

A Bayesian Analysis of Some Nonparametric Problems

Thomas S. Ferguson

The Annals of Statistics

Vol. 1, No. 2 (Mar., 1973), pp. 209-230

Bayesian statisticians were traditionally silent in nonparametric problems until this pioneer work by Tom Ferguson published in 1973. Basically, we need to construct a workable prior on a space of distribution functions. As pointed out by the author: "There are two desirable properties of a prior distribution for nonparametric problems. (I) The support of the prior distribution should be large--with respect to some suitable topology on the space of probability distributions on the sample space. (II) Posterior distributions given a sample of observations from the true probability distribution should be manageable analytically. ... This paper presents a class of prior distributions, called Dirichlet process (DP) priors, broad in the sense of (I), for which (II) is realized, and for which treatment of many nonparametric statistical problems may be carried out, yielding results that are comparable to the classical theory." In application, the author illustrates the DP process on nonparametric estimation for a distribution function, a mean, a quantile, a variance, or a covariance. He further applies the DP prior to a two-sample problem, quantal bio-assay, empirical Bayes and discrimination problems.

This article stimulates a lot of interests, not only in statistics, but also in machine learning community. It is generally considered to be difficult for graduate students to read on their own for the first time. I will make the basic idea in the article simple and easy. So it will be easier for you to read. Looking back, I still consider the paper to be one of the best papers I have read. I am glad I read it while I was in graduate school.

Notes taken by Lynn Kuo