View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

Sonia Petrone

An example and some questions for Bayesian nonparametric statistics, from a predictive point of view

2001/11

UNIVERSITÀ DELL'INSUBRIA
FACOLTÀ DI ECONOMIA

http://eco.uninsubria.it

In questi quaderni vengono pubblicati i lavori dei docenti della Facoltà di Economia dell'Università dell'Insubria. La pubblicazione di contributi di altri studiosi, che abbiano un rapporto didattico o scientifico stabile con la Facoltà, può essere proposta da un professore della Facoltà, dopo che il contributo sia stato discusso pubblicamente. Il nome del proponente è riportato in nota all'articolo. I punti di vista espressi nei quaderni della Facoltà di Economia riflettono unicamente le opinioni degli autori, e non rispecchiano necessariamente quelli della Facoltà di Economia dell'Università dell'Insubria.

These Working papers collect the work of the Faculty of Economics of the University of Insubria. The publication of work by other Authors can be proposed by a member of the Faculty, provided that the paper has been presented in public. The name of the proposer is reported in a footnote. The views expressed in the Working papers reflect the opinions of the Authors only, and not necessarily the ones of the Economics Faculty of the University of Insubria.

© Copyright Sonia Petrone
Printed in Italy in May 2001
Università degli Studi dell'Insubria
Via Ravasi 2, 21100 Varese, Italy

All rights reserved. No part of this paper may be reproduced in any form without permission of the Author.

An example and some questions for Bayesian nonparametric statistics, from a predictive point of view *

Sonia Petrone †

May 10, 2001

Abstract

In this note we raise some questions for Bayesian nonparametric statistics starting from an example. The problem is described by Coram and Diaconis (2001) and regards studying, by probabilistic techniques, the correspondence between the eigenvalues of random unitary matrices and the complex zeros of Riemann's zeta function. This is an intriguing problem, and we will only underline some of the questions which might arise if we look at it from a (nonparametric) Bayesian point of view.

1 Introduction

This note is a discussion of Professor Hjort's paper on "Topics in nonparametric Bayesian statistics", written for the book "Highly Structured Stochastic Systems", edited by Peter Green, Nils Hjort and Sylvia Richardson (to be published by Oxford University Press).

Highly Structured Stochastic Systems (HSSS) is the name of an initiative that has been running since 1993 with funding from the European Science Foundation. HSSS refers to a field of Statistics that is characterised by building complex stochastic models from simple localised components. Examples are in image analysis (simple localised descriptions of individual pixels as dependent on neighbouring pixels, that allow complex global image analysis), expert systems (simple conditional relationships between parents and children in graphical models, that build into highly complex knowledge representations) and hierarchical models (where each layer of the hierarchy depends on the next through a simple structure). The HSSS initiative was stimulated by the recognition that these three areas, that had previously been studied independently, had many common features.

*This work has been partially supported by grants from the University of Insubria.

†Università dell'Insubria, Dipartimento di Economia, Via Ravasi 2, 21100, Varese, Italy. Email: spetrone@eco.uninsubria.it

Since 1993, HSSS has brought together researchers in these and other areas, to explore models and computational tools of common interest. Numerous workshops have been held on particular topics, as well as two international conferences.

The end of the HSSS initiative will then be marked in the previously cited book, which has the aim of presenting the current state of the field in a comprehensive but accessible form.

2 An example

In this note I will try to raise some questions for Bayesian nonparametric statistics starting from an example. The problem is described by Coram and Diaconis (2001) and regards studying, by probabilistic techniques, the correspondence between the eigenvalues of random unitary matrices and the complex zeros of Riemann's zeta function. This is an intriguing problem, and of course I have no pretence to enter into details. I will only underline some of the questions which might arise if we look at it from a (nonparametric) Bayesian point of view.

Evidence exists of a close connection between the zeta zeros and the eigenvalues of typical unitary matrices. The unitary group U_n is the group of the $n \times n$ complex matrices M such that $MM^* = I$. Let us sample matrices from Haar measure on U_n. In the literature, there are results on the distribution of the n eigenvalues of a matrix M drawn from U_n and of some statistics related to them. Let us fix $n = 42$. Then, Coram and Diaconis consider the data set given by the $50,000$ consecutive zeta zeros starting around the $10^{20} + 271959460$'th zero. For comparison with the eigenvalues of random matrices, they put the zeros onto the unit circle following a "wrapping" procedure. In this way, the data are 1190 realizations of 42-points on a circle, and the problem is to test whether this data set could have come from the Haar measure. As a first test, Coram and Diaconis compare the distribution of the trace of a random matrix to the "trace" based on the zeta zeros. Results in the literature show that the norm of the trace of a random matrix is remarkably well approximated by an exponential distribution. Let us compute an analogous "norm squared trace" for each of the 1190 blocks of 42 zeta zeros in the circle, and denote it by $W_i, i = 1, \ldots, 1190$. The W's would be distributed according to the exponential distribution with mean 1 to very good approximation if the they were actually formed from the trace of random matrices U_n. Therefore, we are faced with a goodness-of-fit problem. Can we approach the problem from a (nonparametric) Bayesian point of view? I will briefly discuss some questions which arise (at least, in my mind).

2.1 Some preliminary questions

Perhaps, one contribution that a Bayesian approach might provide is the emphasis on the predictive aspects of the problem; in this sense, the interest would be on "predicting the next zero" of the zeta function and one might ask whether the knowledge about unitary matrices can suggest a good predictive rule. For predictive purposes, we have to consider the data as dependent, and the simplest dependence assumption is exchangeability. Let us assume that $W_i, i = 1, \ldots, 1190$, are part of an infinite sequence of exchangeable random variables. Then, from de Finetti representation theorem, exchangeability is equivalent to assume that the data are conditionally i.i.d. according to a distribution function F, which has prior probability law π. With no further assumption besides exchangeability, the prior is nonparametric, i.e. its support is the class of all the distribution functions on the sample space. We can find several proposals of nonparametric priors in the literature, and some of them are presented in sections 2 to 4 of Hjort's chapter; there are connections with proposals by Dubins and Freedman (1966) and Ishwaran and Zarepour (2000). Many nonparametric priors (such as the Dirichlet process, or Polya trees) are based on a partition of the sample space. However, if in the problem of the zeta zeros we are interested in the n eigenvalues rather than in the W statistic, the sample space would have dimension 42, so partitioning it might be quite problematic. One of the challenges for research in Bayesian nonparametrics is to deal with high dimensional data sets.

It is worth mentioning an alternative way of constructing a nonparametric prior, which is based on mixtures of parametric distributions. Early proposals in this direction go back to the 80's, but encountered computational difficulties which delayed their development. Still, Bayesian nonparametric literature is mainly focused on applications of mixtures in density estimation or regression, and a careful study of a nonparametric "mixture prior" seems missing. Theoretical aspects might be studied by exploiting the approximation properties of the family of mixture models on which the prior is based. For example, the approximation properties of Bernstein polynomials, and therefore of mixtures of betas, are used by Petrone (1999*a*) and Petrone and Wasserman (2001) for proving the properties of the so called Bernstein prior, defined for exchangeable data with values in [0,1]. These results can be extended to more general sample spaces, and used for studying consistency of general mixture priors.

Are these priors well suited for expressing the prior information about the distribution of the W's, or, more generally, of the zeta zeros? It is important that the structure of the prior remains fairly simple, so that the quantities on which one has to express a probability law have a physical interpretation. Nevertheless, the Bayesian approach, requiring to express a prior on a class of probability measures, still appears complicated, and many researchers feel uncomfortable with that; there must be a way of make it simpler. The predictive approach might be fruitful. The representation theorem establishes a correspondence between the prior and the probability law P of the sequence $\{W_k, k \geq 1\}$ (and consequently the sequence of predictive distributions). Therefore, we might choose or even characterize the prior through predictive assumptions (see for example Cifarelli *et al.* 2000). This is an interesting line of research, which however has been pursued mainly for exchangeable sequences; it would be of interest to extend the study to more complex dependence structures.

2.2 Goodness-of-fit

For its many properties, the Dirichlet process provides the central class of nonparametric priors. However, as a consequence of its discrete nature, it is not appropriate as a prior on the alternative hypothesis in a goodness-of-fit problem (Ferguson 1973). Therefore, different priors have been used, usually centered on the null model. Nonparametric envelopes around parametric models are discussed in section 10 of Hjort's chapter, and are regarded as a way of robustifying parametric procedures. However, caution is needed. For example, the peculiar probability of ties among the data implied by the Dirichlet process prior can lead to unexpected consequences of the posterior, so that the latter is not "robust" (see for example Petrone and Raftery 1997). It would be interesting to study the probability law of ties for the generalizations of the Dirichlet process presented by Hjort in section 2.

Bayesian procedures for goodness of fit are still to be fully developed. As already mentioned, the problem might be studied from a predictive point of view, so that a model would be evaluated for its predictive properties, according to a loss function. *A posteriori*, different (Bayesian) procedures might be scored according to the accuracy of the resulting forecasts (possibly, in some cross-validation way).

Lindley paradox for a precise null hypothesis might arise in a nonparametric context too.

2.3 Computational issues and asymptotic behavior

Computational aspects remain an open issue, especially if great accuracy is required, for example for distinguishing between very close hypotheses. Hjort presents interesting results, particularly regarding random Dirichlet means. These methods can also be applied for studying the distribution of functionals of the kind $\int \phi(x; a, \theta) dG(\theta)$, where ϕ is a kernel probability density and G is a Dirichlet process. This possibility is also mentioned by Hjort in section 9.

In the zeta zeros problem, the sample size is 1190. Given the large sample size, would Bayesian and frequentist answers agree? There are Glivenko-Cantelli like results for exchangeable sequences which guarantee that the distance between the predictive and the empirical distribution functions converges to zero as the sample size increases, almost surely with respect to the probability law P of the exchangeable sequence (Berti and Rigo, 1997). This approach is different from the notion of consistency discussed in section 5 of Hjort's chapter; in particular, empirical frequencies are considered in place of the true distribution; however, the results hold almost surely, so there can be exceptional null sets. Besides consistency, it would be interesting to study when we can obtain a nonparametric extension of the Bernstein-von Mises-Laplace theorem, regarding asymptotic normality of the posterior.

2.4 Problems of dependence

Let us consider again the zeta zeros problem, and the data consisting in the traces $W_1, \ldots, W_N$. Coram and Diaconis find that the sequence $\{W_i\}$ shows a negative serial correlation. In a Bayesian framework, this means that the assumption of exchangeability is too restrictive. Unfortunately, most of the literature in Bayesian nonparametrics considers the case of exchangeable data. Extending the analysis to more complex dependence structures is one of the main open areas of research.

Hjort touches problems of regression, survival analysis, inference of random shapes, yet many ideas are still to be developed. For the case of partially exchangeable data, Guglielmi and Melilli (2000) show that any prior can be approximated by mixtures of products of Dirichlet processes. Mixtures of products of Dirichlet processes have been used in regression problems, where the partition of the data in groups is induced by some covariates (see Petrone and Muliere, 1993 and references therein). A general class of priors for dependent data are Dependent Dirichlet processes (MacEachern, 2000). The mixture priors briefly discussed in paragraph 2.1 might be extended in several directions. In particular, multidimensional mixture priors are of interest for application to stochastic regression. Furthermore, the distribution of the mixture weights might depend on covariates, or include a temporal or spatial structure.

References

Berti, P. and Rigo, P. (1997) A Glivenko-Cantelli theorem for exchangeable random variables. *Statistics and Probability Letters*, **32**, 385–391.

Cifarelli, D. M., Muliere, P. and Secchi, P. (2000) Urn schemes for constructing priors. *Technical Report n.429/P*, Politecnico di Milano.

Coram, M. and Diaconis, P. (2001) New tests for the correspondence between unitary eigenvalues and the zeros of Riemann's zeta function. *Technical Report*, Department of Statistics, Stanford University.

Dubins, L. and Freedman, D. (1966) Random distribution functions. *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability*, **2**,183–214, Univ. of California Press.

Guglielmi, A. and Melilli, E. (2000) Approximating de Finetti's measures for partially exchangeable sequences. *Statistics and Probability Letters*, **48**, 309–315.

Ishwaran, H. and Zarepour, M. (2000) Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika*, **87**, 371–390.

MacEachern, S. N. (2000) Dependent Dirichlet Processes. *Technical Report*, Ohio State University.

Muliere, P. and Petrone, S. (1993) A Bayesian predictive approach to sequential search for an optimal dose: parametric and nonparametric models. *Journal of the Italian Statistical Society*, **3**, 349–364.

Petrone, S. and Raftery, A. E. (1997) A note on the Dirichlet process prior in Bayesian nonparametric inference with partial exchangeability. *Statistics and Probability Letters*, **36**, 69–83.

Petrone, S. and Wasserman, L. (2001) Consistency of Bernstein posterior (under revision for the *Journal of the Royal Statistical Society*, B).