Nonparametric Bayesian data analysis
Peter Mueller
UT Austin, USA
July 6-8, 2020 2-4pm (EST)

All models are wrong, but some are useful. Many statisticians know and appreciate G.E.P. Box' comment on statistical modeling. Often the choice of the final inference model is a compromise of an accurate representation of the experimental conditions, a preference for parsimony and the need for a practicable implementation. The competing goals are not always honestly spelled out, and the resulting uncertainties are not fully described. Over the last 20 years a powerful inference approach that allows to mitigate some of these limitations has become increasingly popular. Bayesian nonparametric (BNP) inference allows to acknowledge uncertainty about an assumed sampling model while maintaining a practically feasible inference approach. We could take this feature as a pragmatic characterization of BNP as flexible prior probability models that generalize traditional models by allowing for positive prior probability for a very wide range of alternative models, while centering the prior around a parsimonious traditional model. A formal definition of BNP is as probability models on infinite dimensional parameter spaces. A typical application of BNP is to density estimation.

In this shortcourse we review some of the popular models, including Dirichlet process (DP) models, Polya tree models, DP mixtures and dependent DP (DDP) models. We will review some of the general modeling principles, including species sampling models, stick breaking priors, product partition models for random partition and normalized random measures with indpendent increments. We will briefly discuss some of the main computational algorithms and available software. The discussion will be illustrated by applications to problems in biostatistics and bioinformatics.

Topics covered:

- Definition of BNP and introduction
- Density estimation: Dirichlet process (DP), Stick breaking, DP mixtures, DP clustering, DP mixtures: posterior simulation, Polya trees (PT)
- Regression: BNP survival regression, Dependent DP (DDP), Anova DDP, Weighted mixture of DP, Kernel stick breaking process, Gaussian process priors
- Hierarchical priors: Hierarchical DP, Nested DP
- Mixed effects models: Random effects distributions Multiple subpopulations & classification

Target audience & prerequisites: Anyone with an appreciation for data analysis, and basic knowledge of Bayesian inference. At the level of, for example, Hoff (2009), *A first course in Bayesian statistical models.* Or any other basic text in Bayesian inference.

References: Items #2-4 are free on-line, #1 available as free pdf from McGill's library (same for Hoff (2009))

1. The course will follow the book: Müller, P., Quintana, F., Jara, A., and Hanson, T. (2015), *Bayesian Nonparametric Data Analysis*, Springer.
2. Maybe read (before the course) P. Müller and R. Mitra (2013), "Bayesian Nonparametric Inference – Why and How", *Bayesian Analysis*, 8, 269-302.
3. Lecture notes of a similar course: Müller, P. and Rodriguez, A., (2012) *Nonparametric Bayesian Inference*, IMS Lecture Notes, free at `https://projecteuclid.org/euclid.cbms/1362163742`
4. Excellent notes by Peter Orbantz, at `http://www.gatsby.ucl.ac.uk/~porbanz/papers/porbanz_BNP_draft.pdf`