Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Statistics & Probability Letters 72 (2005) 153–162
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

On choosing the centering distribution in Dirichlet process mixture models

Timothy Hanson[a,*,1], Jayaram Sethuraman[b,2], Ling Xu[a,1]

[a]*Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131, USA*
[b]*Department of Statistics, Florida State University, Tallahassee, FL 32306, USA*

Received 23 August 2002; received in revised form 22 November 2004; accepted 20 December 2004
Available online 2 February 2005

Abstract

We present two results that pertain to the choosing of the centering distribution in a Bayesian setup with a Dirichlet process mixture prior based on Gaussian kernels. Our results indicate that for such kernels, one can choose the centering measure for the Dirichlet process mixture model exactly as one would in the analogous simpler Dirichlet process model.

Keywords: Bayesian; Gaussian; Nonparametric

1. Introduction

Dirichlet process (DP) mixtures have enjoyed much success in the Bayesian semiparametric literature (for classic examples see Escobar, 1994; Escobar and West, 1995; Müller et al., 1996; Kuo and Mallick, 1997; Müller and Rosner, 1997. Despite this activity, little has been written on choosing a centering distribution G_0 for these models. The purpose of this paper is to provide some guidance on choosing the centering distribution G_0 in certain DP mixture models.

*Corresponding author.
E-mail address: hanson@math.unm.edu (T. Hanson).
[1]Research of T. Hanson and Ling Xu was supported by N.S.F. Grant NSF-HRD 0206200.
[2]Research of J. Sethuraman was supported by Army Research Office Grant W911NF-04-1-0333.

0167-7152/$ - see front matter
doi:10.1016/j.spl.2004.12.008

Let the data consist of random variables $V_1, V_2, \ldots, V_n$ which are i.i.d. with common distribution function G. A Bayesian model will place a prior distribution on G. The pure DP model assumes that G has a Dirichlet distribution with parameter α; this model is usually described as follows:

$$G \sim \mathscr{D}_\alpha,$$

$$V_1, \ldots, V_{n+1} | G \sim \text{i.i.d. } G,$$

where V_{n+1} will stand for a future observation. Here the α is a positive measure on the real line R. The total mass is $\alpha(R) > 0$ and the distribution function (df)

$$G_0(x) = \frac{\alpha((-\infty, x])}{\alpha(R)}$$

is called the *centering distribution*. Some authors have incorrectly characterized $\alpha(R)$ as the prior number of observations; this is questioned in Sethuraman and Tiwari (1982). The Dirichlet distribution $\mathscr{D}_\alpha$ and its properties were established in Ferguson (1973). See Sethuraman (1994) for a simpler constructive definition and proof of its properties. From these and other sources, it is well known that

$$\text{the marginal distribution of } V_1 \text{ is } G_0, \tag{1}$$

the conditional distribution of V_i given $V_1, \ldots, V_{i-1}$ is

$$\left(\alpha(R) G_0(\cdot) + \sum_1^{i-1} \delta_{V_j}(\cdot) \right) \Big/ (\alpha(R) + i - 1), \quad i = 2, 3, \ldots, \tag{2}$$

where $\delta_a(x)$ is the distribution function corresponding to a random variable degenerate at a. The *predictive distribution* $P(V_{n+1} | V_1 = v_1, \ldots, V_n = v_n)$ is therefore given by

$$P(V_{n+1} \leqslant x | V_1 = v_1, \ldots, V_n = v_n) = \left(\alpha(R) G_0(x) + \sum_1^{n} \delta_{v_j}(x) \right) \Big/ (\alpha(R) + n). \tag{3}$$

This gives some guidance on choosing the centering distribution G_0.

The purpose of this paper is to examine if results like (1)–(3) hold in the more general DP mixture model described below. Let $K(x, u)$ be a kernel, i.e., for each u, $K(x, u)$ is a df in x and for each x, $K(x, u)$ is measurable in u. For any df G, the mixture kernel $K(x, G)$, defined as $\int K(x, u) \, \mathrm{d}G(u)$, is a df. The DP mixture model with kernel $K(\cdot, \cdot)$ may be described as follows:

$$G \sim \mathscr{D}_\alpha,$$

$$V_1, \ldots, V_{n+1} | G \sim \text{i.i.d. } K(\cdot, G),$$

where as before, V_{n+1} is a future observation, and α is a positive measure with $\alpha(R) > 0$. The df

$$G_0(x) = \frac{\alpha((-\infty, x])}{\alpha(R)}$$

will again be called the centering distribution.