International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Robustness in Bayesian nonparametrics ☆

Sudip Bose

Department of Statistics, The George Washington University, Washington, DC 20052, USA

ARTICLE INFO

Article history:
Received 1 March 2016
Received in revised form 24 November 2016
Accepted 1 December 2016
Available online xxxx

Keywords:
Dirichlet priors
Robustness
Optimality
Minimax posterior regret
Linear functionals
Ratio-linear quantities

ABSTRACT

We consider Bayesian robustness in the context of Bayesian Nonparametrics, and specifically for the Dirichlet Process prior. We show how to find an optimal procedure, based on $\mathcal{C}$-minimax posterior regret (CMPR) for a class of priors $\mathcal{C}$. We consider regret based on squared error loss. The neighborhood classes considered are the density ratio (DR) class and the epsilon-contamination class.

1. Introduction

1.1. Dirichlet process prior

The Dirichlet Process prior [11,12] and its variants are widely used in Bayesian Nonparametrics. Let X be a real random variable. Let $\alpha > 0$ and P_0 a probability measure on the sample space of X. Under the Dirichlet process prior $\mathcal{D}(\alpha P_0)$, the prior distribution of $(P(X \in B_1), \dots P(X \in B_k))$ for any measurable partition $B_1, B_2, \dots B_k$ of the sample space, is a Dirichlet distribution with parameters $(\alpha P_0(B_1), \dots, \alpha P_0(B_k))$. As pointed out in Ferguson [11,12], the posterior distribution is also Dirichlet.

1.2. Bayesian robustness

The idea of considering uncertainty in the prior distribution was an important part of the statistical philosophy of I.J. Good; see for instance Good [13,14]. The work of Berger [3,5] was also influential in the development of Bayesian robustness. Berger [5] focuses on sensitivity to the prior, recognizing that in practice, one cannot exactly specify one's prior. The robust Bayesian approach replaces a single prior by a class of priors.

In a parametric Bayesian analysis, where $X \sim f(x|\theta)$, with real θ, to specify the prior $\pi(\theta)$, one has to *exactly* specify a countably infinite number of probabilities $P(\theta \in A_n)$. In the BNP (Bayesian Nonparametric) setting with Dirichlet process priors, the view point is slightly different. With the Dirichlet process prior, the probability distribution for X can come from a large class of distributions, whereas in a parametric Bayes analysis, the probability distribution of X comes from within a parametric family $f(x|\theta)$. Specifying a single Dirichlet process prior (for real-valued X and the Borel σ-field)

☆ This paper is part of the virtual special issue on Bayesian Nonparametrics, edited by Antonio Lijoi, Antonietta Mira and Alessio Benavoli.
E-mail address: sudip@gwu.edu.

http://dx.doi.org/10.1016/j.ijar.2016.12.001
0888-613X/© 2016 Published by Elsevier Inc.

means that one (implicitly) specifies exactly the probability distribution for the probabilities $P(X \in A)$ for Borel sets A (or at least for countably infinitely many intervals A_n). One should be concerned about the robustness of the analysis to the choice of the Dirichlet process prior. We will study robustness of the analysis via the range of posterior expectations of parametric functions of interest as the prior varies over a class of priors. Related work on robustness with classes of Dirichlet Process (DP) priors has been done in Ruggeri [18–20] and Benavoli et al. [2], which we discuss in Section 2. Walley [22] discusses imprecision in probabilities, including the concept of "near-ignorance priors". Walley's work appears to have influenced many subsequent works, including those of Ruggeri [18,19] and Benavoli et al. [2], cited above. Somewhat related is Section 5.3 of Walley [22], where he discusses an imprecise Beta model.

Arguments against the need for robustness studies. The argument is sometimes made that nonparametric Bayesian analyses have a built-in robustness, since unlike parametric situations, the prior chooses a distribution from an infinite-dimensional class. In particular, for the Dirichlet process prior, the distribution of X is discrete with probability one. Since one can approximate any measure on the Borel subsets of the real line by a discrete measure, a BNP analysis has a special flexibility, and for a large sample of data, one can come close to the true process generating the data. (In other words, the argument is in part, that one is not limited to 'parametric' probability measures from a family with density $f(x|\theta)$.) We point out that, in a nonparametric Bayesian analysis, concerns could well remain about samples that are not large.

Another argument is that a subjective Bayesian may argue that her prior distribution represents her belief and thus there is no need to worry about robustness with respect to the prior. However, many Bayesian analyses are done with a prior that is convenient, rather than based on a careful and deep consideration of one's prior belief. Presumably, the convenience of DPPs (Dirichlet Process Priors) has contributed to their widespread use.

Various approaches to robustness. A popular approach has been to study limiting behavior as one gets more data, and prove posterior consistency under suitable conditions. In the 'parametric' Bayesian setting, Kadane and Chuang [16] discussed the concept of "stability". The essential idea is that "small" changes in the inputs of a decision problem result in a "small" difference in the optimal risk. Ghoshal [15] reviews posterior asymptotics for the Dirichlet process and related priors, and presents some posterior consistency results.

Another approach, in parametric Bayes, has been to search for inherently robust priors. Berger [6] points out that there is some evidence that flat-tailed priors are inherently robust. While this is interesting, we take a different approach. Recognizing that it is very difficult, arguably unrealistic to specify one's prior exactly, we consider a class of prior distributions.

There is a substantial literature on Bayesian robustness with respect to variations in the prior. Much of this literature is concerned with finding the range of posterior expectation of parametric functions as the prior varies over a class of prior distributions. Common examples of parametric functions of interest are the posterior mean and posterior probabilities. The review paper of Berger [5] discusses the central role of ranges of posterior expectations in Bayesian robustness investigations. If the range of posterior expectations of parametric function(s) of interest is (are) small, one has robustness. However, one may be interested in the optimal decision rule, taking into account the uncertainty about π, i.e. based on only knowing that $\pi \in \mathcal{C}$. The essence of our approach in this paper is to make use of the range of posterior expectations to find a procedure that is the most robust according to a criterion, namely minimax posterior regret.

$\mathcal{C}$-Minimax posterior regret. For a given loss function, the minimax decision is a popular choice. It has the best worst-case performance, and is thus appealing. However, under fairly general conditions, the minimax rule is also maximin in the sense that it is achieved for the case where one has worst best-case performance. Essentially, one ends up in the case where one is guaranteed the largest possible (minimum) loss. As an alternative to minimax risk estimators one may prefer the use of minimax posterior regret estimators.

To understand posterior regret, consider a Bayesian setting where the prior $\pi \in \mathcal{C}$. The regret of a decision δ, for a given prior π is the additional expected loss incurred with δ compared to the Bayes decision δ_π for that π. The idea is that one is certain to incur PEL (posterior expected loss) at least $E^\pi[L(\theta, \delta_\pi)|x]$. How much additional PEL does one incur by using δ instead of δ_π? That additional amount, $E^\pi[L(\theta, \delta)|x] - E^\pi[L(\theta, \delta_\pi)|x]$, is the posterior regret. (This is akin to the concept of opportunity cost or expected opportunity loss in Economics.)

For squared error loss, $L(\theta, a) = (\theta - a)^2$, it is well known that the Bayes decision is the posterior expectation of θ, i.e. $\delta_\pi(x) = E^\pi(\theta|x)$. Therefore the minimum PEL (posterior expected loss) under prior π is the PEL of δ_π, which equals

$$E^\pi[(\theta - E^\pi(\theta|x))^2|x] = Var^\pi(\theta|x),$$

the posterior variance of θ under prior π.

For a decision rule δ, the posterior regret under π is the additional posterior expected loss or additional posterior risk incurred by using δ in place of δ_π, the latter being the Bayes rule under π. This equals $E^\pi[(\theta - \delta(x))^2|x] - E^\pi[(\theta - \delta_\pi(x))^2|x]$, which reduces to

$$[\delta(x) - E^\pi(\theta|x)]^2 \tag{1}$$

Parametric function $\psi(\theta)$. A similar result holds for a real-valued parametric function $\psi(\theta)$. If one has squared error (relative to $\psi(\theta)$) loss function $(\psi(\theta) - a)^2$, the posterior regret of δ under π equals

$$[\delta(x) - E^\pi(\psi(\theta)|x)]^2 \tag{2}$$

Please cite this article in press as: S. Bose, Robustness in Bayesian nonparametrics, Int. J. Approx. Reason. (2017), http://dx.doi.org/10.1016/j.ijar.2016.12.001