Environmental Modelling & Software 62 (2014) 300–315

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

The effect of ambiguous prior knowledge on Bayesian model parameter inference and prediction

Simon L. Rinderknecht [a, b, *], Carlo Albert [a], Mark E. Borsuk [c], Nele Schuwirth [a], Hans R. Künsch [d], Peter Reichert [a, b, *]

[a] Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dept. of Systems Analysis, Integrated Assessment and Modelling, CH-8600 Dübendorf, Switzerland
[b] Institute of Biogeochemistry and Pollutant Dynamics (IBP), ETH Zurich, CH-8092 Zürich, Switzerland
[c] Dartmouth College, Thayer School of Engineering, Hanover NH 03755-8000, USA
[d] Seminar for Statistics (SfS), ETH Zurich, CH-8092 Zürich, Switzerland

A R T I C L E I N F O

Article history:
Received 13 August 2014
Accepted 15 August 2014
Available online

Keywords:
Intersubjective knowledge
Imprecise probabilities
Interval probabilities
Robust Bayesian analysis
Density Ratio Class
Bayesian inference
Marginalization and prediction

A B S T R A C T

Environmental modeling often requires combining prior knowledge with information obtained from data. The robust Bayesian approach makes it possible to consider ambiguity in this prior knowledge. Describing such ambiguity using sets of probability distributions defined by the Density Ratio Class has important conceptual advantages over alternative robust formulations. Earlier studies showed that the Density Ratio Class is invariant under Bayesian inference and marginalization. We prove that (i) the Density Ratio Class is also invariant under propagation through deterministic models, whereas (ii) predictions of a stochastic model with parameters defined by a Density Ratio Class are embedded in a Density Ratio Class. These invariance properties make it possible to describe sequential learning and prediction under a unified framework. We developed numerical algorithms to minimize the additional computational burden relative to the use of single priors. Practical feasibility of these methods is demonstrated by their application to a simple ecological model.

1. Introduction

Risk analysis and prediction for environmental management often makes it necessary to combine prior knowledge with information obtained from data. Bayesian statistical inference offers a mathematical framework to do this and even describes an iterative learning process by using the resulting posterior knowledge as prior knowledge for a next updating step with new data (Box and Tiao, 1973; de Finetti, 1974; Howson and Urbach, 1989; Gelman et al., 2003). In this framework, prior knowledge is typically formulated by a single probability distribution to describe either the subjective belief of an individual expert or the intersubjective belief of several experts about the values of specified variables or model parameters. In practice, however, such belief statements are often ambiguous (Einhorn and Hogarth, 1985; Camerer and Weber, 1992; Clemen and Winkler, 1999). This is particularly the case if the intersubjective belief of multiple experts is being used to represent the current state of knowledge of the scientific community (Gillies, 1991; Rinderknecht et al., 2012). To account for this ambiguity, it is important to analyze the influence of the prior on the posterior. This can be done by performing a sensitivity analysis of the results with respect to the parameters of the prior. However, such analyses are quite limited as the priors still remain in the same parametric family and thus do not account for ambiguity resulting from the use of different types of distributions. For this reason, it is more consequent to replace a single prior probability distribution by a non-parametric set of distributions that span the range of appropriate distributions.

Many specifications of such sets of probability distributions over continuous variables, so-called *classes of distributions* or *imprecise probabilities*, have been proposed (Walley, 1991; Caselton and Luo, 1992; Berger, 1994; Ríos Insua et al., 2000, http://www.sipta.org). (Note that the term *imprecise* refers here to the specification of a probability distribution, not to its width.) Despite this theoretical development, the concept of imprecise probabilities, which leads to

* Corresponding authors. Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dept. of Systems Analysis, Integrated Assessment and Modelling, CH-8600 Dübendorf, Switzerland.
E-mail addresses: simon.rinderknecht@nagra.ch (S.L. Rinderknecht), peter.reichert@eawag.ch (P. Reichert).

http://dx.doi.org/10.1016/j.envsoft.2014.08.020
1364-8152/

a robustification of probability statements, is still rarely applied. A reason for this may be that imprecise probabilities are claimed to be over-cautious (O'Hagan (2012) in Krueger et al. (2012)) or are felt to be too difficult to implement. Difficulties could occur during elicitation, when updating priors with data, or when propagating classes of distributions to predictions through deterministic or stochastic models.

We argue that these difficulties can be resolved to a large degree and that Bayesian robustness becomes feasible in many applications if we work with the set of probability distributions from a *Density Ratio Class*, which is also attractive from a conceptual point of view. In Rinderknecht et al. (2011) we developed an elicitation technique for such classes which was then applied to several case studies (Arreaza, 2011 in Scholten et al., 2013, Rinderknecht et al., 2012), to demonstrate that a wide range of ambiguity can occur in practical applications. In the present paper, we address the remaining potential obstacles by showing how Bayesian inference, marginalization, and uncertainty propagation through models can be implemented conceptually and numerically based on the set of probability distributions that are defined by a *Density Ratio Class*. The ease of these implementations relies on the properties of the *Density Ratio Class*. To provide a comprehensive set of important properties and algorithms, this paper combines a review of previously published results with some new results, in particular regarding the propagation of *Density Ratio Classes* through deterministic or stochastic models and with respect to algorithms.

The paper is structured as follows. Section 2 is dedicated to the methodological development. Subsection 2.1 briefly reviews the *Density Ratio Class*. Next, we show in Subsection 2.2 how the *Density Ratio Class* can be used for Bayesian inference, in Subsection 2.3 how it can be marginalized, and in Subsection 2.4 how it can be propagated through a model to quantify prediction uncertainty. Section 3 discusses the numerical implementation of these tasks. In Section 4 we demonstrate the suitability of the approach through application to a simple empirical river periphyton model. Finally we draw our conclusions in Section 5.

2. Methods

In this section, we briefly review the formulation of ambiguous knowledge with *Density Ratio Classes*, we prove the invariance of *Density Ratio Classes* under Bayesian updating and marginalization, and we derive results for model predictions based on the propagation of *Density Ratio Classes* through deterministic or stochastic models.

2.1. Formulation of ambiguous prior knowledge as a Density Ratio Class

DeRobertis and Hartigan (1981) introduced the *Density Ratio Class* under the name of *Intervals of Measures*, whereas Berger (1990) later called the class the *Density Ratio Class*. Wasserman (1992) asserted that, under mild regularity conditions, it is the only probability class to be invariant under Bayesian updating and marginalization. Update invariance is an important property, as it allows for the representation of sequential learning within a common framework. This gives an important advantage to the *Density Ratio Class* relative to other representations of imprecise probabilities. The *Density Ratio Class* also has the ability to accommodate a variety of density function shapes, while limiting 'unreasonable' shapes such as sharp peaks or point masses that might not be deemed reasonable by an expert (Rinderknecht et al., 2011). (Depending on the size of the class, weakly or even strongly expressed multi-modality is still possible.)

For uncertain continuous parameters $\boldsymbol{\theta}\in M\subset\mathbb{R}^n$, the *Density Ratio Class* with lower bound $l\geq 0$ and upper bound $u\geq l$ is defined as the set of probability density functions

$$\Gamma^{DR}_{l,u} := \left\{ \widehat{f}(\boldsymbol{\theta}) = \frac{f(\boldsymbol{\theta})}{\int f(\boldsymbol{\theta}')\,\mathrm{d}\boldsymbol{\theta}'} \;\middle|\; l(\boldsymbol{\theta}) \leq f(\boldsymbol{\theta}) \leq u(\boldsymbol{\theta}) \;\forall\, \boldsymbol{\theta} \right\}, \tag{1}$$

where we assume that $0<\int l(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}\leq\int u(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}<\infty$. The non-normalized densities l and u bound the shapes of the non-normalized probability densities in the class. The class then consists of the normalized densities that fulfill these shape restrictions. In this paper, we shall exclude improper densities since we consider their interpretation questionable (Rinderknecht et al., 2011). Note that the *Density Ratio Class* has the following property:

$$\Gamma^{DR}_{l,u} = \Gamma^{DR}_{\lambda l,\lambda u} \;\forall\; \lambda > 0\,. \tag{2}$$

This implies that one of the "non-normalized" densities, l or u, can still be chosen to be normalized.

Following from (1), the lower and upper probabilities, $\underline{P}$ and $\overline{P}$, of a random variable characterized by the *Density Ratio Class*, $\Gamma^{\mathrm{DR}}_{l,u}$, taking a value within a subset A of its domain are given by

$$\underline{P}(A) = \inf_{\widehat{f}\in\Gamma^{DR}_{l,u}} \int_A \widehat{f}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = \frac{\int_A l(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}}{\int_A l(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} + \int_{A^c} u(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}} \tag{3}$$

and

$$\overline{P}(A) = \sup_{\widehat{f}\in\Gamma^{DR}_{l,u}} \int_A \widehat{f}(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = \frac{\int_A u(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}}{\int_A u(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} + \int_{A^c} l(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}} \tag{4}$$

where A^c is the complement of A. The first of these equations follows from the fact that for any $\widehat{f}\in\Gamma^{DR}_{l,u}$, $\int_A\widehat{f}\,\mathrm{d}\boldsymbol{\theta}$ can be written in the form $\int_A f\,\mathrm{d}\boldsymbol{\theta}/(\int_A f\,\mathrm{d}\boldsymbol{\theta}+\int_{A^c} f\,\mathrm{d}\boldsymbol{\theta})$ and $x/(x+y)$ is decreasing in y for fixed $x>0$ and increasing in x for fixed $y>0$. Note that the equation is obviously also true if either $\int_A l(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}=0$ or $\int_{A^c} u(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}=0$, and the integrals cannot both be zero because of the condition $\int l\,\mathrm{d}\boldsymbol{\theta}>0$. The second equation follows analogously.

In the following three subsections we elaborate important properties of Bayesian inference, marginalization and prediction with the *Density Ratio Class*.

2.2. Bayesian parameter inference with Density Ratio Class priors

The first property we discuss is the invariance of the *Density Ratio Class* under Bayesian inference, or *updating*. The *likelihood function*, $L(\boldsymbol{y}|\boldsymbol{\theta})=p(\boldsymbol{y}|\boldsymbol{\theta})$, is the probability density of model results (or also predictand), $\boldsymbol{y}$, given the model parameters, $\boldsymbol{\theta}$. For statistical inference, we substitute observations for the argument $\boldsymbol{y}$ and are interested in the dependence of L on the parameters. For this reason, we simplify the notation in the following sections to $L(\boldsymbol{\theta})$ and do not explicitly indicate the dependence on the observations, $\boldsymbol{y}$, which in the context of inference are assumed to be fixed. We will return to the full notation in Subsection 2.4., $p(\boldsymbol{z}|\boldsymbol{\theta})$, where $\boldsymbol{y}$ is replaced by $\boldsymbol{z}$ to clarify that, in the context of probabilistic prediction, it is not the observations $\boldsymbol{y}$ that are substituted for the argument of the probability density function.