. 75 metres.

In the current paper we consider the prospective analysis of joint measurements of a similar type. We assume we know the statistical model, including its parameters specified like above with the constant mean and the covariance parameters. A set of 52 critical tunneling locations near depth 250 metres have been selected. The decision to add support at any of these locations comes with the cost of bolting equipment and labor, but ensures that rock fall will not occur at this location. Without the added support at a location we assume the cost of rock fall depends on the uncertain joint intensity at that location. We use VOI analysis to evaluate which borehole information would be more informative in such a decision situation.

According to what was described above, let C_s be the cost of adding support at location s and $R_s E_x[\exp(d_s x_s)]$ be the expected loss associated with rock fall at the same location when we do not add support. Note that C_s, R_s and d_s will depend on rock mechanical properties, fluid composition, geometric considerations, cost of rock mass transport, and other engineering inputs. For simplicity, these input variables have been set to $C_s = C = 20,000$ and $R_s = R = 100$ money units (\$), and $d_s = d = 3$. The prior value for this decision problem becomes $\text{PV}(\mathbb{S}) = \sum_{s\in\mathbb{S}} \max\{-C, -RE_x[\exp(dx_s)]\}$.

Letting $\boldsymbol{y}$ denote a generic joint count data set acquired according to a specific spatial design S, the posterior value is given by $\text{PoV}(\mathbb{S}|S) = \mathsf{E}_y \sum_{s\in\mathbb{S}} \max\{-C, -RE_x[\exp(dx_s)|\boldsymbol{y}]\}$. By similar arguments as in Section 5.1, we get

$$\text{VOI}(\mathbb{S}|S) = \mathsf{E}_y \sum_{s\in\mathbb{S}} \max\{0, RE_x[\exp(dx_s)|\boldsymbol{y}] - C\} - \sum_{s\in\mathbb{S}} \max\{0, RE_x \exp(dx_s) - C\},$$

which can be approximated using the methods presented in Section 3.

In this example the mean for the joint intensity is relatively large, and the prior decision is to add support at all locations. By collecting borehole data we will pull these decisions more clearly towards added support, or towards avoiding support when the neighboring joint count observations are small, indicating that more support is likely not necessary. The VOI depends on the spatial acquisition design S. We compare the VOI of gathering the entire set of 1615 borehole data against the three partial designs mentioned in Table 1. The boreholes for the partial designs were chosen randomly but in a way that samples from smaller designs consisted of a subset of samples from larger ones. The VOI for each design considered is also shown in Table 1. The VOI decreases when we collect less data, but the decrease is slower than one would expect from the fractional splitting of the data. Moreover, the spatial dependence clearly influences the VOI since the strategy with more boreholes and coarser core samples of joint counts has a much higher VOI, even though the data size is the same for the last two options.

These VOI results must be compared with the price levels of the different data acquisition schemes. We compare the option defined by a quarter of the boreholes with that of every second observation in half of the boreholes (Half-Half). The number of data is then

Design	Data size	VOI ($)
All boreholes	1615	216,000
Half of the boreholes	768	165,000
Every second observation in half of the boreholes	383	159,000
A quarter of the boreholes	383	96,000

Table 1: Comparison of different designs for the mining example.

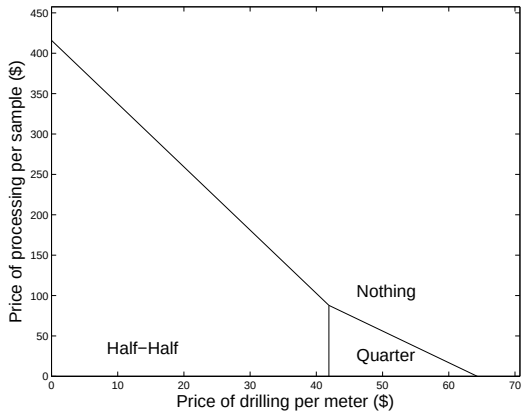


Figure 6: Decision regions for two possible data acquisition schemes in the mining joint example.

the same, so the processing of joint counts data is assumed equal for the two schemes, but the cost of drilling is twice as large for the Half-Half option. The notion of decision regions for data collection here relies on selecting the largest option as follows:

$$\text{Decision} = \text{argmax}\left(\text{VOI}_{\text{Half-Half}} - \text{Price}_{\text{Half-Half}}, \text{VOI}_{\text{Quarter}} - \text{Price}_{\text{Quarter}}, 0\right).$$

Figure 6 shows the decision regions as a function of the price of drilling per metre (first axis) and the price of processing per sample (second axis). When the drilling cost increases, the Quarter option is better. If the prices become very large, the decision is to purchase no data.

6 Discussion

In this paper we derive approximations to the value of information for the generalised linear mixed model with correlated random effects, with particular focus to the spatial case. Our method consists of a mix of Laplace approximation techniques and matrix approximations, together with an approximation to the logistic-normal integral for the binomial model. Under certain conditions on the sample size the approximation is comparable to, and significantly faster than Monte-Carlo integration. In fact, we find that the Monte-Carlo method exhibits larger error when the sample size is large which is in contrast with the error of our analytical approximation.

The VOI can be seen as a design criterion where the objective is expressed in monetary units. In this case budget constraints can be used naturally within a design framework. We illustrated elements of this for the bovine tuberculosis and the mining stability examples.

Because of the inherent correlation in the response variable, the outcome from one site can provide information about the outcome at other sites and indeed in our computations we find that the stronger the correlation the higher the value. We observe strong sensitivity to the variance of the random effects, but the variability in the response is less influential. This suggests that misspecification of the variance may lead to incorrect estimates of the true VOI. A methodology which incorporates parameter uncertainty into the approximation would be useful in addressing this point for example by integrating them out in a Bayesian approach as in Diggle and Lophaven (2006). A second open question is incorporating parameter learning as data arrive sequentially, and to phrase the overall problem as a sequential decision problem.

7 Appendix

7.1 Some preliminary results

Lemma 1.

$$\int_A^\infty \exp(\chi z)\phi(z)\,\mathrm{d}z = \exp(\frac{\chi^2}{2})\Phi(\chi - A).$$

Proof.

$$\begin{aligned}
\int_A^\infty \exp(\chi z)\phi(z)\,\mathrm{d}z &= \int_A^\infty \exp(-\frac{1}{2}z^2 + \chi z)\frac{1}{\sqrt{2\pi}}\,\mathrm{d}z \\
&= \int_A^\infty \exp\{-\frac{1}{2}(z^2 - 2\chi z + \chi^2 - \chi^2)\}\frac{1}{\sqrt{2\pi}}\,\mathrm{d}z \\
&= \exp(\frac{\chi^2}{2})\int_A^\infty \exp\{-\frac{1}{2}(z-\chi)^2\}\frac{1}{\sqrt{2\pi}}\,\mathrm{d}z \\
&= \exp(\frac{\chi^2}{2})\{1 - \Phi(A - \chi)\} \\
&= \exp(\frac{\chi^2}{2})\Phi(\chi - A).
\end{aligned}$$

□

Lemma 2. *If* $z \sim N(\mu, \sigma^2)$*, then*

$$\mathsf{E}_z[\exp(-z)(1+\exp(z))^2] = 2 + \exp(-\mu + \sigma^2/2) + \exp(\mu + \sigma^2/2).$$

Proof.

$$\begin{aligned}\mathsf{E}_z[\exp(-z)(1+\exp(z))^2] &= \int (2+\exp(-\sigma u-\mu)+\exp(\sigma u+\mu))\phi(u)\,\mathrm{d}u \\ &= 2+e^{-\mu}\int \exp(-\sigma u)\phi(u)\,\mathrm{d}u + e^{\mu}\int \exp(\sigma u)\phi(u)\,\mathrm{d}u \\ &= 2+\exp(-\mu+\sigma^2/2)+\exp(\mu+\sigma^2/2).\end{aligned}$$

□

7.2 Approximation to the logistic-normal integral

Let $g(x) = (1+e^{-x})^{-1}$ and consider

$$\Lambda(\mu,\sigma^2) := \int_{-\infty}^{\infty} g(x)\sigma^{-1}\phi\left(\frac{x-\mu}{\sigma}\right)\,\mathrm{d}x, \tag{17}$$

$$\Lambda_a(\mu,\sigma^2) := \int_{a}^{\infty} g(x)\sigma^{-1}\phi\left(\frac{x-\mu}{\sigma}\right)\,\mathrm{d}x. \tag{18}$$

The above integrals do not have a closed-form solution. The one in (17) is known as the logistic-normal integral. Demidenko (2004) discusses different approximations to it in Section 7.1.2. The one in (18) will be referred to as the incomplete logistic-normal integral. By an application of the dominated convergence theorem, both integrals converge to 0 as μ tends to $-\infty$ and to 1 as μ tends to $+\infty$ with σ constant.

One can approximate the logistic function $g(x)$ by the Gaussian CDF $\Phi(\alpha x)$ for an appropriate $\alpha > 0$. Depending on the criterion, $\alpha = \sqrt{\pi/8}$ and $\alpha = 16/(\pi\sqrt{75})$ are two choices mentioned in Demidenko (2004). Let us assume that an appropriate α is chosen. Then, define the approximations to (17) and (18)

$$\begin{aligned}\hat{\Lambda}(\mu,\sigma^2;\alpha) &:= \Phi\left(\frac{\alpha\mu}{\sqrt{1+\alpha^2\sigma^2}}\right) \approx g\left(\frac{\mu}{\sqrt{1+\alpha^2\sigma^2}}\right), \\ \hat{\Lambda}_a(\mu,\sigma^2;\alpha) &:= \int_a^{\infty} \Phi(\alpha x)\times\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\,\mathrm{d}x \\ &= \Phi\left(\frac{\mu-a}{\sigma}\right) - \Pr\left(Z_1 < \frac{\mu-a}{\sigma}, Z_2 < \alpha\sigma Z_1 - \alpha\mu\right) \\ &= \Phi\left(\frac{\mu-a}{\sigma}\right) - \Phi_2\left(\frac{\mu-a}{\sigma}, -\frac{\alpha\mu}{\sqrt{1+\alpha^2\sigma^2}}; -\frac{\alpha\sigma}{\sqrt{1+\alpha^2\sigma^2}}\right),\end{aligned}$$

where Z_1 and Z_2 are independent standard normal random variables and $\Phi_2(x,y;r)$ denotes the bivariate standard normal CDF with correlation r.

7.3 Details on the spatiotemporal model

In this section we describe the spatiotemporal model used in the BTB example.

For the spatial component of the model, let $N = 90$ be the total number of cells and let κ_i denote the number of neighbours of the cell s_i, $i = 1, \ldots, N$. In our model, each cell has at most four neighbours. Next, define the $N \times N$ matrices $P = \text{diag}\{\kappa_i\}$ and B where the (i, j) element of B equals 1 if s_i and s_j are neighbours and 0 otherwise. Then, the joint distribution of $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)$ is set to

$$\boldsymbol{\eta} \sim N_N(0, u^2\Xi), \qquad \Xi = (P - hB)^{-1},$$

where h is a scalar parameter.

A similar definition applies for the temporal component of the model. Here the notion of neighbour corresponds to consecutive time points in $\{1, 2, \ldots, t\}$ we define the diagonal matrix Q and the adjacency matrix C such that the joint distribution of $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_t)$ is

$$\boldsymbol{\epsilon} \sim N_t(0, v^2\Upsilon), \qquad \Upsilon = (Q - kC)^{-1},$$

with scalar parameter k.

For the purposes of this example we fix $h = 0.25$ and $k = 0.5$ and use the available data to estimate β_0, u^2 and v^2. These parameters are estimated by maximum likelihood given all available data after integrating out $\boldsymbol{\eta}$ and $\boldsymbol{\epsilon}$ by Laplace approximation. Given the parameter estimates we derive a plug-in predictive distribution to $\boldsymbol{x}_t$ as discussed in Evangelou et al. (2011), i.e.

$$\boldsymbol{x}_t \sim N_N(\hat{\boldsymbol{\mu}}_t, \hat{\Sigma}_t), \tag{19}$$

The mean and variance of (19) are the required mean and variance terms for computing the VOI.

References

Bhattacharjya, D., Eidsvik, J., and Mukerji, T. (2013). The value of information in portfolio problems with dependent projects. *Decision Analysis*, 10(4):341–351.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York.

Demidenko, E. (2004). *Mixed Models: Theory and Applications*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc.

Diggle, P. and Lophaven, S. (2006). Bayesian geostatistical design. *Scandinavian Journal of Statistics*, 33(1):53–64.

Diggle, P., Zheng, P., and Durr, P. (2005). Nonparametric estimation of spatial segregation in a multivariate point process: bovine tuberculosis in Cornwall, UK. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):645–658.

Ellefmo, S. L. and Eidsvik, J. (2009). Local and spatial joint frequency uncertainty and its application to rock mass characterisation. *Rock mechanics and rock engineering*, 42(4):667–688.

Evangelou, E. and Zhu, Z. (2012). Optimal predictive design augmentation for spatial generalised linear mixed models. *Journal of Statistical Planning and Inference*, 142(12):3242–3253.

Evangelou, E., Zhu, Z., and Smith, R. L. (2011). Estimation and prediction for spatial generalized linear mixed models using high order laplace approximation. *Journal of Statistical Planning and Inference*, 141(11):3564–3577.

Fuentes, M., Chaudhuri, A., and Holland, D. M. (2007). Bayesian entropy for spatial sampling design of environmental data. *Environmental and Ecological Statistics*, 14(3):323–340.

Howard, R. A. and Abbas, A. E. (2015). *Foundations of Decision Analysis*. Prentice Hall. To appear.

McCullagh, P. and Nelder, J. A. (1999). *Generalized Linear Models*. Chapman & Hall Ltd.

Moore, A. L. and McCarthy, M. A. (2010). On valuing information in adaptive-management models. *Conservation Biology*, 24(4):984–993.

Moore, J. L. and Runge, M. C. (2012). Combining structured decision making and value-of-information analyses to identify robust management strategies. *Conservation Biology*, 26(5):810–820.

Morgan, P. B. and Cressie, N. (1997). A comparison of the cost-efficiencies of the sequential, group-sequential, and variable-sample-size-sequential probability ratio tests. *Scandinavian Journal of Statistics*, 24(2):181–200.

Nilsen, B., Shrestha, G. L., Panthi, K. K., Holmoy, K. H., and Olsen, V. (2003). RMR vs Q vs RMi. *Tunnels & Tunnelling International*, 35(5):45–48.

Peyrard, N., Sabbadin, R., Spring, D., Brook, B., and Mac Nally, R. (2013). Model-based adaptive spatial sampling for occurrence map construction. *Statistics and Computing*, 23(1):29–42.

Shun, Z. and McCullagh, P. (1995). Laplace approximation of high dimensional integrals. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 749–760.

Welton, N. J., Ades, A. E., Caldwell, D. M., and Peters, T. J. (2008). Research prioritization based on expected value of partial perfect information: A case-study on interventions to increase uptake of breast cancer screening. *Journal of the Royal Statistical Society. Series A: (Statistics in Society)*, 171(4):807–834.

Willan, A. R. and Pinto, E. M. (2005). The value of information and optimal clinical trial design. *Statistics in Medicine*, 24(12):1791–1806.