Reliability Engineering and System Safety 167 (2017) 351–361

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Prior elicitation for Bayesian generalised linear models with application to risk control option assessment

Geoffrey R. Hosack [a,*], Keith R. Hayes [a], Simon C. Barry [b]

[a] CSIRO, GPO Box 1538, Hobart, Tasmania 7001, Australia
[b] CSIRO, GPO Box 664, Acton, ACT 2601, Australia

ARTICLE INFO

Keywords:
Expert opinion
Model structure uncertainty
Probabilistic safety analysis
Quantitative risk

ABSTRACT

A pragmatic approach to prior elicitation was developed to elicit the parameters and model structure for Bayesian generalised linear models. Predictive elicitation of subjective probability distributions was used to evaluate Risk Control Option (RCO) effectiveness for reducing the risk of ship collisions in Australia's Territorial Sea and Exclusive Economic Zone. The RCOs considered were pilotage, Vessel Traffic Services (VTS) and Ships' Routeing Systems (SRS). Predictive relationships with key covariates were documented. Distance from the Territorial Sea Baseline was important for all RCOs, and aggregate measures of shipping traffic patterns such as volume and the distribution of course over ground headings were related to the effectiveness of both VTS and SRS. A synergistic interaction between pilotage and VTS effectiveness was predicted. The elicitation method enabled a practical approach to eliciting subjective probability distributions while accounting for the complexity and myriad factors that contribute to challenging problems. The approach supports coherent updating given new information, and so can be used to support evidence based decision making.

1. Introduction

Safety analysis is concerned with making decisions while uncertain about the relationship between possible actions and outcomes. Bayesian decision theory provides a coherent approach to decision making under uncertainty [1]. Subjective probability is the language used to capture uncertainty [2] and accommodates both aleatory and epistemic uncertainty [3]. If empirical data are available, then new information can coherently be incorporated via statistical models, that is, Bayesian learning can occur. This process ideally begins by documenting the available expert knowledge and uncertainty. The elicitation of expert opinion using subjective probability distributions rather than eliciting scores or point estimates (such as a mean) is a desirable goal when assessing quantitative risk [4], and substantial effort has been devoted to the elicitation of subjective probability distributions (as opposed to point estimates) of unknown parameters over the years [5,6].

The contribution of subjective probability distributions by experts forms a foundation for evidence based decision making. Some examples of application domains are shipping safety [7], reliability growth assessment for the development of modular software or hardware systems [8], public policy [9], estimating costs and risks for space systems [10,11], nuclear plant safety [12], hydrogeology and water treatment [13]. Well-suited domain experts often have busy schedules that are demanding of their valuable expertise and skill set. Yet complex problems often have many risk factors to consider. A successful elicitation method that allows for Bayesian updating of unknown parameters must permit non-statisticians to contribute their judgements and also mitigate the required elicitation workload to a manageable level.

For example, the assessment of risk for shipping requires the consideration of multiple risk factors such as ship characteristics, vessel traffic patterns, physical factors and the effectiveness of any deployed Risk Control Options designed to reduce risk [14]. A popular method for eliciting expert opinion in the risk analysis of shipping safety is Bayesian networks [15,16]. The expert subjective probability assessments are usually restricted to point estimates, which likely give an incomplete summary of the underlying uncertainty for a given scenario. Many elicitation targets for shipping risk would be more naturally represented by continuous distributions, which may be approximated within a discretised Bayesian network with increasingly fine partitions, but at the cost of complicating the Bayesian network structure and hence elicitation load [17]. Moreover, assessing uncertain model structure for Bayesian networks is time-consuming and difficult [18], and rarely completed. Nevertheless, despite the typical simplifications of point estimates and an assumed known model structure, the increasing use of Bayesian networks has led to recognition of the pressing need to manage the elicitation workload for shipping risk applications [16]. One alternative is to

* Corresponding author.
E-mail address: Geoff.Hosack@csiro.au (G.R. Hosack).

http://dx.doi.org/10.1016/j.ress.2017.06.011
Received 10 June 2016; Received in revised form 19 April 2017; Accepted 11 June 2017
Available online 15 June 2017

develop a prior for unknown coefficients in a regression function that relates risk factors to the exposure of interest. For shipping risk, one approach is suggested by Merrick and van Dorp [19], where point estimates of the relative likelihoods of collision were elicited for various scenarios that form the design points in a regression function.

The uncertainty that experts have under a given scenario may be relevant and so eliciting subjective probability distributions may be preferred to point estimates [4]. Eliciting expert subjective probability distributions has a recognised and important role in the risk analysis of shipping safety [7]. On the other hand, directly eliciting subjective probability distributions for regression coefficients is a difficult task compared to observable quantities [6,20]. A multivariate normal prior is a typical choice for linear and generalised linear models [21], but targeting a multivariate normal prior would necessitate the elicitation of means, variances and covariances. Eliciting these statistical moments is a difficult task for experts [1,6] even without the added complication of introducing known covariates into a regression model.

The assessment of linear model parameters by eliciting from experts their subjective probability distributions of predicted responses, as opposed to point estimates or directly targeting the regression parameters, began with the normal linear model proposed by Kadane et al. [22]. This approach requires the expert to consider responses conditional not only on covariates but also realisations of the response at different design points. A similar approach extends to logistic regression [23] and other generalised linear models [24]. This latter approach allows for piecewise linear models but assumes independence among parameters that correspond to different covariates.

Another predictive approach was elaborated for generalised linear models using conditional mean priors [25]. The expert in this case contributes a distribution, perhaps based on elicited moments or modes, conditional on the values of the covariates at a design point. The form of the prior may depend on the chosen link function and likelihood. However, as mentioned above, a typical choice of prior is multivariate normal independent of the choice of likelihood and link function (e.g., [24,26]). Such multivariate normal priors for generalised linear models are the focus of this paper.

A technique is proposed that elicits subjective probability distributions conditional on scenarios or design points. The proposed approach is a conditional mean prior technique, but differs from [25] in that experts are always asked to provide fractiles (equivalently quantiles or percentiles) of predictive subjective probability distributions. A more important differentiation is the form of the induced prior, which unlike [25] is multivariate normal no matter the choice of link and likelihood function. The approach allows for dependence among all unknown parameters of the linear model including continuous and categorical predictors and all possible interactions, in short, any model that can be represented as a GLM with a linear combination of basis functions. In some cases, the model structure is unknown a priori and must also be assessed from the expert elicitation, and the proposed technique is robust to this situation. The elicitation can accommodate new sources of information, whether from new elicitation or from empirical data, by Bayesian updating and so forms a foundation for evidence based decision making. An example is given for the assessment of Risk Control Options for shipping collision risk in the Territorial Sea and Exclusive Economic Zone that surrounds mainland Australia.

2. Generalised linear model

The generalised linear model (GLM) has three components [27]. The first component specifies an observation model, $p(y_i|\theta_i, \xi)$, for data y_i conditional on the expected response $\mathbb{E}[y_i] = \theta_i$ and any additional parameters ξ used by the observation model chosen from the exponential family. The second component is the *linear predictor,* η_i, that is linked to the expected value of the response. The linear predictor depends on the $p \times 1$ vector x_i of the known covariates evaluated at the design point (or "scenario"), and the $p \times 1$ vector of unknown parameters β,

$$\eta_i = x_i^\top \beta, \tag{1}$$

where the x_i may encode continuous covariates, factors, polynomials or other choice of basis functions. The third component is an invertible *link function*, $g(\theta_i) = \eta_i$, that links the expected response to the linear predictor.

In a Bayesian GLM, a normal prior is often chosen for the parameters β. A normal prior is the conjugate prior for the normal linear model [28], which is a GLM with a Gaussian observation model and identity link function. Although a conjugate prior is sometimes available, this is not the case for all choices of Bayesian GLMs. However, the posterior distribution will under some regularity conditions be approximated by a multivariate normal for large sample sizes [29]. Moreover, a normal prior includes commonly applied basis function models and Gaussian process models for the linear predictor in GLMs [21,30]. The normal prior specification can also ease comparison with other prior structures that may be contributed in the form of hierarchical models, where random effects are often modelled through a Gaussian process prior to capture spatial or temporal dependence [31] or by a multilevel normal prior in the case of exchangeable random effects [21]. A normal prior derived from expert opinion is therefore sought for the unknown coefficients β in Bayesian GLMs.

3. Conditional elicitation step

3.1. Elicitation of subjective conditional probabilities

The target of the elicitation will be the unknown θ. Typically θ is defined such that it has an direct interpretation relative to potential observables and expert knowledge. For example, in a Bernoulli response model where the y_i correspond to successes and failures, $\theta_i = \mathbb{E}[y_i]$ is interpreted as a proportion or probability of an event occurring; in a Poisson response model where the y_i correspond to counts, θ_i is interpreted as an intensity that may relate to the rate of an event occurring over a given time interval, and so on.

The expert (or experts if participating in a group elicitation) are asked to provide their subjective probability estimates for θ_i conditional on the design point x_i. This is the independent conditional mean prior approach where it is useful to assume that expert contributed responses are conditionally independent given the covariates [25]. Another approach is to ask questions about conditional distributions to elicit dependence among observations at different design points (e.g.,[22]). This approach is intractable for many forms of generalised linear models [25]. Moreover, eliciting dependencies requires additional questions, which is an important consideration given the already limited time and statistical resources available for a given elicitation or relevant domain experts (Section 1). Such approaches therefore may prompt experts by making typical proposals based on the model structure, which may run the risk of anchoring an expert to the proposed value rather than eliciting their true opinion. The selection of design points for an independent conditional mean prior elicitation is an experimental design question discussed further in Section 4.2.

A method that elicits fractiles is the preferred approach over other approaches that ask for moments [5,6]. Experts have difficulty assessing how sensitive the mean of a distribution is to its tails, whereas the probability statements encoded by fractiles are easier to elicit [1]. We adopt a normal model for $\eta_i|x_i$ with the conditional mean and variance formed into the parameter vector $\phi_i = [m_i, v_i]^\top$. In the generalised linear model framework, the vectorised link function $g(\,\cdot\,)$ is monotonic and preserves fractiles. Therefore, conditional on a design point x_i, an elicitation of fractiles for the target $\theta_i = g^{-1}(\eta_i)$ directly translates into fractiles for the conditional normal distribution of the linear predictor, $g(\theta_i|\eta_i) = \eta_i|x_i \sim N(m_i, v_i)$. The parameters ϕ_i are chosen to minimise the information lost by approximating the elicited fractiles by a normal distribution as follows.