

Report on the Programme: Bayesian Nonparametric Regression

Basic theme and background

Bayesian nonparametric inference is a relatively young research area, with the earliest papers appearing in the mid 1970's, with greatly increasing activity over the last 10 years. The group of researchers working in this field is still moderately small and there was a need for a focus opportunity, which the BNR programme provided, to survey the field, exchange ideas, identify gaps in the literature, and coordinate research efforts. Specifically in the area of regression, where there is a lot of current activity with many possible avenues of investigation.

Thus the purpose of the BNR programme was to review the current state of nonparametric Bayesian research, to foster collaborations between researchers with different focus areas, and to identify important open problems. One prominent example is the need to combine research concerned with the construction of random probability measures with another direction of research concerned with random mean functions in a regression problem.

There is a need to build and strengthen connections between emerging areas of nonparametric Bayesian inference. Besides the already mentioned connection of random distributions and regression functions, other examples are the relevance of asymptotic results to model choice and inference, generalizations of predictive probability functions and random clustering models. We were particularly pleased for a number of persons from the Machine Learning community to attend the Programme, who highlighted practical problems to be solved.

The four organizers met early in 2005 for an initial planning meeting when we agreed on the broad focus, on invitation lists and priorities. We also decided at that time to have the 2nd week workshop, and to leave maximum flexibility and opportunity for informal interaction for the remaining Programme.

Structure

In weeks 1, 3 and 4, the programme organization was very informal. We organized two informal talks each day. The nature of these talks varied, with some presentations being descriptions of current research ideas, some

describing work in progress, and some that reported on recently completed research. The format of these talks was chosen to encourage interaction and “lively” discussions.

For week 2 (6 to 10 August, 2007) the Programme organizers were responsible for a Workshop on Bayesian Nonparametric Regression. The first day was dedicated to 4 tutorials, with speakers Subhashis Ghosal, Antonio Lijoi, Yee Teh and David Dunson. The remaining four days were regular talks. There were about 90 participants at the Workshop and over 40 speakers.

The first day of the workshop consisted of four tutorials, which provided an excellent survey of methods in nonparametric Bayesian inference. The tutorials were extremely successful. For similar conferences, usually only part of the audience attends introductory tutorials. But for this workshop we noticed that 80% and more of the participants stayed for all four tutorials.

The remaining days of the workshop were research talks, including some talks on innovative applications of nonparametric Bayesian inference (Mukherjee, Ruggiero, Laud, Herring, Popova, House, Williams, Mena) novel methods for survival analysis (Johnson, Guglielmi, De Iorio, Yin), issues of prior choice (Cox, MacEachern), new non-parametric models (Quintana, Griffin, Arjas, Pruenster), constructions of families of dependent random probability measures (Spano, Petrone, Steel, Dunson, Basu), and asymptotic results (Choi, De Blasi, Lee).

Outcome and achievements

The Programme successfully surveyed the field and pointed towards specific directions for future applications of research. It generated interest in alternative models that generalize the most traditional ones which currently dominate practical nonparametric Bayesian methods. At the same time, several talks highlighted and discussed distinguishing features of these traditional models.

Some new models under discussion aimed at introducing dependence into Pólya trees; one of the main methods of creating nonparametric Bayesian models. One formulation of the problem is to drive the dependence with a collection of stochastic processes introduced as a piece of the model. A second formulation is to drive the dependence through shared portions of the trees.

Other ideas included the exploration of random effects models that preserve selected features of a set of distributions as covariate values change.

The feature to preserve in this instance is the expected probability of an event. With the asymmetry of the mapping from the probability scale to the logit scale, this requires a random effects distribution which changes with the covariate in a very particular fashion.

The dependent Dirichlet process (and a bit more generally dependent nonparametric processes) provide a framework which encompass many of the nonparametric Bayesian models that are currently being proposed and fit. The relationship between the models becomes clearer when set in the framework of dependent Dirichlet processes. Setting the models in this context also points the direction to refinement of the models, as there are generally effective strategies that allow one to capture specific types of behavior.

Many other ideas were discussed, such as modelling dependence via partial exchangeability; sampling methods and models for tree and time-structured dependent Dirichlet processes; hierarchical mixtures and general consistency issues.

Most of the talks and presentations in the informal part of the Programme addressed the question that initially motivated the organization of the Workshop, namely the combination of random probability measures and random regression mean functions to construct real nonparametric regression models. Several talks focused on the construction of families of random probability measures indexed with a covariate. This formally provides a constructive definition of the desired combinations.

Aside from the organized collection of participants at the talks; feedback from participants made it quite clear than many collaborations are under way and many new projects initiated as a consequence of the intensity of the Programme.

A book is planned with the CUP as publishers. The contents would be the material from the 4 tutorials, given by Antonio Lijoi, Subhashis Ghosal, Yee Teh and David Dunson, with each of the organizers providing a chapter.

The organizers would like to express their deep gratitude to the INI for their outstanding hospitality and for accepting us as guests for a month.