Bayesian Robustness
IMS Lecture Notes - Monograph Series (1996) Volume 29

MARKOV RANDOM FIELD PRIORS FOR UNIVARIATE DENSITY ESTIMATION[1]

BY ROBERT L. WOLPERT AND MICHAEL LAVINE
Duke University Institute of Statistics & Decision Sciences

We model the unknown distribution function F of a sequence of independent real-valued random variables by partitioning the real line into intervals $\{I_i\}$ and modeling the vector $p = \{p_i\}$ of probabilities assigned to the intervals using Markov random field priors (MRFPs). We argue and illustrate that many commonly-expressed prior opinions about the shape and form of F can be expressed as statements about the joint distribution of neighboring p_i's, leading to simple MRFP expressions for prior beliefs that are awkward to express in other models. In particular, we will show how to model beliefs about continuity, monotonicity, log concavity, and unimodality of a density function f for F. The posterior distributions of the p_i's in our models (and hence the approximate predictive distributions for subsequent observations) are readily computed using Markov chain Monte-Carlo methods.

1. Introduction. We consider the problem of making inferences about or predictions of observations from an unknown probability distribution $F(\cdot)$, on the basis of expressed prior belief or opinion about the nature of $F(\cdot)$ and also of some number $n \geq 0$ of independent observations $\mathbf{X}_n = \{x_1, \ldots, x_n\}$ from the distribution.

1.1. *Conventional Approaches.* Under the assumption that $F(\cdot)$ has a probability density function (pdf) $f(\cdot)$ with respect to Lebesgue measure, with $f \in \{f_\theta : \theta \in \Theta\}$ for some parametric family, predictive inference might be based on the "plug-in" predictive density $f_{\hat{\theta}_n}(x)$, the parametric pdf evaluated at the maximum likelihood estimator $\hat{\theta}_n$; or on a Bayesian predictive distribution $f_\pi(x|\mathbf{X}_n) \propto \int_\Theta f_\theta(x) e^{\ell_n(\theta)} \pi(d\theta)$; even without the parametric assumption, predictive inference might be based on the (degenerate) empirical distribution $\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}(x)$, with mass $1/n$ at each of the n observed points $\{x_i\}$, or a kernel density estimate $\hat{f}_n^\epsilon(x) = \frac{1}{n}\sum_{i=1}^n k_\epsilon(x - x_i) = \hat{f}_n * k_\epsilon(x)$, the convolution of the empirical

[1] Research partially supported by Environmental Protection Agency grant CR822047-01-0 and National Science Foundation grants DMS-9305699 and DMS-9300137.

Key words and phrases. Nonparametric Bayesian; Hammersley-Clifford theorem; Markov chain Monte-Carlo; Metropolis.